University of Groningen

Laryngeal contrast and phonetic voicing

Jansen, Wouter

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2004

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch*. [Thesis fully internal (DIV), University of Groningen]. s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 23-04-2025

Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology Approach to English, Hungarian, and Dutch

Wouter Jansen

The work in this thesis was supported by grants from the Behavioral and Cognitive Neurosciences (BCN) research school Groningen, the Netherlands Organisation for Scientific Research (NWO; grant number 200-50-068), and the European Union's Training and Mobility of Researchers (TMR) programme through the Learning Computational Grammars (LCG) project (grant ERBFM-RXCT980237, principal investigator: John Nerbonne)

Document prepared with LaTeX 2$_{\varepsilon}$ and typeset by pdfTeX
Printed by PrintPartners Ipskamp Enschede

Groningen Dissertations in Linguistics 47
ISSN 0928-0030

RIJKSUNIVERSITEIT GRONINGEN

Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology Approach to English, Hungarian, and Dutch

Proefschrift

ter verkrijging van het doctoraat in de
Letteren
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
donderdag 27 mei 2004
om 13.15 uur

door

Wouter Jansen

geboren op 26 januari 1973
te Zuidhorn

Promotor : Prof. dr. ir. J. Nerbonne

Copromotor : Dr. D.G. Gilbers

Beoordelingscommissie : Prof. dr. C. Gussenhoven
Prof. dr. J. Hoeksema
Prof. dr. V. van Heuven

Preface

The work presented in this dissertation has greatly benefited from the generosity and patience of a number of people, and I am pleased to be able to repay some of my debt to them, finally.

First, much of the work presented here was supported financially by NWO (Gebied Geesteswetenschappen), in the form of grant 200-50-068. This grant, and the travel grant from NWO that allowed me to spend nearly six weeks at Ohio State University during the summer of 2000, are gratefully acknowledged. Two further sources of financial and infrastructural support should be acknowledged here, and with the same gratitude. First, the first four months of my PhD project were funded by a grant from the BCN graduate school of the University of Groningen. Second, towards the end of the project I was able to further develop and broaden my work at the Department of Linguistics of the University of Tübingen through generous financial support from the *Learning Computational Grammars* (LCG) project, which was funded by the European Union's *Training and Mobility of Researchers* (TMR) programme (grant ERBFMRXCT980237, principal investigator: John Nerbonne).

Next, I would like to thank my supervisors, Dicky Gilbers and John Nerbonne for helping me start a PhD project back in 1997, for encouragement, feedback, support in many practical matters, and above all for great amounts of patience in the face of ever-shifting topics and a major disappearing act. Thanks are also due to the members of my committee, Carlos Gussenhoven, Jack Hoeksema, and Vincent van Heuven, for their comments on the original manuscript. Thanks, furthermore, to Ilse van Gemert for all her help in getting this document printed.

None of the experimental part of this book would have been realised without the help of the volunteer subjects who took part in my experiments; I owe them many thanks for their time and effort.

Moreover, I am indebted to the following people for helpful feedback on parts of the work presented below as well as on my thinking on the subject matter in a more general way. They are listed here not in order of importance, nor alphabetically, but in a quasi-chronological order: John Harris, Moira Yip and the other members of the London phonology reading group;

Sue Barry, Stefan Ploch, Peter Sherwood, Mária Gósy, Péter Siptár, Miklós Törkenczy, Szilárd Szentgyörgyi, Sylvia Blaho, Catherine Ringen, Jim Scobbie, Patrick Honeybone, Wiebke Brockhaus, Heinz Giegerich. I am especially grateful to Gert Heise at Groningen and Bernard Howard at the School of Oriental and African Studies in London for their technical support during the recording of the speech material discussed in chapters 2, 4, and in particular 5, 6 and 7.

I've been very lucky over the years to work alongside, and sometimes even with, a number of people whose academic specialisms are different from my own to a greater or lesser extent, but whose views, tastes, and passions –and working hours of course– made me feel at home in academia more than anywhere else. I'll single out Ivelin Stoianov and Stasinos Konstantopoulos (formerly) at Groningen, as well as Michelle Gregory, Jason Brenier, Katja Jasinskaja, and Stefan Benus.

Many thanks to my other friends and my family, Nynke, Flip, Inge, Etienne, Maarten, Richard, Margaret, Kate, Al, for moral and financial support, for kindness, good humour, welcome distractions and more generosity in many other respects than would fit these pages. And I mean this to include well-timed supplies of fresh socks, and evil red-eyed bears, by the way.

Finally, and most importantly, thank you Zoë, for your unwavering support, encouragement, and help in every possible respect with the completion of this book. Far above all else, thank you for walking down Wilmslow Road with me that day in Manchester, and for everything that has come of it.

Dublin, 13th May 2004

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A considerable number of languages use phonetic voicing, the low frequency periodic energy in the speech signal that is produced by vocal fold vibration, to signal a two way lexical distinction between obstruents. For example, Dutch contrasts the voiceless plosives in [pɔl], *tussock (of grass)*, and [tɔl], *spinning top*, with the voiced initial plosives of [bɔl], *round, spherical*, and [dɔl], *crazy (about), foolish*. The two lexical categories identified by voicing in these languages are often described as phonologically *voiceless* vs. *voiced*, but such labels obscure the fact that voicing virtually always acts as part of a cluster of phonetic features when it is used to cue lexical contrast. For example, ceteris paribus, contrastively voiceless (aspirated) obstruents are usually relatively long, preceded by somewhat shortened vowels, and cause a slight increase in the F_0 and F_1 of flanking vowels. This is one of the reasons why I will refer to ‘phonologically voiceless’ obstruents as *fortis, tense*, or [+tense], and to their ‘phonologically voiced’ counterparts as *lenis, lax*, or [-tense].

Not all languages that have a [tense] contrast in this sense use the same voicing categories to cue fortis and lenis obstruents. One type of language contrasts voiceless aspirated fortis plosives (e.g., [p^h, t^h, c^h, k^h, q^h]) with passively voiced lenis plosives in word-initial and word-medial contexts. If the latter appear utterance initially or after another obstruent, they are generally realised as voiceless and unaspirated, e.g., [b̥, d̥, ɟ̊, ɡ̊, ɢ̥], but after a vowel or sonorant consonant they are commonly more or less voiced. I will refer to this type of language, which is exemplified by (standard varieties of) English and German as *aspirating*. A second type of language contrasts plain voiceless fortis plosives ([p, t, c, k, q]) with lenis plosives that are generally prevoiced across phonetic contexts ([b, d, ɟ, ɡ, ɢ]), and will be referred to as *voicing*. Southern and Western varieties of Dutch as well as French and Hungarian are typical voicing languages.

Crucially, the two types of language are consistent in the mapping of [±tense] into durational distinctions and spectral cues other than voicing. For example despite their differences in (utterance and post-obstruent) voicing the lenis stops of both voicing and aspirating languages are shorter than the corresponding stops, have longer preceding vowels, and act as F_0/F_1 depressors. This justifies the use of the four (and perhaps more) gross phonetic categories introduced in the previous paragraph to describe tense and lax obstruents, rather than the two or three sometimes suggested by the phonological literature: aspirated fortis ([pʰ]), plain voiceless fortis ([p]), passively voiced lenis ([b̥] utterance initially or after another obstruent), and actively voiced lenis ([b]).

This dissertation investigates the formal and phonetic properties of fortis and lenis obstruents, with a descriptive focus on the Germanic languages and Hungarian. It argues that these properties are best understood in terms of the nature of human speech production and perception, and is therefore broadly *functionalist* in outlook. The following paragraphs outline how the argument is built up.

1.1 Synopsis

Chapter 2 starts with a description of the production of voicing distinctions in obstruents and defines the notions *active* and *passive* devoicing in terms of the aerodynamic constraints that supraglottal articulatory settings impose on the initiation, continuation, and termination of vocal fold vibration. The second part of this chapter reviews the literature on the production and perception of the complex of cues that signals [tense] in stops and fricatives and the role of voicing within this complex.

Chapter 3 shifts the focus from the phonetic expression of [tense] to its neutralisation in the form of dynamic ‘final devoicing’ as well as at the lexical level. It discusses two issues that can be regarded as independent, but both of which are important to models of laryngeal neutralisation. The first of these issues is the nature of [tense] neutralisation itself. A long-standing and popular approach is to treat [tense] neutralisation processes as instances of phonological *fortition* or *lenition*, i.e. asymmetric rules that targets lax obstruents only and convert them into their respective tense counterparts, or vice versa. An alternative view regards the neutralisation of fortis-lenis distinctions as a symmetric phenomenon that derives a phonologically and phonetically distinct third category of [0tense] obstruents. From the available phonetic evidence it appears that [tense] neutralisation may not be a phonetically homogeneous phenomenon. Data from different languages and processes is sometimes consistent with the first view, sometimes with the second, and sometimes seems to support a third approach that essentially treats neutralisation as an (extreme) case of contrast *reduction* which leaves residual cues to lexical fortis-lenis contrast.

Neutralisation of [tense] contrasts is not equally probable across contexts and types of contrast-bearing sound. The second part of chapter 3 identifies the factors behind neutralisation asymmetries and contrasts formalist approaches to these asymmetries with perceptibility-driven functionalist accounts of the type developed by Steriade (1997). Formalist models tend to concentrate on neutralisation asymmetries induced by neighboring sounds and the position of target obstruents within morphemes or words, and propose that such asymmetries be explained in terms of syllabic and/or higher-order prosodic conditions on phonological rules. Cue-based accounts on the other hand, claim that neutralisation is more likely to occur in contexts where the contrast between tense and lax obstruents is relatively imperceptible, and less likely where it is relatively salient. One of the crucial predictions that distinguishes syllable-driven formalist models from a cue-based approach is that in languages with word-final [tense] neutralisation should also suspend the tense-lax distinction in word internal obstruent + sonorant sequences straddling a syllable boundary. Consequently, the observation that the occurrence of laryngeal neutralisation in obstruent + sonorant sequences is neither constrained by syllabification nor by the presence vs. absence of word-final neutralisation constitutes evidence in favour of the cue-based account.

Furthermore, although the scope of the cue-based model proposed by Steriade (1997) appears to be similar to that of syllable-driven formalist models, I will argue that, at least in principle, it extends naturally to the asymmetry between word-initial and word-final contexts and asymmetries between different types of obstruents. Provided that the hypothesised segmental, positional, and stress-based asymmetries in perceptibility are real, this means that a cue-based model is able to account for a range of neutralisation phenomena in terms of a single mechanism, which would make it far superior to any formalist model available in the literature.

Chapter 4 deals with the various forms of voicing assimilation that can be found in the Germanic group and beyond. Drawing on proposals for the analysis of sandhi phenomena more in general, it establishes criteria for distinguishing phonological from coarticulation-based forms of voicing assimilation and then uses these criteria to classify a number of assimilation processes as they are described in the literature. One of the most important generalisations that plays a role in this exercise is the observation that lenis stops only appear to trigger regressive voicing assimilation (*RVA*) under word sandhi if they belong to the actively prevoiced type, i.e., lenis stops only trigger RVA in voicing languages. This suggests that RVA at word boundaries is a coarticulatory phenomenon, or at least (diachronically) rooted in coarticulation processes. By contrast, voicing assimilation phenomena in morphological paradigms, such as the well-known past tense paradigms of English and Dutch are not phonetically conditioned in this way, an therefore appear to act as phonological rules.

Chapters 5, 6 and 7 report on three experiments designed to test whether RVA across word boundaries is indeed properly regarded as a coarticulation process. The first two experiments examine the phonetic behaviour of velar stop + alveolar consonant sequences in an aspirating variety of English (chapter 5) and Hungarian, a voicing language (chapter 6). Neither of these two languages neutralises [tense] in word-final context and therefore they represent an ideal testing ground for assimilation models. The results of the first experiment indicate that English has a purely coarticulatory form of regressive voicing assimilation at word boundaries. English passively voiced /d/ does not trigger assimilation in a preceding plosive, in contrast to actively voiced /z/, and to a lesser extent to (actively devoiced) /t/ and /s/. Moreover, assimilation only affects the duration of the voiced interval of the velar plosives, but not the duration of their closed phase or the length of the preceding vowels, which is again in full agreement with the predictions of a phonetic approach to RVA.

The results of the second experiment are more complicated. They indicate that as in English, Hungarian RVA is not a phonologically neutralising process. However, unlike the English data, the Hungarian data shows (near-) neutralisation of vowel length distinctions before some obstruent clusters. Although the observed patterns cannot be seen as assimilatory in a straightforward fashion, they contradict a purely articulation-based account and suggest that Hungarian RVA may be (partially) phonologised.

Chapter 7 discusses the results of the third experiment, which was designed to assess the assimilatory effect of Dutch word-initial /p, t, b, d, m, h, V(owel)/ on a preceding /ps/ cluster. The results of this experiment indicate that, as in English, Dutch RVA affects phonetic voicing but not duration features (or F_0) and thus support a phonetic account of RVA in Dutch. Moreover, the data from this experiment calls for a revision of the standard conception of Dutch RVA as a [tense]-asymmetric process triggered by lenis but not fortis obstruents: both the lax plosives /b, d/ and the tense plosives /p, t/ cause statistically significant changes in the duration of the voiced interval of preceding /ks/ and /ps/ clusters vis-à-vis /m/. This finding is consistent with the phonetic underspecification approach to Dutch word-final neutralisation proposed by Ernestus (2000).

Chapters 2 through 7 reject the general thrust of formalist approaches to laryngeal phonology and phonetics in favour of an auditory model of laryngeal neutralisation and an articulatory model of RVA. This argument is mainly founded on the distribution of laryngeal contrast and the phonetic manifestation of regressive voicing assimilation. Some might argue that these are insufficient grounds for an outright rejection of formalist approaches because such approaches still have a role to play, for example in defining the set of laryngeal neutralisation and assimilation rules that the human mind is able to represent. More specifically, they might point out that most of the predictive power of current generative models resides in the detail of modality-neutral segmental representations, and that these models are therefore capable of narrowing the range of phenomena that have to be explained in auditory, articulatory, or other functional terms.

The role of formalist models as a possible source of metaconstraints on functional explanations is investigated in the sixth and final chapter. Two general designs are discussed here: [tense]-based models along the lines of (Lombardi 1994 et seq.), and the VOT-based models proposed by e.g., Harris (1994) and Iverson & Salmons (1995, 1999). Both models are found seriously wanting, because the predicted connections among laryngeal neutralisation rules, regressive assimilation processes, the behaviour of the ‘Germanic’ past tense paradigm, and other phenomena are not borne out by the data. Moreover, under a strict interpretation of monovalent feature representation, both models undergenerate, in particular with regard to the ‘phonologically active’ nature of plain voiceless fortis obstruents. Neither of these models can therefore be regarded as in any sense complimentary or prerequisite to the functional accounts of RVA and [tense] neutralisation developed in the preceding chapters. The failure of the formalist enterprise is further underlined by the observation that representationally richer frameworks are successful to the extent that they approximate continuously-valued feature systems constrained by grammar-external (functional) principles.

The remainder of this chapter outlines the phonological and phonetic transcription conventions used in this study (section 1.2), and more importantly, the descriptive model underpinning chapters 2-8.

1.2 Notes on transcription

Lexical contrasts are transcribed with slanted brackets, e.g., /p, b/, whilst the physical/perceptual manifestations of phonological categories are symbolised using square brackets, e.g., [pʰ, p, b̥, b]. Orthographic forms appear in angular brackets (<, >) and in running text, glosses of non-English words are italicised.

With three exceptions, all impressionistic data from the literature are transcribed as in the sources. The same applies to data from specific regional varieties of Dutch and English. However, phonetic data concerning standard Dutch is represented according to my own pronunciation of the standard language: i.e. with [χ] for /x/ and /ɣ/ in all contexts, with dipthongised long mid vowels [eʲː, øʸː, oʷː], [ɑu] for the back diphthong /ɔu/, and [r] and [ɹ] for onset and coda /r/ respectively. I have transcribed the lax front rounded vowel that is often analysed as /œ/ with [ʏ]. In the transcription of Dutch underlying forms the IPA diacritic [ː] for long sounds is used to represent the set of ’tense’ or phonotactically long vowels, even though the high tense vowels are phonetically short in Dutch standard (and my own) pronunciation (so /iː, yː, uː/ for phonetic [i, y, u]).[1]

Where regional variation is not an issue I have chosen the (southern) British, non-rhotic variety that forms the basis for the pronunciation dictionary of Wells (2000) to represent English data. Apart from the absence of coda /r/ the most notable feature of this variety is a phonemic distinction between the low vowels [æ], [ɑː], [ɒ]. Finally, standard German data are transcribed according to the conventions in Drosdowski & Eisenberg (1995).

1.3 The descriptive framework

1.3.1 Linguistic and extralinguistic speech processing

Few phonologists would disagree with the idea that there are peripheral stages in the production and perception of speech that are independent of any form of linguistic knowledge. Take for example the pulsing of the vocal cords during voicing. It is universally accepted that the individual pulses of the glottis do not result from individual instructions (nerve firings) to the vocal folds. Instead, the musculature of the larynx is more or less static during the production of vocal fold vibration (barring changes in pitch or movements of the larynx as a whole), forcing the glottis to be closed but not too tightly adducted. Glottal pulsing then arises through the aerodynamic-myoelastic effects of pushing air from the lungs through the closed glottis (van den Berg, 1958). Similarly, no one would want to describe mechanical interactions between the movement of the tongue root and tongue tip, or the fact that the physiology of the inner ear warps the incoming acoustic signal in various ways, as linguistic knowledge.

In addition, certain short-term adaptations in articulator movements appear to be beyond what most researchers regard as *linguistic* control. For instance, if the closing gesture of the lower jaw is suddenly interrupted during the production of a bilabial constriction, speakers compensate with increased movement of the upper and lower lips. The lag between the interruption of the lower jaw gesture and the onset of compensatory articulations (often $\leq$ 30 ms) cannot be attributed to any sort of mechanical linkage. Given that reaction times (to linguistic tasks) typically run in the hundreds of milliseconds it is not plausible either that short term adjustments of this kind are orchestrated at any level

---

[1] Dutch [i, y, u] share the phonotactics of long vowels such as [aː] rather than ‘true’ short vowels such as [ɪ, ʏ]. Thus, they can appear in open monosyllables (e.g., [ku], *cow*) and open final syllables. In these contexts they can only be closed by a single consonant (modulo the same exceptions that apply to the other long vowels) whilst they can only occur in open non-final syllables. Characterising [i, y, u] as simply *long* is not wholly unproblematic however, because standard Dutch does allow phonetically long high vowels in loans such as [analiːze], *analysis*. This has created near-minimal pairs such as [zun], *kiss* vs. [zuːm], *zoom*. However, in the absence of an agreed IPA diacritic for ‘tenseness’ I have opted to appropriate the length diacritic to mark the class of phonotactically long vowels in Dutch underlying representations.

of (linguistic) planning. Moreover, similar short-term adaptations have been observed in other, non-linguistic forms of motor behaviour, such as hand and finger movements. Consequently, they are normally treated as reflex-like behaviour triggered by proprioceptive feedback (see Saltzman & Munhall 1989 for an overview and references).

Note that all these ‘physical’ and otherwise extralinguistic aspects of speech processing are roughly what is modelled by the articulatory synthesis models of Ishizaka & Flanagan (1972), Boersma (1998), the *task-dynamic model* implementing the *gestural scores* of (Browman & Goldstein 1986 et seq.), or the cochlear model of Lyons (1982).

There cannot be many phonologists either, who would dispute the claim that the information that is exchanged at the interface between the extralinguistic levels of speech processing and linguistic competence is discretised at anything near the granularity of lexical phonological features (this information corresponds to the input and output parameters respectively of the models mentioned in the previous paragraph). Speakers are able to vary the position of their tongue, the pitch of their voice, their speaking rate, and many other speech features on what for all practical purposes are continuous scales. Similarly, although the mechanics of the inner ear (and pre-cortical processing) introduce various nonlinearities in the signal, and although e.g., the frequency resolution of the human auditory signal is far from infinite, this resolution is again greater than that of virtually all phonological feature systems. For instance, Boersma (1998) estimates that (cardinal) [i] and [u] are 12 *Just Notable Differences* (JNDs) apart in auditory (F_1-F_2) space, but no known language has 11 intermediate vowels between [i] and [u] along the front-back dimension (i.e. vowels with the same auditory F_1).

The scalar nature of the information that is exchanged between linguistic competence and the peripheral physical systems is also evinced by the observation that languages that according to ‘broad’ descriptions share sounds or sound inventories, often display subtle but reliable phonetic differences between seemingly equivalent sounds. It is well-known for example, that Danish /i/ is somewhat higher and fronter, on average, than English /i/ (Disner, 1983), and Bradlow (1995) finds similar differences between English and Spanish vowels. As Pierrehumbert et al. (2000) point out, there is no reason to assume that this sort of crosslinguistic variation is constrained in terms of points on a discrete scale, and so it must be concluded that linguistic competence includes knowledge that is best represented on continuous scales.

Although the topic of gradient but linguistic processing has come to the fore in recent years, its existence is acknowledged by Chomsky & Halle (1968), who conceive of lexical representation in binary terms, but allow features to acquire scalar values at the final stages of a derivation. Lexical Phonology also allows features with scalar values, at least at the postlexical level (Kaisse & Shaw, 1985; Mohanan, 1986). Other models seek to model all linguistic processing in terms of discrete representations and therefore try to dispense with the 'systematic' or 'linguistic' phonetic level as a significant level of representation (Pierrehumbert & Beckman, 1988; Kaye, 1989; Coleman, 1992; Harris & Lindsey, 1995). But to the extent that they are intended as (partial) models of human speech production and perception, such frameworks cannot go without a module that translates between discrete feature structures and the continuously-valued information that is supplied and required by the relevant peripheral physical systems.[2]

From here on, I will refer to the collective aspects of speech processing that are guided by linguistic competence, i.e. both categorical and gradient processes, as the *phonetic grammar*. Similarly, following Kingston & Diehl (1994) I will refer to the part of linguistic competence that guides this collection of processes as *phonetic knowledge*. The next sections are devoted to the assumptions this study makes about the organisation of the phonetic grammar.

1.3.2 Phonology and phonetics

A conception of the phonetic grammar that is associated with a lot of (early) work in laboratory phonology holds that categorical and gradient processes operate in two separate modules and on fundamentally different feature structures (Keating, 1990a; Gussenhoven, 1996). According to this view, the *phonology* is the module that deals with lexical representations and categorical rules, whereas the (linguistic) *phonetics* takes care of the subsequent gradient processes. At the interface between the two levels, discrete phonological representations are translated into continuously-valued structures of the sort that are produced/understood by the physical levels.

In this type of framework, the Dutch vowel /i/ is represented by the phonology as [+high, -low, -back, -round], or some equivalent (autosegmental) structure. This discrete structure is translated from an auditory representation with F_1 and F_2 values of, say, 3.5 and 14 Bark (339 and 2357 Hz) for a male (Dutch) speaker by the phonetics-phonology interface (or rather from normalised values that filter out the effect of speaker size). The articulatory component of the phonology-phonetics interface translates the discrete representation of /i/ into the corresponding instructions or *targets* for the physical system. In a language

---

[2] Proponents of the latter type of model often subscribe to a 'denotational' view of the relationship between phonetics and phonology. On this view, phonological structures denote real-world articulatory and acoustic 'events'. Note that, despite appearances to the contrary, this view does not entail that all linguistic structure is discrete: it is technically possible to take a denotational view of a systematic phonetic representation consisting of scalar features. It is difficult to see however, how a denotational phonology-phonetics interface could be embodied by real human language users, whose knowledge about articulatory and acoustic events is mediated by peripheral processing of acoustic, visual, and proprioceptive feedback (cf. Pierrehumbert et al. 2000).

in which /i/ is subject to a categorical labial harmony process, the phonology first converts its structure into [+high, -low, -back, +round] (equivalent to lexical /y/), and the phonology-phonetics interface translates between this structure and the appropriate perceptual (say, 12.5 Bark, 1884 Hz or again, the appropriate value on a normalised scale) and articulatory scales.

An alternative approach, which is embodied in the framework of *Articulatory Phonology* (e.g., Browman & Goldstein 1986 et seq.; Byrd 1996a) and adopted by Flemming (2001) and Pierrehumbert et al. (2000), is to dispense with the phonology as an independent model and represent both categorical and gradient processes in terms of continuously-valued feature structures. On this view, the lexical representation of Dutch /i/ consists simply of F_1, (value: 3.5 Bark) F_2 (value: 14 Bark), other relevant spectral and durational parameters, and the corresponding articulatory targets. The labial harmony rule referred to above is assumed to act directly on these parameters, changing the F_2 value of /i/ into 12.5 Bark. Since this 'is' the lexical F_2 value of /y/, the harmony rule acts as a categorical, neutralising process even though it is stated in terms of gradient features. In other words, the single-module approach capitalises on the fact continuously-valued features can encode categorical processes (and thus eliminates the duplication of information at the phonetics-phonology interface: see further below).

Strictly speaking these two conceptions of the phonetic grammar are independent from the choice between a formalist or functionalist view on the origin of phonological and phonetic constraints. However, practically speaking, formalist models that aim to explain the nature of phonological constraints depend on a separation of phonetics and phonology. As discussed in 1.4 below, formalist models usually derive the set of possible phonological rules from an alphabet of representational primitives and a severely restricted set of combinatory principles. However, if phonetic knowledge is encoded in terms of continuous representations the number of possible natural classes is infinite or (stipulating that all categories must be at least 1 JND apart) at least very large. Consequently, the number of possible rules that can be derived according to the formalist logic grows very large as well, and the resulting grammars are almost guaranteed to be massively overgenerating.

Therefore, the non-modular[3] conception of the phonetic grammar more or less implies a (partially) functionalist perspective on the origin of phonological and phonetic constraints. Functionalist models derive the set of possible, or rather *probable*, rules from external, 'ecological', factors, such as need for robustly perceptible cues to phonological distinctions. Consequently, functionalist

---

[3] I call this view of the phonetic grammar *non-modular* because it does not distinguish separate phonological and linguistic phonetic modules. However, strictly speaking it is still modularised, because it consists of articulatory and auditory (as well as other perceptual) components.

models are able to rule out e.g., processes that change the F_2 value of a vowel upwards by the equivalent of 1 Hz: the effects of such a process would simply be imperceptible.

Although I ultimately subscribe to the single-module conception of the phonetic grammar, I will refer to *phonological* (categorical) vs. *phonetic* (linguistic gradient) rules, mainly for expository reasons, and using the diagnostics identified by Myers (2000) to distinguish between the two types of processes. For the same reasons, I will refer to (lexical) *phonological categories* and their *phonetic interpretation*. In transcriptions, the former will be indicated by slanted, and the latter by square, brackets. For instance, the contrastively voiced labial and alveolar stops of Dutch will be referred to as phonologically [-tense] and symbolised as /b, d/ if their lexical status is at issue, but (outside neutralisation contexts) as [b, d] where their phonetic properties are relevant to the discussion.

However, because these labels merely serve descriptive convenience, I will not make any specific assumptions about the 'nature of phonological representation'. [±tense] is used to represent lexical laryngeal contrast rather than the more familiar [±voice] to keep track of the essential distinction between phonetic voicing and what is often known as 'phonological' voicing, but nothing of importance hinges on this. Where it is relevant, the representations used by others will be described in what I hope is sufficient detail. Furthermore, I will use the terms *rule*, *process* or *constraint* in a purely descriptive way, without committing to a procedural (derivational) or declarative interpretation.[4]

1.3.3 Phonetic rules and representations

Johnson et al. (1993) describe the production of an utterance as a two-stage process. As they subscribe to a modular theory of the phonetic grammar, the first step maps discretely-valued phonological features into phonetic features with values drawn from continuous articulatory scales. The second step modifies the initial values of these features to derive the variation in the realisation of phonological categories that is observable in speech. Because Johnson et al. describe the second step as a mapping between 'parametric phonetic' representations, I will assume that they are identical in nature, i.e. that they consist of the same set of features that range over the same scales of values. Moreover, I will assume that the output of the second step represents the instructions to the physical articulatory system, and thus that Johnson et al.'s 'parametric-to-parametric' mapping encompasses the full body of linguistic phonetic (i.e. gradient) rules in the sense defined above.

---

[4] Models that maintain separate phonological and linguistic phonetic components are inherently procedural at the interface, where discrete representations are converted into gradient ones. Single-module phonetic grammars however, can be modelled in terms of declarative constraints (Flemming, 2001).

The central topic of Johnson et al.'s paper is the nature of the initial values of the parametric phonetic representations, in a sense the 'underlying' phonetic values. They argue that these values are chosen to optimise auditory contrast among phonological categories within the available phonetic space. Using terminology associated with similar theory of speech production proposed by Lindblom (1990), they label these optimally spaced points *hyper(articulated) targets*. The second, phonetic rule, stage in the production process either maintains these hypertargets, or modifies them in a way that generally speaking results in a diminished amount of contrast between phonetic categories in the resulting utterance. Following Lindblom's theory, I will refer to the latter phenomenon as *hypoarticulation* but it is essentially similar to the idea of target *undershoot*.

The variable realisation of vowels serves as a simple illustration of Johnson et al.'s model. The phonology-phonetics interface is assumed to assign the same 'peripheral' auditory F_1/F_2 formant values to the 4 vowels /i/ ([+high, -low, -back, -round]) , /æ/ ([-high, + low, -back, -round]), /ɑ/ ([-high, +low, +back, -round]), and /u/ ([+high, -low, +back, +round]) regardless of the phonetic context, the degree of stress or of speech rate and register. These hypertargets are indicated by the black dots in figure 1.1. More centralised vowel realisations, which are found in unstressed syllables for example, are derived by the subsequent application of phonetic rules, and so are effects of segmental context, such as the fronting of back vowels before coronal consonants (cf. Flemming 2001).

The hypothesis that auditory dispersion or contrast optimisation more generally plays a role in the structuring of (phonetic) sound inventories is not new to the model of Johnson et al. (1993). But they are among the first the present direct evidence for the idea that hyperarticulated targets play a role in speech processing. They report 3 *Method Of Adjustment* (MOA) experiments in which test subjects were asked to adjust the settings of a vowel synthesiser until the output matched what they perceived as the vowels in a list of visually presented stimulus words. The same set of subjects were asked to read the stimulus words in normal 'citation' and hyperarticulated forms (the latter elicited by way of feedback from the experimenter). The responses to the MOA task show that the subjects systematically selected sounds with more extreme formant values than the values they produced in the citation form reading task, even when a number of potential confounds (e.g., phonetic training of the test subjects) were eliminated: see figure 1.1. The vowel space of the hyperarticulated readings corresponds more closely to the boundaries found in the MOA experiments. Johnson et al.'s tentative conclusion from the observation that the test subjects treated the hyperarticulated vowels as representative for the stimulus words is that hypertargets are primary to reduced forms in speech production.

Johnson et al. (1993) do not offer a formal implementation of the mapping between parametric phonetic representations: such implementations are pro-

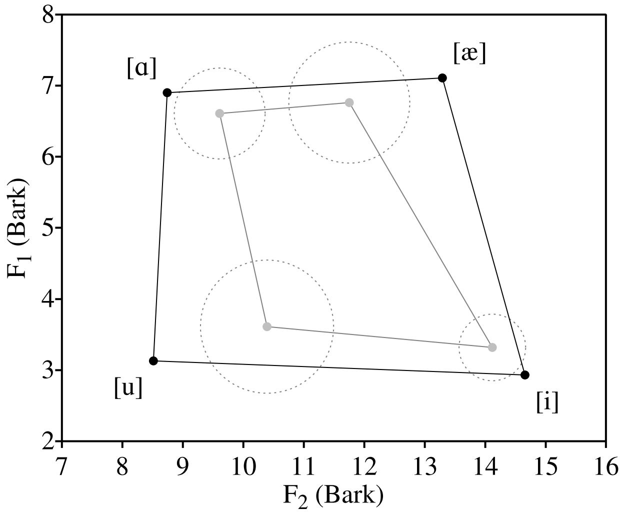

Figure 1.1: The *hyperspace effect* according to Johnson et al. (1993). The hypertargets in black correspond to the MOA results for /i, æ, ɑ, u/ of Johnson et al.'s experiment 1, converted to Bark; the grey dots represent the average values for the citation form readings of the same vowels in the same experiment (Johnson et al. 1993:520).

vided by Articulatory Phonology (Browman & Goldstein 1986 et seq.; Byrd 1996a) and the *Window Model* of coarticulation (Keating 1990a; see also e.g., Huffman 1993). What matters however is their notion of hypoarticulation as increased variability. For example, in contexts which allow little or no hyperarticulation, the physical articulatory system is instructed to produce the English vowels /i, æ, ɑ, u/ with articulatory gestures that match the black dots in figure 1.1 very precisely. But in hypoarticulation contexts, the F_1-F_2 (and corresponding articulatory) values vary across wider ranges of values, e.g., those indicated by the dotted grey circles in figure 1.1 in a way that is determined by the effort put in by the speaker and the phonetic context.

This study assumes that hypoarticulation, conceived as a relaxation of auditory specificity, and implemented as a reduction of articulatory effort, is the driving force behind many phonetic rules. This auditory-articulatory take on hypoarticulation phenomena is borrowed from the *Hyperarticulation and Hypoarticulation* (H & H) theory of Lindblom (1990) and similar hybrid accounts (Boersma, 1998; Flemming, 2001) and contrasts with the purely articulation-driven views espoused by Articulatory Phonology and Kirchner (1998).[5]

---

[5] For an overview of coarticulation phenomena and models, see Farnetani (1997), and for def-

Two broad classes of phonetic rule that can be construed in terms of hypoarticulation are *coarticulation* and (phonetic) *reduction/lenition*. The former term will be used to refer to situations in which the realisation of a particular sound is influenced by that of a temporally close second sound with which it shares one or more (mechanically linked) articulators. For example, the precise constriction location of intervocalic /k/ in English and other languages depends on the quality of the flanking vowels: it is slightly fronter between front vowels but somewhat backed between back vowels. Conversely, the place of articulation (F_2 in rough acoustic/perceptual terms) of vowels is often influenced by the neighbouring consonants: back vowels /u/ tends to be somewhat fronter between coronal than between velar consonants (Lindblom, 1963; Flemming, 2001). Although interactions between vowel place and consonant place of articulation have been phonologised by numerous languages, e.g., in terms of velar fronting (palatalisation) or the alternation between front and back velars that is common in the Turkic languages, they also occur as gradient processes.

These gradient processes can be understood in hypoarticulatory terms as follows. Vowels and non-labial consonants share an active articulator: the tongue. This means that if e.g., a front vowel and a back consonant are produced in sequence, the tongue has to be retracted a certain amount within a relatively short time span. The amount of retraction that is needed depends on the hypertargets for the vowel and consonant and the degree in which the realised targets are allowed to deviate from these hypertargets, i.e. the degree of hypoarticulation. All else being equal, if there is a low degree of hypoarticulation, both the vowel and consonant have to be realised with targets that are relatively 'faithful' to their hypertargets, which results in a relatively great amount of tongue retraction and hence a limited amount of observable place (F_2) coarticulation. If the realised targets are allowed to deviate from the hypertargets to a greater extent, the front vowel *can* be realised with a less front articulation and the back consonant with more fronting, which means a smaller amount of tongue retraction. The reason that the vowel and consonant *are* realised with relatively close constriction locations is, according to many, that the smaller tongue displacement saves articulatory energy (e.g., Lindblom 1990; Boersma 1998; Kirchner 1998; Flemming 2001). Phrased in more general terms, relaxing constraints on the realisation of targets in auditory space allows for smaller transitions in articulator movements, and hence for a lower articulatory energy expenditure.

---

initions of articulatory effort, Boersma (1998) and Kirchner (1998). Note that the theory of coarticulation summarised in these paragraphs does not entail that speakers compute the articulatory energy involved in the realisation of an utterance (and a range of alternatives) before they produce it, as is suggested by the work of Boersma, Kirchner, and also Flemming (2001). Effort considerations could equally well enter the phonetic grammar if speakers receive some form of feedback about the energy consumed by the production of utterances with a given phonetic makeup, and simply learned from this experience to avoid overly difficult forms. See further below.

I will use the terms *reduction* and *lenition* for realisations of segments that deviate from their hypertargets in a way that can be described as a decrease in the overall magnitude (and speed) of the relevant articulator movements. A hypoarticulation-based account attributes lenition in this sense to exactly the same mechanism as coarticulation, although it involves the somewhat problematic notion of a *neutral vocal tract configuration*. The basic idea is that relaxation of auditory constraints on the realisation of a given target not only allows for deviations to accommodate the implementation of neighbouring sounds, but also a reduction in the magnitude of articulatory gestures with regard to some equilibrium point. This equilibrium point is often defined as the vocal tract configuration for schwa. For vowels, a gradient reduction in gesture size results in gradient centralisation whilst for stops it leads to shortening (to the extent that the duration of a stop is due to the magnitude of the closing gesture), affrication, spirantisation, or gliding, depending on the amount of gestural weakening. Because both phenomena are seen as reflexes of the same mechanism, the prediction of a hyperarticulation-based theory of coarticulation and reduction/lenition is that reduced sounds should always show an increased amount of coarticulation with neighbouring sounds, and vice versa.[6]

1.3.4 Hypoarticulation and prosody

Judging by the behaviour of reduction and coarticulation phenomena the degree of hypoarticulation varies at both global and local levels. Globally it varies with speech rate and register. For example, Moon & Lindblom (1994) show how vowel reduction and consonant-vowel coarticulation increase with decreasing clarity of speech, where clarity is defined in terms of the instructions given to the test subjects. Fast speech is generally considered to be conducive to hypoarticulation, and the evidence in the literature broadly supports this view. Studies such as Lindblom (1963), Engstrand (1988), Byrd & Tan (1996), Kessinger & Blumstein (1997), record increased undershoot and coarticulation of targets for pitch, VOT, and place for vowels and consonants. On the other hand, speech in noisy environments has often been claimed to be hyperarticulated: this phenomenon is also known as the *Lombard reflex*. The review by Junqua (1996) of work on speech in noisy environments notes several features also found in clear speech elicited by different methods, though there is also evidence for speaker-dependent and more fine-tuned adaptation of speech to specific types of noise.

A number of factors seem to condition more local fluctuations in hypoartic-

---

[6] Note that the view of lenition/reduction described here is similar to the conception of (phonological) lenition as the loss of phonologically marked structure that is developed by Harris (1994), especially if the resulting unmarked configurations are interpreted in terms of *phonetic underspecification* (see 1.3.5 below). An interpretation in these terms appears to be suggested by Harris & Lindsey (1995).

ulation. Since a lot of work on local hypoarticulation has focused on its articulatory reflexes, such fluctuations are now commonly referred to as *articulatory strengthening* and *weakening* (e.g., Pierrehumbert & Talkin 1992; De Jong 1995; Jun 1995; Gordon 1996; Byrd & Saltzman 1998; Hsu & Jun 1998; Keating et al. 1998; Fougeron 1999). The factors involved include (lexical) stress, morphosyntax, and information structure. The effects of the latter two variables are often assumed to be mediated by a *prosodic phrase structure* (Halliday, 1960; Selkirk, 1986; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Ladd, 1996) and since lexical stress is part of prosody structure by virtually all definitions of the term, I will refer to their collective effects on phonetic realisation as *prosodic*.

Prosody introduces two major hypoarticulation asymmetries: one between (lexically) stressed and unstressed contexts, and a second one between constituent-initial and constituent-final contexts. Stressed syllables and constituent-initial positions are relatively resistant to reduction and coarticulation, and under the theory sketched in the previous section these environments should therefore be considered local hypoarticulation minima. Unstressed medial, and final contexts on the other hand, often exhibit consonant lenition, vowel reduction, and increased levels of coarticulation, and might therefore be regarded as local hypoarticulation maxima.

Observations about the relation between prosody and segmental realisation have been made both before, and outside the context of, recent experimental work explicitly couched in terms of articulatory strengthening. For example, Jones (1956) as well as Kahn (1976) highlight the role of lexical stress in the realisation of English fortis stops, which have more aspiration in the onsets of stressed syllables than elsewhere. The various lenition processes that affect English /t/ outside strengthening contexts is documented and analysed by Harris (1994). However, instrumental studies on articulatory strengthening have both quantified these and other phenomena, and demonstrated that they are much more general than might be gleaned from impressionistic descriptions of vowel reduction and consonant lenition. For example, Turk (1992) shows that, like alveolar stops, English labial and velar stops are subject to shortening in intervocalic contexts, even if the consequences are less perceptible than those of flapping.

Instrumental studies have also uncovered evidence indicating that the asymmetry between initial and final contexts is not restricted to the (prosodic) word level, but holds across higher levels of prosodic phrasing as well, in a way that is sensitive to juncture strength. For example, in a survey of 4 languages Keating et al. (1998) find that the amount of peak linguopalatal contact and seal duration (the duration of full oral tract constriction) in constituent-initial /t, n/ increases with the strength of the preceding juncture, and thus that within a given constituent, peak contact and seal duration are greater in initial than in medial contexts. For example their EPG data for two French speakers show a mean maximal contact of $>$ 60% of the measurement area for Intonation Phrase (*IP*)-initial /t/, which drops to just above 50% for IP-medial word-initial /t/. There is some evidence that the strengthening effects of stress and position are mutually reinforcing (i.e. initial stressed syllables are less hypoarticulated than stressed noninitial ones) but the effect is not simply additive (Lavoie, 2001). In addition, instrumental studies have established a correlation between the amount of segment-to-segment coarticulation and prosody. Work by De Jong et al. (1992) and De Jong (1995) shows that segments in syllables bearing lexical stress are less coarticulated than similar sequences in unstressed syllables.

1.3.5 Absent targets: phonetic underspecification

There is a considerable amount of data to suggest that where a given phonological contrast is neutralised, the resulting sounds sometimes lack targets for the phonetic parameters that signal that contrast in other, non-neutralisation environments. For example, chapter 3 discusses evidence adduced by Ernestus (2000) that the word-final laryngeal neutralisation (‘final devoicing’) of Dutch obstruents produces stops and fricatives without targets for phonetic voicing, segmental duration and other cues to [±tense]. Of course final obstruents in Dutch have voiced and voiceless intervals of definite lengths, but Ernestus claims that this voicing is completely derived from coarticulation. The oral tract configurations for stops and fricatives militate against the continuation of voicing after the offset of the preceding vowel or sonorant beyond certain (aerodynamically determined) points (see section 2.1 below), and utterance finally, this ‘segment-internal’ coarticulation results in the eponymous final devoicing. However, utterance medially, coarticulation with flanking (voiced) sonorant sounds and especially actively voiced lenis obstruents ([b, d]) is predicted to result in a greater amount of voicing for neutralised obstruents, and, as highlighted by experimental data in chapter 7, this is exactly what is observed.[7]

The analysis of final obstruent neutralisation in Dutch defended by Ernestus (2000) is an instance of a more general descriptive tool that gained popularity in the early years of laboratory phonology and is commonly known as *surface*, or *phonetic*, *underspecification* (Pierrehumbert & Beckman, 1988; Keating, 1988). Note that phonetic underspecification is effectively the limiting case of hypoarticulation in the sense defined above: it describes sounds that allow the maximal amount of variability (that is physically possible) with regard to the underspecified phonetic dimension. So whilst the [+tense] obstruents of Dutch are specified

---

[7] Ernestus’s analysis of Dutch final obstruent neutralisation is discussed in more detail in chapter 3 below.

as mostly voiceless and at least its [-tense] plosives as voiced for the larger part of their durations, the voicing of neutralised final obstruents is allowed to range across the whole continuum from fully voiceless to fully voiced. Phonetic underspecification of the complex of phonetic cues that signal [tense] therefore defines a third category of [0tense] (neutralised) obstruents in addition to the [±tense] stops and fricatives that occur in non-neutralisation contexts. This contradicts standard analyses of final laryngeal neutralisation in Dutch, which hold that neutralised obstruents are [+tense] and therefore phonetically indistinguishable from [+tense] obstruents in environments where the [tense] contrast is not suspended.

Th account of Japanese tonal phonology in Pierrehumbert & Beckman (1988) is one of the original studies that developed phonetic underspecification in an area where full specification had been the explicit norm, and thus serves as a good illustration of the mechanics of the device. Japanese is a pitch accent language in which the presence and place of a tonal accent in a word is lexically contrastive, but not the shape of the tonal melodies of accented and unaccented words. Nevertheless, many earlier accounts claim that all syllables in Japanese are phonologically and hence phonetically specified for tone and so they represent the phonological melody of /moriya-no mawari-no o *ma*warisan/, *the Forests-neighbourhood policeman*, where the italicised segments indicate the sole accented syllable, approximately as in (1a). The most natural interpretation of this melody assigns high pitch targets to H and low targets to L and therefore derives a rise from /mo/ to /ri/ followed by a high plateau and a relatively abrupt fall between H-toned /no/ and /o/:

(1) Specification of Japanese pitch contours (after Pierrehumbert & Beckman 1988)

a. Full specification

L	H	H	H	H	H	H	H	L	H	L	L	L	L
mo	ri	ya	no	ma	wa	ri	no	o	ma	wa	ri	sa	n

b. Phonetic underspecification

L	H							L	H	L			L
mo	ri	ya	no	ma	wa	ri	no	o	ma	wa	ri	sa	n

However, Pierrehumbert and Beckman find that the pitch contours represented by this melody show a gradual fall from a peak corresponding to the H tone on /ri/ to the L on /o/. Moreover, systematic manipulation of the number of moras between the initial LH sequence and second L and varying the phonological length of the syllable carrying the second L shows that the slope of this contour is an approximately linear function interpolating between the pitch values of the first H and the second L. They conclude that the syllables in the /ya…no/ interval cannot be assigned pitch targets like the Hs on /ri/ and /ma/

or e.g., the L on /wa/, which correspond to clear local highs and lows in the pitch contour, but form a third distinct category of syllables with regard to phonetic interpretation in not bearing a pitch target. There are no phonological grounds for retaining the Hs on these syllables, because they do not mark lexical contrast directly or indirectly by conditioning the distribution of other features, and therefore Pierrehumbert & Beckman (1988) represent them as underspecified for tone targets, as in (1b).[8]

1.4 Formalism vs. functionalism

> Since language is not, in its essence, a means for transmitting such [cognitive] information – though no one denies that we constantly use language for this very purpose – then it is hardly surprising to find in languages much ambiguity and redundancy, as well as other properties that are obviously undesirable in a good communication code. In sum, the theme of language as a game opens up perspectives that are by no means unattractive, so that others might wish to explore them further. (Halle 1975:528)

> We may say that a living body or organ is well designed if it has attributes that an intelligent and knowledgeable engineer might have built into it in order to achieve some sensible purpose, such as flying, swimming, seeing, eating, reproducing, or more generally promoting the survival and replication of the organism's genes. It is not necessary to suppose that the design of a body or organ is the best that an engineer could conceive of. Often the best that one engineer can do is, in any case, exceeded by the best that another engineer can do, especially another who lives later in the history of technology. But any engineer can recognise an object that has been designed, even poorly designed, for a purpose,, and he can usually work out what that purpose is just by looking at the structure of the object. (Dawkins 1988:21)

Formalism and *functionalism* are labels for hypotheses about the origins of the rules in the phonetic grammar. Formalism, which is normally only concerned with phonological processes, claims that such rules are motivated by a small number of grammar-internal principles that are essentially arbitrary with regard to the use of speech as a communication tool. This arbitrariness is highlighted

---

[8] See (Pierrehumbert & Beckman 1988:chapter 2) for the arguments against the idea that the contour between phrase-initial Highs and following Lows is a result of full tonal specification interacting with independent pitch range modification (i.e., declination).

by Halle's analogy between phonology and a mathematical game: the rules of the latter only exist for the sake of the game itself. Moreover, they can take any conceivable shape, as long as a limited number of basic constraints on the system as a whole (e.g., consistency) are respected. Functionalism, on the other hand, hypothesises that phonetic grammars are organised in ways that benefit speech perception, grammatical segmentation, lexical access, as well as speech production. In other words, functionalism claims that phonological and phonetic rules are designed to be communication tools.

As '-isms', formalism and functionalism represent claims about the phonetic (or more specifically the phonological) grammar as a whole. But testable formalist and functionalist hypotheses can be formulated for specific phenomena, and at least in specific cases, the controversy between the two paradigms can be resolved on empirical grounds. This section explores the types of prediction that are derived from formalist and functionalist theories, and goes on to argue for a 'diachronic' version of functionalism which holds that functional considerations enter the grammar in a stepwise fashion during language acquisition and change. One of the main advantages of this theory over 'synchronic' functionalism is that it can account for so-called *crazy rules*, as long as such rules can be decomposed into a diachronic series of small changes, each of which is functionally motivated.

1.4.1 Radical formalism

Taken to its logical conclusion, formalism predicts that the relation between phonological categories and their phonetic exponents is completely arbitrary. Foley (1977) and latterly Hale & Reiss (2000a,b) provide perhaps the closest approximations of this position. It entails that a set *a* of phonetic segments [p, t, k, f, s, x] should be equally likely to form a *phonological* natural class as a set *b* consisting of [p, ɭ̥, ð, ɰ, ɗ, ʘ]. In other words both sets are predicted to be equally probable phonetic interpretations of the phonological categories /p, t, k, f, s, x/. It follows that there should languages in which the sounds in *b* exhibit what might be regarded as normal obstruent phonology: the ability to precede sonorants in syllable onsets, to trigger place assimilation in a preceding nasal, or to form [±tense] pairs (say, with [b, d, ɡ, v, z, ɣ]) that are subject to neutralisation in word-final contexts.

The predictions of this radical formalism are falsified by the simple observation that phonological natural classes generally (although not always precisely) correspond to phonetic natural classes. Consequently, most models that would be counted in the formalist camp in the context of recent debates about the issue, are in fact hybrid frameworks, which incorporate notions such as articulatory/auditory *enhancement* (Stevens & Keyser, 1989). With the exception of Archangeli & Pulleyblank (1994), few of these models adhere to a well-defined policy concerning the range of phonological phenomena that should be regarded as phonetically *grounded*, and it often seems to be a matter of common sense or the scope of formalist machinery. Nevertheless, even the ostensibly anti-functionalist Kaye (1989) maintains that phonological rules have an ultimate purpose as aids to grammatical segmentation and lexical access.[9]

The only feasible way of rescuing radical formalism is to claim that, as mental objects, phonological grammars operate on substance-free structures, but that language acquisition filters out those grammar-phonetic interpretation pairs that are impossible to use. This position, which borrows heavily from the functionalist theory of phonological change developed by Ohala (1981, 1993), is taken by Hale & Reiss (2000a). It implies that grammars treating sets *a* and *b* above as /p, t, k, f, s, x/ are identical at the level of phonological representation, and, as far as that representation is concerned, equally likely to occur. Because language learners have problems in acquiring this system of obstruents when it is paired with the sounds in *b*, it will only ever be interpreted in terms of *a* or a phonetically similar set of sounds such as [pʰ, t̺ʰ, qʰ, ɸ, s̺, χ].

Hale & Reiss (2000a) then define the discipline *phonology* as the study of phonetically arbitrary systems that can be mentally represented, rather than as the study of phonetic systems that are selected by language learners. The advantage of this position is that it exempts their version of radical formalism from the all-too obvious objections sketched above. But as they relinquish most of the (usual) predictions about the gross phonetic shapes of spoken language, it is unclear how models constructed according to this logic can be tested. Hale & Reiss (2000a,b) may be interested in 'I-phonology' (an abstract level of mental representation), but 'E-Phonology' (observations about speech production and perception) is the only available data.[10] Conversely, any descriptive generalisation concerning an E-phonological phenomenon can only be attributed to an I-phonological mechanism with some confidence if an external (acquisition-driven) explanation can be categorically ruled out. It would therefore appear that Hale and Reiss's research program invests rather heavily in the potential (or perceived) limitations of acquisition-driven (functionalist) explanations of phonetic inventories and rules.

Worse, in theory it is possible that there are 'latent' principles of I-phonology that will never emerge in spoken language, because they are impossible to acquire and use for human speakers, whatever their phonetic exponence. In a sense therefore, the conception of phonology adopted by Hale & Reiss (2000a,b) is

---

[9] Perhaps the position of Kaye and other proponents of Government Phonology is better summed up as 'opposed to *articulatory* phonetic explanations of sound patterns'. See Harris & Lindsey (1995, 2000).

[10] Hale & Reiss's use of the term (phonological) *computation* in these two papers does not seem to refer to online language processing, and should probably be understood at the more abstract level of *computational theory* in the sense of Marr (1982).

comparable to a form of theoretical genetics investigating the space of 'possible species' as constrained by hypothetical 'syntactic' restrictions on nucleotide sequences, but without access to the chemistry that would enable it to test its claims.

1.4.2 Synchronic functionalism

Perhaps the most radical form of functionalism is represented by the 'synchronically functional' models of Boersma (1998), Kirchner (1998), Flemming (2001) and, in a slightly different way, Steriade (1997). These models imply that all rules of the phonetic grammar (both phonological and phonetic) are motivated on grounds of speech perception, ease of articulation, and other usage-based considerations. Moreover, they imply that the relative utility of a given utterance with respect to these functional considerations is computed online during speech production. For example, Steriade (1997) notes that the contrast between alveolar and retroflex consonants is less stable in word-initial and postconsonantal than in postvocalic contexts: if a language allows it in the former context it also maintains it in the latter, but the reverse does not hold. Steriade explains this contextual asymmetry in terms of the relative *perceptibility* of the contrast in question, i.e., the perceptual distance between corresponding alveolar and retroflex consonants. There is a marked difference in the F_3 and F_4 transitions for alveolar and retroflex consonants at the V-C boundary but not at the C-V boundary, and consequently it seems safe to assume that the perceptual distance between alveolars is greater after a vowel than after a consonant, where there are no F_3 and F_4 transitions. The *licensing-by-cue* model of Steriade (1997) suggests that information about the context-dependent relative perceptibility of the alveolar-retroflex contrast is encoded as such in speakers' phonetic knowledge, and forms the basis for a cascade of phonological constraints on the distribution of retroflexes.

The claim that speakers make online judgments about the perceptual and articulatory disadvantages of phonetic forms (given the background noise in their immediate environment) is more explicit in Boersma (1998), Kirchner (1998), and Flemming (2001). The last of these presents an elegant model of consonant-to-vowel coarticulation that calculates the realised F_2 (locus) targets for sequences of consonants and vowels as a function of their faithfulness to the relevant hypertargets, the effort involved in realising an F_2 transition of a given size, and the importance speakers attach to these factors at a given time. The fact that this model can compute the relative amount of effort involved in any possible F_2 transition strongly effectively entails that speakers are able to the same during online speech production.

Models that propose to model all of the phonetic grammar in these terms invariably founder on the observation that a number of well-documented phonological rules, and even some *productive* phonological rules, lack synchronic motivation in terms of perceptibility, ease of articulation, or other usage-based considerations (e.g., Bach & Harms 1969; Anderson 1981; Gussenhoven 1996). Interest in such *unnatural* or *crazy* rules and their implications for phonological and phonetic models seems to be tied to (perceived) paradigm shifts in the field, and thus seems to emerge cyclically.

As a first example of a crazy rule, consider the dialectology of velar softening in Faroese. Velar softening, a change of a velar stop [k, g] to a palatoalveolar affricate [t͡ʃ, d͡ʒ] before nonlow front vowels is in itself a fully motivated process. As Flemming (1995) points out, velar obstruents can become fronted to palatals by coarticulation with vowels involving a coronal gesture. The resulting palatals are then likely to be reanalysed as palatoalveolar affricates because releasing a dorso-palatal occlusion tends to create a relatively high amount of friction. However, Hellberg (1980) demonstrates how the morphonology of Faroese tends to retain the reflexes of velar softening, even if vowel change has removed the original conditioning environment. Consequently, it is impossible to describe this phenomenon in a synchronic functional grammar of the type proposed by Boersma (1998), Kirchner (1998) and others, unless it is encoded directly into lexical forms and treated as inert debris of language change that is somehow left untouched by usage-based mechanisms.

(2) Faroese velar softening (data from Hellberg 1980)

Orthography	Phonology	Gloss
<koma>	/koma/	come-INF.
<kemur>	/t͡ʃemur/	come-2/3.SING.PRES.
<gav>	/gav/	give-PRET.
<geva>	/t͡ʃeva/	give-INF.
<bøkur>	/bøkur/	book-NOM./ACC.PL.INDEF.
<bókin>	/bout͡ʃin/	book-NOM.SING.INDEF.
<egg>	/ɛgː/	egg-NOM./ACC.SING.INDEF.
<eggið>	/ɛd͡ʒːið/	egg-NOM./ACC.SING.DEF.

The examples in (2) represent a relatively abstract (orthography-driven) analysis of Faroese velar softening, illustrating how palatoalveolar fricatives before nonlow front vowels alternate with velar stops elsewhere (Hellberg's [č] and [ǰ] have been replaced by [t͡ʃ] and [d͡ʒ]. These examples suggest that the process is synchronically motivated along the lines described by Flemming (1995). However, the transparency of the velar stop/ palatoalveolar affricate alternation in (2) is deceptive, because relatively recent sound changes in many modern dialects of Faroese have distorted the mapping between vowel quality and the place of articulation of dorsal stops. Thus, in a northern dialect described by Hellberg (1980), <tóku> *took*-PL. is realised with a nonlow and front suffix vowel but nevertheless retains the velar stop that was motivated by the original high back suffix vowel: [touke]. In other words, in spite of the presence of the triggering environment in the surface form, the velar softening rule does not apply. Conversely, in the dialect of the island of Suðuroy, <fiskin> *fish*-ACC.SING.DEF. is realised with a back or centralised rounded suffix vowel but velar softening nevertheless applies: [fist͡ʃɵn]. Note that the same dialect realises <fiskum> as [fiskɵn].

In contrast to the northern and Suðuroy (southern) dialects of Faroese, a number of varieties retain the distinction between /i/ and /u/ in suffixes but have redistributed them. This redistribution process again obscures the relation between velar softening and surface vowel quality. For instance, velar stops are preserved before [-ir]-NOM.PL: cf. <røkur>, [røːkir] *rock ledges*, <vikur> vi[viːkir], *weeks*, <lungur> [luŋir], *lungs*. On the other hand, palatoalveolar affricates appear before [-ɵr]-2/3.PRES.SING. (historical [-ir]): <vakir> [veːt͡ʃɵr], *is awake*, <tekir> [teːt͡ʃɵr], *covers*. Note that where the present day quality of the suffixal vowel corresponds to its original value, velar softening is transparent: <sangir> [saɳd͡ʒir], *songs*, and <leggur> [lɛg̊ːɵr], *puts*.

(3) Limburg Dutch diminutive formation (data from Gussenhoven 1996)

UR	Phonetic form	Gloss
/duːm/ +/kə/	[dyːmkə]	thumb
/voːt/ +/kə/	[vøːcə]	foot
/kraːx/ +/kə/	[kreːçskə]	collar
/snɔr/ +/kə/	[snʏrkə]	moustache
/bɑl/ +/kə/	[bɛlkə]	ball

Gussenhoven (1996) provides a crazy rule from Limburg Dutch that proves even more problematic for theories of synchronic functionalism. In this group of dialects the suffixation of diminutive /kə/ causes the last stressed back vowel of a stem to front, and, if it is a low vowel, to raise (cf. the examples in 3). At one stage, this umlaut rule was a regular palatal harmony process triggered by a high front /i/ in the diminutive suffix. Although the phonetic grounding of vowel harmony is not fully understood, it seems likely that the process is rooted in vowel-to-vowel coarticulation and related (compensatory) perceptual processes (Fowler, 1981; Busá & Ohala, 1999). At a later stage, Limburg Dutch reduced the suffix vowel to /ə/. Although (centralising) vowel reduction is itself an uncontroversially natural process, in this instance it removed the trigger for the umlaut rule, rendering it phonetically opaque in synchronic terms. Nevertheless the Limburg dialects retained the process as part of their morphology, and according to Gussenhoven it is synchronically productive. It is this productivity that is especially problematic for synchronic functional models since it indicates that synchronically unmotivated patterns are not necessarily inert, but are at some level recognised and applied as rules by speakers.

1.4.3 Diachronic functionalism

The existence of crazy rules is sometimes touted as proof that phonological grammars are built around a non-functional core and are, to an extent, a mathematical game after all. However, the sorts of crazy rules that are documented in the literature merely seem to falsify synchronic versions of functionalism, but not an alternative theory, which I will label *diachronic* or *evolutionary* functionalism. This form of functionalism is central to the theories of language change pursued by Ohala (1981, 1993) and (Blevins, to appear), underpins several recent attempts to simulate language evolution (de Boer, 1999, 2001; Kirby, 1999; Briscoe, 2000; Kochetov, 2003), and is endorsed by Hale & Reiss (2000a), albeit not as part of what they consider the study of phonology to be about.

Rather than claiming that speakers are able to make online judgments about the effort involved in the production of an utterance and its precise perceptual consequences, diachronic functionalism views most phonetic behaviour as simply learned. Language learners are assumed to be fundamentally conservative in striving to copy the patterns they encounter in their speech community as faithfully as possible.[11] However, speech transmission is an inherently noisy process, both in the literal sense of ‘affected by background noise’ and because speech perception and production are not perfect, error-free, processes. The noise in the speech transmission chain is likely to introduce copying errors of various sorts. Although Ohala (1981, 1993) seems to assume that these copying errors are necessarily discrete at the level of lexical phonological contrast, given that the peripheral auditory and articulatory systems process continuously-valued representations, this assumption is unfounded. For example, on encountering a certain number of (partially) devoiced word-final [-tense] obstruents, a learner of English might conclude that voicing distinctions do not cue [±tense] in this environment. But if additional phonetic distinctions between [±tense] obstruents in terms of segmental duration, F_0/F_1 perturbation, and release characteristics are sufficiently salient, there is little ground for this learner to decide that there is no phonological contrast at all, and to include a rule of word-final laryngeal neutralisation in his/her developing phonetic grammar.

The central claim of diachronic functionalism is that various forms of *feedback* received by language learners create a form of selectional pressure that determines whether the copying errors survive in their mature grammars as innovations with a chance of being passed on to the next generation of learners. One form of feedback is supplied by the learners’ own perceptual systems and

---

[11] Evolutionary functionalism does not require that the language acquisition process be fully inductive. In fact, both Kirby (1999) and Briscoe (2000) investigate scenarios in which language and an emergent UG co-evolve.

provides information e.g., about the relative amount of effort spent in producing an utterance (proprioceptive feedback) with a given phonetic make-up. The second form of feedback consists of the responses of the speech community to forms produced by the learner, which provides a measure of the communicative utility of an utterance with a particular phonetic make-up. This second type of feedback comes in a variety of linguistic and non-linguistic forms, and includes information both about the efficacy of a form in conveying the intended message, and its social status.

The probability of survival of a given phonetic form or pattern depends on the net amount of positive feedback received by the learner. Forms that incur a low amount of positive feedback are likely to be discarded whilst in favour of alternative phonetic encodings of the same message that receive a higher amount of positive feedback. On the assumption that effective communication (construed in the broadest sense possible) is the main goal of speaking and hence that feedback from the speech community receives considerable more weight than proprioceptive feedback, this selection process creates a bias towards forms that are easy to parse by listeners and are easy to produce by speakers to the extent that this does not interfere with parsing.[12] Thus, usage-based constraints can enter the phonetic grammar without speakers being able to assess their utility for various purposes in explicit terms, and diachronic functionalism removes all undesirable 'teleology' (Ohala, 1993) from the phonetic grammar.

The idea that functional considerations enter the phonetic grammar during acquisition has a number of important ramifications. First of all, because function-driven change is cumulative (successive generations each add their own innovations), diachronic functionalism predicts the existence of crazy rules, as long as they can be decomposed into a sequence of changes that are in themselves motivated by parsing or production considerations. Judging by the literature on the topic, this is at least typically the case: in fact, authors such as Bach & Harms (1969), Anderson (1981), and Gussenhoven (1996) make a point of demonstrating how crazy rules emerge from the aggregation of phonetically motivated changes. Note that this observation contradicts radical formalism (barring the version espoused by Hale & Reiss 2000a,b), which predicts that individual changes need not be functionally motivated and hence that crazy rules do not necessarily decompose in terms of such motivated changes. The latter position implies that a pattern along the lines of the Limburg Dutch diminutive illustrated in (3) could arise without an intermediate stage in which the suffix contains a high front vowel.

Second, as hinted above, evolutionary functionalism derives the presence of language usage-based constraints in the phonetic grammar as an epiphenomenon

---

[12] In this context, the term *parsing* should be understood as the totality of sound processing operations performed by a listener to decode a message.

of the language learning process. This entails, for instance, that retroflexes are not avoided in initial and postconsonantal contexts because speakers know that they are hard to distinguish from alveolars there, but simply because (a) learners fail to perceive a contrast between alveolars and retroflexes in these contexts and reanalyse all stops as alveolar; or (b) learners 'inventing' a contrast in these contexts (e.g., by reanalysing coarticulation differences between following rhotic and non-rhotic sounds) do not get sufficient positive feedback from their speech community (i.e., because there are no advantages from a parsing point of view).

Similarly, feedback-driven selection of innovations arising out of copying errors is able to account for the instability of, or gaps corresponding to, phonetically voiced [ɡ] in [p, t, k, b, d, (ɡ)] systems (offered by Boersma 1998 as an example of true teleology in language change). It seems probable that learners trying to produce voiced [ɡ] occasionally stumble on nearby sounds in articulatory space such as voiced [ɣ, ŋ], voiceless fortis [x], or voiceless lenis [ɣ̊], all of which are somewhat easier to produce because they do not involve trying to maintain voicing behind a back constriction that allows for only limited oral cavity expansion (cf. chapter 2). All these sounds retain an important property of [ɡ], i.e., its place of articulation, and compared to e.g., [c, q, t͡ʃ, ɟ̊, ɢ̥, d̥͡ʒ̊, ʝ, ʁ, ɲ, ɴ] (ignoring the effects of flanking vowels), they are therefore relatively likely to be tolerated as substitutions by the speech community. Consequently, it seems safe to assume that they receive a relatively high amount of positive feedback, and the (correct) prediction follows that they are the most likely alternative candidates beside [ɡ] to take on its structural role in a /p, t, k, b, d, ɡ/ inventory.

Two further candidates that retain the place cues of [ɡ] whilst being easier to produce in terms of voicing are voiceless fortis [k] and voiceless lenis [ɡ̊]. Substitution of the former leads to neutralisation in production and perception of the [tense] contrast for the velar place of articulation, which may be tolerated by the speech community under certain circumstances. Substitution of [ɡ̊] on the other hand, does not lead to full neutralisation, but depending on the other cues involved in the phonetic expression of [tense], may reduce the amount of contrast with /k/, which is in turn predicted to raise the chance of misperception and neutralisation by the next generation of learners. Thus evolutionary functionalism is able to handle both cases of apparent goal-driven behaviour by speakers (pace Boersma 1998) as well as gradient sound change (see above: pace Ohala 1993).

1.4.4 The emergence of structure

From the point of view of the nonmodular phonetic grammar model described in section 1.3.2 above, a very important consequence of evolutionary functionalism is that it derives phonetic (and hence phonological) categories in continuous articulatory and perceptual space. This point is perhaps best illustrated by a brief summary of the simulations carried out by de Boer (1999, 2001).

The architecture of de Boer's model consists of a population of 20 *agents* representing human language users. Every agent is endowed with an (initially empty) inventory of paired articulatory and auditory vowel targets, a vowel synthesiser (articulation model) and a vowel recogniser (perception model). Both articulatory and auditory space are modelled in continuous terms: there is no level of discrete representations that would hardwire category formation into the model. Articulatory targets are represented in terms of height, position and rounding whilst auditory targets are represented as a set of co-ordinates in F_1-F_2 space expressed on the Bark scale. The second formant is calculated as the perceptual F_2 or $F_2{}'$ (F_2-prime), which takes on board the contribution of higher formants in the acoustic spectrum to the perceived frequency of the second resonance peak (cf. Chistovich & Lublinskaya 1979).

Simulations consist of a series of *imitation games* between pairs of agents. Each game starts with an initiator transmitting a vowel sound generated from a randomly selected articulatory target in its inventory. The receiver, or *imitator* classifies this signal in terms of the perceptually nearest vowel in its own system, synthesises the corresponding articulatory target and sends it back to the initiator. An imitation game is labelled as successful if the response signal is classified as identical to the stimulus by the initiator, and the success or failure is relayed to the imitator in terms of a 'non-verbal' feedback signal. This feedback signal and the longer term communicative effectiveness of a vowel category (defined as the ratio between the number of times a vowel is used and the number of successful uses) determine how the vowel inventory of the agents is updated after every game. Vowel targets can be shifted in articulatory and auditory space, and vowel categories can be introduced, merged, or discarded. The mapping between feedback (history) and specific update operations introduce a bias in the model towards a vowel system that is shared by all members of the population: it favours high communicative effectiveness indices for all individual vowel targets in the inventories of all individual agents and the sum of these indices is maximal if all agents share the same inventory.

Two further properties of the model developed by de Boer (1999, 2001) are crucial to cumulative effect of the imitation games on the vowel inventories of the agents. The first is the addition of noise to the vowel signals transmitted between the agents. Technically speaking, this noise consists of transforming the signals in the F_1 and F_2' domains randomly, but within fixed bounds that represent the 'noise level'. This means that vowel targets with overlapping noise ranges run the risk of being confused during imitation games. Given the communicative pressure described in the previous paragraph, noise addition therefore creates a bias towards auditory dispersion. Secondly, and essentially to keep lexical pressure on the model, a random vowel is added to the inventory of an agent with a probability of .01 per game. This pushes the model away from a shared inventory with a highly effective single vowel.

After a certain number of imitation games, the model starts to converge on a relatively steady state in which the agents have highly similar inventories, with the spacing of vowels (and consequently their number) roughly inversely proportional to the level of noise in the transmission process. Every individual agent has a finite number of vowel targets with more or less stable co-ordinates that approximate the configurations of vowel targets in the rest of the artificial speech community, and can therefore be said to have developed a set of vowel *categories*. As a collective, the agents converge on clusters of targets in articulatory and auditory space that are similar to the phonetic clusters that realise lexical contrasts in human speech production.

Elsewhere (Jansen, 2001b) I have criticised some aspects of de Boer's methodology and the details of his interpretation of the simulation results. But these criticisms by no means undermine his basic conclusion that it is possible to generate vowel categories in continuous phonetic space on the basis of a noisy speech transmission chain and selection on the basis of feedback from a speech community, the two most important ingredients of diachronic functionalism. Intuitively speaking, the logic of this approach is perhaps easiest to apply to the development of vowel categories, but in principle, it is capable of generating categories in any sort of multidimensional space without the intervention of a discretely-valued level of representation, i.e. a separate phonological module.

For example, an extended version of de Boer's model should be capable of accounting for the phonetic properties associated with the [tense] contrast. As pointed out in chapter 2, there are good grounds to believe that the multiple cues many languages associate with the lexical contrast between /p, t, c, k, q/ and /b, d, ɟ, g, ɢ/ are organised in a mutually enhancing fashion. Under a diachronic functional theory this organisation would arise without the need for an explicit categorical [±tense] feature. Speakers would simply 'discover' the observed configurations of phonetic features by trial and error during the acquisition process: the combination of voicelessness with a short segmental duration for example would incur less positive feedback than the (commonly observed) combinations of (active) phonetic voicing with a short obstruent duration and (active) devoicing with long segmental duration. The same line of reasoning can be applied to the emergence of prosodic hierarchies in the vein of Nespor & Vogel (1986) or Pierrehumbert & Beckman (1988), which can of course not be explicitly encoded in a nonmodular phonetic grammar.

1.4.5 Perceptibility

The theory of phonological change defended by Ohala (1981, 1993) revolves around the effects of perception errors during language learning. In one of two possible scenarios, a learner fails to detect a phonological contrast in the speech of the surrounding speech community and therefore neutralises it in his/her developing grammar. In the second scenario, the learner interprets gradient context-dependent variation, due to e.g., coarticulation, as a reflex of a lexical phonological contrast and grammaticalises it as such, even it is a gradient phonetic rule in the speech of older speakers. Feedback-driven selection then determines whether these types of innovative neutralisation and phonologisation survive in the adult grammar of the learners and subsequent generalisations.

Thus, Ohala's model, and diachronic functionalism more generally relies relatively heavily on the notion of *relative perceptibility* or *salience*: the assumption seems warranted that contrasts are more likely to escape detection by learners when they are relatively imperceptible, and conversely, that relatively salient forms of gradient variation are more likely to be phonologised. Consequently, this notion deserves to be made a little more precise.

First, the relative perceptibility of a phonetic (hence phonological) contrast between two sounds can be defined in terms of the likelihood that two sounds are confused with each other by listeners. Studies of perceptual confusion such as Miller & Nicely (1955) show that this likelihood is far from the same for every possible pairing of sounds. For example, the voiced lenis fricatives of English are more likely to be confused with each other and lenis stops, than the corresponding voiceless fortis fricatives. Likewise, the relative perceptibility of a given phonetic category in a given context can be defined in terms of the frequency with which it is identified correctly by listeners. Mielke (2001), for instance, demonstrates how [h] is less perceptible at the end of an utterance (i.e., is identified correctly in a lower number of instances) than before a vowel.

Second, it appears that the relative perceptibility of a given contrast or a given sound in a particular context depends on a number of factors, including the number of available cues and their interaction with (e.g., masking by) the phonetic context they appear in, and the native language of a listener. The roles of both of these factors is demonstrated by the experiments reported in Mielke (2001), which show both language-specific effects in the perceptibility of [h], and crosslinguistic effects based on the availability of specific cues. Mielke's data shows how native speakers of Turkish and Arabic, languages in which [h] and similar sounds have a relatively wide distribution, are better at perceiving this sound across phonetic contexts than native speakers of English, in which [h] only occurs before stressed vowels, and French, which lacks contrastive [h] altogether. Despite these differences in overall identification levels, the effects of phonetic context are remarkably similar across languages. Thus, for all 4 languages, the lowest proportions of correct [h] (and non-[h]) identifications occur before voiceless obstruents and utterance finally. The most likely cause of this effect is the absence in this set of environments of the voicing/F_0 onset that signals (the end of) [h] before voiced sounds. This mechanism is probably reinforced by the low salience of consonantal onset cues vis-à-vis offset cues, which has been demonstrated independently by Raphael (1981).

1.5 Conclusion: the phonetics-phonology interface revisited

Figure 1.2 depicts the model of speech production and perception described in the previous sections. The 'underlying' representations of this model do not consist of abstract phonological features, but of hyperarticulated articulatory and auditory targets represented in terms of continuously-valued features. These parametric representations have the same structure as the interface representations supplied/used by the peripheral perceptual and articulation systems. Articulatory representations can be conceived of as gestural scores in the Articulatory Phonology sense (Browman & Goldstein, 1986), *articulator windows* in the fashion of Keating (1990b), or the speech motor goals of Perkell et al. (1995). Irrespective of the choice of framework however, interface level articulatory representations specify all aspects of articulation that cannot be attributed to coarticulation, the anatomy of the vocal tract, or low-level reflexes.

Auditory forms encode the linguistic aspects of the acoustic form of speech sounds which is initially delivered by the peripheral auditory system. There is ample evidence that linguistic auditory processing imposes various forms of normalisation on the raw input signal and integrates individual acoustic cues into more abstract objects: as reviewed in chapter 2 for example, voicing, F_0, and F_1 cues to [tense] may all be integrated into a single 'low frequency' feature. On the other hand, since native speakers of different languages (i.e., voicing and aspirating languages) respond differently to the presence vs. absence of the voicing component of this higher level perceptual feature, it must be assumed that some or more of the individual acoustic cues are differentiated at some stages of linguistic auditory processing.[13]

In the production of an utterance, articulatory hyperforms are filtered through a set of categorical and gradient rules. The former change (clusters of) hypertarget values in discrete steps, or 'remove' targets altogether, that is, phonetically underspecify sounds for one or more phonetic features. The latter set of rules acts in a continuous rather than discrete fashion, but since they operate on the same parametric representations, gradient rules may occasionally have the same effects as discrete, phonological rules. These rule blocks can be

---

[13] For expository reasons I have omitted the role of other sensory modalities, notably vision, in speech perception. Nothing crucial hinges on this. For a detailed discussion of the role of visual information in speech perception, and its integration with auditory information, see Massaro (1998).

interpreted in procedural or declarative terms, and (under the latter interpretation) may be regarded as generalisations over ‘clouds’ of stored exemplars (with hyperforms assuming some special status) or as devices used to construct parts of linguistic phonetic forms on the fly during speech production. Which of these interpretations is the most suitable for which (sub)sets of rules depends on data this study is not specifically concerned with (see e.g., Levelt 1989).

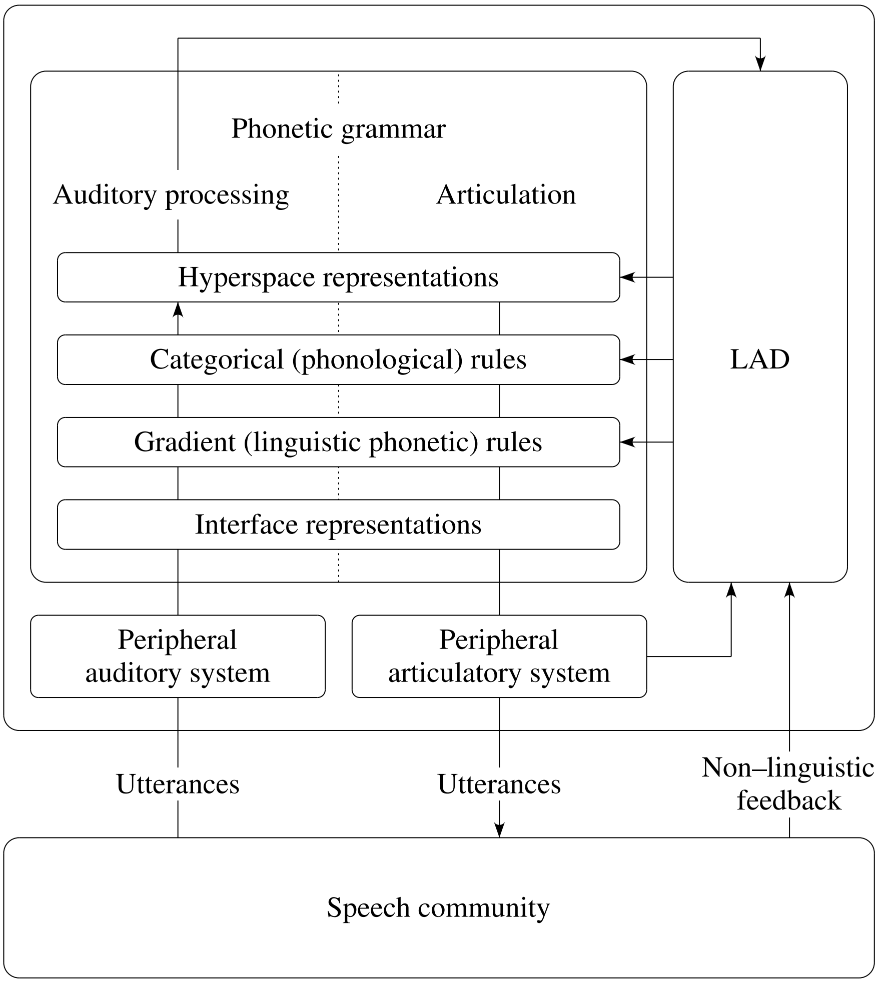

Figure 1.2: The model of speech production and perception adopted in this study.

The peripheral articulatory system responds to the instructions provided by the phonetic grammar by producing utterances, often to some sort of human audience. The label *speech community* in figure 1.2 generalises over all possible forms of audience that are capable of providing some sort of feedback to the producer of the utterance. Any form of spoken feedback is processed by the original speaker's peripheral auditory system and delivered to the linguistic system, which maps it onto a grammatical form and ultimately some sort of meaning. No one would (still) claim that this mapping proceeds in a strict bottom-up fashion, reconstructing the hypothesised stages in the production process in a step-by-step fashion, and there is a reasonable amount of evidence to suppose that knowledge of phonetic and phonological rules aids this process. For instance, several researchers have found that reflexes of coarticulation or phonological assimilation do little or nothing to impede lexical access, whilst some have even suggested that the presence of context effects improves the sound-to-meaning mapping (Elman & McClelland, 1986; Gaskell & Marslen-Wilson, 1996, 1998; Quené & Krull, 1999). Similarly, Aylett (2000) reports psycholinguistic data which indicates that listeners benefit from the 'hypoarticulation contour' imposed on utterances by prosodic strengthening and weakening at constituent edges. It is for these reasons that the rule blocks straddle the auditory-articulatory divide in figure 1.2.

As argued in section 1.4.3, phonetic and phonological rules are not constructed on the basis of grammar internal formal templates or functional principles such as effort minimisation, but on the basis of learning, error, and feedback. Using traditional terminology, I have labelled the module responsible for (re)structuring the phonetic grammar on the basis of incoming information *Language Acquisition Device* (LAD). The use of this term highlights the role of the acquisition process in generating linguistic change and the incorporation of functional mechanisms, but does not imply that (re)structuring of the grammar ceases completely after the offset of the famous 'critical period' for language acquisition. The LAD receives data from a variety of sources, some of which are indicated in figure 1.2.

Because errors in perception and production, selectively incorporated into the phonetic grammar by the LAD, drive the form of phonological and phonetic rules, the formal statement of those rules becomes arbitrary. Phonological rules might be stated using the formalism adopted by Chomsky & Halle (1968) or in autosegmental terms, with (distinctive) features serving as notational shorthands for clusters of phonetic features, but as long as both frameworks are able to capture the relevant generalisations there are no empirical grounds for deciding between them. Phrased in more general terms, the framework adopted in this dissertation renders all empirical arguments for or against particular formalisms void, whether they concern, e.g., the advantages of autosegmental feature lattices over feature bundles, monovalent over bivalent feature representation, or declarative over procedural grammars.

Despite its differences with models of the phonology and the phonology interface typically encountered in the theoretical phonology literature (at least until recently), the model illustrated in figure 1.2 reconstructs a number of properties found in more traditional frameworks. By way of conclusion to this section and this chapter it is perhaps useful to point out some of the more important parallels.

First and foremost, as pointed out by Johnson et al. (1993) there in an important parallel between the hyperform-interface mapping and non-monotonic lexical-to-surface mappings in traditional phonological grammars: both lead to the loss or distortion of (lexical) information. For example, many generative models, including most current versions of Optimality Theory, in principle allow a lexical contrast between /i, e, a, o, u/ to be neutralised to phonological and phonetic [ə] on the surface by removing and/or replacing the relevant features. This mapping involves a loss of information in the sense that it is impossible to reconstruct the underlying vowel contrast on the basis of the forms exhibiting a reduction schwa. Similarly, phonetic vowel reduction can reduce a [i, e, a, o, u] distinction in hyperspace to, say, [ɘ, ɵ, ɐ], and whilst this process is incompletely neutralising, it does not allow for the original phonetic values to be reconstructed. For example, [ɘ] might correspond to hyperspace [i, e] or even [ɪ], but without additional (e.g., paradigmatic) information it is impossible to determine the underlying phonetic category.

Second, whilst the framework adopted here abolishes phonology as a separate, representationally distinct level of representation, it does not dispense with the notion of phonological contrast as a discontinuity in phonetic space. Although some, such as Port (1996) have implied that rules operating in phonetically discrete fashions do not exist, there is clear experimental evidence to the contrary (see Zsiga 1997 and chapter 2 below). Therefore, the diachronic functional model in figure 1.2 retains a set of phonological rules as opposed to a set of gradient phonetic rules, even if both types of rule operate on the same parametric phonetic representations. Which of the rules described in descriptive grammars or the theoretical literature as categorical indeed belong to this class, is simply an empirical matter.

Third, it is precisely the absence of a phonology-phonetics interface in the sense of e.g., Keating (1990a) that renders the framework in figure 1.2 similar in some ways to the monostratal models of Pierrehumbert & Beckman (1988) and Harris & Lindsey (1995, 2000). For example, the latter state that individual phonological elements, and consequently the lexical, intermediate and ‘surface’ forms composed of them are always phonetically interpretable. Phonological rules manipulate elements but do not transform them into (approximations of) interface representations. Thus, occurring on its own, the element *A* , is interpreted as a vowel with a low first and high second resonance, i.e., an unrounded low vowel. This view contradicts the position of Chomsky & Halle (1968), restated more recently by Bromberger & Halle (1989), which holds that the purpose of phonological and phonetic rules is to progressively convert abstract underlying forms into structures that are understood at the interface levels. The model adopted here sides with Harris & Lindsey (1995, 2000) in the sense that hyperforms can be understood by the peripheral systems, in spite of the fact that the auditory/articulatory values encoded in hyperforms are not typical of interface forms encountered in speech production.

Fourth and finally, the LAD as conceived here corresponds to (certain versions of) *H-EVAL* in OT, albeit in a fairly abstract sense. The LAD evaluates forms produced by the speaker with respect to several forms of feedback (‘constraints’), preferring forms that receive a certain amount of positive feedback (‘a certain number of violation marks’) over those that incur less positive feedback (‘more violation marks’). The crucial difference is that the LAD processes feedback to forms that have been produced at a particular place and in the presence of a particular audience whereas H-EVAL is normally viewed as a device that determines which forms can (and will) be produced in the first place. Nevertheless, the basic idea that phonetic grammars are shaped by competing factors selecting optimal candidates from an array of alternatives (generated by *GEN* or errors in production and perception) is central to both standard OT models and the framework adopted here.

Chapter 2

The phonetics of the fortis-lenis contrast

In chapter 1 I defined the fortis-lenis contrast as a lexical contrast between obstruents that is realised in part in terms of phonetic voicing distinctions. This definition is the same as definition of the phonological feature [voice] employed by Keating (1984), which is in turn based on the definition of Lieberman (1970, 1977). From the perspective of this study, this definition serves the sole purpose of descriptive taxonomy: it offers a coherent framework to discuss the types of two term 'laryngeal' contrast that are typical of the obstruent inventories of the languages spoken in Europe (and are widespread elsewhere, too), and the roles of laryngeal neutralisation and voicing assimilation in those obstruent inventories. The use of the terms fortis/tense, lenis/lax, [±tense] is emphatically not intended to assert any sort of deep phonetic unity within the two sets of obstruents they defined in terms of 'articulatory force' or similar notions. My main objective in choosing [±tense] over [±voice] is simple to maintain a distinction between a phonological distinction that is signalled by a complex of cues, and phonetic voicing, which one of those cues.

This chapter provides a survey of the phonetic characteristics of tense and lax obstruents on the basis of a literature review. As a vast amount of work has been done on the phonetic reflexes of laryngeal contrast in obstruents, this survey does not pretend to be in any way comprehensive. It does represent an attempt at a fair summary of the current state of knowledge in the fields that are of most relevance to the subsequent chapters.

Section 2.1 introduces the aerodynamics of voice production, which plays an important role in the definition of the notions of active and passive (de)voicing. Next, 2.2 is devoted to the phonetic manifestation of [tense] in plain oral stops. This section describes the distinction between aspirating and voicing languages in greater detail and reviews the other components of the complex of (possibly mutually enhancing cues) that is associated with the fortis-lenis contrast. Section 2.3 provides a similar, if much shorter description of the phonetics of [tense] in fricatives and lexical affricates. Section 2.4 finally, provides a brief summary and conclusions.

2.1 The production of voicing in obstruents

Voicing, the acoustic result of vocal fold vibration, has an important extralinguistic function in speech as (the predominant) sound source or 'carrier signal'. For a considerable number of sounds in the inventories of the world's languages this is the only function of voicing. For example, the fact that in English voiced (as opposed to whispered) speech [n] is usually produced with voicing has in itself no linguistic significance: it merely acts as a carrier of the spectral modulations that identify [n]. The spectral signature of [n] may be better audible with a voiced than with a whispered source, but this does not mean that voicing is in any way part of the phonetic target for the alveolar nasal or similar sounds. In other sounds, the presence or absence, fundamental frequency, and/or quality of voicing does fulfill

Vocal fold vibration is produced by pushing air from the lungs through a closed but not tightly compressed glottis, which requires the air pressure to be lower above the glottis than below it. The minimum transglottal pressure difference that is sufficient to keep the glottal cycle in motion has been estimated at 200 Pa (2000 dyne/cm^2). Due to the inertia of the vocal folds the pressure differential needed to *initiate* voicing is about twice as large (Baer, 1975). These physical preconditions for vocal fold vibration form the basis under the notions of *passive*, or *spontaneous*, vs. *active* voicing and devoicing. Sounds or parts of sounds are said to be passively voiced if a closed equilibrium position of the vocal folds and normal subglottal pressure (according to Stevens 1998, 8000 dyne/cm^2/ 800 Pa is typical) are sufficient to initiate or maintain the physical conditions for vocal fold vibration. Sonorants are typical examples of passively voiced sounds: because their supralaryngeal articulations allow air to escape freely from the supraglottal vocal tract (either through the oral or nasal tract or both) the supraglottal pressure during these sounds remains approximately equal to atmospheric pressure.

Sounds are said to be passively *devoiced* if a closed equilibrium position of the vocal folds and normal subglottal pressure are *in*sufficient to initiate or maintain the physical conditions for vocal fold vibration. Passive devoicing is typical of the closure phase of plosives, during which the supraglottal vocal tract is sealed off and continued airflow from the lungs leads to an increase in supraglottal air pressure. An increase in supraglottal pressure results in a reduced transglottal pressure differential, which slips below the critical threshold for maintained voicing if oral closure lasts long enough, or fails to reach the critical level for vocal fold vibration to commence.

Figure 2.1 provides an example of the passive effects of supraglottal articulation on vocal fold vibration. It displays the broad band spectrogram (left) and electroglottograph (*EGG*) trace (right) of an apical trill at the beginning of an utterance of Dutch /rɔbən/ *seals*, as produced by the author. The section marked /r/ on the spectrogram shows 3 periods of decreased energy across frequency bands, representing 3 apical taps, each lasting approximately 15-20 ms. The EGG trace shows a slight decrease or terrace in the globally rising amplitude and fundamental frequency contours for each of these taps, with local minima lagging slight behind minima in the overall acoustic energy. Since Dutch /r/ is not contrastively voiced, the conclusion must be that this ripple in the amplitude and F_0 contours of the EGG trace is due to the aerodynamic coupling of the oral tract and the vocal folds. Each brief closure of the oral tract causes a slight increase in intraoral pressure and thereby a decrease in the transglottal pressure drop, which causes a decrease in the amplitude and rate of vocal fold vibration. Because the closure intervals are relatively short, the periodic rise of intraoral pressure is relatively limited, and therefore insufficient to prevent voicing at any point during the trill.[1]

Active voicing refers to a situation in which passive devoicing is overcome or postponed by adding a number of articulatory gestures to the basic mechanisms of expiration and glottal closure. In plosives, such gestures typically aim at enlarging the oral tract volume, which slows down the build up of supraglottal pressure during oral closure. Likewise, *active devoicing*, although perhaps a less common term, may be used to refer to situations in which sounds which would be passively voiced by imposing their supralaryngeal articulatory settings on a voiced source configuration, are devoiced by means of articulatory

---

[1] EGG traces represent the strength of an electrical signal led across a speaker's glottis by means of two electrodes placed on either side of the larynx. When the vocal folds are open, the tissue paths connecting the two electrodes are relatively long, which results in a relatively high impedance (i.e., because the specific resistance of the tissue is nonzero) and consequently a relatively weak signal. When the vocal folds are closed, on the other hand, the electrodes are connected by the shortest possible tissue path (a straight line), which results in a relatively low impedance and a relatively strong signal. Thus, the strength of the signal is an indicator of the relative size of the glottis.

The English EGG/audio examples used in this chapter were produced by a female speaker of a southern British variety, whilst the Dutch examples were produced by the author. Recordings were made in single sessions in a sound-proofed room. The audio signal was recorded straight onto computer disc as one channel of a stereo file with a sample rate of 44.1 kHz, using a Brüel and Kjær condenser microphone (Type 4165) and measuring amplifier (Type 2609). The equipment used to record the EGG signal was built by Laryngograph Ltd, London, UK (UK patent 1533112). Its output signal was recorded as the second channel of the stereo file used to record the audio signal. Afterwards, both the audio and EGG signal were low-pass filtered and resampled at 22.5 kHz.

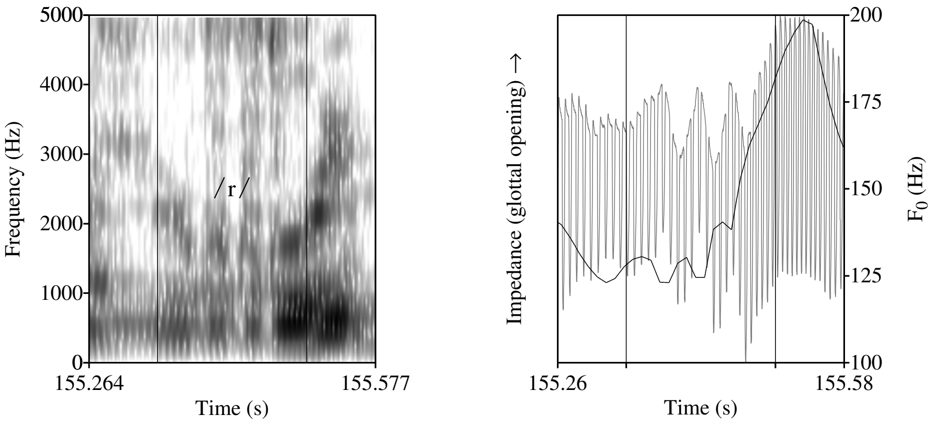

Figure 2.1: Broad band spectrogram (left) and EGG trace (right) with superimposed F_0 contour (in black) of an apical trill in an utterance of Dutch /rɔbən/.

readjustments.

Voicing and devoicing in plosives A number of aerodynamic simulation studies have investigated the limits on passive voicing in plosives in different phonetic environments (Ohala 1983; Westbury & Keating 1986; Hayes & Stivers 1996; Stevens 1998; see also Boersma 1998; Kirchner 1998). These studies indicate that postvocalic (but utterance-medial) plosives are subject to passive devoicing 25-100 ms after the onset of oral occlusion, depending on the subglottal pressure, the tenseness of the tissue lining the vocal tract, and the place of the constriction. Note that simulations by Boersma (1998) and Stevens (1998) show that in addition to lowering the transglottal pressure drop, increasing intraoral pressure exerts a lateral force on the vocal folds that may push them apart and thereby help to terminate vocal fold vibration.

The tenseness of the vocal tract walls is a factor in the timing of passive devoicing because it determines the amount of passive expansion the vocal tract behind the constriction can undergo, and thereby the speed with which the rising intraoral pressure (and hence the transglottal pressure differential) reaches the critical level. The place of constriction similarly influences passive devoicing: Westbury & Keating (1986) estimate that the duration of the voicing tail (i.e., voicing continued from a preceding sound) in velar stops may be as much as 30% shorter as that in the corresponding bilabials, due to the fact that the size of the cavity behind the constriction is smaller in the former than in the latter and lined with (proportionally) less tissue that can expand in response to rising pressure. All else being equal, this entails a more rapid increase of intraoral pressure behind posterior than behind more anterior constrictions.

A special utterance-medial environment with respect to passive voicing is the *postnasal* context. Using the same aerodynamic model as Westbury & Keating (1986), Hayes & Stivers (1996) demonstrate that a preceding nasal can increase the amount of passive voicing in a stop. Coarticulation between a nasal and a following stop will lead to some degree of nasal leakage during the initial portion of the stop, depending on how rapidly the velopharyngeal port is closed at the end of the nasal. Under most circumstances, the amount of coarticulatory velopharyngeal leakage is insufficient to produce audible nasalization, but does act as a temporary pressure valve. It prevents the intraoral pressure from rising or at least slows down the increase in pressure until the velopharyngeal port is fully closed. Hayes & Stivers (1996) suggest that the velic raising gesture that spills over into the stop during nasal-to-stop coarticulation further facilitates voicing by expanding the volume behind the stop constriction.

Westbury & Keating (1986) argue that initially and finally in breath groups (and therefore utterance initially and finally), plosives are likely to have less passive voicing due to lower subglottal pressure (and hence transglottal pressure). First consider the initial context. To initiate voicing in an utterance-initial stop, the transglottal pressure drop has to exceed the higher level of 400 Pa, and then stay above 200 Pa to allow sustained voicing during the occlusion phase. According to Westbury & Keating's model, the inertia of the lungs slows the build up of subglottal pressure, which means that the transglottal pressure is does not exceed the initiation level until well after oral closure, and rapidly drops below the lower threshold afterwards, even if the glottis is (initially) closed to prevent an instantaneous rise in intraoral pressure through leakage. In breath group-final position, plosives are predicted to have a smaller amount of spontaneous voicing than medially with the same preceding context, because in the former environment the expiratory force of the lungs will be counterbalanced by the onset of an inspiration gesture. The result of this form of 'respiratory coarticulation' is a decrease in subglottal pressure that causes the transglottal pressure difference to fall below the critical threshold as early as 30 ms after oral closure.

Ladefoged (1973) and Stevens (1998) list a number of articulatory measures that are in principle able to counteract passive voicing and devoicing in plosives. Most of these gestures have been observed in the production of plosives with contrastive (de)voicing (cf. Perkell 1969; Svirsky et al. 1997), or have been inferred from their acoustics. Passive devoicing can be slowed down by allowing the vocal tract behind the constriction to expand passively, or to expand it actively, by (e.g.) lowering the larynx, raising the soft palate and advancing the tongue root, and thus to slow down the increase of intraoral pressure. Other strategies that contribute to voicing in plosives are lowering the tension of the vocal folds (which lowers the pressure thresholds for vibration) and nasalization, which acts as a pressure vent. The latter is not often observed with contrastively voiced plosives that are otherwise described as plain, possibly because of its distinct auditory effect (though cf. Jones 2001).

Passive voicing of plosives (e.g., postvocalically or postnasally) on the other hand, can be counteracted by tensing the vocal tract walls to reduce the amount of passive expansion, and by actively decreasing the size of the cavity behind the oral constriction, for example by raising the larynx. In addition, tensing and/or medial compression (glottalisation), both of which raise the critical pressure thresholds for vocal fold vibration, or glottal abduction, which removes one of the basic preconditions for voicing, can be used as devoicing strategies.

Voicing and devoicing in (strident) fricatives The high-intensity turbulence noise that is typical of (strident) fricatives requires a relatively high volume velocity airflow through a narrow constriction in the oral tract. This in turn requires an at least equally high airflow across the glottis. Thus, on aerodynamical grounds the ideal glottal configuration for the production of a fricative is widely adducted, and articulation data show that contrastively voiceless fricatives are indeed produced with a glottal abduction gesture similar in size to that of voiceless aspirated plosives (Löfqvist & Yoshioka, 1980; Löfqvist, 1981; Yoshioka et al., 1981, 1982). Stevens et al. (1992) and Stevens (1998) show that if the cross-sectional area of the glottal opening is larger than the opening at the oral constriction, transglottal airflow exceeds the airflow through the oral constriction, which causes an increase in intraoral pressure (to a level equal to the subglottal pressure if the oral constriction is held long enough) with a concomitant decrease in the transglottal pressure differential. Since it removes both of the basic conditions for vocal fold vibration, any fricative that is produced with a substantial amount of glottal abduction therefore becomes devoiced during most if not all of the frication interval. Judging by the models of Stevens et al. (1992) and Stevens (1998) it would appear that additional active devoicing strategies are hardly necessary to produce a voiceless fricative.

To maintain vocal fold vibration throughout a fricative, an uneasy balance has to be struck between the aerodynamic requirements for voicing and those for the production of a turbulent noise source. The vocal folds are partially adducted to allow both vibration and a relatively high transglottal airflow. To keep the transglottal pressure drop from falling below the critical thresholds for voicing, the cross-sectional areas of the glottal and oral valves can be set so that their sizes are exactly equal, or the size of the cavity behind the constriction can be expanded passively and/or actively using the articulatory measures mentioned above (Stevens et al., 1992; Stevens, 1998). Perhaps because the former option involves very precise control of the glottal opening (a small increase in glottal opening causes a relatively large increase in intraoral pressure), Stevens (1998) appears to assume that speakers normally opt for the latter strategy.

2.2 Plosives

2.2.1 Voicing targets for (utterance-)initial fortis and lenis plosives

The classic typology of contrastive voicing in word-initial stops is due to Lisker & Abramson (1964). In a survey of 23 languages they found that such stops fall into three broad categories that show little crosslinguistic variation: (1) a prevoiced or negative VOT category in which voicing starts well before the release of the plosive, (2) a zero or short lag VOT category in which voicing starts at or shortly after the stop release, and (3) a long lag VOT category in which voicing starts more than around 35 ms after the release of the stop. The cut-off point between categories 2 and 3 is conventionally put at 35 ms (Keating, 1984). If a language has a single series of stops, these belong almost always to the second category: 49 out of 50 languages with a single series of oral stops sampled by Maddieson (1984) have a short lag VOT. Aleut (cf. Cho & Ladefoged 1999), which has a single series of long lag VOT stops is one of the few well-documented exceptions. Other languages, such as Thai, employ all three VOT categories in the signalling of lexical contrast. The languages that represent the focus of this study select two neighbouring VOT categories.

Voicing languages contrast prevoiced lenis plosives with zero to short lag VOT plosives utterance initially and after another obstruent. This type of language dominates in eastern and southern Europe, comprising virtually all varieties of Romance and Slavonic as well as the Baltic languages and Hungarian. Prevoicing varieties of Germanic are Afrikaans, (southern and western dialects of) Dutch, (West) Frisian, Yiddish, Scottish English and Rhineland German. The second type of fortis-lenis language contrasts zero to short lag VOT lenis stops with their long lag VOT counterparts. The articulatory mechanism involved in the production of the latter, at least in the languages under consideration here, is *aspiration*, i.e., a large glottal abduction that peaks around the oral release of a stop. Apart from along lag VOT the result of this abduction gesture is an interval of [h]-like noise (generated at the glottis) following the release of a stop. Languages belonging to this second type are referred to in this study as *aspirating languages*. Danish, Faroese, Icelandic, Norwegian, Swedish, and standard varieties of English and German are all aspirating languages, and outside Germanic this type seems to be common in e.g., the Turkic group.[2]

---

[2] See Wissing (1991) on Afrikaans; e.g., Slis & Cohen (1969a), Cohen et al. (1972) On Dutch; Tiersma (1985) on (West) Frisian; Birnbaum (1979) and Katz (1987) on Yiddish; Wells (1982a) and Kohler (1979) on Scottish English; Kohler (1979) on Rhineland German; e.g., Fischer-Jørgensen (1968) and Hutters (1985) on Danish; Thráinsson (1978), Kress (1982) on Icelandic; Vanvik (1972) and Kristoffersen (2000) on Norwegian; Moulton (1962), Jessen (1998) and a host

Authors pursuing a unified phonetic conception of the fortis-lenis distinction (or some equivalent), sometimes play down the (perceptual) importance of voicing distinctions (e.g., Keating 1984; Kohler 1984; Kingston & Diehl 1994, 1995) as secondary to other cues to [±tense]. The behaviour of such cues clearly differentiates zero to short lag lenis stops from zero to short lag fortis stops and groups them with their respective prevoiced lenis and aspirated fortis counterparts. It has also been suggested that the VOT continuum does not constitute a unitary perceptual dimension but that the identification of long lag stops depends at least as much on (the nature of) the following aspiration noise as on the timing of voicing onset itself (Boersma, 1998; Jessen, 1998). That the presence of aspiration noise can play an independent and central role in cueing lexical distinctions among stops is indicated by perception data from languages that cross-classify closure voicing and postaspiration (with ‘breathy voiced’ [b^{ɦ}, d^{ɦ}] etc. as a fourth series: Schiefer 1992)

Furthermore, it is well-known that there are differences in voicing between the lenis stops of aspirating languages and the fortis stops of voicing languages in postsonorant contexts (cf. Keating 1984 and below). In the light of all this evidence that Lisker and Abramson’s three term taxonomy does not yield an exhaustive phonetic characterisation of laryngeal contrast even in systems that do employ VOT it is not surprising that its very basis has been called into question. For example, Raphael et al. (1995) propose that Lisker & Abramsons short lag class should be subdivided into two categories, one for the lenis stops of aspirating languages, and one for the fortis stops of voicing languages, with a slightly longer VOT. Cho & Ladefoged (1999) (see also Cho & Ladefoged 1997; Ladefoged & Cho 2000) even identify 4 degrees of positive VOT.

Nevertheless, Lisker & Abramson’s typology remains a useful descriptive tool, especially inasmuch as it highlights the fact that in terms of voicing, two kinds of fortis (i.e., long lag and short lag VOT) and two kinds of lenis stops (short lag and negative VOT) are found in word-initial contexts. It seems more than likely that two terms of this four way typology, viz. prevoiced lenis and long lag VOT fortis plosives, are subject to active voicing and active devoicing respectively. As pointed out in 2.1 above, utterance-initial and post-obstruent stops are subject to near-complete passive devoicing on aerodynamic grounds and therefore have to be enhanced by a number of cavity-expanding and other measures if they are to be produced with a substantial amount of closure voicing. Similarly, active devoicing is required to produce a voiceless interval > 35 ms

---

of references in the latter on German; e.g., Wells (1982a,b), Docherty (1992), and Gimson (1994) on English (dialects); and the descriptions König & van der Auwera (1994), although the contribution on Dutch erroneously describes the lenis plosives of this language as “less voiced” than their English counterparts. Maddieson (1984) appears to assign German to the voicing group, but this is incorrect for most varieties. For descriptions and references regarding the obstruent systems of the Turkic languages, cf. Johanson & Csató (1998).

after the release of a plosive. Conventional auditory descriptions as well as instrumental studies confirm that one of the main measures involved is vocal fold abduction or aspiration (Löfqvist & Yoshioka, 1984; Löfqvist, 1981; Yoshioka et al., 1981). The peak (i.e., maximal extent) of the glottal abduction gesture in aspirated fortis stops is typically timed to coincide with the oral release.

Articulatory data by gathered by Yoshioka et al. (1982) indicate that Dutch unaspirated fortis stops are also produced with what looks like a glottal abduction gesture, but this gesture is smaller and peaks during oral closure, and therefore does not result in a long lag VOT or aspiration noise. Whether the glottal abduction observed by Yoshioka et al. (1982) reflect active devoicing or vocal fold abduction as a (passive) result of increased intraoral pressure is not entirely clear, but the difference in voicing between fortis and lenis short lag stops that can be observed in postsonorant contexts, suggests the former. For example, if no pause intervenes, the /b/ in English /ɪn/ + /bɛd/, *in bed*, tends to have audibly more voicing than the /p/ of Dutch /ən/ + /pɛt/ *a cap (i.e., head wear)*.

The greater amount of voicing in the English labial stop might of course be attributed to a (weak) prevoicing target. However, there is considerable speech production evidence to support the idea that any prevoicing observed in lenis plosives in English and other aspirating languages reflects passive voicing rather than a (weakly) voiced phonetic target. Flege (1982) reports that around 75% of the utterance-initial /b/s produced by his 10 test subjects were produced with vocal fold adduction well before the oral release, which means that one of the two basic preconditions for voicing was satisfied. However, prevoicing occurred in only 117 out of 200 tokens, which is a considerably lower frequency than what is typically found for voicing languages such as Polish (Keating, 1984), French, or the Thai series of plosives that is usually regarded as prevoiced (e.g., Kessinger & Blumstein 1997). Flege also reports considerable intraspeaker variation in the production of prevoiced and short lag lenis stops: 3 speakers exclusively produced short lag stops, 4 speakers produced only prevoiced stops and the remaining 3 vacillated between the two types of stop. Given that labial stops are more prone to spontaneous voicing than stops with a more posterior constriction, it is therefore difficult to see this result as an indication that English lenis stops have a (weakly) prevoiced target.[3]

In addition, utterance-initial and post-obstruent English lenis stops tend to have a small positive mean VOT across speaking rates, instead of increased prevoicing at slower (and presumably more hyperarticulated) rates, which might be interpreted as evidence for a weak prevoicing target that is only realised in hyperspeech. (Miller et al., 1986; Kessinger & Blumstein, 1997; Magloire &

---

[3] Unfortunately, Flege (1982) does not provide data on the VOT of /p/ when produced by the same speakers. This could have clarified whether the group producing exclusively prevoiced /b/ were using a voicing system, contrasting [b] with unaspirated [p].

Green, 1999). In this respect English short lag stops behave virtually identically to the short lag fortis stops of French, Spanish, Thai, and Dutch[4] In contrast, lenis stops in French, Spanish, and Dutch, as well as the prevoiced members of the Thai three term distinction have a negative VOT across speaking rates and do show an increased VOT at slower rates (Kessinger & Blumstein, 1997; Magloire & Green, 1999).

Finally, the contribution of the three term VOT distinction identified by Lisker & Abramson (1964) to the perception of fortis and lenis plosives has been amply documented, and provides a strong argument for distinguishing two types of lenis and two types of fortis stop. Even if prevoicing in initial lenis plosives is perceptually integrated with other low frequency cues such as F_0 and F_1 perturbations (Kingston & Diehl 1995: see further below) the effect of voicing is strong enough to make 'language-specific' categorical perception of the VOT continuum the stock and trade introductions to speech perception (cf. Clark & Yallop 1995 section 8.5). For instance, Lisker & Abramson (1970) and Abramson & Lisker (1972) report on identification experiments with synthesised stops which show that native speakers of Spanish and American English place the category boundaries between fortis and lenis stops at different places along the VOT continuum. The Spanish subjects put the category boundary (defined as the 50% crossover points of the identification curves) between /d/ and /t/ at a VOT of 22 ms, whereas the English speakers placed this boundary at 35 ms.

A similar result is obtained by Slis & Cohen (1969b), who find that speakers of (prevoicing) Dutch always identify a stop as fortis if the voice bar is removed from a resynthesised lenis stop, even if other cues are left intact. Öhman (1962) (cited by Slis and Cohen) on the other hand reports that when similar stimuli are presented to (aspirating) Swedish listeners, this does not result in significantly more fortis responses. Even though the latter results are based on intervocalic rather than initial plosives, they still illustrate a difference in the perceptual relevance of stop closure voicing: if voicing played no role in the identification of fortis and lenis stops speakers of voicing and aspirating languages are expected to behave identically.

2.2.2 Positional variation in stop voicing: word-medial and word-final contexts

Within the framework introduced in the previous chapter, there are three possible sources for positional variation in stop voicing. The first is passive (de)voicing as a result of passive or active (de)voicing in a neighbouring sound. The potential role of passive voicing was already hinted at in the discussion of short lag stops above. The second source is lenition in prosodically weak environments.

---

[4] For data on the effects of speaking rate on VOT in Dutch, see chapter 7.

Lenition reduces the magnitude of active (de)voicing articulations associated with stops, and might thereby reinforce the effects of passive (de)voicing. Since word-initial, and especially utterance-initial, contexts are relatively invulnerable to lenition (cf. chapter 1) I ignored its potential effects on word-initial voicing distinctions in the previous section.[5]

Passive (de)voicing, either as a simple byproduct of (a choice of carrier sound and) vocal tract aerodynamics or reinforced by lenition would be expected to have an influence on the voicing of stops even if they have the same targets across contexts. Imagine a language with actively (pre)voiced lenis plosives [b, d, ɟ, ɡ] word initially. Even if the final lenis plosives of this language are assigned a set of active voicing gestures that is identical to that of initial stops and identically timed with respect to the points of oral closure and release, attenuation of these gestures by lenition and the ‘respiratory coarticulation’ hypothesised by Westbury & Keating (1986), might still result in a partially devoiced realisation utterance finally (i.e., by pushing the transglottal pressure difference below the critical threshold before oral release).

At least in theory however, it is also possible that non-initial stops are assigned different voicing targets than their initial counterparts. An example of positional variation that appears not to result (synchronically) from vocal tract mechanics or lenition is the realisation of the Dutch rhotic liquid. In a number of dialects, this sound is produced with different (active and passive) articulator in different environments, but such differences are never lexically distinctive. Prevocalically (within the same word) an apical alveolar tap [ɾ] (less often a full trill [r]) or an uvular trill [ʀ] are common, whilst elsewhere dialects select from a range of sound including the aforementioned rhotics and [ɹ], [ʁ̞]. Thus, some western speakers use a uvular sound prevocalically but an alveolar approximant elsewhere. As the articulatory reduction of a uvular trill does not yield an alveolar approximant, it seems safe to assume that these two sounds represent different phonetic targets, in spite of the fact that they ‘represent the same phoneme /r/’.

It seems perfectly conceivable that the phonetic interpretation of the fortis-lenis contrast exhibits similar positional asymmetries in the assignment of targets for voicing and other cues. Unfortunately, voicing distinctions in (word-)medial and in particular word-final stops have received far less attention in the (experimental) literature than initial VOT contrast. The following paragraphs contain a brief review of the information that is available.

---

[5] Note that lenition is sometimes defined in terms of increased sonority and (hence) voicing (see Lavoie 2001 for a recent attempt). Under the present definition (cf.1.3.3), lenition only leads to (increased) obstruent voicing if the aerodynamic conditions for passive voicing are met, e.g., post-nasally or intervocalically.

Medial prevocalic stops A number of studies provide quantitative acoustic data on the (de)voicing of word-medial stops occurring between sonorants (including vowels) and vowels (e.g., Slis & Cohen 1969a; Keating 1984; Hutters 1985; Lisker 1986; Docherty 1992; Jessen 1998; Helgason 1998 et seq.; Lavoie 2001). This work indicates that in this type of environment the voicing targets for fortis and lenis stops are highly similar to the targets found word initially: voicing languages distinguish actively voiced stops from plain voiceless (actively devoiced) fortis ones whilst in aspirating languages, passively voiced lenis stops often contrast with voiceless aspirated (long lag VOT) fortis stops. A notable exception to this pattern is that (geminate) fortis stops in Icelandic, Faroese, and (dialectally) also in Swedish and Norwegian tend to be *pre*aspirated instead of, or as well as, postaspirated (Helgason 2001: see further below). In the onset of stressed syllables, fortis stops generally retain the amount of aspiration/VOT they have word initially, whilst they are deaspirated to various extents in reduced syllables (cf. Kahn 1976; Lavoie 2001 confirms experimentally that the effect of stress is far more limited word initially). Lenition in unstressed syllables leads to the shortening of oral closure intervals, which may conspire with deaspiration to produce fortis stops with relatively long voiced intervals or even full voicing. The most famous example of this is the English flapping rule, which leads to the (near-)merger of /t/ and /d/ to voiced [ɾ] in intervocalic contexts (Kahn, 1976; Fox & Terbeek, 1977).

An apparent exception to the claim that word-initial voicing categories are maintained word medially between sonorants and vowels is that the lenis stops of aspirating languages are often partially or fully voiced in the latter context. This is illustrated at the left in figure 2.2, which depicts a broad band spectrogram (top) and EGG trace (bottom) of the medial stop in English /keɪbl̩/ produced with a falling nuclear accent. Low-amplitude voicing, represented by a voice bar in the spectrogram and periodic pulsing in the EGG trace, continues throughout the closure phase of this stop.

The observation that the lenis stops of aspirating languages are usually voiceless utterance initially but partially or wholly voiced between a sonorant and a vowel has led some to the conclusion that in aspirating languages /b, d, ɟ, g, ɢ/ are assigned different voicing targets depending on the phonetic context. For instance, Keating (1984, 1990a) proposes that the underlying phonetic categories for [-tense] in aspirating languages are [-voice, -aspirated] utterance initially (as well as after another obstruent, and [+voice -aspirated] between a sonorant and vowel (as opposed to [+voice, -aspirated] across environments in voicing languages). Since Keating's interpretation rules are intended to capture *linguistic* aspects of phonetic realisation, this is tantamount to the claim that in aspirating languages, medial lenis stops are actively voiced.

However, there are good grounds to assume that the voicing targets for lenis

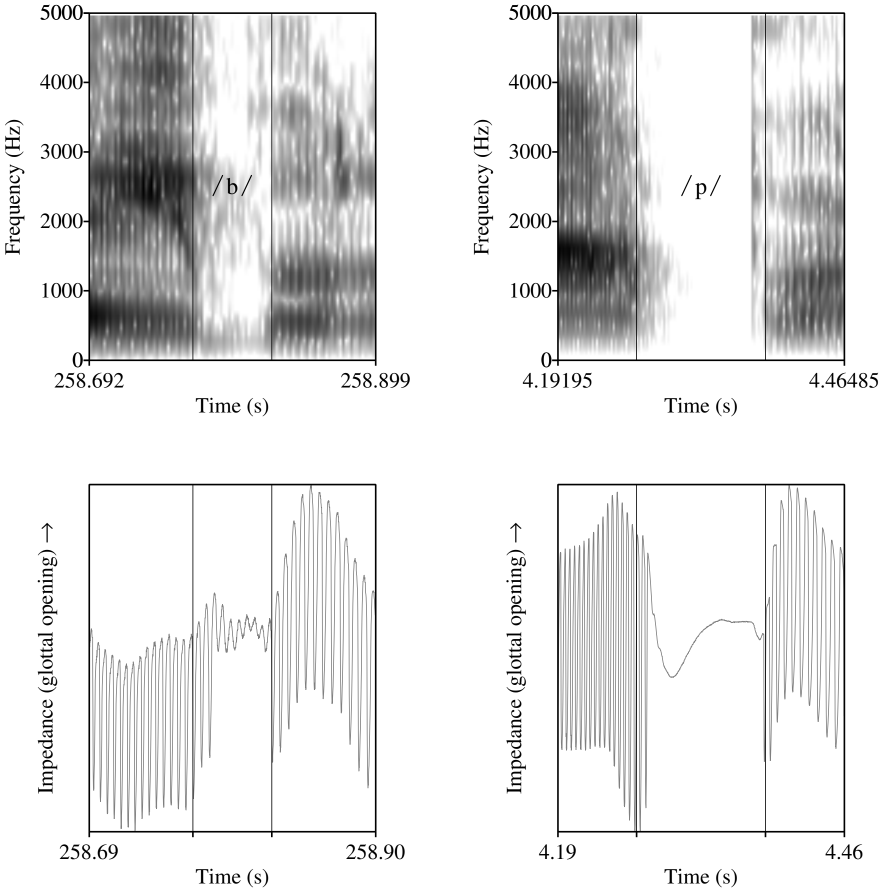

Figure 2.2: Broad band spectrograms (top) and EGG traces (bottom) of the medial labial stops in utterances of English /keibl̩/ (left) and Dutch /kaːpər/ (right).

stops are identical across initial and medial contexts in aspirating languages, and that any observed voicing is passive rather than active. First and foremost, the fortis stops of aspirating languages retain their long lag VOT intervocalically which implies that the voicing contrast with lenis stops is preserved even if the latter are partially or wholly voiceless (which they may be). In other words, the voicing of medial lenis stops of aspirating languages fails to trigger the VOT trade-off that seems so evident in utterance-initial contexts, where prevoiced stops contrast with short lag plosives rather than with aspirated ones.

Second, recall that Flege (1982) reports that 75% of English utterance-initial /b/s are produced with adducted vocal cords. Vocal fold adduction alone is not necessarily sufficient to generate voicing utterance initially, but the aerodynamic modelling studies mentioned above indicate that after a vowel or other sonorant, when the vocal folds are already in motion, vibration may continue up to 100 ms after oral closure, if the rest position of the vocal folds remains closed. Given the relatively short closure durations of singleton lenis stops, less would suffice to render a significant proportion of such sounds fully voiced. Both observations raise the suspicion that the voicing of intervocalic lenis stops in aspirating languages is passive rather than produced deliberately, and this conclusion is consistent with results of a study by Öhman (1962) who find that Swedish listeners do not use closure voicing to identify lenis stops in this context.

Finally, in 2.2 I mentioned that word initially fortis stops of voicing languages and the lenis stops of aspirating languages differ in terms of voicing when preceded by a vowel or sonorant consonant, despite their similar voicing/VOT utterance initially and after another obstruent. The same applies to word-medial postsonorant and prevocalic contexts, where ‘voicing’ /p, t, c, k, q/ are audibly less voiced than their lenis counterparts in aspirating languages. This is illustrated at the right in figure 2.2, which represents a broad-band spectrogram and EGG trace for the medial stop in Dutch /kapər/, *buccaneer*, also produced on a falling nuclear accent. In contrast to the medial stop of /keɪbl̩/, vocal fold vibration terminates after only two cycles in this plosive, and the remainder of its closure phase is voiceless. From a cue trading perspective, the mostly voiceless realisation of medial fortis stops in voicing languages is advantageous in maintaining the phonetic contrast with fully voiced lenis stops. Given the aerodynamic simulations of Ohala (1983), Westbury & Keating (1986) and others (though see Boersma 1998), it seems likely that whilst the lenis stops of aspirating languages lack voicing targets, the fortis stops of voicing languages are actively devoiced. In any case, the difference in voicing between ‘aspirating’ /b, d, ɟ, g, ɢ/ and ‘voicing’ /p, t, c, k, q/ indicates that they have *different* voicing targets, and hence constitutes an argument against models that treat them as a single, passively voiced, category (see chapter 8).

Preconsonantal and final stops Laryngeal distinctions in obstruents are relatively rare before consonants, especially other obstruents, and word finally. There is evidence that obstruents in neutralisation contexts are phonetically distinct from both their fortis and lenis counterparts (see chapter 3) and therefore this section only deals with the voicing of final and preconsonantal stops in contexts where the fortis-lenis contrast is maintained.

The rarity of laryngeal contrast in word-final contexts is perhaps a reason that there are no phonetic surveys of voicing distinctions in final stops to match the typology established by Lisker & Abramson (1964) and later work for word-initial stops. Consequently, it is hard to judge whether, and if so, how, the distinction between voicing and aspirating languages extends to word-final and preconsonantal contexts. It has been noted that French (a voicing language) and (aspirating) English use different acoustic cues to mark (and recover) [tense] in word-final stops (Mack, 1982; Flege & Hillenbrand, 1987). The former seems to rely relatively heavily on the properties of the release burst, contrasting released voiceless fortis stops with lenis stops that have voiced closure and release phases (sometimes described as an 'embryonic vowel'). In English, on the other hand, distinctions in vowel (and sonorant consonant) length play a more prominent role in signalling [±tense] word finally. Speakers of this language often partially devoice (utterance-final) word-final lenis stops (which suggests that they are passively voiced) and may leave both these and their fortis counterparts unreleased.

The VOT-based autosegmental models discussed in chapter 8 predict that aspirating and voicing languages employ different voicing targets to mark [tense] on word-final obstruents. More specifically, these models predict that in voicing languages, but not in aspirating languages, lenis stops are actively voiced (as they are in other contexts) whilst fortis stops are actively devoiced in aspirating languages, but not in voicing languages. The behaviour of word-final lenis stops in English and French is consistent with this prediction and so is data on English and Hungarian velar stops discussed in chapters 5 and 6. An observation that suggests some degree of consistency across contexts in the use of (de)voicing *strategies* is that preaspiration, which affects final and medial stops in various (and different) configurations in Icelandic, Faroese, and (dialectally) in Norwegian, Swedish, and English (on the latter, see Docherty & Foulkes 1999), typically occurs in languages that postaspirate their word-initial fortis stops (Ladefoged & Maddieson, 1996). In other words, none of the relatively few languages that exhibit fortis stop preaspiration in medial and final contexts distinguish prevoiced lenis and short lag VOT fortis stops word initially.

Although the tendency of preaspiration to 'alternate with' postaspiration/long lag VOT suggests that the terms *voicing language* and *aspirating language* are meaningful across positions in the word, it also highlights the possibility that voicing targets exhibit genuine positional variation. Specifically, the occurrence of fortis stop preaspiration suggests that in word-medial and word-final postvocalic contexts *V(oice) T(ermination) T(ime)*, i.e., the relative timing of the onset of an obstruent and the offset of voicing, can be used as an alternative to, or in addition to, VOT as a cue to [tense] (cf. Steriade 1997; Helgason 1999). An analysis in these terms of preaspiration in Germanic and European languages with other genetic affiliations (curiously all but ignored by Steriade 1997), seems far from straightforward and is clearly beyond the scope of this study.[6] However, a few relatively simple facts about English stops indicate that

---

[6] Thráinsson (1978) provides a number of arguments against treating Icelandic preaspiration as (synchronically) 'inverted' postaspiration. Nevertheless, Icelandic does not provide exceptions to

voicing distinctions in word-final and (some) preconsonantal contexts may indeed be organised in terms of VTT rather than VOT.

Note, first of all, that word-final fortis stops in English are generally deaspirated, even though impressionistic descriptions (in the phonological literature) would sometimes seem to imply that they are not. Glottal articulation data gathered by Yoshioka et al. (1981) show that word-final stops in American English are accompanied by at best rudimentary glottal abduction gestures, and under the definition of aspiration used here, this makes them unaspirated. Under the wider (and unhelpful) definition of aspiration as any form of aperiodic noise following the oral release of stops, they are still different from word-initial and word-medial prevocalic fortis stops in English, which are characterised by a substantial amount of glottal abduction, timed to peak around the time of oral release. The absence of aspiration suggests that aspiration noise and (in the case of a following vowel) positive VOT do not play a major role in signalling fortis stops word finally.

Moreover, word-final and (some) word-medial fortis stops are frequently preglottalised or even weakly ejective in English, i.e., they are produced with leading or simultaneous glottal *compression* rather than abduction. Gimson (1994) states that preglottalisation

> occurs in syllable-final position where a vowel, nasal, or lateral precedes and where a pause or consonant follows. [Preglottalisation] is more likely to occur at the end of an accented syllable.(Gimson 1994:155)

Views differ on the nature of the preglottalisation process. Since word-initial and word-medial stressed syllable-initial stops are never preglottalised, Harris (1994) treats it as a lenition process. However, this view is tied to a relatively abstract, phonological, notion of lenition, because in purely phonetic terms preglottalised stops can only be related to aspirated plosives by gesture substitution (swapping glottal abduction for tight adduction), and not by gesture loss or weakening, which results in deaspiration and, depending on the phonetic context, passive voicing. An alternative view, which is implicit in terms such as *glottal reinforcement*, [±tense] is realised in word-final and word-medial preconsonantal contexts in terms of early (fortis) vs. late (lenis) VTT targets rather than as long vs. short lag VOT (e.g., Westbury & Keating 1986). On this view, preglottalisation serves to stop vocal fold vibration from continuing after the fortis stop onset to the point where the critical 2000 dyne/cm_2/200 Pa threshold is reached, and thus to create a voicing distinction with lenis stops, which have a longer voice tail.

---

the striking generalisation that preaspiration only occurs in contexts where a contrast is maintained between (historically) fortis and/or geminate and lenis and/or singleton stops (Helgason, 1999).

The idea that preglottalisation is used to implement an early VTT target makes sense in purely mechanical terms, since medial compression of the glottis is an effective way to stop vocal fold vibration. But the underlying thought that voicing targets in final and preconsonantal contexts are organised with respect to obstruent onsets rather than offsets, would gain credibility from complementary evidence (apart from preaspiration data) showing that voicing is actively manipulated at sonorant-to-obstruent transitions (as well as from perceptual evidence). Jones (2001) provides one such complementary observation. He discusses the postnasalisation of lenis stops in some Lancashire dialects of English, which are recorded in Wright (1952) and Orton & Halliday (1963). The postnasalisation process inserts a 'parasitic nasal' at the end of a phrase-final lenis stop, as in [uːz wɛdⁿ], *she's married*. Jones proposes that this process arose as a strategy to prolong the voiced intervals in lenis stops through the (very effective) mechanism of nasal venting.[7]

Although the English data is largely impressionistic and possibly exceptional, it does raise the possibility that VTT distinctions are used to signal [tense] contrasts between preconsonantal and final stops, and therefore that there is genuine positional variation in the voicing targets associated with [±tense], and also in the articulatory strategies to implement those targets.

2.2.3 Other correlates of [tense] in plosives

In the introduction to this chapter, I mentioned that one of the reasons for distinguishing fortis and lenis as separate dimensions from the various voicing/VOT categories they map into, is that they define sets of obstruents that share a number phonetic properties regardless of their voicing characteristics. For example, both long lag and short lag VOT fortis stops raise the F_0 on a following vowel, whilst both actively and passively voiced lenis stops act as F_0 depressors. A number of excellent survey articles and book chapters document the common phonetic features of fortis and of lenis stops, notably Kohler (1984), Lisker (1986), and Kingston & Diehl (1994). As there are no major disagreements concerning the features involved, there is no need to recapitulate the detail of these studies here, and so this section presents only a brief review of the main correlates of [±tense] across voicing categories.

---

[7] A potential problem for this interpretation of Lancashire nasalization is that both lenis and fortis affricates are nasalised. It seems clear that voicing enhancement cannot be the goal of nasalising the fortis affricate /t͡ʃ/. Hahn (1998) notes a pattern that reinforces the impression that preaspiration and preglottalisation are parallel strategies for the implementation of early VTT. In Tuvan and Tofa, two closely related Turkic languages spoken in southern Siberia and northern China, plain voiceless stops contrast with preglottalised voiceless stops word finally, whilst related Yellow Uyghur and more distantly related Salar, both spoken in northern China, contrast plain voiceless and voiceless preaspirated plosives in this environment. Thus, Tuvan [ot] *fire* vs. [oˀt] *grass*, and Yellow Uyghur [ot] *fire* vs. [oʰt] *grass*.

Release burst features Both in aspirating and prevoicing languages, lenis stops have weaker release bursts than fortis stops, both in terms of duration and amplitude (e.g., Fischer-Jørgensen 1954; Halle et al. 1957; Slis & Cohen 1969a; Zue 1976). Lavoie (2001) shows that in English and Spanish, burst duration is subject to prosodic strengthening: the release bursts of fortis and lenis stops are longer than elsewhere in word-initial and prestress contexts. Perception experiments with French speakers indicate that at least in this language the quality of the release burst plays an important role in the identification of (word-final) stops as fortis or lenis (Wajskop & Sweerts 1973; van Dommelen 1983a, 1985; see also Kohler 1985).

Low frequency spectral features Most other things being equal, the F_0 and F_1 of a voiced vowel following a fortis stop start somewhat higher than the fundamental frequency and first formant of a vowel following a lenis stop. The effect is much stronger following than preceding stops, and it decays over time, so that F_0/F_1 differences are normally maximal at the time of voicing onset. The precise magnitude of the maximal F_0 difference varies across studies and languages, but rarely seems to exceed 30 Hz for female speakers (cf. House & Fairbanks 1953; Ohde 1984; Kingston & Diehl 1994; Jessen 1998, and references there). Although this may appear to so small as to be inaudible when superimposed on more dramatic variations due to intonation and vowel quality contrasts, perception experiments indicate that realistic variations in both F_0 and F_1 contribute to the identification of stops as fortis or lenis (e.g., Haggard et al. 1970; Kingston & Diehl 1995). F_0 microprosody due to [tense] or similar oppositions is therefore generally seen as a source of tonogenesis (Hombert et al., 1979).

Various attempts have been made to relate the effects of [tense] on F_0 and F_1 to each other and to the presence of closure voicing, which is also a low-frequency 'event'. Ladefoged (1973) and Stevens (1998) speculate that F_0 differences arise as a byproduct of active devoicing strategies. Relaxing the vocal cords lowers the critical thresholds for voicing somewhat and would therefore benefit the production of actively voiced stops but also lower F_0; conversely, tensing the vocal cords is an effective devoicing tactic but raises F_0, too. In addition, larynx lowering during lenis stops has been claimed to result in reduced tension in the vocal folds and hence in F_0 lowering. F_1 lowering in the vicinity of lenis stops has been similarly related to expansion of the pharyngeal cavity, which improves the aerodynamic conditions for voicing, but also results in a lower first formant (cf. Stevens 1998).

The prediction of this essentially mechanistic account is that the presence of F_0 and F_1 lowering implies the presence of (active) voicing. This prediction is plainly contradicted by the findings of Kingston & Diehl (1994), who show convincingly that no such implication exists: both the passively (de)voiced lenis stops of aspirating languages, and the actively prevoiced lenis stops of voicing languages act as F_0/F_1 depressors. Interestingly, fundamental frequency lowering by aspirated stops does not imply that they are (actively) voiced either. Whilst ('breathy') voiced aspirates do often act as F_0 depressors and can trigger the development of phonologically low tone (as in, e.g., Punjabi), Downing & Gick (2001) list a number of instrumental studies reporting F_0 lowering in the vicinity of voiceless aspirated plosives in several languages. Downing and Gick offer Botswana Kalang'a, a Southern Bantu language, as an instance of a language where such voiceless aspirates have developed into depressors of phonological tone.[8]

The reverse implication on the other hand, does seem to hold: actively voiced stops are typically accompanied by F_0/F_1 lowering. This observations is consistent with the alternative account of the cooccurrence of F_0/F_1 and (often) voicing distinctions provided by Kingston & Diehl (1994, 1995). They claim that the three phonetic properties in question are integrated into a single *low frequency* feature during auditory processing and are therefore treated by listeners (hence by speakers) as manifestations of the same phenomenon. On this view, the observation that active voicing is accompanied by F_0/F_1 *lowering* rather than raising follows from the fact that all three features causes a downward shift the spectral balance of the speech signal. Conversely, the absence of closure voicing and a raised F_0/F_1 all contribute to a shift of energy to higher frequencies in the spectrum.

Kingston & Diehl's theory is consistent with experimental data on cue trading in the perception of fortis and lenis stops. This data indicates, first, that (adult) listeners are able to use F_0 and/or F_1 cues to compensate for the presence or absence of closure voicing and vice versa in speech and non-speech stimuli. Second, it shows that stimuli with 'convergent' F_0, F_1, and voicing properties (i.e., low F_0/F_1 combined with voiced stop closure and high F_0/F_1 combined with voiceless stop closure), are rated as the 'best' (easiest to classify) members of their categories (Haggard et al., 1970; Kingston & Diehl, 1995). Note that the results of these studies are not necessarily at variance with the trade-off between prevoicing and long lag VOT that was identified in stop production and perception by Lisker and Abramson and others. It seems quite plausible that when parsing normal speech, native speakers of aspirating languages rely less heavily on low frequency information than speakers of voicing languages (because they have an extra cue in fortis stop aspiration), but that they are nevertheless able to integrate the presence or absence of voicing with F_0/F_1 cues.

---

[8] It is crucial to the argument by Downing & Gick (2001) that they are able to show that the voiceless aspirates of Botswana Kalang'a are not derived diachronically from earlier voiced obstruents.

Segmental duration In most languages for which data is available, including most of the Germanic languages, the closure phase of medial and final lenis stops is shorter than that of the corresponding fortis stops, whilst the preceding vowel (and sonorant, if present) is longer (Chen 1970; Kluender et al. 1988 and references there). House & Fairbanks (1953) and Lehiste (1970) report that in English, the duration of vowels before lenis stops and sonorant consonants is roughly equal. This suggests that the distinction in vowel duration before fortis and lenis stops is due to an asymmetric shortening process before the former, rather than to a symmetric process that lengthens vowels before lenis stops as well. Consequently, the vowel length effect is sometimes referred to as *pre-fortis clipping* (e.g., Harris 1994).

Perception experiments with English subjects indicate that the pattern found in production corresponds to listeners' expectations. Longer (voiceless) closure intervals result in an increase in the number of fortis responses, whereas increasing the length of a preceding vowel leads to an increased number of lenis responses (Denes, 1955; Liberman et al, 1961). Slis & Cohen (1969a) report an experiment in which Dutch speakers were asked to adjust the duration of vowels preceding fortis and lenis obstruents. The responses show a mean duration difference of 25 ms in the expected direction. More recent studies confirm the results of these early studies (e.g., Port & Dalby 1982; Luce & Charles-Luce 1985).

Most accounts of the relation between [tense] and segmental duration treat the effects on vowel length and consonantal closure length as intrinsically related, and thus in some ways analogous to fixed (syllable rhyme) quantity constraints in languages with distinctive length in both their vowel and consonant inventories. However, as any notion of isochrony in phonetics, the idea that speakers assign a fixed amount of time to a vowel (and sonorant) + obstruent sequence is not in itself an explanation.[9] As Kluender et al. (1988) point out, most production-based accounts founder on this observation: even if there is a mechanistic reason for vowel lengthening before lenis obstruents (Chomsky & Halle, 1968), or vowel shortening before fortis obstruents (e.g., Belasco 1953), or both, there is no a priori reason for speakers not to compensate these effects by active adjustments of vowel length.

Many perception-driven accounts on the other hand, avoid this objection because they derive the inverse patterning of [±tense] obstruent length and preceding vowel duration as a form of auditory enhancement, much as Kingston & Diehl (1994, 1995) construe the interactions between F_0 and F_1 voicing in terms of mutual enhancement (Port & Dalby 1982; Massaro & Cohen 1983, and

---

[9] Note that this fixed amount of time would be have to be set at a fairly abstract level of phonetic timing to allow for the effects of e.g., intrinsic vowel and consonant duration on observed acoustic durations. See Dauer (1983) on the related issue of stress-timing in English.

notably Kluender et al. 1988). The central idea of this approach is that increased vowel duration makes the duration of a following obstruent appear shorter, and conversely that a decrease in vowel duration increases the perceived duration of a following obstruent, and that vowel duration and obstruent duration are therefore integrated into a single percept. This hypothesis is largely supported by experimental evidence (Parker et al., 1986; Kluender et al., 1988). Moreover, Javkin (1976) and Parker et al. (1986) suggest that this account of duration phenomena should be embedded in a wider auditory theory of [tense] signalling (see alsoKingston & Diehl 1994). They find that the presence of voicing (and possibly the related F_0 and F_1 cues) during the constriction phase of an obstruent results in a shorter perceived duration, which prompts the conclusion that voicing/low frequency and durational cues are mutually enhancing too.[10]

Although the auditory enhancement theory of [tense]-driven segmental duration effects provides an explanation of why vowels and consonants should engage in duration trading relations, a remaining a problem is why these relations should virtually always be established in V-C rather than C-V sequences. One conceivable solution to this problem is that V-C (and particularly V-obstruent) transitions are less prominent than V-C transitions (Raphael 1981: see chapter 3) and that this causes V-C sequences to be parsed as chunks by the auditory system.

A problem of a different order is that effects of [tense] on vowel and obstruent closure duration are robust in citation forms and in purposefully designed carrier phrases, but appear to be considerably weaker and/or bound to specific contexts in more spontaneous forms of speech (Klatt 1975; Crystal & House 1982 et seq.). It may therefore be that the importance of segmental duration in signalling [tense] distinctions has been somewhat overstated on the basis of laboratory studies.

2.3 Fricatives and affricates

2.3.1 Voicing targets for tense and lax fricatives

Across languages and across contexts, the two basic distinctive voicing categories for fricatives appear to be voiceless unaspirated and voiced. Any qualifications that can be made to this generalisation are anecdotal and should therefore be treated with some caution.

---

[10] Any observation of mutual enhancement between two (clusters) of cues suggests that they may be traded against each other in speech perception. Consequently, the work of Javkin (1976) and Parker et al. (1986) implies that speakers of voicing languages may rely less on durational cues than speakers of aspirating languages. The study of English and French word-final stops by Flege & Hillenbrand (1987) quoted above suggests that this is indeed the case, but this topic deserves further investigation.

The differences in distinctive voicing between fricatives and plosives are perhaps nowhere more evident than in the rarity of aspirated fricatives, which only seem to occur in languages that already have distinctively voiced and plain voiceless fricatives (e.g., Burmese, Mazahua: Maddieson 1984; Ladefoged & Maddieson 1996). Although there are claims that lax fricatives are not equally voiced across different languages, there is little quantitative evidence to back up such claims.

Thus, it would appear that languages that maintain a fortis-lenis distinction in their fricative inventories generally oppose voiceless unaspirated fortis to voiced lenis fricatives, irrespective of whether they implement the fortis-lenis contrast between stop series as plain voiceless vs. prevoiced or voiceless aspirated vs. passively voiced. The Germanic group of languages certainly seems to conform to this generalisation. In (aspirating) English, and (marginally) German, voiceless unaspirated fortis [f, s, ʃ] contrast with voiced lenis [v, z, (ʒ)] whilst a phonetically similar contrast is found in (voicing) Yiddish and varieties of Dutch that maintain tense-lax distinctions between fricatives. A perception experiments reported by Stevens et al. (1992) shows that fricative voicing is used as a cue to [tense] by listeners in at least one aspirating language (American English). They find that word-initial lenis fricatives need at least 20 ms of voicing to be robustly categorised as such by listeners, while final lenis fricatives require 30 ms of voicing during the friction phase. In a similar vein, Slis & Cohen (1969a), quoting Forrez (1966), report that adding a low frequency periodic component to a synthetic alveolar fricative increases the number of /z/ classifications by Dutch test subjects. It appears therefore, that the voicing-aspirating distinction that is so useful to classify the realisation of (non-final) fortis and lenis stops does not extend to fricatives.[11]

Nevertheless, some phonetic descriptions suggest that there are crosslinguistic differences in the realisation of lax fricatives. For example, Jones (1956) claims that French [-tense] fricatives are produced with more or stronger voicing than their English counterparts. Zwaardemaker & Eijkman (1928) even make a three way distinction between the VOTs of French, English, and Dutch and German voiced fricatives, which they represent as [zzz] vs. [z̥zz] vs. [z̥z̥z]. But because these and similar (early) descriptive studies also tend to identify multiple distinctions in (negative) VOT that are not backed up by instrumental phonetic work, their assertions about fricative voicing cannot be taken for granted.

Experimental data on the voicing of English (N. Thorsen, 1971; Haggard, 1978; Docherty, 1992; Stevens et al., 1992; Smith, 1996) and German (Jessen, 1998) lenis fricatives reinforce impressionistic observations suggesting that such

---

[11] It is impossible to test the generality of this claim against phonetic inventory databases such as the various incarnations of the *UCLA Phonetic Segment Inventory Database* (UPSID) because such databases do not draw sufficiently fine-grained distinctions between phonetic voicing categories. See 3.4 below for details

fricatives tend to be produced as partially or fully devoiced, depending on the phonetic context. For example, Docherty (1992) finds that on average 89.3% , 100.0%, and 84.8% respectively of the friction intervals of utterance-initial /v, ð, z/ is produced with voicing. The percentages established by Haggard (1978) are somewhat lower, but show the same pattern with proportionally the longest voiced spans for /ð/ and the shortest for /z/. Jessen (1998) also reports a somewhat lower figure of 76% voicing for the friction intervals of German [v, z]. Observations of this kind lead Haggard (1978) to the claim that "[in English fricatives] production of the voicing and manner features is organised for the benefit of perception of the place feature(Haggard 1978:98)". However, in the absence of quantitative data on the realisation of the allegedly 'fully voiced' lax fricatives of French and similar languages, it is impossible to say whether the partial devoicing of English and German lax fricatives represents a distinct voicing target or just a general tendency of distinctively voiced fricatives.

2.3.2 Other features of tense and lax fricatives

It seems generally accepted that the other phonetic cues to [tense] in fricatives are similar to those found in plosives. Lax fricatives have shorter frication intervals and are preceded by longer vowels than their tense counterparts. Stevens et al. (1992) find a 30 ms frication duration difference between (longer) fortis and (shorter) lenis fricatives in English, whilst Crystal & House (1988b) report a mean difference of 39 ms. Differences of 50 and 59 ms respectively have been reported for Dutch and German /f, s, v, z/ (Slis & Cohen, 1969a; Jessen, 1998). Furthermore, Stevens et al. (1992) find that in English, vowels preceding word-final but utterance-medial lenis fricatives are 24 ms longer than those preceding a fortis fricative. In utterance-final position this difference increases to 41 ms. Crystal & House (1988a) present English data suggesting that this phenomenon is robust only for lexically long vowels. Finally, Slis & Cohen (1969a) show that in Dutch (laboratory speech) medial lenis fricatives cause a preceding vowel to lengthen too, by 25 ms on average.

Stevens et al. (1992) have suggested that the frication duration differences between English fortis and lenis fricatives are mechanical by-products of the voicing distinctions between them. If the duration of fricatives is defined in terms of the interval between the onset and offset of F_1 transitions in flanking vowels instead of in terms of frication duration, any significant difference in the duration of /s/ and /z/ disappears. This indicates that they are produced with oral constriction gestures of identical duration, and consequently that the difference in frication duration is not a function of oral tract control. Stevens et al. (1992) suggest instead that voiced fricatives have shorter frication intervals because they are produced with a smaller glottal abduction gesture, which fulfils the aerodynamic requirements for turbulence noise generation for a relatively short interval in comparison to the large abduction gesture that accompanies voiceless fricatives (cf. 2.1). An aerodynamic source of the frication duration difference of course does not exclude the possibilities that it is used as a cue by listeners.

In addition, lax fricatives are signalled by a lower relative frication intensity (Strevens, 1960; Balise & Diehl , 1994) and lower F_0/F_1 offsets and onsets of preceding and following vowels respectively (House & Fairbanks, 1953; Stevens et al., 1992). The similar clusters of cues associated with [±tense] in stops and fricatives suggests that they are shaped by similar perceptual integration and (hence) cue trading mechanisms (cf. 2.2.3 above). There is some evidence that, as for stops, F_0 lowering does not imply voicing in fricatives. Downing & Gick (2001) describe two voiceless labiodental fricatives in Nambya, a Southern Bantu language, one of which acts as a tone depressor. Although this type of sound is sometimes described as 'breathy voiced' (cf. Ladefoged & Maddieson 1996) Downing and Gick demonstrate that the depressor fricative is both phonetically voiceless and has longer frication intervals than the non-depressor. The proto-Bantu precursor of Nambya depressor [f] has been reconstructed as a voiceless [t] + [u̜] sequence, which makes it unlikely that its synchronic effect on tone stems from phonetic F_0 lowering by a voiced obstruent. Downing & Gick (2001) therefore treat it analogously to the Botswana Kalang'a depressor aspirates, as stemming from a voiceless F_0 depressor.

2.3.3 A note on tense vs. lax in affricates

It is a common complaint that (voicing distinctions in) fricatives are underresearched in comparison to plain stops. The same could be said of lexical affricates. From phonetic descriptions it appears that, unlike fricatives, fortis-lenis distinctions on affricates are generally cued as they are on the plain stops of the language in question. Thus, the lexical affricates /t͡ʃ, t͡s, d͡ʒ, (d͡z)/ of Hungarian, a voicing language, are realised as (actively) voiced vs. plain voiceless (Kenesei et al., 1998), and the same applies to the lexical affricates found in the Slavonic and Baltic languages. Similarly, the lexical affricates /t͡ʃ, d͡ʒ/ of English are usually described in roughly the same aspirated vs. unaspirated (with partial voicing in certain environments) terms as its plain stops (Jones, 1956; Gimson, 1994). The same sources assign the durational and spectral characteristics to the /t͡ʃ//-/d͡ʒ/ that would be expected from the behaviour of plain stops and fricatives. Thus, /t͡ʃ// is described as possessing a longer and louder release stage and preceded by a shorter vowel than /d͡ʒ//

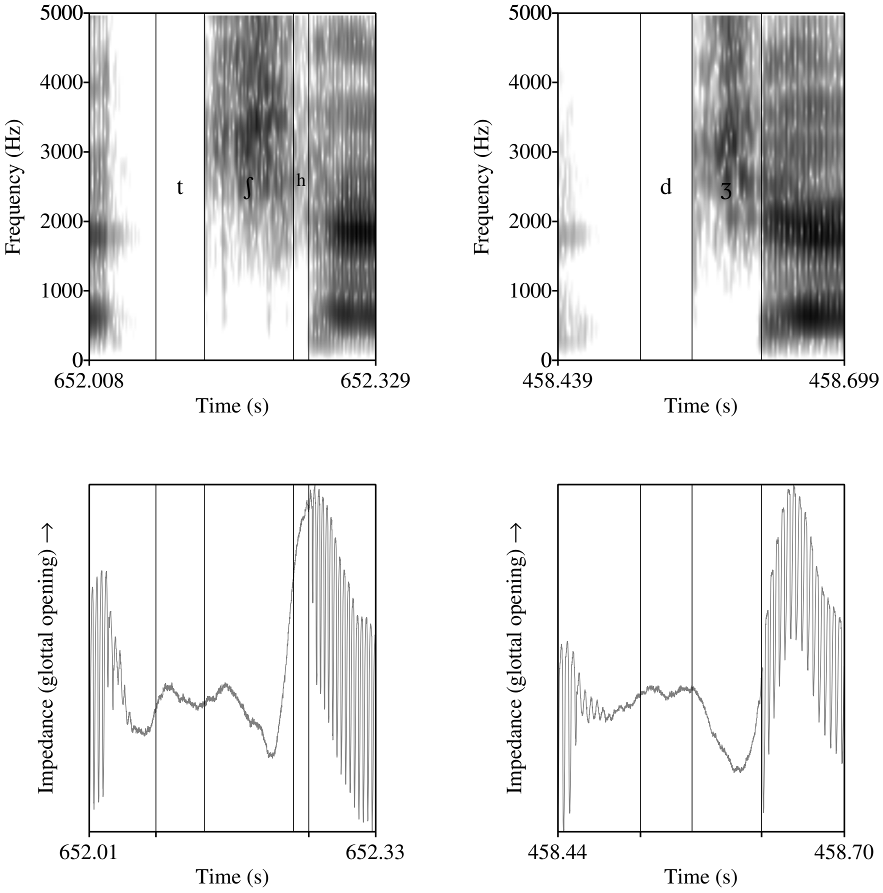

Figure 2.3: Broad band spectrograms and EGG traces of English word-initial affricates

However, as illustrated in figure 2.3 the release stage of the English lax affricate /d͡ʒ/ is markedly longer than that of the corresponding plain stops, and the aspiration of /t͡ʃ/ can be partially or fully overlapped by its release stage (c.f. Jones 1956). Figure 2.3 represents the broad band spectrograms and EGG traces of the initial affricates of English /t͡ʃɛɹɪ/ and /d͡ʒɛlɪ/, produced (with a nuclear accent on the first syllable) in the carrier sentence *please say* _ *again*. According to the standard definition, the VOT of /t͡ʃ/ is 117 ms, and composed of the release stage (marked by ʃ: 100 ms) and a brief period of aspiration noise (marked by ʰ: 17 ms). This is fairly long, but by no means out of bounds for a plain tense stop. On the other hand, the 63 ms VOT of /d͡ʒ/ is well within the (>35 ms) ‘long lag’ bracket and highly untypical of plain lax stops in the same prosodic and segmental phonetic contexts. Note that if VOT is measured from the oral release *offset*, the 17 ms difference between /t͡ʃ/ and /d͡ʒ/ is not what would be expected of an aspirating language either.[12]

It appears therefore that VOT is at best secondary to the duration and quality of the fricative release in signalling [±tense] affricates in the relevant set of contexts and/or that /d͡ʒ/ is signalled by an independent VOT category. In any case, the phonetic properties of English fortis and lenis affricates reinforce the conclusion drawn in the previous section that [tense] is not necessarily cued in the same way across obstruent manner classes within the same language.

2.4 Summary and remaining issues

The first and principal aim of this chapter was to develop a phonetic taxonomy of fortis and lenis obstruents and voicing assimilation on the basis of a literature review. After a brief review of the aerodynamics of voicing production, I defined a four-term phonetic distinction between prevoiced lenis, passively voiced lenis, actively devoiced (plain voiceless) fortis and actively devoiced aspirated fortis stops. The two term distinction between fortis and lenis stops is based on phonetic features other than voicing, such as segmental duration, release burst characteristics and formant perturbations. The clustering of these cues (and their relative values) around a single lexical distinction may well be explained in terms of their perceptual synergy. An even stronger argument can be advanced at a somewhat lower level for the perceptual integration of voicing and low frequency spectral cues in a single low frequency (or spectral balance) feature.

The four-term voicing distinction harks back to the original three term VOT distinction for word-initial stops as established by Lisker & Abramson (1964), but takes on board differences between lenis and fortis instantiations of their zero-to-short lag category. The former typically has a short lag VOT in utterance-initial and post-obstruent contexts, but can acquire voicing from a preceding sonorant, which indicates that they are passively voiced. The latter remain mostly voiceless across phonetic environments, which is an indication that they belong to a distinct, actively devoiced category. Articulatory evidence and aerodynamic modelling suggest that actively voiced and devoiced stops are accompanied by a range of articulatory measures aimed at manipulating the size of the oral tract by the occlusion and the critical transglottal pressure threshold for vocal fold vibration.

In section 2.2 I argued that the four way phonetic classification of tense and lax stops holds for word-initial and prevocalic medial stops, but that word-final

---

[12] The /d/s preceding both affricates in figure 2.3 are unreleased and therefore the timing of the boundaries marking the onset of the latter is arbitrary.

(and preconsonantal) stops seem to show some genuine positional variation in voicing targets. At least in aspirating languages of Germanic, and in particular English, postaspiration seems to play a minor role in signalling [tense] in this context. As shown in 2.3 the four-term phonetic taxonomy does not extend to fricatives either: although there is some anecdotal evidence of a finer-grained organisation of voicing distinctions, the basic phonetic categories for tense and lax fricatives are plain voiceless vs. (partially) voiced.

The lexical affricates /t͡ʃ, d͡ʒ/ of English finally, do not appear to fit the plain stop template perfectly either, because the standard definition of VOT seems inappropriate for these sounds, which may (instead) be heavily cued by their release characteristics.

Chapter 3

Laryngeal neutralisation

To phonologists familiar with the Germanic or Slavonic languages, the principal example of laryngeal neutralisation is likely to be the ‘dynamic’ process of *final laryngeal neutralisation*, generally referred to as *final devoicing*. This phenomenon is illustrated with examples from Dutch in 4. In Dutch, a lexical contrast between word-final fortis and lenis obstruents is realised in the regular past tense (as discussed above) and before certain vowel-initial suffixes such as the //-ən plural suffix, but not in unsuffixed forms or before a variety of other suffixes, including the diminutive suffix. Similar processes are found in Frisian (Tiersma, 1985), German, and many of the Slavonic and Turkic languages. Terms such as *final devoicing* or (German) *Auslautverhärtung* originate in the realisation of the neutralised series as mostly voiceless in citation forms and the identification of devoicing with fortition (i.e., a change from [-tense] to [+tense]).

(4) Final neutralisation in Dutch

UR	Plural	Citation	Diminutive	Gloss
/xrɑp/	[χrɑpə(n)]	[χrɑp]	[χrɑpjə]	joke
/krɑb/	[krɑbə(n)]	[krɑp]	[krɑpjə]	crab
/ɣraːt/	[χraːtə(n)]	[χraːt]	[χraːtjə]	fishbone
/ɣraːd/	[χraːdə(n)]	[χraːt]	[χraːtjə]	degree

However, laryngeal neutralisation also occurs in the form of ‘static’ constraints on lexical items. In Thai, which maintains a three term laryngeal contrast in word-initial stops, only a single series of stops (commonly described as phonetically plain voiceless) is present word finally at the lexical level. In other words, Thai lacks alternations between non-neutralised and neutralised forms of the type illustrated in (4). Similar lexical constraints against the marking of laryngeal contrast word finally occur in several other (south)east Asian languages including Khmer (or Cambodian: Mon-Khmer), Manipuri (or Meithei), Tibetan (both Tibeto-Burman) as well as Korean (Henderson, 1952; Bhat & Ningomba, 1997; Chang & Shefts, 1964; Cho, 1990a/1999). Furthermore, even languages that impose few or no restrictions on the marking of laryngeal contrast on single obstruents often restrict obstruent clusters to a single laryngeal specification, or suspend the marking laryngeal contrast on such sequences altogether. Finally, in many languages that mark laryngeal contrast on stops across contexts, the scope for marking the same contrast on fricatives is often more limited. For instance, Frisian lacks a contrast between fortis and lenis obstruents word initially, whilst the North Germanic languages neutralise the distinction across the board for sibilants, or, if /v/-type sounds are treated as sonorants (as in Vanvik 1972), for all fricatives.

Although these lexical, 'static', constraints on the occurrence of laryngeal contrast tend to receive less attention in the generative literature than their dynamic counterparts, they sometimes play an important role in theory formation. For example, arguments for and against syllable-driven analyses of (final) laryngeal neutralisation are ultimately based on the behaviour of word internal non-alternating obstruent + sonorant clusters.

The goal of this chapter is to present a number of observations about neutralisation phenomena and to assess these observations in the light of formalist and functionalist, cue-driven ideas about neutralisation phenomena. Although voicing assimilation rather than laryngeal neutralisation is the focus of the experimental part of this study, the survey presented here is important as a backdrop to the investigation of Hungarian RVA in chapter 6 and the results of the Dutch experiment reported in chapter 7. The conclusions of this chapter are also central to the critique of current autosegmental models of assimilation and neutralisation that is the topic of chapter 8.

Section 3.1 starts with the recognition that theories of laryngeal neutralisation consist of a component describing the nature of the assimilation process, and a second component detailing the set of contexts in which neutralisation occurs. In this regard, *context* should be understood in a broad sense so as to include the phonetic features of the obstruents targeted by neutralisation as well as prosody. This section proceeds to outline the two rivaling conceptions of the first component that are relevant to the wider purposes of this study and contrasts the relatively fragmented formalist view of the second component with the potential of a unified cue-based model.

The remaining sections examine to what extent the contrasting views of each component are supported by the available data. Thus, section 3.2 finds some evidence for the idea that laryngeal neutralisation results in phonetic underspecification and discusses experiments suggesting that laryngeal neutralisation is often phonetically incomplete. Sections 3.3 and 3.4 assess the effects of obstruent features on the stability of laryngeal contrast, discussing the evidence for an asymmetry between plosives and fricatives but not between voicing and aspirating languages. There is some evidence that the instability of laryngeal contrasts in fricative inventories is due to the poor realisation of place cues in voiced fricatives and are therefore amenable to an explanation in terms of perceptibility. Next, 3.5 reviews the arguments levelled by Steriade (1997) against syllable-driven models of final neutralisation, which crucially fail in their predictions about the behaviour of nonfinal obstruent + sonorant clusters. By contrast, cue-based models give an accurate account of the observations that neutralisation in these clusters occurs in tautosyllabic as well as heterosyllabic sequences and does not necessarily coincide with word-final neutralisation.

Section 3.6 then tries to extend the cue-based models to the relatively uniform behaviour of laryngeal distinctions in obstruents appearing in (stable) word-initial and (unstable) word-final positions across left and right-hand phonetic environments. The main argument for this idea is that it is likely that prosodic strengthening and weakening have a levelling effect on the perceptibility of laryngeal contrast at word edges. If this is indeed the case, the extended cue-based model is preferable to both word-based formalist models and the *Paradigm Uniformity* of Steriade (1997) account on grounds of generality. Section 3.7 finally, presents a summary of the chapter and highlights some of the issues that are left unexplored.

3.1 Theories of (final) laryngeal neutralisation

Theories of laryngeal neutralisation consist of two broad components. The first specifies the nature of the neutralisation itself, i.e., its input (in the case of dynamic neutralisation) and the phonological and phonetic status of its output. The second component defines the phonetic environments which are likely to trigger neutralisation in substantive phonetic and/or prosodic and/or morphosyntactic terms. In this regard, the term *environments* should be taken in a broad sense, so as to include the phonetic features of sounds targeted by neutralisation. For instance, fricatives are more vulnerable to neutralisation than stops, and this effect of obstruent manner should be predicted by any complete theory of laryngeal neutralisation. The range of theories and models that have been proposed in the literature is far too broad to attempt a comprehensive survey here. Instead this section outlines a number of assumptions and predictions that are common to (semi-)formalist approaches, and contrasts these with the cue-based theory of laryngeal neutralisation proposed by Steriade (1997)

3.1.1 The nature of neutralisation processes

A pervasive assumption about the nature of final laryngeal neutralisation in obstruent inventories split by a fortis-lenis-type contrast is that it is a *fortition* process converting [-tense] obstruents into their [+tense] counterparts. In such theories, the [-tense] final obstruent of /krɑb/ in 4 changes into its [+tense] counterpart in the diminutive and citation forms, whereas neutralisation does not apply to the underlying /p/ of /xrɑp/, which already is [+tense].

In a more general sense too, accounts of laryngeal neutralisation often rest on the assumptions (a) that neutralisation is fundamentally asymmetric (b) that the output of neutralisation is identical to a series that is lexically present. These assumptions often play a role in (generative) analyses of final neutralisation in languages with richer laryngeal systems (e.g., Thai), which tend to identify the result of neutralisation with the lexical series realised as plain voiceless. They are often visible in accounts of other forms dynamic neutralisation: postnasal voicing, for instance, is often assumed to result in obstruents that are identical to actively voiced (lenis) stops (cf. Pater 1996/1999). Finally, assumptions (a) and (b) tend to be (implicitly) applied in the representation of lexical neutralisation patterns. and to the treatment of lexical neutralisation For example, English [s] + obstruent clusters are analysed as lexically [+tense] by Iverson & Salmons (1999) (see chapter 8). The key prediction of a fortition analysis of laryngeal neutralisation is that neutralised obstruents should behave as fortis obstruents both phonetically and phonologically.

A number of researchers propose an alternative view of laryngeal neutralisation that harks back to older, ‘archiphonemic’, and therefore symmetric conceptions of laryngeal neutralisation. According to this alternative view, neutralised obstruents form a class that is distinct from both their fortis and lenis counterparts and is therefore predicted to behave in a distinct fashion both phonetically and phonologically. Consequently, all lexically contrastive series are affected in the case of a dynamic neutralisation process: both the final /b/ of /krɑb/ and the final /p/ of /xrɑp/ in 4 transform into a third class of obstruents in the citation and diminutive forms. For convenience I will refer to this class of obstruents as [0tense], and denote individual members using capitals (/P,T,K,S/), following structuralist notational conventions.

In recent work, this symmetric conception of (laryngeal) neutralisation is typically tied to the concept of *phonetic* or *surface* underspecification in the sense of Keating (1988) or Pierrehumbert & Beckman (1988). Whilst formally similar to older archiphonemic analysis, the notion of phonetic underspecification critically refers to the phonetic behaviour of a sound rather than the (mere) suspension of phonological contrast (cf. chapter 1). This means that for a sound to be analysed as phonetically underspecified for a certain feature, it should behave passively with regard to all phonetic correlates of that feature. For example, a [0tense] obstruent should be completely passively voiced in the sense defined in chapter 2, and as such be distinct from [+tense] obstruents that are actively devoiced as well as from [-tense] obstruents that are actively voiced.[1] Underspecification models of laryngeal neutralisation are usually associated with phonetic studies (Hsu, 1996; Ernestus, 2000) or functional models (Steriade, 1997), but do not strictly speaking pre-empt the choice for a theory of neutralisation contexts.

3.1.2 Syllable-driven approaches to laryngeal neutralisation

It has been common in the phonological literature to characterise the environments in which laryngeal neutralisation processes tend (not) to occur in terms of syllable structure. Two broad varieties of syllabic theory can be distinguished: one that tries to capture the set of contexts which are prone to neutralisation (treating sites resistant to neutralisation as the elsewhere environment); and one that tries to capture the set of environments in which laryngeal contrast is relatively robust (treating sites that are prone to neutralisation as the elsewhere case).

The first variety, exemplified by Mascaró (1987/1995), Cho (1990a/1999) treats final neutralisation as an operation targeting obstruents that appear in a syllable rhyme or (if it is recognised as separate entity) coda, but ignoring all other contexts. Thus, this approach tries to define in positive terms where laryngeal contrast is neutralised and treats the set of contexts where contrast is maintained as the elsewhere environment. Word-final obstruents are subject to the process because they are assumed to be parsed as rhyme constituents.

The second approach, adopted by, e.g., Gussman (1992) and central to the work of Linda Lombardi, instead tries to define the set of contexts in which contrast is *maintained* in positive terms, and specifies the set of contexts in which neutralisation operates as the complement of this environment. Under this analysis, word-final obstruents are targeted by neutralisation because they appear in an elsewhere context. For example, the (parametric) *Laryngeal Constraint* of Lombardi (1994, 1995a,b) states that laryngeal features are only licensed when they appear in the configuration in figure 3.1: the LAR class node that obligatorily dominates the substantive laryngeal features [voice], [asp], and [gl] should be dominated by a *root* node adjacent to a tautosyllabic sonorant (the formulation in Lombardi 1994 et seq. usually omits the root node). In languages with the Laryngeal Constraint switched on, lexically present LAR nodes appearing

---

[1] Whilst the [0tense] category identified by Ernestus (2000) and others and the lenis plosives of aspirating languages as analysed in chapter 2 both lack targets for voicing they are still distinct in that the latter but not the former category has targets for the other correlates of tense. Further research is needed to establish whether speakers and listeners can indeed distinguish between these two categories.

in any configuration that is different (such as word-final obstruents) are deleted, leaving a laryngeally unmarked obstruent, which in Lombardi's framework is equivalent to [+tense] (see chapter 8 for details).

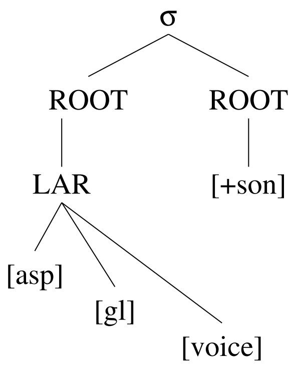

Figure 3.1: Licensing configuration for laryngeal contrast according to the *Laryngeal Constraint* (adapted from Lombardi 1994 with ROOT nodes added)

The differences between the two broad types of syllabic theory reside in matters of overall model architecture and in (alleged) formal efficiency or elegance. Generally speaking, the *licensing* approach envisaged by Lombardi (1994, 1995a,b) is part of a recent trend in generative phonology towards filter-cum-repair-rule or fully declarative models, whilst the 'coda neutralisation' theory stems from older rewrite rule-based phonological frameworks. In broad terms, the prediction of any syllabic theory is that the full set of neutralisation sites in a given language can be defined as a natural class in terms of syllable structure, and syllabic theories can only be distinguished from word or cue-based alternatives if the set of predicted neutralisation sites is in some way different from those emerging from morphosyntactic (or higher-level prosodic) or cueing considerations. Consequently, nearly everything in the defense of a syllabic theory of laryngeal neutralisation contexts hinges on an independently motivated theory of syllabification.

Models that combine a fortition approach with a syllable-driven account of neutralisation contexts are fairly frequent in the literature and may perhaps be regarded as the, 'default', or zero hypothesis, choice for generative models. The survey of voicing assimilation and laryngeal neutralisation phenomena presented by Mascaró & Wetzels (2001) is a case in point as it questions neither the validity of a fortition analysis nor the claim that final neutralisation can be syllable-driven, whereas it defends the case for a theoretically separate phenomenon of *word*-final neutralisation. However, claims regarding the nature of laryngeal neutralisation and hypotheses about contexts in which neutralisation occurs are separate components of a theory and can therefore be combined freely. Thus, Trommelen & Zonneveld (1979) propose a fortition account of Dutch final neutralisation claiming that the process occurs finally at the end of words, or more precisely, at the end of constituents separated by analytical morphological boundaries. Gussman (1992) on the other hand analyses Polish neutralisation as an essentially symmetric process deleting [-voice] (i.e., [+tense]) as well as [+voice] ([-tense]) specifications operating in a set of contexts defined in terms of syllable structure.[2]

3.1.3 Cue-based approaches to laryngeal neutralisation

Steriade (1997) proposes a cue-based theory of neutralisation contexts as an alternative to, and improvement over, syllable-driven models, concentrating on the (universal aspects of the) role of phonetic context in determining laryngeal neutralisation in stops. The prediction of this theory, in line with cue-based approaches more generally (c.f. chapter 1), is that neutralisation of laryngeal contrast is more likely in environments where it is relatively hard to recover from the speech signal. The general thrust of Steriade's approach is consistent with both 'diachronic' and 'synchronic' theories of the relation between perception and phonology, but the specific model she proposes belongs to the 'synchronic school'. Consequently, Steriade (1997) derives hard synchronic implications from relative perceptibility such that the availability of laryngeal contrast in a context with perceptibility x implies the existence of laryngeal contrast in all phonetic contexts where its perceptibility is greater than x.

Early on in the paper, Steriade (1997) suggests the hierarchy in (5) to capture the relative perceptibility of fortis-lenis type laryngeal contrast in stops, where 'V' represents a vowel, '$>$' indicates greater perceptibility, and '$\|$' is used to represent a physical pause. This hierarchy is expanded and refined in the course of the argument, but it suffices to illustrate the basic idea of Steriade's theory. It states, for example, that ceteris paribus, the fortis-lenis contrast is more perceptible between a vowel and a following sonorant (including vowels) than between a vowel and another obstruent. According to the 'diachronic' version of functional phonology this means that learners of a language that has contrasts in both contexts are most likely to misperceive it as non-existent in the latter and neutralise it in their own grammars. In Steriade's 'synchronic' model it means that constraints on laryngeal contrast in V__[-son] contexts outrank those that ban contrast in V__[+son] environments, and the prediction follows that a language

---

[2] Gussman uses a 'late' default rule to insert [-voice] on [0tense] obstruents in environments where they are phonetically voiceless.

that maintains contrast in V__[-son] also maintains it in V__[-son] (and in the intervening environment V__‖).

(5) Perceptibility hierarchy for laryngeal contrast in stops, according to (Steriade 1997:12)

V__[+son] > V__‖ > V__[-son] > {[-son]__[-son], [-son]__‖, ‖__[-son]}

The hierarchy in (5) is mainly based on the *number* of cues that are potentially available in each context. Thus, in V__[+son] plosive *onset* cues (preceding vowel duration, V-to-C formant transitions, preaspiration or preglottalisation), *internal* cues (presence and quality of voicing, duration), as well as *offset* cues (duration and relative amplitude of the release burst, postaspiration or postglottalisation, C-to-V formant transitions) are available to signal the distinction between fortis and lenis plosives. The V__[-son] context is somewhat poorer in cueing potential in lacking at least the V-to-C formant transitions and depending on manner of articulation of [-son] and the amount of coarticulation, the release and postaspiration cues too (a refined version of the hierarchy would need to separate the various options here). It might be surmised from this that the fortis-lenis distinction is less salient in V__[-son] than in V__[+son] environments. In a number of instances the rankings on Steriade's perceptibility hierarchy are directly backed up by experimental evidence. For example, it Raphael (1981) shows how plosive onset cues are perceptually less salient than offset cues: in case of conflicting information, listeners give priority to the latter. This means that laryngeal contrast in stops lacking offset cues is less perceptible in those that do not, even if the different *numbers* of cues themselves should be demonstrated to have no effect on perceptibility.

Although inferences made on the basis of numbers of (potential) cues are not ultimately valid indicators of relative perceptibility, the set of individual rankings that make up the hierarchy in (5) all amount to testable hypotheses. For instance, the claim that a given laryngeal contrast, realised by a particular set of phonetic cues, is less perceptible (universally and/or to native speakers of a particular language) in the context [-son]__[-son] than in V__[-son], can be tested using the type of experimental methodology employed by, e.g., Mielke (2001). Thus, both the perceptibility hierarchy in (5) and the predictions derived from it can be falsified independently of each other. On the other hand, as long as they are interpreted as autonomous grammatical mechanisms, the Laryngeal Constraint and similar devices remain stipulations, even if to some extent they share the spirit of the hierarchy in (5) (i.e., in recognising that laryngeal contrast is relatively stable in the presence of a following sonorant).

Moreover, a cue-based theory of laryngeal neutralisation bears the promise of unifying the description of all laryngeal neutralisation asymmetries in terms of a single mechanism. This would be an achievement beyond anything formalist models seem to be able to deliver. Recall from chapter 1 that the relative perceptibility of a contrast is not just a function of the immediate phonetic environment of the sound bearing it, but also of its own acoustic properties, prosody, and phonetic knowledge of the listener (as gained from exposure during language acquisition). For example, the asymmetry between word-initial and word-final contexts in terms of their propensity for laryngeal neutralisation might be related to the phonetics of initial strengthening and its (presumed) effects on relative perceptibility (c.f. section 3.6). Similarly, given the observation that place cues are harder to detect in voiced than in voiceless sibilants, a cue-based theory predicts that contrastively voiced sibilant series are likely to have fewer distinctive places of articulation than their voiceless counterparts and more likely to be absent altogether from a given language (see section 3.4). In other words, it is predicted that laryngeal contrast is relatively unstable in sibilant fricatives. This prediction is a simple instantiation of the basic hypothesis that neutralisation is a function of relative perceptibility and does not require any additional mechanisms. Although there are (semi-)formalist models (Steriade 1992, 1993; Vaux 1998: see chapter 8) that address the relative instability of laryngeal contrast in (sibilant) fricatives, I am not aware of any such models that employ the same formal mechanism to account for segment internal and context biases in laryngeal neutralisation.

3.2 The phonetics of laryngeal neutralisation

This section summarises a two sets of observations concerning the phonetics of laryngeal neutralisation. First it reviews phonetic evidence indicating that laryngeal neutralisation can lead to phonetic underspecification of [tense]. It then considers a second body of phonetic data, which suggests that processes traditionally described as categorical neutralisation processes are in fact phonetically incomplete. The first type of evidence data is problematic for a fortition analysis of laryngeal neutralisation because it requires neutralised obstruents to be classified as distinct from fortis as well as lenis obstruents. The second type of evidence may eventually undermine a more fundamental assumption of generative analyses, namely the idea that every surface form is derived from a single underlying representation.

3.2.1 Laryngeal neutralisation as phonetic underspecification

Whereas fortition accounts and more generally lexical feature accounts of laryngeal neutralisation predict that neutralised obstruents should be phonologically and phonetically identical, phonetic underspecification models predict that phonetically speaking they are distinct from all lexically contrastive series. One proponent of the phonetic underspecification approach to laryngeal neutralisation is Ernestus (2000), who argues that the behaviour of word-final obstruents in Dutch is consistent with the absence of voicing targets. For example, Dutch final obstruents are voiceless utterance finally but often audibly voiced before vowel-initial enclitics, i.e., at weak prosodic boundaries before unstressed syllables (cf. Gussenhoven 1986): [hɛbɪk] for /hɛb + ɪk/ *have I* as opposed to [ɪkhɛp] *I have*, whilst /dɑtɪk/ *that I* is often realised as [dɑdɪk]. As noted in chapter 2, fortis plosives occurring in word internal intervocalic environments are generally mostly voiceless in Dutch. Ernestus (2000) reports a listening experiment showing that phonetically trained native speakers of Dutch are able to draw a three way distinction among word-medial fortis and lenis and word-final plosives, which suggests that observed voicing distinctions between unambiguously fortis obstruents and word-final obstruents indeed reflect a difference in underlying targets. In addition, the fact that word-final stops, as opposed to fortis-initial plosives, are realised with a negligible amount of glottal abduction in Dutch (Yoshioka et al., 1982) is also consonant with the idea that they have no phonetic targets for [tense].[3] On similar grounds Hsu (1996) argues that laryngeal neutralisation of (the three term distinction in) Taiwanese stops results in phonetic underspecification.

To what degree the surface underspecification account of laryngeal neutralisation can be extended beyond the data considered by Hsu (1996) and Ernestus (2000) is an empirical matter. One case where the suspension of laryngeal contrast seems to coincide with the absence of voicing/VOT targets is the realisation of word-initial sibilant + plosive clusters in the aspirating varieties of Germanic. Whereas the [tense] plosives of these languages are realised with aspiration and a long lag VOT word initially, plosives preceded by tautomorphemic [s] are normally voiceless unaspirated and have a short lag VOT. In the light of the fact that all the languages and dialects in question suspend laryngeal contrast in word-initial sibilant + plosive clusters ([sp, st, sk, ʃp, ʃt] never contrast with, e.g., [zb, zd, zg, ʒb, ʒd]) might be interpreted as evidence that they are [0tense] rather than [+tense] and phonetically underspecified for [tense] correlates. On this analysis the plain voiceless realisation of sibilant + stop clusters follows directly from passive (de)voicing. As argued in chapter 2 it is plausible that the initial sibilant is subject to passive devoicing because a large amount of glottal abduction represents the ideal configuration for the production of high-intensity noise (measurements reported by Yoshioka et al. 1981 indicate that the initial sibilant is indeed produced with a large amount of glottal abduction). The

---

[3] Yoshioka et al. (1982) suggest that the lack of glottal abduction in Dutch word-final stops is due to glottalisation as in English. However, in English and other languages with glottalisation, irregular voicing occurs during and in the vicinity of stops affected by the process, whereas no such effect has been reported for Dutch.

voiceless realisation of the initial sibilant removes any source of passive voicing for the adjacent plosive which will therefore be produced without any voicing even in the absence of active devoicing measures. If sibilant + plosive clusters are produced without active (de)voicing measures, their short lag VOT is also predictable on aerodynamic grounds: voicing sets in as soon as the transglottal pressure difference allows for it, i.e., shortly after the oral release of the plosive.

The hypothesis that word-initial sibilant + plosive clusters are [0tense] of course demands that they should be phonetically distinct from both their [+tense] and [-tense] counterparts with regard to correlates of the fortis-lenis distinction, that is, unless it can be shown that a phonetic feature of one of the latter two classes reflects passive behaviour as well (the lack of voicing targets for lenis stops in aspirating languages being a case in point). For example, if sibilant + plosive clusters are [0tense], they should have different effects on the F_0 of a following vowel than the corresponding singleton [+tense] and [-tense] plosives. Unfortunately, data on F_0 trajectories is inconclusive. Caisse (1982) and Ohde (1984) find similar pitch perturbations after fortis stops and [s] + stop clusters, which suggests they should be classified together. On the other hand, for 4 out of the 5 speakers investigated by Kingston & Diehl (1994), the points on the F_0 trajectories after sibilant + plosive clusters are roughly intermediate between the pitch values following singleton fortis and lenis stops, as predicted by the phonetic underspecification hypothesis.

It does seem clear, meanwhile, that not all cases of laryngeal neutralisation lead to phonetic underspecification, since there are several cases in which (lexical) neutralisation produces obstruents that appear to have voicing targets. The word-initial alveolar sibilant of German is realised as [z], despite the fact that it does not contrast with a voiceless fricative at the same place of articulation. A similar phenomenon occurs in eastern dialects of Dutch, whilst the standard variety has [z] in sibilant + [ʋ] clusters (e.g., [zwaːn], *swan*; [zʋeʲvən] *to hover, glide*; [zʋuːɹt], *rind*). An example from outside the Germanic group is Western Aleut, which has a single series of oral stops with word-initial VOTs ranging between 76 and 92 ms (Cho & Ladefoged , 1997), i.e., values well within the $>$ 35 ms bracket that is normally labelled as aspirated or long lag VOT.

It should be clear from the description of obstruent aerodynamics in chapter 2 that neither the voiced realisation of the German and Dutch fricatives nor the long lag VOT of Western Aleut plosives can be attributed to passive voicing. Assuming that neutralised sibilant fricatives are optimised for noise generation, which involves a wide glottal abduction, it is difficult for them to acquire any voicing by passive means, unless they are subject to lenition and occur intervocalically. Neither of these two conditions have been described as necessary for the Dutch and German sibilants in question to be pronounced as [z]. With regard to passively devoiced oral stops, aerodynamics dictates that voicing commences as soon as the two basic preconditions (closed vocal folds and sufficient subglottal pressure) are fulfilled. For a plosive-vowel sequence produced on a voiced carrier signal, the time lag from oral release within which these conditions are met is likely to be considerably shorter than the VOTs observed for Western Aleut (c.f. Westbury & Keating 1986; Stevens 1998). Consequently, German and Dutch (dialectal) word-initial [z] must be regarded as specified for phonetic voicing, whilst Western Aleut stops can only be analysed as actively devoiced (aspirated). Without further evidence it is impossible to tell whether these sounds share other phonetic properties normally associated with [-tense] and [+tense] respectively.

It is perhaps important to emphasise that evidence for the phonetic underspecification of neutralised obstruents does not strictly speaking preclude a fortition analysis of laryngeal neutralisation, which must consider phonological as well as phonetic data. A case in which phonological information must be brought to bear on the analysis is the lexical neutralisation of [tense] for velar stops in Dutch. Whilst Dutch opposes /p-b/ and /t-d/ there is only a single velar stop, which is usually transcribed [k]. A reason to represent this stop as lexically fortis across contexts might be that it groups with /p, t/ in the regular past tense paradigm: /raːk/ + /də/ yields [raːktə], *hit (a target)*, just as /raːp/ + /də/ surfaces as [raːptə], *gathered, was picking up*.

However, technically speaking, a fortition analysis of dynamic neutralisation can be maintained even in the absence of this sort of phonological information. In any model that conceives of the mapping between lexical forms and representations at the physical (auditory and articulatory) interface level in derivational terms, it is possible to say that laryngeal neutralisation is a process that converts [-tense] obstruents into their [+tense] counterparts. The distinct phonetic interpretation of neutralised obstruents can then be relegated to a separate mechanism that implements neutralised [+tense] obstruents differently from contrastively [+tense] obstruents. However, for this approach to work, the second mechanism has to operate in exactly the same set of environments as the fortition rule, and to a large degree therefore duplicates it. Consequently, in the absence of evidence that a given set of laryngeally neutralised and phonetically underspecified plosives and fricatives act as fortis obstruents in phonological processes, the phonetically opaque version of the fortition hypothesis cannot be justified.

3.2.2 Incomplete laryngeal neutralisation and its implications

Whilst the phonetic data uncovered by Hsu (1996) and Ernestus (2000) undermines the asymmetric aspect of the fortition analysis of (final) laryngeal neutralisation, its *categorical* nature has also been challenged on phonetic grounds. An ever-growing series of production and perception studies reveals that speakers and listeners are able to make subtle but statistically significant phonetic distinctions between fortis and lenis obstruents in neutralisation contexts, which suggests that laryngeal neutralisation is phonetically incomplete. Mitleb (1981), Port et al. (1981), O'Dell & Port (1983), Charles-Luce (1985), Port & Crawford (1989), Piroth et al. (1991) find evidence that laryngeal neutralisation in German leaves some residual cues to the lexical status of obstruents. Similar results have been reported for Catalan (Dinnsen & Charles-Luce, 1984; Charles-Luce, 1993), Polish (Slowiaczek & Dinnsen, 1985; Slowiazcek & Szymanska, 1989), Romanian (Steriade & Zhang, 2001), and Dutch (Ernestus & Baayen, 2003).

In contrast, Fourakis & Iverson (1984) and Kahlen-Halstenbach (1990), and Jassem & Richter (1989) find phonetically complete neutralisation for German and Polish respectively. In addition, experiments reported by Jongman et al. (1992), Baumann (1995), Ernestus (2000), and Kopkalli (1993) indicate that Dutch and Turkish final neutralisation erases all phonetic distinctions between word-final fortis and lenis obstruents. Finally, an effect of juncture strength on the degree of neutralisation emerges from acoustic data gathered by Piroth et al. (1991), who finds incomplete neutralisation at word boundaries in German, but complete neutralisation before word internal morpheme boundaries.

What is at issue in the incomplete neutralisation debate that was sparked by the earlier of these studies is not so much the (replicability of their) bare results, but rather the nature of phonetic knowledge and the experimental methodology that lends the best insight into that nature. For instance, Fourakis & Iverson (1984) claim that the incomplete neutralisation of final obstruents in German observed by Mitleb (1981), Port et al. (1981), and O'Dell & Port (1983), is due to a special mode of pronunciation that bypasses phonological rules used in normal communication and is directly driven by the orthography. They support this analysis with the results of two experiments, one of which examines the realisation of word-final obstruents in the responses to a reading task. Two out of the 4 subjects involved in this experiment, which is designed to replicate the effects found in the earlier studies, produced residual cues to the underlying distinction between fortis and lenis obstruents. However the same speakers did not distinguish underlying word-final fortis and lenis obstruents in the third person preterite indicative singular forms of strong verbs such as /laːd(ən)/ *(to) load* in a verb conjugation task that relied on orally presented stimuli. Thus, the pair of experiments conducted by Fourakis & Iverson (1984) appears to establish a connection between orthographically presented stimuli and the incomplete neutralisation effect.

Fourakis and Iverson use the observation that in the reading experiment 3 out of 4 their subjects distinguish the members of the minimal pair <Weck> *breakfast roll (dial.)* vs. <weg> 'away' (by means of vowel duration) as their trump card in this argument. Since <weg> never inflects, learners of German have no auditory evidence that its underlying form is distinct from that of <Weck> and the only route to a lenis pronunciation of the final velar is through the orthography. Note, however, that this argument only goes through on the assumption that orthographic information has no bearing on underlying phonological forms (and see Giegerich 1994 for evidence that this assumption is not necessarily valid).

Baumann (1995) echoes the idea that incomplete neutralisation reflects a special, spelling driven, mechanism that has little relevance outside the laboratory. She notes a number of further experimental factors that seem to reduce the theoretical import of incomplete neutralisation data. For instance, some studies (Port & O'Dell, 1985; Charles-Luce, 1985) use very low frequency words as stimuli, and many of the test subjects probably were fairly advanced speakers of English, which suggests the possibility of second language interference, especially given the fact that most of the studies listed above were carried out in laboratories in the United States. In addition, the use of minimal pairs may have revealed the purpose of the experiments to the subjects and thereby prompted the observed response patterns.

However, it is hard to find a systematic connection between the possible confounds listed by Baumann (1995) and the designs of the various studies reporting incomplete neutralisation effects. For instance, Catalan orthography does not represent the contrast between word-final lenis and fortis obstruents, yet Dinnsen & Charles-Luce (1984) and Charles-Luce (1993) find that the distinction between the two types may be incompletely neutralised, especially if the words acting as carriers are semantically nonredundant. The Dutch subjects used by Jongman et al. (1992) and Baumann (1995) were probably fairly proficient in one (English) and possibly in a second (French) non-neutralising language. Furthermore, Charles-Luce (1993) argues that in at least one respect the design used by Fourakis & Iverson (1984) was no less 'artificial' or biased than some of the experiments that established incomplete neutralisation: the conjugation task removes all semantic context, which is hardly representative of the average conversational situation.

As the causes of incomplete neutralisation effects remain controversial, there is no agreement about their analysis. One possible account, suggested by Dinnsen & Charles-Luce (1984) among others, effectively treats incomplete neutralisation as the result of gradient phonetic processes rather than a categorical phonological ones. According to this type of account, the final obstruent of, say, Dutch /krɑb/ would be [-tense] at both the lexical and surface phonological levels, just as its English cognate /kræb/. The Dutch plosive would then be converted into a series of phonetic targets that are very close to those for the final obstruent of /ɣrɑp/, whereas the phonetic interpretation of English word-final /b/ would maintain a much more perceptible phonetic contrast with /p/. This approach is perfectly feasible within a generative model, even if it may involve the admission that the phonetic (or 'scalar phonological') component of the grammar is larger than previously thought. From a functionalist perspective on the other hand, it raises the awkward questions why and how speakers maintain and acquire contrasts that appear to be so near the absolute thresholds of perceptibility.

An alternative approach, favoured by, e.g., Ernestus & Baayen (2003), views incomplete neutralisation effects as products of intraparadigmatic interference in speech production and perception. This approach is founded on a view of lexical organisation and lexical access that is different from the one adopted in much generative work. According to the latter, there is little redundancy in the lexicon, most roots are represented only once, and morphologically complex forms are derived on-the-fly from simplex forms. According to the former view, the mental lexicon stores both bare roots and many morphologically complex forms (e.g., Bybee 2001). The activation of a specific instance of a given root during lexical access is then assumed to co-activate (to a somewhat lesser extent) the phonological representation of related forms. Where there are phonological differences among the various manifestations of the root in question, this in turn leads to interference in the production and perception of specific instances of that root because all activated entries (can) feed into the production and perception systems.

On this second type of account, the residual phonetic contrast between the final labial stops of [kɹɑp] and [χɹɑp] would arise from interference from the full-blown phonetic contrast between the labial stops in the corresponding plural forms [kɹɑbə(n)] and [χɹɑpə(n)]. In other words, the pronunciation of the final labial stop in [kɹɑp] would tend to be slightly more [b]-like than that in [χɹɑp] because in producing the former, speakers would coactivate the lexical entry for [kɹɑbə(n)], whereas the final stop of the latter would be 'biased' by the /p/ of [χɹɑpə(n)].

Generative phonological theory can accommodate accounts of incomplete neutralisation that are based on intraparadigmatic interference, but only as grammar-external, 'performance' mechanisms, since it does not allow for notions such as (spreading) activation. It would be possible to say, for example, that at some abstract level speakers 'know' that the pairs [kɹɑp]-[kɹɑbə(n)] and [χɹɑp]-[χɹɑpən] each derive from a single underlying root, but that the way this knowledge is implemented in the minds/brains of speakers gives rise to intraparadigmatic interference and, consequently, incomplete neutralisation effects. However, this argument only holds as long as there is an independent argument for differentiating competence from performance; and the latter argument would seem to be weakened by the viability of non-modular models of the phonetics-phonology interface (cf. 1.3.2).

From a functionalist perspective, accounts of incomplete neutralisation that are based on interference from alternate representations are more attractive than an account based on phonetic rules that are acquired directly from auditory information. Under an interference account, speakers would not need to extract any subtle contrast between the labials in [kɹɑp] and [χɹɑp] from the signal in order to be able to produce it, at least as long as they do extract the contrast between the corresponding plural forms. Acquisition of the latter contrast would automatically trigger interference and hence incomplete neutralisation. Note however, that this line of reasoning does not in itself provide functional grounding of lexical redundancy and coactivation.

3.3 Obstruent feature asymmetries: voicing vs. aspirating languages

A cue-based theory predicts that the probabilities of neutralisation affecting two particular forms of laryngeal contrast differ if the sets of phonetic cues associated with each contrast differ in overall relative salience. For example, it might be that the phonetic difference between (actively devoiced) plain voiceless stops such as [p, t, c, k, q] and their ejective counterparts [p’, t’, c’, k’ q’] is less salient to listeners than the difference between the former series and their aspirated counterparts [p^{h}, t^{h}, c^{h}, k^{h}, q^{h}]. If this is indeed the case, a cue-based theory would predict that the voiceless-ejective contrast is more prone to neutralisation than the voiceless-aspiration contrast.

It is beyond doubt that asymmetries of this sort do not appear in categorical fashion: laryngeal contrasts of all phonetic types are more or less subject to (dynamic) neutralisation (e.g., Lombardi 1994). Meanwhile, there is no clear evidence, let alone statistically reliable generalisations, about different tendencies towards neutralisation for phonetically different types of contrast. There is some anecdotal data suggesting that closure voicing contrast is more robust in languages that cross-classify voicing and aspiration, such as Sanskrit, Bangla, or Khasi which are sometimes described as de*aspirating* before obstruents and/or word finally (cf. Kostic & Das 1972; Nagajara 1990; Steriade 1997), rather than de*voicing*. In other words the suggestion is that neutralisation leads to an opposition between e.g., [p, t, k] vs. [b, d, ɡ] rather than [p^{h}, t^{h}, k^{h}] and (say) [b̥^{ɦ}, d̥^{ɦ}, ɡ̊^{ɦ}]. However, recall from the previous section that Thai drops its prevoiced rather than aspirated series before sonorants, and given the lack of (typological) data on the production and perception of four-term systems mixing voicing and aspiration, it is utterly impossible to draw any conclusions from impressionistic and anecdotal data like this.

The only reliable generalisation that can be made about laryngeal neutralisation in the two types of fortis-lenis languages that are the focus of this study is that (dynamical) final neutralisation occurs frequently in both voicing and aspirating languages. The Germanic group itself exhibits the full paradigm of voicing languages with final neutralisation (standard Dutch, Frisian), aspirating languages with neutralisation (standard German), voicing languages without neutralisation (Yiddish), and aspirating languages without neutralisation (standard varieties of English, Norwegian, Swedish). Outside Germanic, dynamic final neutralisation is common in Slavonic (which is generally voicing) and Turkic (which is generally aspirating: c.f. Johanson & Csató 1998). Interestingly, incomplete final neutralisation effects have been found for both voicing (Polish: Slowiazcek & Szymanska 1989) and aspirating (German: Mitleb 1981 and later studies cited in 3.2.2) languages, and the same applies to experiments showing phonetically complete neutralisation (voicing Dutch vs. aspirating Turkish: Baumann 1995; Kopkalli 1993). A final indication that (the development of) final neutralisation is at least not highly sensitive to the voicing-aspirating distinction (or vice versa) is that voicing dialects of aspirating standard languages (Scottish English and Rhineland German) seem to follow the standard in terms of final neutralisation, and the same applies to aspirating dialects of voicing standards (north-eastern varieties of Dutch).

The safest approach to (final) laryngeal neutralisation phenomena is therefore to analyse them along the same lines as the ‘Germanic past tense paradigm’ discussed in chapter 4, by treating them as independent from the voicing-aspirating distinction. In the absence of data about the relative perceptual efficacy of voicing vs. aspirating fortis-lenis distinctions (or rather the phonetic contrasts between final stops: see chapter 2) there is no ground for establishing whether this assumption clashes with a cue-based model.

3.4 Obstruent feature asymmetries: plosives vs. fricatives

Much research on the marking of laryngeal contrast in obstruent inventories notes a relative scarcity of voiced obstruents. Maddieson (1984) observes that 291 out of 317 languages (91.8%) in the 1984 version of the UCLA Phonetic Segment Inventories Database (henceforth UPSID$_{317}$) have a plain voiceless stop series whereas 212 (66.9%) have a series of plain voiced stops, where a *series* is defined as the presence of at least one stop of a particular type in an inventory. An even stronger preference can be observed in the same database for voiceless over voiced fricatives and especially voiceless over voiced sibilants. Balise & Diehl (1994) find that 74.5% of all sibilants in UPSID$_{317}$ are voiceless as opposed to 61.9% of nonsibilant fricatives and 61.3% of stops (the overall figure for fricatives is 68.9%). They report that a similar generalisation emerges from the survey by Ruhlen (1975) of 706 languages.

In the light of a potentially critical lack of phonetic detail in UPSID$_{317}$ (see below in this section), it is probably safer to consider the frequencies of certain types of *contrast* instead of the frequencies of phonetic categories themselves. This approach yields an even stronger asymmetry in laryngeal marking between fricatives and stops. For example, the number of languages that have a contrast between unaffricated plosives labelled by UPSID$_{317}$ as 'voiced' and/or 'voiceless' and/or 'voiceless unaspirated', i.e., some sort of VOT distinction, in at least one of the labial, coronal (dental through retroflex, excluding postalveolars) or dorsal (palatal through uvular, including postalveolars) regions, is 236 (74.4%). By contrast, the number of languages that has a contrast between fricatives labelled as 'voiceless' and 'voiced' in one of three regions that can be referred to as *anterior nonsibilant*, *sibilant*, and *posterior nonsibilant* is 119. This amounts to 40.5 % of the languages that have fricatives and 37.5% of the total number of languages in UPSID$_{317}$. The number of languages that have at least 1 sibilant labelled as 'voiceless' and 1 sibilant that is labelled as 'voiced' is even smaller at 101 (34.4 and 31.9%). This suggests that laryngeal contrast (supported by voicing distinctions) is rarer or less stable in fricatives than in plain stops.[4]

In addition, there is evidence of a (thus far) more anecdotal nature for the relative instability of laryngeal contrast in fricative inventories. For example, whereas the plosive series of all Germanic languages are largely split by a fortis-lenis contrast, the North Germanic group lacks laryngeal distinctions between sibilants, and between fricative series altogether if weak (sonorant-like) sounds such as [v/ʋ, ʝ, ɣ] are excluded. A lexical constraint against the marking of laryngeal contrast in English fricative + sonorant clusters, discussed in section 3.5 below, also evinces the relatively weak capacity of fricatives to support (voicing-based) laryngeal contrast. A more restricted version of this constraint is active in Dutch, which has only [s] before /m, n, l/ and as mentioned in 3.2.1 above,

---

[4] For the purpose of these counts I used the label *anterior nonsibilant* for fricatives classified by UPSID$_{317}$ as 'nonsibilant' and bilabial through retroflex in terms of articulation place, but excluding postalveolar sounds; *posterior nonsibilants* refers to 'nonsibilant' postalveolar and palatal sounds through to uvulars. Note that there is evidence for the treatment (in UPSID$_{317}$ and here) of sibilancy and place of articulation as separate dimensions in perceptual phonetic space (Choo & Huckvale, 1998).

Pharyngeals, epiglottals and glottals were excluded from both plosive and fricative counts as were geminates and sounds marked with the superscripts *2* (unassimilated loans), *3* (posited underlying segment), *4* (segment possibly derivable from others), or *5* (particularly vague or contradictory description) in Maddieson (1984), but note that this has little effect on the resulting numbers and percentages. The languages in UPSID$_{317}$ (Burmese, Karen, Mazahua) that have aspirated fricatives invariably have 'voiceless' and 'voiced' fricatives at the same place of articulation, and the inclusion of this third phonetic category therefore has no bearing on the results. Finally, the definition of 'aspiration'/'voicing' contrast in terms of broad place of articulation used here is more restrictive than the definition implied by the criteria for a *series* in Maddieson (1984). According to the latter, an inventory with three plain plosives labelled as [b, t, k] (e.g., Seneca) would exhibit laryngeal contrast. Under the former definition this inventory does not show laryngeal contrast because it there is no pair of plosives with distinct voicing in at least one of the labial, coronal or dorsal regions.

[z] preceding /ʋ/.

Ohala (1983) and many others have sought to attribute the apparent bias against voicing distinctions in obstruent systems, and fricative inventories in particular, to the relative amount of articulatory effort involved in the production of actively voiced obstruents. This hypothesis recognises that, under circumstances defined in chapter 2, obstruent voicing requires some form of active enhancement whereas voicelessness can be realised without articulatory intervention. From this it is inferred that voiced obstruents are more costly in terms of articulatory effort, and in a functional model built on effort avoidance it follows that voiced obstruents tend to be rejected in favour of easier to produce voiceless obstruents. Taking this line of reasoning a step further, Ohala (1983) and Vallée et al. (2002) suggest that voiced fricatives are disfavoured more strongly than (pre)voiced plosives because the precise co-ordination between glottal and supraglottal constrictions that is critical in the production of the former requires more effort than articulation of the latter (cf. chapter 2).

Balise & Diehl (1994) on the other hand, propose that the typological bias against (voicing-supported) laryngeal contrast in fricative systems derives at least in part from the fact that the presence of voicing interferes with the perception of place cues in fricatives, at least for speakers of English. According to this theory, voicing-based laryngeal contrasts in fricative inventories tend to be avoided (or neutralised diachronically) because it is relatively hard to recover their place cues. It is supported by two types of observation: (1) the presence of voicing in a fricative reduces the amplitude of frication noise, which is an important carrier of place information; (2) studies of consonant confusions indicate that across various signal-to-noise ratios, voiceless fricatives are identified correctly more often than their voiced counterparts. In addition, Balise & Diehl (1994) report data suggesting that the latter effect is stronger for sibilants than for nonsibilants, which might account for the (slightly) greater tendency towards laryngeal neutralisation in sibilant inventories.[5]

The findings reported by Balise & Diehl (1994) and their (admittedly tentative) interpretation of those findings are highly interesting from the perspective of a cue-based theory of laryngeal neutralisation because they suggest that laryngeal neutralisation asymmetries related to adjacent sounds and the effects of phonetic features of the target obstruents themselves can be explained in terms of a single factor: relative perceptibility. In other words, the propensity for laryngeal neutralisation in pre-obstruent contexts (see section 3.5 below) and the observation that fricative inventories are prone to laryngeal neutralisation might be accounted for by the low perceptibility of lexical contrasts in both phonetic

---

[5] Interestingly, the perception data indicates that confusion along the place axis is far greater than along the voicing axis. For example, [v] is more likely to be confused with [ð] than with [f]. This observation would seem to argue against the proposal by Steriade (1992, 1993) that fricatives are less suited for the expression of laryngeal contrast.

‘contexts’. This goal seems unattainable to formalist models, which seem to be forced to rely on separate mechanisms to account for context effects (e.g., the Laryngeal Constraint in figure 3.1) and asymmetries between different obstruent types (i.e., if they deal with observations of the latter type at all: cf. chapter 8).

This is not to say that there is anything near a sufficiently fleshed out perceptual account of plosive-fricative asymmetries in laryngeal neutralisation. Such an account would have to be part of a broader theory of obstruent internal phonetic effects on laryngeal neutralisation, which would require a great amount of additional data. For instance, as indicated in the previous section, it is not clear whether laryngeal contrasts in stops with different phonetic expressions exhibit neutralisation asymmetries. Second, the low perceptibility of place contrast in the voiced fricative series highlighted by Balise & Diehl (1994) might account for neutralisation of voicing-assisted laryngeal contrast in fricative inventories, but it does not necessarily explain the higher incidence of neutralisation in fricative inventories than in stop inventories. It has been suggested that voicing affects the perception of place contrast in stops as well (e.g., Boersma 1998), and according to the theory described here that would mean that stop inventories would also gravitate towards an exclusively voiceless state. Thirdly, because it essentially treats laryngeal neutralisation of fricative oppositions as an epiphenomenon of place neutralisation in the voiced series, Balise and Diehl’s theory predicts that the process always yields a voiceless (actively devoiced) fricative. This prediction is contradicted by the voicing of word-initial alveolar sibilants in German and similar cases of neutralised fricative voicing (see sections 3.2.1 and 3.5).

The biggest caveat associated with both phonetic-substance based theories of neutralisation in obstruent inventories mentioned in this section, however, is the reliability of the typological generalisations they seek to account for, and especially of those relating to plosives. UPSID$_{317}$ and similar databases, as many of their sources, partially or wholly conflate the distinction between voicing as a phonetic feature and the phonological contrasts it is used to support. Thus, UPSID$_{317}$ reduces the four way phonetic voicing distinction recognised in this study among actively devoiced aspirated (fortis), actively devoiced unaspirated (fortis), passively voiced (lenis or neutralised) and actively voiced (lenis) to a three term taxonomy of voiceless aspirated, plain voiceless, and plain voiced, and does so in a way that leaves the true phonetic voicing of plosives partially opaque. For example, it represents the fortis-lenis distinction in both German and Bulgarian plosives as plain voiced vs. voiceless aspirated, even though it is clear that German is an aspirating language (Moulton 1962; Jessen 1998 and a host of references in the latter) and Bulgarian a voicing language (Ternes & Vladimirova-Buhtz, 1999). This means that the fact that both languages exhibit a VOT-based (fortis-lenis) distinction is correctly encoded, but the difference between German (word-initial) [b̥, d̥, g̊] and Bulgarian [b, d, ɡ(, bʲ, dʲ, ɡʲ)] is obscured, as is the phonetic distinction between German (word-initial, prestress) [pʰ, tʰ, kʰ] and Bulgarian [p, t, k(, pʲ, tʲ, kʲ)]. Consequently, generalisations drawn from UPSID$_{317}$ about the relative frequency of ‘voiced’ stops are in part about passively voiced lenis stops, and the total set of ‘voiceless aspirated’ stops includes at least some (actively devoiced) plain voiceless items.[6]

It is tempting to attach a greater amount of phonetic realism to the labels *voiceless fricative* and *voiced fricative* in UPSID$_{317}$. First, it seems that the phonetic typology of laryngeal contrast is much simpler in fricatives than in stops, and this constrains the space for error (i.e., given the set of labels used by UPSID$_{317}$ to mark laryngeal distinctions in fricatives). In Germanic, fortis-lenis distinctions between fricatives are always supported by voicing, regardless of the use of voicing/VOT to cue stops. Furthermore, the number of fricative inventories divided by more than 2 contrastive ‘laryngeal’ dimensions seems genuinely low in comparison to the number of stop inventories with this property. This may well be another indication of the relatively restricted phonetic means of expressing laryngeal contrast in fricatives: note that the continuous high airflow across an oral constriction required for the production of fricative noise puts inherent limitations on the number of laryngeal actions and configurations that are available. In other words, it seems likely that in many instances where UPSID$_{317}$ represents a fricative as voiced, it is indeed produced with some amount of active voicing. I have certainly not found any glaring problems of the German/Bulgarian type mentioned in the previous paragraph. However, the possibility remains that some of the fricatives represented by UPSID$_{317}$ as ‘voiced’ or ‘voiceless’ should be reanalysed (e.g., as breathy voiced or voiceless depressor fricatives: cf. Downing & Gick 2001 and chapter 2). Consequently, any generalisations drawn from UPSID$_{317}$ about the behaviour of voiced or voiceless fricatives should be treated with caution, if not the same amount of caution that should be observed when dealing with generalisations about VOT/voicing in stops.

I strongly suspect that the lack of phonetic discrimination with respect to obstruent voicing in UPSID$_{317}$ is related to the ambivalence of terms such as *voicing* and *aspiration* in phonetic descriptions. Regardless of their ultimate source, the phonetic approximations in UPSID$_{317}$ and elsewhere are sometimes insufficient to test theories grounded in phonetic substance. For instance the effort-based theory proposed by Ohala (1983) and others makes predictions about the phonetic voicing of preferred obstruents and can therefore not be tested against a database that occasionally glosses over the difference between actively voiced

---

[6] Interestingly, a more recent edition of UPSID containing 451 languages (UPSID$_{451}$) reclassifies the fortis stops of both languages as plain voiceless, whilst maintaining a voiced-voiceless aspirated system for Norwegian.

and passively voiced (lenis) obstruents, or the distinction between plain voiceless and voiceless aspirated obstruents. The difference between the passively voiced word-initial [b̥, d̥, g̊] of German and Bulgarian [b, d, ɡ(, bʲ, dʲ, ɡʲ)] is critical to the effort-based theory since it predicts that the former but not the latter are disfavoured by languages. The distinction between passively voiced and *actively devoiced* obstruents is equally critical under an effort-based theory, which predicts that, as far as voicing is concerned, the passively voiced lenis stops of aspirating languages are preferred over the actively devoiced stops of voicing languages. Similarly, the theory espoused by Balise & Diehl (1994) relies crucially on the presence of phonetic voicing, not on other (clusters of) phonetic features or structural properties. However, since distinctions between active and passive (de)voicing are not represented (consistently) in UPSID$_{317}$, it does not allow theories based on phonetic voicing to be tested. Finally note that, conversely, the significance of the high frequency of ‘plain voiceless obstruents’ in UPSID$_{317}$ is severely reduced by the lack of phonetic specificity of the label.[7]

3.5 Context asymmetries: right and left-adjacent sounds

The behaviour of laryngeal contrast in consonant clusters, and in particular in obstruent + sonorant clusters, provides the critical data for comparing syllabic theories and cue-based accounts of laryngeal neutralisation (strictly word-based theories have nothing to say about word-initial and medial clusters). The former predict that phonetically similar sequences exhibit differences in neutralisation if they are syllabified differently. For instance, for C_1C_2 sequences where C_1 is an obstruent and C_2 a sonorant, syllable-driven theories predict that laryngeal neutralisation is much more likely to affect the obstruent if C_1 and C_2 are heterosyllabic than if they group together in the same syllable. Linear models

---

[7] Westbury & Keating (1986) list a number of additional gaps and indeterminacies in the available data that make it hard to test effort-based models of laryngeal neutralisation. The assignment of a very specific physical (or perceptual) interpretation to phonetic symbols in the absence of instrumental data is problematic for substance-based theories more generally. For example, the common use of the symbols [ʌ] and [ɔ] to refer to the vowels in (American) English <done> and <dawn> might be interpreted by assigning these vowels a similar acoustic or perceptual vowel height (F_1) and backness (F_2 or F_2-F_1), because the symbols are paired as open-mid and back in the IPA (cardinal) vowel chart. Schwartz et al. (1997a,b) essentially generalise this procedure to the set of vowel symbols used in UPSID$_{317}$. Yet acoustic data indicates that the practical use of IPA symbols provides only very rough approximations of the orientation of vowels in F_1-F_2 space. Measurements by Peterson & Barney (1952) show that American English [ʌ] is both considerably lower and fronter (has higher values for both F_1 and F_2) than [ɔ]: in terms of height it groups with [æ] and [ɑ] rather than with [ɔ]. Therefore, evidence from UPSID$_{317}$ or similar sources for dispersion or symmetry (Boersma 1998: chapter 16) in vowel inventories, can only be interpreted in relatively global or abstract (phonological) fashion.

on the other hand predict that, all else (e.g., position in the word) being equal, there should be no such difference. This section does not offer the sort of data survey that would allow anything near a definitive conclusion in this matter. It does argue that the available evidence favours linear and cue-based rather than syllabic accounts, specifically because laryngeal distinctions may be maintained in heterosyllabic obstruent-sonorant sequences whilst they may be suspended in tautosyllabic clusters.

There can be little doubt that laryngeal neutralisation is common in obstruent clusters. Examining a sample of 104 languages, Greenberg (1978) found a strong preference for clusters that are homogeneous in terms of “voicing”, which might be treated as an indication that there is a typological tendency not to allow marking of laryngeal contrast on individual members of obstruent clusters. There is an additional preference for such clusters to be phonetically voiceless.[8] The Germanic languages fit this typological pattern by suspending the marking of laryngeal contrast on most obstruent clusters in monomorphemic forms. For instance, with the exception of Yiddish (Birnbaum, 1979) and to a lesser extent English, Germanic does not seem to allow laryngeal contrast between the individual members of tautomorphemic clusters (oppositions between e.g., /bt/ and /pt/ or /pt/ and /pd/) and bars all contrast between word-initial sibilant + obstruent clusters. The exceptions to the ban on cluster-internal contrast in English involve words with the Latinate prefix <ab> (<absent, absurd, abstract>) and a few other forms (e.g., <magpie>) which seem likely to have been reanalysed as monomorphemic, as well as clusters that might be argued to stem from synchronic vowel elision (notably <medicine>). Zonneveld (1983) shows how in Dutch, laryngeal contrast between clusters is extremely limited word medially and finally, where the overwhelming majority of obstruent sequences is phonetically voiceless. Exceptions such as /maːɣd/ ’virgin’ vs. /mɑxt/ *power, might*, and /vɔxt/ ’moisture’ vs. /voːɣd/, *guardian, custodian*, are far and few between. This generalisation extends to German (Brockhaus, 1995; Wiese, 1996), Norwegian (Kristoffersen, 2000), and to a lesser extent English.

In addition, there is at least one case in which laryngeal neutralisation in obstruent sequences conforms to the predictions of a strict ‘implicational’ interpretation of the hierarchy in (5): it appears that languages with word-final, and therefore utterance-final (prepausal), neutralisation of singleton obstruents never retain laryngeal contrast in the prefinal positions of word-final obstruent sequences.[9] As illustrated in (6), the lexical distinctions between the final obstruents of the Dutch stems in the leftmost columns of (4) are neutralised not only utterance finally but also before participial -/d/: /ɣə/ + /krɑb/ + /d/ *(has)*

---

[8] Unfortunately the data presented by Greenberg (1978) do not allow for any robust generalisations about the marking of laryngeal contrast in obstruent clusters vis-à-vis other contexts.

[9] On the treatment of *word-final* contexts in a cue-based theory, see section 3.6 below.

scratched, and /ɣə/ + /ɣrɑp/ + /d/ *(has) joked*, surface with phonetically identical [pt] clusters rather than as [χəkrɑbt], [χəχrɑpt]. Similar generalisations apply to German and Frisian as well as to Lithuanian (Steriade, 1997). Whilst the presence of the pattern in (4) thus appears to imply the pattern in (6) below, the Norwegian 'regressive devoicing' data discussed in chapter 4 strongly suggests that the reverse is not true: underlying /tryg/ *secure* surfaces as [tɾy kt] when suffixed with /t/, but retains its lenis features in unsuffixed forms: [tɾyg].[10]

(6) Neutralisation in Dutch participles

UR	Participle	Gloss
/xrɑp/	[χəχrɑpt]	joke
/krɑb/	[χəkrɑpt]	scratch (an itch)
/ɣraːz/	[χəχraːst]	graze (of animals)
/krɑs/	[χəkrɑst]	scrape, scratch (a smooth surface)

Note that whereas the implicational relationship between the patterns in (4) and (6) follows automatically from (a specific interpretation of) the hierarchy in (5), any (word-based) formalist model trying to account for the pattern in (6) using an assimilation rule would have to stipulate that this rule is always present in languages with final neutralisation. However, the behaviour of obstruent sequences does not critically distinguish cue-based theories from *syllabic* accounts, since both approaches predict that prefinal positions in such clusters are prime neutralisation sites. Laryngeal contrast in obstruents is relatively difficult to perceive before another obstruent because it removes the carrier for a some important cues, notably formant perturbations and positive VOT distinctions. Depending on the amount of coarticulation between obstruents in a given language, a following obstruent may also partially or wholly obscure the release bursts of plosives (Henderson & Repp, 1982) and the constriction duration of both plosives and obstruents. In languages where this is the case, laryngeal contrast would be even harder to perceive than in languages that clearly delimit obstruents in sequence. Syllabic accounts predict that prefinal positions in obstruent sequences are neutralisation targets because they are in a coda position or otherwise structurally different from the sort of onsets that typically preserve laryngeal contrast. This applies to word-initial obstruent clusters as well as to medial and final sequences since they are often treated as heterosyllabic (Kaye, 1992; Harris, 1994) or (mainly because they violate sonority sequencing) otherwise different from normal complex onsets (e.g., obstruent + sonorant sequences: cf. Blevins 1995). Thus, the fact that speakers of Dutch pronounce the acronym <ABVA> as [ɑpfa] (e.g., Booij 1995) does not necessarily mean that laryngeal neutralisation in this language is constrained by syllabic structure.

---

[10] Both the Dutch and the Norwegian patterns are assumed to be neutralising here.

Moreover, cue-based and syllabic theories make similar predictions for laryngeal marking in the final positions of obstruent sequences and the behaviour as clusters of a whole. Virtually all syllabification algorithms parse the final elements of obstruent clusters and the corresponding singleton obstruents (e.g., the alveolar plosives in English <top> and <stop>) identically and so it follows that they should behave identically in terms of laryngeal marking. Similarly, if sequences of obstruents as a whole are allowed to inherit the properties of their final members (cf. the account of the Dutch past tense rule presented by Lombardi 1994), e.g., English /steɪt/ should be as likely to contrast with /zdeɪt/ as /deɪt/ with /teɪt/ (i.e., <Tate>, proper name). In cue-based models, cluster-final obstruents would have the benefit of the relatively salient offset cues, whilst the marking of contrast at the level of clusters as a whole could in principle involve the same set of cues that is involved in the realisation of [tense] in singleton obstruents. Consequently, both syllabic and cue-based models would seem to need extra apparatus to distinguish clusters as a whole and final elements on the one hand from prefinal positions.[11]

All this means that the key set of cases that allows syllabic and cue-based theories to be tested against each other, consists of obstruent + sonorant clusters, which can have different parses, depending on the consonants and language involved. Syllabic accounts predict that if the obstruent and sonorant in such sequences group together as a complex onset (or are assigned some equivalent structure), the obstruent is less likely to be targeted by laryngeal neutralisation than if the sequence is split by a syllable boundary, i.e., if the obstruent is in a coda position or part of some equivalent structure. For example, syllabic accounts predict that neutralisation is more likely to affect the contrast between the medial stops in English /mægnəm/, /hæknɪ/, /kædmɪəm/, /pʌtnɪ/ (area of London), /bræd lɪ/ (proper name), [mɒtlɪ], than the medial stops in e.g., /koʊbrə/, /koʊprə/, /pɛtrl̩/, /skwɒdrn̩/, /proʊgræm/, /ækrɪd/. Since plosive + /r/ clusters are allowed word initially in English they are often treated as (at least potentially) tautosyllabic wherever (else) they occur, whilst plosive + nasal clusters do not occur word initially and are therefore generally treated as heterosyllabic where they occur word medially. By contrast, cue-based accounts would predict that to the extent that the cues to [±tense] are equally salient in both contexts, medial plosive + /r/ and plosive + nasal sequences are equally likely targets for laryngeal neutralisation.

Steriade (1997) makes two sets of observations that contradict the predictions of syllable-based approaches. First, in languages with word-final laryngeal neutralisation arguably heterosyllabic obstruent + sonorant clusters may preserve laryngeal contrast. Lithuanian is one of her main examples. This lan-

---

[11] On the problems posed by the Dutch regular past tense paradigm for syllable-driven models, and, to a lesser extent, cue-based approaches, see section 3.7 below

guage suspends all laryngeal contrast word finally, but preserves it word medially before obstruents, witness forms such as [silpnas], *weak* vs. [skobnis], *table*. Word-initial labial plosive + nasal sequences do not occur in Lithuanian, which is an argument for treating them as heterosyllabic. Likewise, some German speakers pronounce the plosives in words like <Adler>, *eagle*, and <ordnen>, *to put in order*, as [aːdlər] and [ɔrdnən], thus maintaining a contrast with the fortis plosives in words such as <partner>, [partnər], *partner* (see Brockhaus 1995 for extensive discussion of the small set of words capable of maintaining this contrast). Note that the speakers who use these forms do neutralise the opposition between fortis and lenis contrasts word finally. Alveolar stop + nasal/lateral clusters do not occur in German other than in medial position, and so again there seems little reason to treat them as tautosyllabic.

Second, laryngeal neutralisation may occur in obstruent + sonorant clusters that are best analysed as tautosyllabic. The behaviour of English fricative + sonorant clusters exemplifies this phenomenon. Whereas /f,v/, /s, z/ and marginally /θ, ð/ and /ʃ, ʒ/, contrast word initially before vowels, a single series of voiceless fricatives occurs before /r, l, w, j/: cf. (7a). Plosives on the other hand, retain the fortis-lenis distinction in this environment (7b). Word medially before sonorants, the contrast between fortis and lenis fricatives is at best limited. A near-minimal pair like <chevron>-<saffron>, [ʃɛvɹən], [sæfɹən] establishes a contrast for the labiodental place of articulation but a similar pair for the opposition /s/-/z/ is hard to find with alveolar sibilants showing a marked tendency to voicing before medial sonorants (7c). On the other hand, only voiceless dentals and postalveolars occur word medially before a sonorant. As the obstruent clusters illustrated here can occur word initially, there seems little reason to treat the laryngeal neutralisation of English fricatives before sonorants as syllable-final.[12]

There is hardly any more reason to treat the English data as exceptional or a phenomenon restricted to fricatives. Steriade (1997) lists several languages which maintain laryngeal oppositions word initially between prevocalic stops but not between stops followed by a sonorant consonant, including the Mon-Khmer languages Pacoh and Sre. Thai maintains a three term opposition between prevoiced, short lag and long lag stops prevocalically but only contrasts short lag and long lag stops before /r, l, w/ (Noss, 1964).

It is perhaps perhaps ironic that German, the textbook example of a language with 'final devoicing' should provide a further instance of this pattern. A number of High German dialects maintains a contrast between tense and lax plosives prevocalically, but suspends it before (tautosyllabic) liquids: interestingly,

---

[12] In American (as well as several other) varieties of English alveolar plosives followed by a high back rounded vowel have no intervening palatal glide. Coronal fricative + high back vowel sequences lack a palatal glide even for many British speakers, but they are included for completeness. On the realisation of word-medial fricatives before obstruents, see chapter 7.

the single series of plosives that occur in this context are generally described as 'voiceless lenis' (i.e, voiceless unaspirated). The phenomenon is known as *binnenhochdeutsche Konsonantenschwächung* in German dialectology and seems fairly well-documented. The Darmstadt (Rhenish Franconian) dialect described by Keller (1961) is a good example. In this dialect, Middle High German /p/ and /b/ have merged before sonorants to yield pronunciations such as [b̥]*latz*, *square* (Standard German [p^{h}l̥ats]), and *A*[b̥]*ril*, *April* (Keller 1961:171).[13]

(7) Laryngeal contrast in (British) English obstruent + sonorant clusters (Jones, 1977; Wells, 2000)

a. Suspension of contrast in word-initial fricative + sonorant fricatives

Orthography	Pronunciation	Orthography	Pronunciation
thrive	[θɹaɪv]	thwart	[θwɔːt]
fright	[fɹaɪt]	swan	[swɒn]
flight	[flaɪt]	Thule	[θjuːlɪ]
slight	[slaɪt]	suit	[sjuːt]
refuse (v.)	[ɹəfjuːz]		

b. Contrast is retained in word-initial plosive + sonorant clusters

Orthography	Pronunciation	Orthography	Pronunciation
plight	[pl̥aɪt]	blight	[b̥laɪt]
try	[tɹ̥aɪ]	dry	[d̥ɹaɪ]
Punic	[p^{h}juːnɪk]	bugle	[b̥juːgl̩]
twelve	[tw̥ɛlv]	dwell	[d̥wɛl]
cute	[k^{h}juːt]	gules (red)	[g̊juːlz]
tune	[t^{h}juːn]	dune	[d̥juːn]

c. Realisation of medial alveolar sibilants followed by a sonorant

Orthography	Pronunciation	Orthography	Pronunciation
osmosis	[ɒzmoʊsɪs]	Bosnia	[b̥ɑznɪə]
Oslo	[ɒsloʊ][ɒzloʊ]	gosling	[g̊ɒzlɪŋ]
Israel	[ɪzrɛl]	Bosworth	[b̥ɒzwɜθ]

In sum, the first set of observations establishes that the occurrence of word-final laryngeal neutralisation does not entail neutralisation in environments that can be regarded as syllable-final or non-syllable-initial, whilst the second set of observations shows that laryngeal neutralisation may occur before sonorants irrespective of syllabic structure. Moreover, the Lithuanian and English fricative data establish the double dissociation of word-final and pre-sonorant laryngeal neutralisation: in the former language word-final neutralisation exists in the absence of neutralisation before sonorants, whilst the latter exhibits pre-sonorant neutralisation but maintains the contrast between fortis and lenis fricatives word

---

[13] Thanks to Wiebke Brockhaus for pointing me to the literature on this topic.

finally.[14] Since the two phenomena can occur independently of each other, their cooccurrence is predicted as one of 4 possible patterns and thereby ceases to be a convincing argument for a syllabic analysis of neutralisation. In other words, the observations that Dutch suspends laryngeal contrast both word finally and word medially before nasals (the medial stops in orthographic <partner>, *partner* and <ordner> appear to be pronounced identically) can be plausibly construed as the (chance) coocurrence of two independent processes rather than reflections of the same process.

3.6 The word-initial vs. word-final asymmetry

Word-initial and word-final contexts have such different propensities for laryngeal neutralisation that the observation is rarely made in these terms, Westbury & Keating (1986) being a notable exception. Usually it is simply implied by statements about (syllable-)final neutralisation, and illustrated with paradigms such as (4) Yet it remains a striking observation that many, if not most, of the languages that neutralise laryngeal contrast word finally, preserve it word initially as well as medially between vowels. Initial neutralisation occurs trivially in languages that lack laryngeal contrast altogether, but it appears to be rare in languages that maintain a contrast elsewhere. Steriade (1997) even seems to claim that word-initial laryngeal neutralisation invariably implies final neutralisation. Westbury & Keating (1986) list Cuna, Efik, Ewondo, and Tamil as languages that lack laryngeal contrast word initially (and finally) but retain it medially.[15] Steriade (1997) adds Lac Simon, an Algonquian language, and Totontepec, which belong to the Mixtecan group. I can add the fricatives of Frisian, which support laryngeal contrast medially between sonorants and vowels, but not word initially or finally (Tiersma, 1985). It would seem therefore, that laryngeal contrast is only marginally less stable word initially than medially (between sonorants).

By contrast, the list of languages that maintain some form of contrast initially (and medially), but not finally remains considerably longer, and although a typological survey of laryngeal neutralisation phenomena is long overdue (cf. Brockhaus 1995), this is an indication that laryngeal contrast is more stable word

---

[14] This dissociation contradicts the strict implicational interpretation of the perceptibility hierarchy in (5) which predicts that laryngeal contrast is always more stable before sonorants than utterance finally.

[15] Tamil is somewhat problematic as an example of this phenomenon. Whether or not it has contrasts initially depends on what status is assigned to the voiced stops it has borrowed from Sanskrit. Similarly the status of the medial contrast referred to by Westbury & Keating (1986) rests on the synchronic analysis of medial stop gemination and voicing (by Caldwell's Law). Thus, another author (Steever 1990:239) is able to state that in Tamil "[v]oiced stops contrast with voiceless stops only in initial position. . . "

initially than finally. This set of languages includes Dutch, Frisian (plosives), and German within Germanic, many of the Slavonic languages, including Bulgarian, Polish, Russian, Slovak (Rubach, 1993), Lithuanian, many of the Turkic languages, including Turkish and Turkmen (Johanson & Csató, 1998; Clark, 1998), Thai, Vietnamese, Zoque, and Basque (Westbury & Keating, 1986).

As pointed out above, syllable-driven models of neutralisation account for the initial-final asymmetry by virtue of the fact that syllabification algorithms typically parse word initial (presonorant) obstruents as onsets and word-final obstruents as codas (or in some other way different from prevocalic obstruents). Other formalist models treat the phenomenon in terms of morphological or suprasyllabic prosodic structure (Trommelen & Zonneveld, 1979; Rubach, 1993; Brockhaus, 1995). I have referred to both of the latter as *word-based* since, if prosodic domains are involved, they are usually of the kind that interacts with word-level (analytical, non-cohering) morphology. Thus, the neutralisation of the lexical contrast between the final stops in Dutch /roːd/ *red*, and /vərɣeːt/ *forget*, before the suffix /ɑxtəɣ/ *-ish*, *-like* (cf. [roːʷtɑχtəχ] and [vəɹχeʲːtɑχtəχ]) can be attributed to morphological structure which indicates that a word (level) boundary separates the suffix from its hosts, or to the fact that host and suffix are parsed as independent prosodic domains. Instances of neutralisation in word internal contexts are clearly outside the scope of either of the two word-based approaches, unless morphological or prosodic structure is used in a wildly diacritic fashion. This is especially true of instances of neutralisation before sonorant consonants, as they cannot be analysed as (static) assimilation of [tense]

For cue-based models on the other hand, word-based asymmetries might seem to be a problem. It is impossible to attribute strong tendencies for laryngeal contrasts to be suspended word finally but to be preserved initially to the influence of adjacent sounds. Unlike *utterance-final* (prepausal) or *utterance-initial* (postpausal), *word-initial* and *word-final* are not phonetic contexts in the way that _[l] is a phonetic context or _[+son] a range of phonetic contexts. Word-final obstruents can precede any sound that is found word initially, whilst their word-initial counterparts can be preceded by anything found word finally, yet word final laryngeal neutralisation tends to behave uniformly across right-hand phonetic contexts, and the preservation of word-initial laryngeal contrast rarely if ever depends on the nature of the preceding sound.[16]

Steriade (1997) concludes that the initial-final asymmetry can not be driven by phonetic context alone. She therefore adopts a *Paradigm Uniformity* constraint to account for the uniform behaviour of word-final neutralisation across contexts, and a positional faithfulness constraint to shield initial contrasts from

---

[16] Especially under the assumption that Dutch post-obstruent (fricative) devoicing and a similar process noted in Basque by Hualde (1991) are phonetic devoicing rather than phonological neutralisation processes.

neutralisation. Paradigm Uniformity (Kenstowicz, 1995) is a family of output-output faithfulness constraints demanding that all the phonological or phonetic forms of a morpheme across a certain paradigm should be identical to the form of an 'attractor' morpheme that occurs at a single designated point in the paradigm. Constraints of this type are used in optimality-theoretic models of paradigmatic levelling and other 'morpheme invariance' effects. The formal property that sets Paradigm Uniformity apart from most other forms of output-output faithfulness is that it evaluates *n-tuplets* instead of *pairs* of forms, where *n* equals the number of positions in the paradigm.

Paradigm Uniformity is brought to bear on the analysis of final neutralisation by extending the notion of *paradigm* to the full set of possible phonetic contexts in which a morpheme can occur, and by employing citation forms of words as attractors. Obstruents at the end of citation forms, i.e., utterance-final obstruents, have a consistent right-hand phonetic context (silence), which might be argued to have relatively poor cueing potential. Although the number of available cues is fairly large, consisting of both onset cues (preceding vowel duration, V-to-C formant transitions, timing and nature of voicing offset), internal cues (closure duration), and some offset cues (amplitude and duration of release bursts), Steriade claims that the absence of several highly salient offset cues (timing and nature of voicing onset C-to-V formant transitions) nevertheless ranks V__‖ contexts (universally) below all __[+son] environments. Consequently, citation form-final obstruents are comparatively vulnerable to final neutralisation, and their attractor status generalises the neutralisation resulting from this vulnerability to all word-final obstruents, some of which occur in contexts with better cueing potential (i.e., before sonorants).[17]

The *Paradigm Uniformity (right edge)* constraint employed by Steriade (1997) is relativised to apply to word-final contexts only so as to exempt word-initial laryngeal distinctions from the effects of cueing limitations on citation form initial (postpausal) obstruents. Nevertheless, initial contrasts need an extra boost to attain uniform and approximately equal stability as word internal intersonorant distinctions: the two environments in question both support the highly salient offset cues as well as internal cues, but the latter adds a stable set of onset cues, which is more or less unavailable to the latter, depending on the preceding context. If neutralisation is derived from cueing potential alone it would follow that the marking of initial laryngeal distinctions is left-hand context sensitive, and less stable than medially between sonorants to the degree that

---

[17] A detailed review of the mechanics of the OT model constructed by Steriade (1997) is of no concern here, but note that it is inessential to designate the citation form as paradigmatic attractor: (high-ranked) *Paradigm Uniformity (right edge)* filters out every nonuniform paradigm regardless of which context is granted attractor status. This is fortunate since there is no independent evidence to motivate a psychologically special status for citation forms as opposed to, say, other relatively hyperarticulated forms of a word.

they are less perceptible. This is not borne out by the available data, and Steriade (1997) obtains uniformity and stability in initial environments by means of a positional faithfulness constraint *Preserve [voice] in #__* which demands that laryngeal distinctions be preserved word initially. *Preserve [voice] in #__* is highly similar to positional faithfulness constraints on the marking of laryngeal contrast employed by Grijzenhout & Krämer (1998) and Lombardi (1999) (see chapter 8)

The adoption of this constraint highlights both the problems that Steriade's model as it stands suffers from and, to my mind at least, its true potential. The principal problem is that the introduction of *Preserve [voice] in #__* means that the uniformity in neutralisation in word-initial and word-final contexts is accounted for in terms of two separate, formally and functionally unrelated, devices. The good news is that in principle it is possible to unify the analysis of word edge effects on neutralisation as a cue-based phenomenon. A crucial assumption underpinning Steriade's model is that the perceptibility of a contrast is solely a function of the sounds adjacent to the carrier, but recall from the discussion of neutralisation in chapter 1 that prosody is also likely to be an important influence on relative perceptibility. Recall too, that a wealth of evidence, in particular from work on articulatory strengthening, indicates that prosody has markedly asymmetric effects at the word level, 'strengthening' word-initial sounds (even in the absence of lexical stress) and weakening (unstressed) segments elsewhere, including word finally. Crucially, although it may affect word sandhi phenomena, prosody itself does not vary with the left-hand and right-hand phonetic context of a word: word-initial segments are strengthened and word-final ones weakened whether they are preceded or followed by an obstruent, sonorant or physical pause.

Now if prosodic strengthening and weakening indeed affect the perceptibility of contrast, this would entail a decrease in the perceptibility of word-final laryngeal distinctions *across phonetic contexts provided by adjacent sounds*. In other words, after the effects of prosody are factored in, the [tense] cues of a word-final (or unstressed medial) obstruent preceding a sonorant would be less perceptible than the [tense] cues of the same obstruent preceding the same sonorant in word-initial (or stressed) position. Under a cue-based theory of neutralisation it follows that laryngeal distinctions are less viable word finally than word initially, again across contexts provided by adjacent sounds. Another way of phrasing this is to say that prosody creates a partially uniform phonetic context at word edges and can therefore be expected to act as a leveller of perceptibility differences and hence neutralisation asymmetries arising from other factors.[18]

---

[18] J. Beckman (1997) suggests that positional faithfulness constraints are ultimately grounded in phonetic factors such as perceptibility. A possible additional source of uniformity in neutralisation at word edges is that the (temporal) organisation of laryngeal cues is generally optimised to work

Whereas the formalist prosodic approaches referred to at the outset of this section claim that neutralisation is directly driven by prosodic phrasing, the cue-based alternative proposed here holds that prosody is simply one of the factors influencing relative perceptibility and affects neutralisation via the same single mechanism that is responsible for neutralisation asymmetries triggered by adjacent sounds and internal phonetic features (e.g., the plosive fricative-asymmetry). In 'diachronic' functional models this single mechanism is misperception by learners, in 'synchronic' models, a ban on contrasts with a perceptual salience below a specified level. There is one sense in which the cue-driven and some formalist theories converge: to the degree that prosody is indeed a system for signalling grammatical boundaries, the former reconstructs the idea espoused by Kaye (1989) and Harris (1994) that phonological rules exist to assist in word recognition and parsing. But since neither formalist word-based models nor the account by Steriade (1997) seem able to unify the analysis of neutralisation patterns triggered by adjacent sounds, internal features and prosody, a fully cue-based approach represents a considerable improvement, at least in principle.

A considerable amount of new phonetic data is needed, first to establish whether a word-final vs. word-initial neutralisation asymmetry can indeed be predicted on the basis of perceptibility, and second to test whether prosodic effects on perceptibility indeed map into neutralisation asymmetries. At least impressionistically, flapping of alveolar stops in (American) English greatly reduces the amount of perceptual contrast between /t/ and /d/ (so much so that it has been treated as a neutralising process) and this might be counted as preliminary evidence that processes directly driven by (or originating in) prosodic weakening are able to reduce the perceptibility of phonological contrasts. Note, incidentally, that the dialects in question lenite word-final alveolar stops across right-hand phonetic contexts, even where this does not lead to flapping (Harris, 1994). Furthermore, it is not inconceivable that (if they are indeed not artefacts of 'spelling pronunciation') the incomplete neutralisation phenomena discussed in 3.2.2 above reflect final weakening, perhaps reinforced by some other mechanism. However, the relative perceptibility of laryngeal contrast in obstruents needs to be examined more systematically, and preferably on the basis of mate-

---

within, rather than across word boundaries. As pointed out in chapter 2, the observation that English fortis plosives are not aspirated (but rather preglottalised) word finally even when they occur before a sonorant, can be interpreted in these terms. From a functional perspective, this type of paradigmatic uniformity in cue organisation might be advantageous in offering a robust cue to laryngeal distinctions across phonetic contexts and as an aid in word segmentation. Note that there is an important and testable difference between this hypothesis and the use of *Paradigm Uniformity* by Steriade (1997). The former predicts paradigm uniformity at the phonetic level in languages that *maintain* laryngeal contrast word finally, and (as a result) at the phonological (neutralisation) level in languages that suspend it. Steriade's model on the other hand proposes that laryngeal *neutralisation* instead of cue organisation is generalised across the paradigm, and thus predicts that uniformity effects only occur at the phonological level.

rial that does not exhibit highly conspicuous lenition processes.

As far as the interaction between prosody and laryngeal neutralisation is concerned, there does not seem to be a large amount of data either, and much of it relates to laryngeal segments rather than laryngeally marked obstruents. The fact that in English, /h/ only occurs prevocalically in word-initial and stressed syllables is well-known (cf. Harris 1994). I am not aware of any examples in which the fortis-lenis distinction is neutralised outside these contexts other than the Danish pattern where originally lenis stops spirantise to leave a single series of stops behind (which is very common, and may in fact explain the absence of fortis-lenis mergers). Prosody-driven neutralisation in nonfinal contexts certainly does not figure in the survey article of Mascaró & Wetzels (2001). However, on the basis of a survey of languages spoken across Europe, north and central Asia, Butskhrikidze (1998) claims that the occurrence of final neutralisation is highly constrained by lexical stress. Her *Universal 2.1* reads "[i]f the fixed accent is on the final syllable, devoicing of the final obstruents does not occur". Because there are languages that, by most accounts, have final stress and final neutralisation, e.g., several Turkic languages, this would seem to be a 'soft' universal expressing a (statistically reliable) tendency, or a statement about the preconditions for the development of final neutralisation. Unfortunately, as the body of her work has not been translated from the Georgian original, I am currently not in a position to evaluate the data underpinning Butskhrikidze's claims.

3.7 Summary and remaining issues

This chapter presented a number of generalisations about laryngeal neutralisation that form a necessary backdrop for some of the experimental work presented in the next two chapters and especially for chapter 8. In addition, the general thrust of formalist approaches to some of these generalisations was compared to a fully cue-based account. This cue-based account attempts to combine the insights of Steriade (1997) with regard to right and left-adjacent context effects with, first, the suggestion by Balise & Diehl (1994) that laryngeal neutralisation of fricatives may be grounded in perception, and second, the hypothesis that positional faithfulness ultimately derives from the effects of prosodic strengthening and weakening on the relative perceptual salience of phonological distinctions (cf. J. Beckman 1996, 1997). Thus, this theory is an instantiation of the strongest version of the functionalist hypothesis described in chapter 1 which states that all phonological neutralisation phenomena derive from the low perceptibility of distinctions.

Section 3.2 discussed data pertaining to the phonetic manifestation of neutralisation. Two kinds of phonetic data emerged here: data suggesting that laryngeal neutralisation results in phonetic underspecification, and evidence that final, ‘dynamic’ neutralisation is sometimes incomplete, leaving residual cues to the underlying values of [tense]. The first observation is troublesome for a fortition analysis of neutralisation in fortis-lenis systems because fortition predicts that the output of neutralisation is an obstruent that is phonetically and phonologically indistinguishable from its unequivocally fortis counterparts.

The next two sections examined the effects of obstruent features on the likelihood that those obstruents are targeted by laryngeal neutralisation. Section 3.3 concluded that on the basis of the available evidence, there is little reason to assume that laryngeal neutralisation behaves differently in voicing and aspirating languages. Section 3.4 discussed the well-known claim that laryngeal contrast is less stable in (sibilant) fricatives than in plosives. This claim is largely supported by the frequencies of laryngeal contrast (as distinct from phonetic voicing) in plosive and fricative inventories in typological databases such as UPSID$_{317}$ as well as by the context-sensitive neutralisation of fricatives in languages such as English and Frisian. In addition, section 3.4 reviewed the possibilities of explaining this asymmetry in terms of articulatory effort or perceptibility. However, whilst there is encouraging experimental support for the latter, both theories are hard to test against available typological databases, which tend to partially conflate phonetic voicing and participation in phonological contrasts such as [±tense]. The generalisations established in this section play an important role in the discussion of autosegmental models in chapter 8.

Section 3.5 engaged with the central generalisations underlying the theory of Steriade (1997). Steriade shows that syllable-driven accounts of final neutralisation fail in their core predictions about neutralisation in obstruent + sonorant clusters. First, they predict that word-final neutralisation should always coincide with neutralisation in heterosyllabic obstruents clusters. Second, they predict that neutralisation in word-medial obstruent + sonorant clusters follows syllabification patterns so that neutralisation occurs in heterosyllabic but not in tautosyllabic sequences. Both claims are demonstrably wrong. In fact, the dissociation between laryngeal neutralisation in word-final contexts and nonfinal obstruent + sonorant sequences is even stronger than Steriade (1997) suggests: both processes can occur independently from one another. A cue-based model is better suited to account for this data because it predicts that obstruent + sonorant sequences should behave in identical fashion irrespective of syllabification (and ceteris paribus).

Finally 3.6 assessed the possible explanations for the uncontroversial generalisation that word-final neutralisation is much more common than initial neutralisation, which seems to hardly occur at all (unless as part of neutralisation at the inventory level). Morphology or prosody-driven formalist accounts simply claim that the right edge of the word is a weak licensor of laryngeal contrast, whilst Steriade (1997) attempts to explain the initial-final asymmetry by generalising neutralisation in the weakly-cued *utterance-final* context across all other phonetic contexts. However, since word-level prosody manifests itself as strengthening and weakening, it seems likely to have a levelling effect on context-induced perceptibility differences in word-final (and unstressed) as well as word-initial (and stressed) environments. If this is indeed the case, a cue-based theory derives the neutralisation asymmetry between word-initial and word-final contexts in exactly the same way as the effects of obstruent internal features and adjacent sounds, and this would represent an improvement over both formalist accounts and Steriade's model. Although a great deal research is needed both to establish the precise predictions of the fully cue-based theory and to test them, I want to stress again that both are possible enterprises. Perceptibility measures can be established using known experimental designs whilst testing them against neutralisation asymmetries does not require any (statistical) methods that are beyond the means of phonologists.

Inevitably, there is a range of interesting and important issues that must remain unexplored here. To conclude this chapter I will briefly mention two topics that would deserve attention in a fuller survey of neutralisation phenomena. First it is often implicitly assumed that dynamic final neutralisation is fully symmetric with regard to manner and place of articulation distinctions. This is certainly the case in Dutch and German, where final neutralisation targets both fricatives and plosives across all places of articulation. However, it is not at all clear to what extent this represents the 'normal', most frequently occurring pattern in languages with established or developing neutralisation rules. Parker (1981), quoting Andersen (1972) and Stevens (1975), observes that in Belorussian final plosives started to neutralise before the corresponding fricatives whilst this pattern was reversed in German. Moreover, final neutralisation in Turkmen, as described by Clark (1998), generally targets stops and affricates but leaves dental nonsibilant fricatives untouched (there is no stable contrast for sibilant fricatives): cf. /duːð/, [duːð], *salt* vs. /nɑːmɨθ/, [nɑːmɨθ], *honour, shame*.

Second, in principle, theories about the phonological and phonetic nature of the neutralisation process are independent from theories about the contexts (in a broad sense) in which it is most likely to occur. This means that it is technically speaking possible to maintain a strictly formalist theory of neutralisation contexts whilst adopting a (phonetic) underspecification view of the neutralisation process. To some extent this is the sort of approach pursued by Ernestus (2000). Conversely, a cue-based model does not of itself rule out that laryngeal neutralisation is fortition. However, from a broader functionalist point of view, the claim by Balise & Diehl (1994) that voiceless fricatives are better carriers of fricative place contrasts implies that there is some pressure on phonetic grammars to employ active devoicing in the realisation of neutralised fricatives not to cue [tense], but to enhance the expression of place contrast. This implication is contradicted by [tense]-neutralised but voiced word-initial [z] in German and in Dutch dialects (which appears in a strengthening context, so that phonetic reduction resulting in passive voicing can be ruled out). But the idea that voiceless and voiceless aspirated (fortis) stops are more salient than (lenis) voiced ones has a long tradition (cf. Parker 1981) and may go some way in explaining the aspiration of laryngeally neutralised stops of Western Aleut and other languages examined by Cho & Ladefoged (1997).

Chapter 4

Voicing assimilation

Perhaps the easiest way to introduce *voicing assimilation*, also commonly referred to as *voice assimilation*, is by a brief discussion of the Dutch examples in (8). Dutch has an extremely well-documented process of regressive voicing assimilation that applies at boundaries between words and between stems and non-cohering suffixes (Zwaardemaker & Eijkman, 1928; Cohen et al., 1972; Trommelen & Zonneveld, 1979; Booij, 1995). In many instances where a word-final (or stem-final) obstruent is followed by a lax plosive /b/ or /d/, this obstruent is realised as voiced. Note that word-final obstruents in Dutch are subject to neutralisation of [±tense]: as a result RVA applies in equal measure to underlying [+tense] and [-tense] obstruents. For example, the final /s/ of the first member of the compound /vɪs/ +/diːfjə is equally likely to be realised as voiced as the underlying /z/ in /rɛiz/ + /duːl/. Elsewhere, and particularly before fortis obstruents and utterance finally, Dutch final obstruents are often realised as voiceless (barring brief voicing tails of preceding sonorants): [ʋeʲːk], [zɑnt], [vɪs], [rɛis].

(8) Regressive voicing assimilation to Dutch lax plosives

UR	Phonetic form	Gloss
/ʋeːk/ + /diːr/	[ʋeʲːgdiːɹ]	mollusc
/zɑnd/ + /bɑnk/	[zɑndbɑnk]	sand bank
/vɪs/ + /diːfjə/	[vɪzdifjə]	common tern (sterna hirundo)
/rɛiz/ + /duːl/	[rɛizdul]	destination

Given the strict distinction between *voicing* and *tense-lax/fortis-lenis* that I have adopted, the term *voicing assimilation* implies that the process only affects this single cue to the tense-lax distinction. However, many descriptions and accounts of voicing assimilation rules assume implicitly or explicitly that they operate directly on [±tense], which implies that the whole cue complex associated with it is affected. This assumption is so pervasive in the literature that it underpins even some of the experimental work attempting to demonstrate that RVA is a non-neutralising, and therefore phonetic process (e.g., Charles-Luce 1993; Burton & Robblee 1997). It is at least implied by the transcriptions in (8), which follow common practice in representing voicing-assimilated /t/ as [d] rather than as e.g., [t̬].[1]

In the discussion below I will nevertheless retain *voicing assimilation* as a cover term for the range of phenomena that have been so labelled in the literature. The purpose of this terminological compromise is not to confuse the reader but to convey the fact that the whole range of observed assimilation rules is so often viewed as a homogeneous set and attributed to a single underlying mechanism. Strictly speaking, *[tense]-assimilation* would do more justice to the nature of that mechanism, but it is an unfamiliar label and ultimately equally inappropriate as a cover term as *voicing assimilation*.

Voicing assimilation in this broad sense then, is an ubiquitous phenomenon. It occurs in morphological paradigms as well as between words in compounds and phrases, and there are few if any languages that maintain a [±tense] distinction that lack voicing assimilation rules altogether. In addition, many models of laryngeal phonology would also count lexical phonotactic restrictions on [tense] in obstruent sequences as instances of voicing assimilation.

This chapter is an attempt to dissect the set of ‘dynamic’ cases in which the voicing (and other relevant phonetic features) of obstruents vary with a varying context, i.e., the type of phenomenon illustrated in (8) above (lexical constraints on [±tense] are discussed in chapter 3). Section 4.1 immediately below identifies two common approaches to sandhi processes in the literature. The first of treats sandhi phenomena as operations at the (lexical) phonological level and is typically associated with recent generative models of voicing assimilation. The second is a phonetic, or more accurately *articulatory* theory of sandhi, which is embodied by models such as Articulatory Phonology (Browman & Goldstein 1986 et seq.). Although phonological models and articulatory phonetic or ‘gestural’ frameworks are often used to provide alternative accounts of the same data, it would seem that neither of them is dispensable and that they really apply to complementary sets of data rather than overlapping ones. The rest of the section spells out the specific predictions with regard to the phonetic manifestations of voicing assimilation that can be inferred from a phonological approach, and those implied by an articulatory analysis. These predictions are different and form the basis for the experimental hypotheses tested in the next chapters.

The following three sections of this chapter then discuss three forms of voicing assimilation and investigates for each of these which of the two theories

---

[1] For the sake of the argument I have assumed here that neutralised final obstruents in Dutch are properly represented as phonologically and phonetically [+tense]. Although this is a common assumption it may note be ultimately viable. See chapters 3 and 7 for more detailed discussion of final neutralisation and Dutch RVA respectively.

provides the most accurate account of its properties. First, voicing assimilation rules that are found at word-internal morpheme boundaries seem to behave as predicted by a phonological account. Second, in many instances, the progressive devoicing of obstruents at word boundaries appears to result from passive devoicing in the sense of section 2.1 above rather than an operation that ought to be expressed at the phonological level. This is hardly a novel or perhaps even controversial claim (cf. Harms 1973). However, section 4.4 proposes that the third form of voicing assimilation, i.e. regressive voicing assimilation across word boundaries, should be approached as an articulatory process driven by the production of voicing distinctions, or at least as diachronically grounded in such a process. The main argument for this proposal is that RVA across word boundaries is conditioned by the voicing categories of the triggering obstruents, at least in the case of lax plosives and affricates. Since this form of voicing assimilation is often put within the scope of generative models, this proposal is more contentious.

4.1 Modelling voicing assimilation

There are two common but distinct ways of thinking about sandhi phenomena, and hence about voicing assimilation. The first treats sandhi in terms of operations on (lexical) phonological features. In recent generative models this approach is typically implemented by means of autosegmental spreading and/or delinking rules, or (in Optimality Theory) *AGREE*-type constraints. The second approach treats sandhi rules as a product of the processes that govern the temporal co-ordination of articulatory gestures in speech production. This type of approach is sometimes simply referred to as *gestural*, perhaps because *Articulatory Phonology* (Browman & Goldstein 1986 et seq.) is one of the predominant formal frameworks in this area. However, since I do not intend to commit to a specific framework in this regard, and in keeping with the discussion in section 1.3.2 above, I will simply continue to refer to *articulation-driven*, *coarticulation(-based)*, or simply *phonetic* models.

The two approaches identified here are different in more than a conceptual sense: they generate different predictions with regard to the type of sandhi phenomena that can occur, and their phonetic manifestations. Consider first the phonological view of external sandhi. According to this view, any phonological feature may be targeted by sandhi rules, whether the feature in question has a consistent articulatory implementation or not. For example, the feature [±sonorant] might be defined phonetically in terms of airflow through the oral tract, spontaneous (de)voicing, or acoustic/perceptual intensity, but it has no consistent articulatory implementation. For instance, the 'sonorancy' of nasal consonants is due to the opening of the velopharyngeal port, whilst [l] and glides are [+sonorant] because of the relatively large size of the passage at the point of maximum oral constriction. Nevertheless, models that adopt [±sonorant] as a phonological feature and do not impose any specific restrictions on its availability for sandhi processes are in principle capable of expressing [sonorant] assimilation rules. Such rules would turn sequences of e.g. /d/ + /l/ into [nl] (leftward spread of [+sonorant) or /p/ + /n/ into [pd] (rightward spread of [-sonorant]).

By the definition provided in section 1.3.2, phonological processes are phonetically discrete. A second prediction of the phonological approach to sandhi processes is therefore that they are phonetically discrete. This means that there are no intermediate phonetic realisations between sequences that have not undergone a sandhi process and those that have, or only a finite number of such realisations. For example, phonological analyses of voicing assimilation generally generate two possible phonological surface forms and therefore two phonetic categories for underlying sequences such as /k/ + /b/ ([-voice][+voice] in underlying phonological representation). The first of these emerges when for some reason voicing assimilation fails to apply and can be symbolised as [kb] ([-voice][+voice] on the surface); the second represents the case of 'total' voicing assimilation, [ɡb] ([+voice][+voice]). Recent autosegmental models are sometimes capable of generating one or more surface forms that are intermediate between these extremes and as a result they provide some handle on cases of 'subphonemic', or partial assimilation (cf. Hayes 1992). However, the number of intermediate categories available is usually very small and always finite, and as a consequence phonological approaches predict that even partial assimilation creates discrete phonetic categories.

An articulation-driven approach to sandhi rules, on the other hand, predicts that only the articulatory gestures involved in the production of those sounds are available to sandhi processes. This means, for example, that the phonological feature [±sonorant] as such cannot play a role in assimilation rules. Instead, sandhi effects of the sounds classified by this feature are predicted to pattern along articulatory lines, which entails that a [n], but not a [l] should be able to turn a preceding [d] into a nasal because the former but not the latter is accompanied by the requisite velopharyngeal opening movement.

Second, since articulatory control (and physical articulation) operate on phonetic scales that can be regarded as continuous for all practical purposes, a phonetic approach predicts that sandhi phenomena are gradient at the phonetic level. This entails that assimilation rules do not produce discrete phonetic categories, but a continuous range of forms between unassimilated and assimilated sound sequences.

Proponents of both approaches would sometimes seem to imply otherwise, but it seems fairly clear that neither of them is capable of accounting for all documented sandhi phenomena. For example, Sardinian has a rule whereby word-final /s/ and /r/ are neutralised to [s] if the following word starts with a voiceless stop, and to [l] if it starts with another consonant (Bolognesi 1998; Ladd & Scobbie forthcoming: examples in 9)

(9) Neutralisation of word final /s/ and /r/ in Sardinian

/sos/ + /puddos/	[sɔspudːɔzɔ]	‘the chickens’
/tres/ + /manos/	[trɛlmanɔzɔ]	‘three hands’
/battor/ + /frades/	[batːɔlfradɛzɛ]	‘four brothers’

This rule cannot be modelled as a coarticulation process because the surface sequences cannot be described in terms of articulatory interference between the sounds in the underlying forms. For example, there is nothing in the articulation of [m] (which consists mainly of a full bilabial constriction accompanied by an opening of the velopharyngeal passage) that would force a preceding [s] to approximate the tongue configuration of a lateral approximant or to lose its wide glottal abduction.

On the other hand, a variety of external sandhi phenomena in a number of languages, and in particular place assimilation rules, have been shown to operate in a gradient fashion and are therefore impossible to describe in phonological terms (e.g. Barry 1992; Nolan 1992). Consequently, the most viable model of the phonology-phonetics interface seems to be one that can accommodate both phonological and coarticulatory sandhi rules (cf. Zsiga 1997).

The remainder of this section attempts to flesh out corresponding sets of predictions regarding the phonetic manifestations of voicing assimilation when analysed as a phonological process and when analysed as a coarticulatory. These predictions form an important basis for the rest of this study, which argues that voicing assimilation occurs both as a(n incompletely neutralising) phonological rule and as a coarticulation process.

4.1.1 Phonological analyses of voicing assimilation

Many recent generative accounts of voicing assimilation phenomena assume that the rules in question apply to the phonological feature that represents the lexical contrast between tense and lax obstruents. Here and below, I will focus on this type of account in discussing phonological approaches to voicing assimilation, but note that the predictions in (10) hold for any model which assumes that voicing assimilation entails a transfer of (lexical) phonological identity from the assimilation trigger to the assimilation target.

The mechanics of the phonological, or perhaps more accurately, *lexical feature*, conception of voicing assimilation adopted by the generative accounts referred to in the previous paragraph are illustrated in figure 4.1, using the autosegmental notation adopted by virtually all recent generative models of laryngeal phonology. The lexical contrast between tense and lax obstruents is represented by a [±tense] feature, or some formal equivalent such as [±voice]. This feature may or may not be dominated a *Laryngeal* (LAR) class node. Voicing assimilation is modelled by spreading the lexically contrastive feature or the dominating LAR node from the assimilation trigger, in this instance the obstruent represented by *Root*$_2$ to a target, here symbolised by *Root*$_1$. Under the most common version of this analysis, the original LAR node and [±tense] specification of *Root*$_1$ are delinked and removed from the representation so that trigger and target share a single lexical laryngeal feature. This spreading-cum-delinking analysis is formally equivalent to more old-fashioned polarity switching 'agreement' rules of the [+tense] → [-tense]/__[-tense] variety.

Under a less common but, in the light of data reported below, more sophisticated approach the underlying LAR node and [±tense] specification of *Root*$_1$ are preserved, and the result of the spreading operation is a 'mixed', doubly articulated or contour segment carrying two [tense] features with distinct values (cf. Hayes 1992).

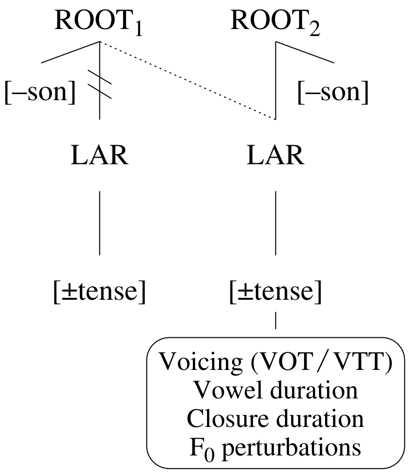

Figure 4.1: An autosegmental lexical feature analysis of voicing assimilation and its phonetic manifestations.

For two reasons, it is hard to pin down the predictions of recent lexical feature analyses regarding the phonetic manifestation of voicing assimilation. First, the lattice-like nature of autosegmental representations defines a greater number of possible structures for the same feature alphabet and number of segments (i.e., root nodes) than the feature bundle model on which SPE was built. This is already evident in the fact that there is a choice of applying spreading with delinking, or without it. Furthermore, there are two ways to represent sequences with homogeneous specifications for [tense] such as [ɡb]: the first is a sequence of two root nodes each carrying its own LAR node with [-tense] feature; the second consists of two root nodes sharing a single LAR node and [-tense] specification. Like the difference between singly and doubly articulated segments employed by Hayes (1992), the difference between these two structures can in principle be employed to represent a distinction between underlyingly homogenous [tense] sequences such as /ɡ/ + /b/ and underlyingly heterogeneous but surface-assimilated clusters such as /k/ + /b/ realised as [ɡb].

The second reason it is difficult to establish the predictions of recent lexical feature analyses is simply that the relation between autosegmental phonological representations and linguistic phonetic forms (i.e., the data structures required at the interfaces with peripheral processing) is not spelt out, or at least not fully. There are two general ways of thinking about this relation. According to the first, autosegmental feature models occupy the phonology component of modular models of the phonology-phonetics interface. The second sees autosegmental representation itself as (the backbone) of linguistic phonetic representation. The latter view is the one adopted by Harris & Lindsey (1995) and Avery & Idsardi (2001), albeit in different ways.

I will not to try to flesh out all the possible predictions about the phonetics of voicing assimilation that can be derived from autosegmental representation and different takes on its relation to linguistic phonetic representation. For the purpose of deriving phonetic predictions I will rather take the sort of spreading-cum-delinking rule illustrated in figure 4.1 as a convenient shorthand for the traditional polarity-switching rules [-tense] → [+tense]__[+tense], [+tense] → [-tense]__[-tense], and (combined) [-αtense] → [αtense]__[αtense]. In addition, I will assume that [tense] is a proper phonological feature, and therefore that voicing assimilation operates in the phonological component of a modular model of the phonology-phonetics interface. Note that prominent representatives of autosegmental lexical feature analyses of voicing assimilation such as Lombardi (1994, 1995a,b, 1996) and Iverson & Salmons (1995, 1999) leave this option completely open. Moreover, the finer distinctions that can be encoded by autosegmental lattices are generally ignored by optimality-theoretic models of laryngeal phonology that use *AGREE*-type constraints (e.g., Lombardi 1995c, 1997, 1999; Grijzenhout & Krämer 1998; Borowsky 2000: see 8.2.5). These accounts effectively treat all surface [αtense][αtense] sequences identically regardless of whether they are derived from homogeneous or heterogeneous underlying clusters.

Given the general absence of explicit models of phonetic interpretation in the (recent) generative literature, I feel that my 'phonemic' reading of its phonetic implications does not do it any gross injustice. This opinion is bolstered by occasional explicit assertions that voicing assimilation processes are neutralising (e.g., Siptár & Törkenczy 2000), or that optional and/or non-categorical phenomena lie without the scope of generative models (Lombardi, 1999). However, any reader who derives specific phonetic predictions from an autosegmental (or any other) model that do not agree with those I will extract in the next few paragraphs is invited to treat the latter as a straw man. This reader is also invited to test his or her own predictions against the data reported in chapters 5, 6, and 7. However, any autosegmental model of voicing assimilation will also have to address the more fundamental issues raised in chapters 1 and 8.

Now, the key property of a polarity-switching lexical feature analysis of voicing assimilation is that sequences that are subject to assimilation surface as phonologically identical to underlyingly homogeneous sequences. Thus, this type of analysis draws no structural distinction between underlying /g/ + /b/ on the one hand and /k/ + /b/ after regressive assimilation on the other: both are represented as identical [-tense][-tense] sequences. Two key predictions follow from this property (cf. 10a and 10b below). First, the fact that underlying [αtense][αtense] sequences and assimilated clusters are phonologically identical means that they are phonetically identical as well, and therefore that voicing assimilation is a phonetically neutralising process. Phonetic interpretation has no access to information about the underlying status of the obstruents in a cluster and can therefore assign only a single phonetic form to any [αtense][αtense] sequence in any (wider) context. Thus, underlyingly homogeneous /g/ + /b/ and regressively assimilated /k/ + /b/ are predicted to surface as identical [gb].

(10) Predictions of a (polarity-switching) lexical feature analysis of voicing assimilation
   a. Voicing assimilation processes are phonetically neutralising (polarity-switching analyses only)
   b. All locally relevant cues to [tense] participate in voicing assimilation processes
   c. The ability of a given tense or lax obstruent to trigger or undergo voicing assimilation is unrelated to the phonetic manifestation of [tense] in that obstruent

Second, the neutralising nature of voicing assimilation under a lexical feature analysis entails that the process affects all cues to [tense] that are relevant in the environment of the assimilation target. Strictly speaking, this means that *voicing assimilation* is a misnomer. For instance, if the /g/ in a /g/ + /b/ sequence is normally distinct from /k/ in /k/ + /p/ clusters in terms of preceding vowel duration, closure duration and voicing, then assimilated /k/ in /k/ + /b/ is predicted to share all of these characteristics. In a sense therefore, all cues signalling the fortis-lenis distinction are predicted to 'spread with' [tense] where voicing assimilation applies. This is symbolised by the cluster of phonetic features suspended from the spread [±tense] feature of Root$_2$ in figure 4.1. Note, incidentally, that this second prediction might also be assigned to a spreading-without-delinking analysis. On this view, voicing assimilation does not completely neutralise [tense] distinctions in target obstruents, but still affects the whole cluster of relevant cues. So prediction (10b) follows from a wider range of lexical feature accounts than (10a).

If the fortis-lenis contrast is universally represented with a single formal feature, e.g., [±tense], a third prediction that follows from a lexical feature analysis is that voicing assimilation is not phonetically conditioned. For example, if the passively and actively voiced lenis stops /b, d, g/ of English and Yiddish respectively are represented by the same feature and value [-tense], both languages are equally likely to have a rule spreading this feature value, despite the phonetic differences between their lax stops (assuming that [tense] spreading rules are a necessary part of the universally available inventory of phonological rules). Along similar lines, a lexical feature analysis built on a single universal feature [tense] (or some equivalent) predicts that if a language has passively voiced lenis stops but actively voiced lenis fricatives, they should behave alike in terms of their assimilatory behaviour. Only if there are two or more distinct features in the universally available pool to represent tense-lax distinctions can these and similar predictions be avoided, because this allows for the (crude) matching of phonological with phonetic categories, and thus for the phonology to have access to phonetic information. Models that take this approach are discussed in chapter 8.3.

4.1.2 Articulation-driven voicing assimilation

According to the logic of articulation-driven models, the phonetic effects of a sandhi rule are purely determined by the articulatory properties of the sounds involved. In order to establish any predictions about the phonetic characteristics of an articulation-based form of voicing assimilation, it is therefore necessary to establish which reflexes of [tense] are ‘available to’ the coarticulation mechanism. It is argued below that the main phonetic features of [tense] that are able to trigger articulatory adjustment in a neighbouring sound are the articulations involved in the production of voicing contrasts.

The key idea behind this theory is probably best illustrated with the behaviour of actively voiced lax stops. Recall from 2.1 above that actively voiced stops are likely to be produced with an array of enhancing gestures aimed at enlarging the size of the oral cavity behind the occlusion, and (to a lesser extent) lowering the threshold transglottal pressure difference for vocal fold vibration. Under virtually all models of articulatory implementation this entails that these gestures are subject to coarticulation. As reviewed in chapter 1, speakers are assumed to set certain limits on the extent of coarticulation that depend in part on their communicative needs and are in part likely to be language-specific, but the underlying mechanism (negotiation between subsequent targets involving the same or mechanically linked articulators) is invariably supposed to be universal. Consequently, the voicing-enhancing gestures associated with an actively voiced stop are predicted to 'spill over' into neighbouring sounds, and mainly the preceding one(s) if there is indeed a bias towards anticipatory over perseveratory coarticulation (cf. Farnetani 1997). Because the gestures that spill over are designed to control the voicing of the trigger obstruent, they will also have an effect on the voicing of neigbouring sounds. For example, if a word-initial /b/ is produced with an advanced tongue root and lowered larynx in order to facilitate voicing, a coarticulated preceding /k/ will be subject to some degree of larynx lowering and tongue root advancement, too. If the effect is sufficiently strong it will improve the conditions for voicing during the /k/ and thereby lengthen the voice tail into this obstruent from a preceding vowel.

Along the same lines, the voicing control measures associated with actively *devoiced* obstruents are predicted to influence the voicing of neighbouring obstruents. Passively voiced plosives, on the other hand can have no coarticulatory effect on the voicing control of neighbouring obstruents: if the word-initial /b/ referred to above is produced without any articulated moves designed for voicing control, there is simply nothing to spill over into flanking sounds. Instead, the voicing control gestures of flanking sounds will have relatively free reign in a passively voiced sound since there is no local voicing control to counterbalance the spill over.

Thus, a coarticulation-based account of voicing assimilation derives the process as the inevitable by-product of two mechanisms that are plausible on independent grounds. As discussed in chapter 1, coarticulation (in a narrow phonetic sense) is normally conceived as a truly universal mechanism that governs the production of any continuous sequence of sounds produced by any speaker at any given time. Although there are differences of opinion about how the mechanism should be modelled exactly, its reflexes in a wide variety of contexts are well-documented, and consequently its existence is motivated independently from any observations concerning (de)voicing in obstruent clusters. Similarly, as pointed out in 2.1 above, the existence of secondary articulations aimed at voicing control is arguable on independent (aerodynamic) grounds grounds, and is not invoked only to deal with voicing patterns in obstruent clusters.

In other words, according to a coarticulation-based theory, RVA at word boundaries occurs automatically as part of a much more general process of sound co-production every time an actively (de)voiced obstruent is juxtaposed with another obstruent: *a specific voicing coarticulation rule does not have to be postulated*. In this respect, a coarticulation-based approach to word boundaries is different from a lexical feature analysis. Rules of [±tense] agreement or spreading-and-delinking adhere to more general formal templates, but formal grammars in which every phonological feature agrees or spreads in every possible context have to my knowledge never been proposed. Consequently, [tense] spreading rules have to be specified explicitly for every language which possesses RVA at word boundaries.

This study is certainly not the first to look at voicing assimilation as a coarticulation process: notably Slis (1985, 1987) and more recently Ernestus (2000) have made similar proposals. They are supported in part by a series of studies of glottal coarticulation (Löfqvist, 1980, 1981; Löfqvist & Yoshioka, 1980, 1984; Yoshioka et al., 1981, 1982), which demonstrate that the glottal abduction gestures associated with contrastively voiceless stops and fricatives vary according to the phonetic context. For example, the results of Yoshioka et al. (1981) indicate that when flanked by sonorants or a pause, English word-final /s/ and word-initial /k/ are accompanied by an opening and closing movement of the glottis with a distinct peak. But when word-final /s/ is combined with word-initial /k/ the results is not two discrete peaks separated by a brief closure of the glottis, but a composite gesture without full medial closure, and in some cases only a faint trace of two separate peaks. In the light of acoustic studies reporting incompletely neutralising RVA, it is interesting that this composite gesture remains distinct in timing, shape, and size from the large single peak found in word-initial /s/ + /k/ sequences. Moreover, the posterior cricoarytenoid muscle, which is largely responsible for glottal abduction, shows two distinct activity peaks in the production of the former but not the latter.

However, none of these studies spell out the precise typological implications of a coarticulation-based theory of voicing assimilation. Some of these predictions appear in (11). First, a coarticulation-based approach predicts that only the cues to [tense] that 'can be coarticulated' should participate in voicing assimilation. As explained above this set includes active (de)voicing because it is supported directly by articulatory gestures. In addition, it includes any phonetic features that are mechanically dependent on voicing gestures such as the duration of frication in fricatives, which varies as a function of the size of (peak) glottal abduction (see 2.3.2 above). The set of features that can be coarticulated potentially also includes F_0 and F_1, but as pointed out in 2.2.3 the articulatory basis of these cues remains unclear.

By contrast, assimilatory effects on the *durational* correlates of [tense] would seem to be excluded by a coarticulation model. The work of Kluender et al. (1988) rules out a mechanic link between voicing and obstruent or preceding vowel duration (cf. 2.2.3), and consequently the coarticulatory spill over of voicing control moves should not be able to influence durational parameters apart from frication duration in fricatives (and perhaps stop releases: see 2.3).

Second, on most accounts, segmental duration does not correspond to 'substantive' articulatory gestures but to the *scaling* of such gestures in phonetic space and/or time. This scaling can be modelled 'extrinsically' by directly assigning (relative) durations to phonetic targets, or 'intrinsically'. Intrinsic timing control consists of aligning specific articulatory or auditory landmarks of subsequent sounds with each other (e.g., the onset of a nasal with the oral release of a preceding plosive) and specifying the excursion size and/or speed of the gestures used in their production (Browman & Goldstein 1986 et seq.; Huffman 1993; Kirchner 1998).[2] But under neither of these analyses does segmental duration 'spread' or 'spill over' into neighbouring sounds. For example, modelling increased duration by assigning greater excursion sizes to the relevant articulators may in some cases force a greater amount of coarticulation on neighbouring sounds (to compensate for longer trajectories between subsequent articulatory targets), but the resulting adjustments would not lead to lengthening of those neighbouring sounds.

Note that this property of coarticulation models is attractive on empirical grounds, because relative phonetic length generally does not spill over between neighbouring sounds: phonetically long vowels, for instance, are not necessarily flanked by long consonants. In this respect coarticulation models clearly mirror recent generative treatments of (contrastive) length as prosodic rather than substantive. Such models encode length distinctions above the *Root* level (cf. figure 4.1) as positions on a dedicated timing (moraic or skeletal) tier, which (given certain basic assumptions) means that in contrast to, e.g., [nasal] or place of articulation, length does not spread between sounds. In this sense, the prediction of a coarticulation-based model that RVA should have no effect on the duration of the target obstruent and the preceding vowel parallels the prediction of generative models that phonological geminates do not induce lengthening in neighbouring vowels.

The second prediction of a coarticulation-based approach is that as any form of coarticulation, coarticulatory voicing assimilation rules should operate as gradient rules that are sensitive to speech rate and style to the degree that such phenomena can be understood in terms of global hypoarticulation. This implies that voicing assimilation can occasionally be phonetically neutralising, but only at the end point of a scale that contains many incompletely neutralised cases. A further implication of the phonetic status of voicing assimilation is that the process is automatically blocked by a sufficiently long physical pause: unlike the effects of phonological feature spreading or agreement, which operate on adjacent points in abstract time, the effects of coarticulation decay as a function of real time. As a result, any stop or fricative that is separated from a neighbouring actively (de)voiced obstruent by a (silent) interval of a certain duration will be

---

[2] The distinction between extrinsic and intrinsic timing is due to Fowler (1980).

‘out of reach’ of the anticipation or perseveration of the relevant (de)voicing-enhancing articulatory moves.

(11) Predictions of a coarticulation-based approach to voicing assimilation

   a. Only those cues to [tense] that are subject to coarticulation in obstruent sequences (mainly voicing) participate in voicing assimilation
   b. Voicing assimilation is a gradient, non-neutralising process
   c. The ability of a given tense or lax obstruent to trigger or undergo voicing assimilation is a function of the phonetic manifestation of [tense] in that obstruent (and primarily its voicing targets)

I already described the third prediction of the phonetic theory of voicing assimilation proposed in this section with specific reference to the behaviour of lax stops: the ability of a given obstruent to trigger or undergo voicing assimilation is intrinsically related to the phonetic manifestation of [tense] in that obstruent. Given that the articulation of voicing distinctions provides the main set of articulatory gestures that are available for coarticulation, it follows that the capacity of a segment for triggering voicing assimilation can be predicted from its voicing target (which can in turn be read off its behaviour in utterance-initial and postsonorant environments).

This general hypothesis brings several interesting predictions with respect to the patterning of specific classes of sounds in specific languages. For instance, English lax obstruents are predicted to behave in a manner-asymmetric fashion with respect to voicing assimilation: actively voiced /v, z, ʒ/ but not (passively voiced) b, d, ɡ, d͡ʒ are expected to trigger voicing assimilation in a preceding obstruent. Furthermore, both types of fortis obstruent defined above (aspirated and plain voiceless) should in principle trigger voicing assimilation in a preceding obstruent, since both categories are arguably actively devoiced. Consequently (regressive) voicing assimilation to fortis but not lenis stops is predicted to occur in aspirating as well as voicing languages.

A final and potentially important prediction of the present theory of coarticulatory voicing assimilation is that sonorant consonants should never trigger RVA, except in languages where they are contrastively (de)voiced. I noted in 2.1 that it is likely that in languages without contrastive voicing on sonorants, the common voicing of this class of sounds is a result of the voiced carrier sound and passive voicing rather than the presence of an active voicing target. Since the phonetic theory described in this section derives all RVA (at word boundaries at least) from the presence of active (de)voicing targets, it follows that noncontrastively voiced sonorants should not be able to trigger voicing assimilation.[3]

---

[3] Many recent generative models make a similar prediction because they treat sonorants without contrastive voicing as unmarked for the [voice] or an equivalent feature.

4.2 Voicing assimilation in morphological paradigms

The Germanic group of languages provides two excellent examples of voicing assimilation rules that appear to be phonological in nature, insofar as they are not conditioned by the phonetics of tense. The first example, illustrated with well-known examples from English and Dutch in (12) consists of the alternation of a [t]-initial suffix after stems ending in a fortis obstruent with a [d]-initial suffix after (most) sonorant-final and lax obstruent-final stems. Where it is associated with past tense morphology I will refer to it as the 'Germanic past tense paradigm' for convenience. What is perhaps the most common analysis of this allomorphic rule treats the [d]-initial variants as underlying and derives the [t]-initial forms by progressive assimilation to stem-final fortis obstruents but the pattern itself does not force this analysis. But irrespective of whether [d] or [t] is regarded as underlying, the essential generalisation remains that the process derives obstruent clusters which are homogeneous in terms of voicing, or perhaps rather the lexical feature [tense]. Crucially for present purposes, it is not phonetically conditioned to the extent that it occurs both in voicing varieties (e.g., standard Dutch, Scottish English, Frisian) and aspirating varieties (e.g., north-eastern dialects of Dutch, standard varieties of English, Swedish, and Norwegian) of Germanic. A highly similar pattern is found in several Turkic languages, where it appears to be equally insensitive to the voicing aspirating distinction, since it occurs both in e.g., (aspirating) Turkish and (voicing) Turkmen.[4]

Quantitative information on the phonetic behaviour of the various incarnations of the pattern in (12) is not available, but it seems fairly safe to assume that it is discrete rather than gradient in all of the languages in which it occurs, and affects all correlates of [tense]. Like final laryngeal neutralisation, it certainly does not exhibit the characteristics that are typical of phonetic rules such as dependency on speech rate and/or style. It is never optional, and in some ways even phonetically abstract. For instance, many Dutch speakers retain a lexical distinction between stem-final [x] and [ɣ] in their marking of the regular past tense (cf. /lɑx/ + /tə/, [lɑχtə], *laughed* vs. /zaːɣ/ + /tə/, [zaːʁdə], *sawed*) without phonetically realising the contrast in any other environment: such speakers produce both /lɑx/ +/ən/ and /zaːɣ/ + /ən/ with a fully voiceless velar or uvular fricative. Thus, the pattern in (12) can be maintained in the absence of any phonetic motivation.

---

[4] Needless to say, the pattern in (12) is not limited to past tense paradigms but also governs, e.g., the English regular plural. The Norwegian version of the rule shows an interesting twist in comparison with its cognates. [d(ə)] suffixes appear after vowels, glides and lenis obstruents including /v/-/ʋ/ whereas [t(ə)] is used after fortis obstruents and (other) sonorant consonants (Kristoffersen, 2000). Baitchura (1975) suggests that like most other Turkic languages, Turkmen is an aspirating language, but the grammar by Clark (1998) describes it as voicing.

(12) Progressive assimilation in the ‘Germanic past tense paradigm’

a. Dutch past tense paradigm

UR	Phonetic form	Gloss
/zɑk/ + /də/	[zɑktə]	failed
/ʋɑs/ + /də/	[ʋɑstə]	washed
/dœb/ + /də/	[dʉbdə]	doubted, wavered
/xraːz/ + /də/	[xraːzdə]	grazed
/krœl/ + /də/	[krʉdə]	curled

b. English past tense paradigm

UR	Phonetic form	Gloss
/lɒp/ + /d/	[lɒpt]	lopped
/lɒk/ + /d/	[lɒkt]	locked
/lɒb/ + /d/	[lɒbd]	lobbed
/lɒg/ + /d/	[lɒgd]	logged
/loːn/ + /d/	[loʊnd]	loaned

The second assimilatory pattern that is likely to be a case of [tense] assimilation rather than coarticulatory voicing assimilation is illustrated in (13) with the behaviour of the Norwegian adjectival agreement marker /t/, which causes a preceding lax obstruent to fully devoice (Kristoffersen, 2000). It can be accounted for in a straightforward manner by spreading a [+tense] feature backwards from the agreement marker. Similar processes are found in morphosyntactically restricted contexts across the Germanic group including (aspirating) Danish (Panzer, 1981), Swedish (Hellberg, 1974), and (voicing) Yiddish (Birnbaum, 1979; Katz, 1987), and depending on how the pattern in (6) in chapter 3 is analysed, (aspirating) German and (voicing) Dutch as well. Thus, Yiddish realises /red/ + /st/, *you*-SG.FAM. *speak* as [retst] and /ʃraib/ + /st/, *you*-SG.FAM. *write* as [ʃraipst] (Katz, 1987).[5]

Again, quantitative evidence demonstrating the phonetically discrete behaviour of this type of rule does not seem to be available, but like instances of the Germanic past tense paradigm, it typically displays all the hallmarks of a phonological rule. Interestingly, Kristoffersen (2000) finds that there is at least one lexical exception to the Norwegian rule in (13): /gløg/, *intelligent* surfaces with a (voiced) lax stop in both its uninflected ([glœg]) and inflected ([glœgt]) forms. Note that regressive [tense]/voicing assimilation in morphological paradigms is by no means restricted to [+tense] spreading/devoicing: regressive assimilation to suffixes with tense or lax initial obstruents (including single /d/) occurs in Hungarian (Kenesei et al., 1998; Siptár & Törkenczy, 2000).

---

[5] The same pattern is found in a small number of (lexicalised) English forms such as the irregular past tense [lɛft] from /liːv/, and [lɒst] from /luːz/.

(13) Norwegian regressive ‘devoicing’ (data from Kristoffersen 2000)

UR	Uninflected	Inflected	Gloss
/tryg/	[tɾyg]	[tɾy kt]	secure
/grov/	[gɾoːʋ]	[gɾɔft]	coarse
/stiv/	[stiːʋ]	[stiːft]	rigid

4.3 Progressive devoicing at word boundaries

Progressive devoicing refers to the situation in which an obstruent that is voiced in (some) other contexts becomes voiceless when preceded by another obstruent. At first sight this might be regarded as a good description of what happens to the underlying /d/ of the ‘Germanic’ past tense suffix in (12), but it has long been recognised that some cases of progressive devoicing are better treated as passive devoicing in the sense of 2.1. For instance, Harms (1973) notes that [mædd̥ɒg] with a voiceless second /d/ is a common realisation of English <mad dog>. Similarly, the final obstruents of <bagged> or <lodged> are often phonetically voiceless, even though they remain distinct from the final sounds of <racked> or <botched> (e.g., Jones 1956). This is illustrated in figure 4.2. A lexical feature analysis of these cases is inappropriate on two grounds. The first is that a [tense] spreading or agreement account predicts that devoicing occurs only after tense obstruents, which is clearly incorrect. As far as English lenis plosives are concerned the second reason is that these sounds are generally voiceless in a wider range of contexts (e.g., utterance initially) that is perfectly predictable once it is assumed that they are only passively voiced (cf. 2.1).

Progressive devoicing at word boundaries is not restricted to passively voiced obstruents such as the lenis stops of English (or German: Drosdowski & Eisenberg 1995). Actively voiced obstruents are also routinely subject to devoicing when preceded by another obstruent. In spite of the fact that the devoicing of actively voiced obstruents can not be attributed to their lack of voicing targets, a number of the cases in question are probably amenable to an explanation in terms of passive devoicing.

Consider the devoicing of lax stops in Dutch as an example. A production study carried out by Slis (1986) shows that the devoicing of Dutch word-initial lax plosives preceded by another obstruent is sensitive to (nuclear) stress: of the lax plosives in the onset of a stress-bearing syllable 21% is devoiced, but of those in the onset of a stressed syllable no less than 57% is voiceless. One possible account of this process treats it in terms of a rule spreading [+tense] from a preceding obstruent (such an approach is available to the many models that treat the single (neutralised) series of word-final obstruents in Dutch as [+tense]).

Alternatively, Dutch lenis plosive devoicing can be viewed as another ‘non-rule’ in the sense of Harms (1973). The key to this second analysis is the obser-

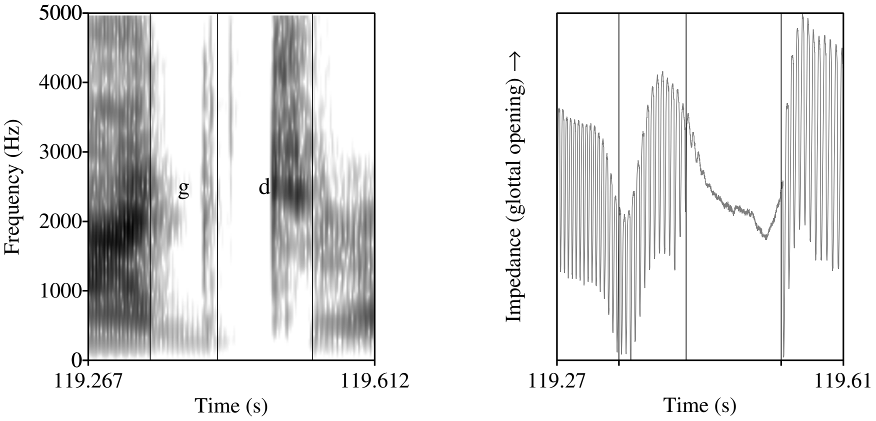

Figure 4.2: Progressive devoicing in English. Broad band spectrogram (left) and EGG trace of English <nagged>, produced in the carrier sentence *with what does __ rhyme?* The stem-final velar stop is voiced throughout, but the suffix /d/ is almost completely voiceless.

vation that the prolongation of an obstruent constriction makes the continuation or initiation of voicing progressively more difficult, even if voicing-enhancing gestures are present (cf. 2.1). Thus, the presence of a preceding obstruent imposes adverse aerodynamic conditions for the voicing of a lax obstruent, which means that articulatory moves aimed at voicing control are less likely to sustain (full) voicing than in phonetic contexts that are less hostile to vocal fold vibration. Within certain physical limits, it should be possible to counteract the aerodynamic conditions imposed by a preceding obstruent by expanding the gestures involved in voicing control. However, the general effect of destressing is is articulatory weakening rather than strengthening, which means that in unstressed contexts the articulatory gestures involved in the production of active voicing are less likely to be sufficient to counteract the effect of a preceding obstruent. In other words, under a passive devoicing account the effect of stress on the devoicing of lax plosives does not have to be stipulated (as it has to in a [+tense]-based analysis), but follows from the independently motivated effect of prosodic structure on articulation (cf. chapter 1).[6]

The partial devoicing of English lenis fricatives, briefly mentioned in 2.3 is a further likely candidate for analysis in terms of passive devoicing. It is unclear, however, whether the same applies to the devoicing of Dutch lax fricatives. This process, which is illustrated in (14), is highly noticeable and generally regarded

---

[6] Ernestus (2000) presents an analysis roughly along these lines of the process that devoices the initial /d/ of several Dutch function words.

as a phonological rule, unlike the processes described in the previous paragraphs (Booij, 1995; Lombardi, 1997; Grijzenhout & Krämer , 1998; Ernestus, 2000). This is reflected in the customary transcription of devoiced /v, z/ as [f, s] rather than [v̥, z̥].

An alternative approach is to treat Dutch lax fricatives as lacking a voicing target, analogously to the lax plosives of aspirating languages. This approach would predict that the contrast between fortis and lenis fricatives is consistently realised as [f, s] vs. [v̥, z̥] utterance initially as well as after obstruents. Ernestus (2000) claims that Dutch /v, z/ may be produced with voicing utterance initially, but there seems to be no quantitative data available to assess the extent of this voicing (and compare it with the voicing of e.g., English lax fricatives). Given the account of regressive voicing assimilation at word boundaries developed below, it would also be predicted that Dutch fricatives do not trigger RVA, which according to Slis (1987) they do only very occasionally. Finally, one of the key predictions of a passive devoicing account of Dutch postobstruent fricative devoicing (and every other passive devoicing analysis) is that /v, z/ retain the other phonetic features of [-tense] in contexts where they are devoiced. A lexical feature account on the other hand, predicts that the devoiced fricatives of /riːt/ + /zɑŋər/ and /rɛɪz/ + /vərhaːl/ are phonetically identical to lexical /f, s/.

(14) Progressive devoicing of lax fricatives in Dutch

UR	Phonetic form	Gloss
/drœk/ + /vɑt/	[drʉkfɑt]	pressure vessel
/riːt/ + /zɑŋər/	[ritsɑŋəɹ]	sedge warbler
/lɔf/ + /zɑŋ/	[lɔfsɑŋ]	eulogy
/klɑs/ + /vərklɛɪnɪŋ/	klɑsfərklɛɪnɪŋ]	class size reduction
/drɛɪv/ + /zɑnd/	[drɛɪfsɑnt]	quicksand
/rɛɪz/ + /vərhaːl/	[rɛɪsfərhaːl]	travel story

4.4 Regressive voicing assimilation at word boundaries

It has been common practice in generative phonology to treat assimilation phenomena of the type described in section 4.2 above on a par with voicing assimilation rules operating word sandhi contexts. This is especially evident in some generative accounts of the typology of voicing assimilation phenomena and in descriptions of assimilation rules that fail to mention whether the processes in question are in any way restricted by the morphology. The attempt by Lombardi (1999) to force the voicing assimilation processes found in Yiddish and Swedish into a single typology defined on purely phonological grounds provides recent examples of this style of analysis. It is striking that Lombardi proceeds with her account despite the fact that by all descriptive accounts (several of which she cites) voicing assimilation in Swedish is restricted to very specific morphosyntactic paradigms (e.g., Hellberg 1974), whereas Yiddish has regular regressive voicing assimilation across word boundaries as well as within morphological paradigms (e.g., Katz 1987).

In the light of a tradition of such analyses, and the lexical feature view of voicing assimilation more generally, it is no surprise to find explicit assertions of the idea that voicing assimilation is not (ever) conditioned by the phonetic realisation of [tense]. Thus, Keating (1984) states that:

> [c]luster voicing assimilation is another common phonological rule which appears to apply generally across phonetic categories. Thus Polish has regressive voicing assimilation [...] and a [voice] contrast of {voiced} vs. {vl. unaspirated} stops; Danish however, has progressive 'voicing' assimilation, but an aspiration contrast in initial position [...].(Keating 1984:294)

There is evidence however, that in contrast to the rules described in 4.2 above, voicing assimilation under word sandhi is, or at least can be, phonetically conditioned as well as phonetically gradient. The principal observation underpinning the first claim is that lax plosives only seem to trigger RVA across word boundaries if they belong to the actively prevoiced category; the most reliable evidence for the second is provided by quantitative studies of RVA in a number of languages. To some extent this evidence is corroborated by descriptive claims that assimilation across word boundaries is incomplete or optional and dependent on speech style or register and blocked by physical pauses. But as such claims often differ from language to language and linguist to linguist (as is shown by a comparison of, e.g., Vago 1980 and Siptár & Törkenczy 2000), it is difficult to gauge their articulatory and/or perceptual reality or their generality.

The observation that word-initial lax stops only trigger RVA if they belong to the actively prevoiced category is made by Kohler (1979) as part of a description of the realisation of [tense] in German (dialects) and French. It appears to hold across languages that maintain a distinction between plosives that is (partially) based on VOT distinctions and is beautifully illustrated by the typology of Germanic languages and dialects. All and only the varieties of this group that are described as employing a prevoiced vs. short lag realisation of fortis and lenis plosives are also reported as exhibiting RVA to lenis plosives (cf. Kohler 1979, 1984): Afrikaans (Wissing, 1991), (Western and Southern) Dutch (e.g., Cohen et al. 1972), Scottish English (Kohler, 1979; Wells, 1982a), (West) Frisian (Riemersma, 1979; Tiersma, 1985), Rhineland German (Kohler, 1979), and (Eastern varieties of) Yiddish (Katz, 1987; Jacobs et al., 1994). The remaining varieties of English and German (that maintain a VOT contrast for their initial plosive series) and the North Germanic languages, all of which belong to the aspirating type, show no RVA to lenis plosives.[7]

Although the assimilatory behaviour of lax stops indicates that RVA at word boundaries can be phonetically *conditioned*, this does not in itself establish the phonetic (gradient, cue-specific) status of the phenomenon. Direct quantitative evidence of gradience is hard to find, but there is ample evidence for *incomplete neutralisation* in RVA at word boundaries, which is a necessary albeit not a sufficient property of coarticulation rules. This evidence provided by a range of quantitative studies: O. Thorsen (1966) on French, N. Thorsen (1971) on English, Charles-Luce (1993) on Catalan, Burton & Robblee (1997) on Russian, and Barry & Teifour (1999) on Syrian Arabic. These studies investigate the effects of the [±tense] specification of a word-initial obstruent (C_2) on the phonetic features of a single preceding (word-final) obstruents (C_1). In spite of their slight methodological differences, all three studies find that assimilation of C_1 to C_2 does not completely erase underlying [tense] distinctions in C_1 position. In other words, in contrast to prediction (10a) of a lexical feature analysis underlying /k/ + /b/ and /g/ + /b/ sequences do not surface as phonetically indistinguishable [gb]. For example, Barry & Teifour (1999) found that the mean duration of the voiced interval of a [+tense] C_1 fricative followed by a [-tense] obstruent (e.g., /s/ +/d/) was shorter (63 ms) than that of a underlyingly [-tense] fricative in the same context (e.g., /z/ +/d/: 88 ms) but considerably longer than that of a [+tense] fricative followed by another [+tense] obstruent (e.g., /s/ +/t/: 11 ms).

In addition, the study of Catalan by Charles-Luce (1993) provides some evidence against prediction (10b) of the phonological approach. Charles-Luce reports that whereas C_1 voicing in Catalan stop + /s/ and stop + /r/ and the duration of the preceding vowel show the expected assimilatory behaviour (albeit not always in a neutralising fashion), this does not extend to the duration of C_1 itself. A lexical feature analysis predicts that C_1 should be longer before /s/ than before /r/, but Charles-Luce (1993) finds exactly the opposite pattern. [8]

---

[7] Myers (2002) provides quantitative data on English which appear to show that (aspirating) English lax obstruents do have the capacity to trigger RVA. However, my own experimental data, reported in chapter 5, strongly indicate that the overall effect of lax obstruents on the voicing of a preceding obstruent found by Myers is an artefact of a manner-specific effect of English lax *fricatives*. Slis (1987) found no difference in RVA between speakers of (voicing) southern/western dialects and (aspirating) north-eastern varieties of Dutch. However, he failed to establish whether his north-eastern subjects indeed used long lag VOT tense stops and passively voiced lax stops during the experiment, or whether they shifted their pronunciation towards the voicing standard language.

[8] Charles-Luce (1993) seems to treat /r/ on a par with lax obstruents (which are excluded from her experiment) as a trigger of voicing assimilation, presumably because Catalan word-final obstruents are regularly voiced before sonorants as well as before lax obstruents. There are good reasons to distinguish between voicing assimilation to sonorants and voicing assimilation to lax obstruents (see below), but they do not alter the conclusion that the results obtained by Charles-Luce (1993) run counter to a lexical feature analysis of RVA.

Finally, there are reports of regressive assimilation to non-contrastively voiced sonorants, e.g., in Krakow Polish, in Catalan, and in Frisian (fricatives only). This observation might seem to contradict prediction (11c) There is an important argument to treat this phenomenon as distinct from regressive assimilation to actively voiced lenis obstruents, however. It seems that only word-final obstruents that are subject to (dynamic) laryngeal neutralisation can assimilate to a following sonorant: note that all the languages mentioned above neutralise the fortis-lenis contrast word finally. Moreover, the only dialects of Hungarian which exhibit RVA to sonorants are the ones that also have final neutralisation, e.g., the variety of the West Dunántúl region (Kiss, 2001). Varieties of Hungarian that maintain the contrast between tense and lax obstruents word finally only show regressive assimilation to obstruents.

This apparent generalisation suggests an alternative account of voicing assimilation to sonorants, which relies on the idea that neutralised obstruents lack targets for voicing and other correlates of [tense]. Evidence for this idea was reviewed in 3.2.1 above. If neutralised obstruents indeed lack voicing targets, they should show a greater degree of voicing between a vowel and a following sonorant than actively devoiced obstruents, simply as a result of the passive continuation of voicing into the constriction phase. It could well be this increased amount of voicing (relative to utterance-final and __[+tense] contexts) that is interpreted by linguists as voicing assimilation. It could also become a source of confusion to listeners, who might reanalyse all presonorant obstruents (along with obstruents preceding a lax obstruent) as [-tense] on the surface, at least in theory (which would in turn lead to pronunciations that are likely to be interpreted as assimilation by linguists).

According to the first variant of this explanation, word-final neutralised obstruents before sonorants should have less voicing than those preceding an actively voiced obstruent: the latter would ‘inherit’ some of the active voicing of the following sound in addition to the passive voicing spilling over from the preceding vowel or sonorant consonant. According to the second variant there should be at least a group of languages for which this description holds, i.e., languages such as Dutch, which have final neutralisation, but on most descriptions lack RVA to sonorants. Chapter 7 reports direct evidence for the prediction that neutralised obstruents have more voicing when followed by a sonorant consonant than before a tense obstruent, but less than when followed by an actively voiced lenis stop.[9]

---

[9] The idea that neutralised stops can undergo passive voicing between voiced sonorants and thus appear to exhibit assimilation of voicing is not new: it is central to the account of English flapping in Harris (1994).

4.5 Summary

The aim of this chapter was to dissect voicing assimilation, a phenomenon that might be regarded by some as one of the better understood phonological processes. Drawing from the theoretical treatment of sandhi processes more in general, I first defined two phonetic templates: one for a (lexical) phonological form of voicing assimilation and one for a purely coarticulatory form. Subsequent chapters then matched descriptions of various assimilation processes against these templates.

The first set of processes I considered involved the voicing assimilation rules found in the morphological paradigms of the Germanic and other languages. These processes appear to be phonological in nature, although as yet no clinching quantitative evidence is available. Second, in many instances, the phonetic characteristics of progressive devoicing rules that operate across word boundaries suggest that they are fully explained in terms of passive devoicing. Third, regressive voicing assimilation across word boundaries is clearly conditioned by the phonetic properties of the trigger obstruents in a way that suggests that they are driven by coarticulation, or at least diachronically rooted in coarticulation. This suggestion is bolstered by quantitative evidence indicating that several RVA processes that operate across word boundaries across word boundaries are incompletely neutralising. The next three chapters represent an attempt to extend the quantitative evidence in this area.

Chapter 5

Experiment 1: regressive voicing assimilation in English

In chapter 4 I described three phonetic characteristics that a voicing assimilation rule would need to be classified as coarticulation-based: (1) since coarticulatory voicing assimilation is likely to be driven mainly by the articulatory gestures involved in the production of voicing distinctions it should be triggered by actively (de)voiced obstruents only; (2) the process should be reflected only by the voicing of target obstruents, and those features that are mechanically linked to voicing, such as frication duration; (3) the process does not categorically erase phonetic distinctions between tense and lax target obstruents (cf. 11 in 4.1.2).

The aim of this chapter is to test the three main predictions of a coarticulation-based view of RVA across word boundaries by means of an acoustic investigation of regressive assimilation in English. Although English is often regarded as a language with little or no voicing assimilation at word boundaries, it does in fact allow for all three predictions to be tested because English possesses both actively voiced and actively devoiced sounds, and because it maintains a contrast between fortis and lenis obstruents in word-final contexts.

The two experimental results reported below broadly support the hypothesis formulated in in 4.4 above which holds that RVA at word boundaries is an articulatory process: Actively voiced English /z/, and to a lesser extent, actively devoiced /t, s/ all appear to trigger some form of RVA, in contrast to passively voiced /d/. The phonetic reflexes of this process are mostly limited to the phonetic voicing of the target obstruent, and consequently the process is non-neutralising.

5.1 Predictions

One of the principal predictions of the phonetic theory of RVA described in 4.1.2 is that the capacity of a sound for triggering assimilation is a function of its voicing target: actively devoiced sounds should trigger devoicing, actively voiced sounds are expected to increase the voicing of a preceding sound if possible, whilst passively voiced sounds (sounds lacking a voicing target) should not affect the voicing of a preceding sound. The specific predictions that can be derived from this theory for aspirating varieties of English are listed in (15). Fortis obstruents /p, t, k, t͡ʃ, f, s, ʃ/ are expected to cause some degree of devoicing in a preceding obstruent, because they are likely to be actively devoiced. Given that the lax fricatives /v, z, ʒ/ are actively voiced (see section 2.3), these sounds should cause an increase in the duration of the voicing interval of a preceding obstruent. The lax plosives /b, d, ɡ, d͡ʒ/ on the other hand, should act as an intermediate, ‘neutral’ environment for a preceding obstruent similar to that provided by a following sonorant. Both the sonorants (cf. 2.1) and the word-initial lax plosives of aspirating varieties of English (2.2.1) are arguably passively voiced, which means that they lack articulatory targets gestures related to the production of voicing distinctions, and consequently they are unable to pass on such gestures to neighbouring sounds by means of coarticulation.

(15) Predictions of a coarticulation-based approach to voicing assimilation regarding obstruent sequences in aspirating varieties of English

a. English obstruents fall into 3 classes in terms of their influence on (the voicing of) a preceding obstruent. (1) fortis obstruents trigger devoicing; (2) lenis *fricatives* cause an increase in the voicing of a preceding obstruent; (3) *lenis* stops behave as an intermediate, ‘neutral’ category, as do sonorants. Cf. (11c)

b. The assimilatory effects of fortis obstruents and lenis fricatives are limited to voicing and features mechanically dependent on the production of voicing distinctions. Cf. (11a)

c. English RVA at word boundaries is a gradient process that is not neutralising in most instances Cf. (11b)

In addition, it follows from the mechanism underpinning RVA in the phonetic theory, first, that the effects of tense obstruents and lax fricatives on a preceding obstruent should be limited to voicing and those phonetic features mechanically dependent on the production of voicing distinctions. Second, the phonetic theory predicts that RVA across word boundaries in English is a gradient process which is non-neutralising on most occasions, even for the phonetic features that it does affect.

By contrast, under the strict interpretation of a lexical feature analysis assumed in 4.1.1, differences in voicing between lax plosives and fricatives are predicted to have no impact on the occurrence of RVA: either the process applies in a manner-symmetric fashion, or it does not apply at all before lax obstruents (see prediction 10c in 4.1.1). Second, even if a lexical feature model can *represent* manner-asymmetric RVA at word boundaries it still predicts that where it applies, the process affects all phonetic cues to [tense] (10a). Third, it predicts that, as a result, even manner-asymmetric RVA will always act in a phonetically neutralising fashion.

The acoustic investigation reported below was specifically designed to test these two contrasting sets of three predictions. The results of this investigation indicate that the behaviour of English velar stop + /t, d, s, z, r/ sequences is as predicted by the phonetic theory on all three counts in (15), and therefore warrant a revision of traditional descriptions of voicing assimilation in English. These traditional descriptions, and some of the relevant experimental literature is reviewed in the next section.

5.2 Previous descriptions of the phonetics of English obstruent clusters

As noted at several points in the previous chapters, impressionistic accounts usually describe standard varieties of English as possessing little or no regressive voicing assimilation across word boundaries. Indeed, Jones (1956) warns native speakers of French and Dutch against making the mistake of applying RVA in English obstruent clusters (e.g., Jones's §851). This point is echoed by Gimson (1994), who does, however, claim that at the boundaries of 'close-knit' groups of words, lenis fricatives (but normally not lenis stops) may devoice completely when preceding a fortis obstruent. Moreover, according to Gimson, this form of RVA may also shorten the preceding vowel, although this phenomenon is judged to be "relatively rare" (Gimson 1994:257). This description of RVA to English fortis fricatives is reminiscent of the account of a regressive devoicing rule found in Yorkshire dialects of English provided by Wells (1982a). According to Wells's description, this rule is triggered by all fortis obstruents and affects voicing as well as durational correlates of [tense].

Instances of regressive assimilation to lax obstruents documented in impressionistic accounts invariably involve function words or word internal contexts. One instance is the voiced pronunciation of plosive + alveolar fricative clusters in words containing orthographic <x> such as <exam>, [ɪgzæm]; <exhibit>, [ɪgzɪbɪt]; <excerpt> [ɛgzɜːpt]. These clusters presumably originate from an older (and invariant) form [ks] which was subsequently affected by the process of [s]-voicing that is also responsible for the pronunciation of e.g., <desire> as [d̥ɪzaɪə]. According to Borowsky (2000) the (historical) voicing of word-medial prestress /s/ in turn triggers voiced realisations of the preceding /k/. The same author discusses a second example of apparent word internal RVA in English: the optional voiced realisation of the final alveolar fricative of prefixal <dis> before lax stops, as in [dɪzɡ aɪz] for <disguise>. This observation seems to directly contravene a phonetic approach to RVA, which predicts that English lax stops should be unable to act as triggers of the process.

Borowsky (2000) grants that the voiced realisations of orthographic <x> and the alveolar fricative of <dis> are optional, but her descriptions nevertheless fail to do justice to nature of voicing patterns in English medial obstruents and obstruent clusters. A number of additional observations cast doubt on her claim that the optional voicing of the medial fricative in <disguise> is due to the same (synchronic) assimilatory mechanism that underlies the (also optional) voicing of the medial /s/ in Dutch /mɪs/ + /daːd/, [mɪzdaːt], *crime*. For example, judging by pronunciation dictionaries such as Windsor Lewis (1972), Jones (1977), and Wells (2000), the English process idiosyncratically affects the final sibilant of <dis> before /ɡ/ and /d/ (e.g., <disdain>) but not before /b/ or /d͡ʒ/ (cf. <disbar, disband, disjoin, disjunct>), although this apparent place of articulation effect may be an artefact of morphological transparency and/or lexical frequency. Furthermore, the final sibilant of <mis-> in e.g., <misguided>, <misgiving> is normally realised as [s] rather than [z]. Dutch RVA by contrast, is not lexically selective in this way.

More importantly, the correct generalisation about the voicing of English orthographic <s> seems to be that it can occur before sonorants as well as lax obstruents, but not before tense obstruents. Thus, the examples in (7c) in the previous chapter match <Osborne>, <Osgood>, <Marsden>, <Neasden>, which all have [z] rather than [s], but e.g., <Oscar, osprey> are normally pronounced with [s]. Similarly, as illustrated in (16), postpausal and postsonorant weak forms of <is> can be realised with [z] before lax obstruents and sonorants, but not before tense obstruents (Lakoff, 1972; Selkirk, 1980).

(16) Voicing of weak <is> in English (examples from Selkirk 1980)

Orthography	Phonetic form
<Is Jack going?>	[(ɪ)zd̥͡ʒ̊ækɡ̊oʊɪŋ]
<Is Will going?>	[(ɪ)zwɪɫɡ̊oʊɪŋ]
<Is Pete going?>	[(ɪ)spʰiːtɡ̊oʊɪŋ]

As argued in chapter 4, observations that (neutralised or weakened) obstruents are voiced before both lax obstruents and sonorants do not necessarily imply that either or both of these classes actively contribute to the effect. Moreover, such observations are consistent with a coarticulation-based approach to RVA as described in 4.1.2 to the extent that the voicing process is motivated independently (e.g., by passive voicing). Given that English alveolar sibilant voicing is indeed motivated independently, an articulation-driven account of its origins would say that the process applies freely before passively voiced lax obstruents and (equally passively voiced) sonorants because they cannot influence the voicing of a preceding sound (cf. prediction 15a), but that it is blocked by assimilation to actively devoiced fortis obstruents, which can. In other words, the observation that words such as <Osborne> and <disguise> are commonly produced with [z] does not necessarily constitute evidence against a phonetic view of regressive voicing assimilation in English.

The picture of English voicing assimilation I have drawn so far is broadly speaking reflected in the generative literature on the topic, which tends to classify English as a language without RVA across word boundaries and sometimes as a language in which only the laryngeal specification of fortis obstruents is visible to the phonology (Harris 1994; cf. chapter 8). Perhaps in part because of this picture, instrumental investigation of English obstruent clusters with mixed underlying [tense] specification has been limited, and in all but one case that I am aware of, does not allow for the specific predictions of the phonetic theory of RVA to be tested against those of a lexical feature analysis. However, the picture emerging from the single study in question is considerably more encouraging for the phonetic theory than the one sketched by impressionistic accounts.

The quantitative (acoustic) study of laryngeal contrast and voicing in American English fricatives conducted by Stevens et al. (1992) shows a clear effect of following context on the voicing of fortis and lenis fricatives. For example, in their corpus lenis fricatives (/v, z/) have on average 29 ms of voicing preceding a fortis fricative (/f, s/), which increases to 58 ms before another lenis fricative or a vowel. However, since Stevens et al. (1992) do not provide separate means for fricative + vowel sequences and homogeneous (tense + tense and lax + lax) clusters, there is no baseline measure to determine whether the shorter voicing intervals in the fortis environments are the result from some form of active devoicing or whether the greater amount of voicing in the lenis contexts is the result of RVA to the lenis fricatives, or both. Furthermore, Stevens et al. (1992) do not provide tests of the statistical significance of the differences in the mean voicing values they observe.

Although statistical tests are provided in an acoustic investigation of English velar obstruents in various contexts by Myers (2002), it does not allow testing of the phonetic theory of RVA either. Whilst Myers's test stimuli contain both /z/ and /ɡ/ as following obstruents, his (statistical) analysis does not distinguish between these two environments. Consequently, it is impossible to determine whether the slight increase in voicing he observes before lax obstruents is due to a symmetric effect of /ɡ/ and /z/ or an asymmetric effect of /z/.

However, an early and all but forgotten study by N. Thorsen (1971) (and one that I have only become aware of after most of the work reported below

Table 5.1: Voicing duration (ms) and ratio of the closed phases of unstressed English /t/ and /d/ followed by a C_2 in the onset of a stressed syllable, as reported by N. Thorsen (1971).

C_1C_2	C_1 voicing		$C_1 + C_2$	C_1 voicing	
	Duration	Ratio		Duration	Ratio
/tl/	35	0.53	/dl/	56	1.00
/tk/	31	0.66	/dk/	44	0.86
/tg/	41	0.87	/dg/	51	0.91
/ts/	33	0.60	/ds/	50	0.79
/tz/	64	0.82	/dz/	62	0.96

had been completed) does shed considerable light on the issues raised by the phonetic theory. This study investigates voicing and other phonetic features of [tense] in English alveolar stop C_1 + consonant C_2 sequences straddling word and morpheme boundaries in three different prosodic contexts, and crucially reports measurements for stop, fricative, and sonorant C_2s separately. Some of the mean values reported by Thorsen are represented in table 5.1.

Although the match between this voicing data and the three term classification in (15a) is imperfect, it is striking how the absolute durations of the voiced intervals of /t/ and /d/ are longer before /z/ than before /g/. Moreover, the mean C_1 voicing value for /d/ +/l/ is more or less intermediate between those for /d/ + /z/ on the one hand and /d/ + /s/ and /d/ +/k/ on the other, although /d/ +/g/ would be expected to group with /d/ +/l/ rather than with /d/ +/s/. Absolute voicing duration in /t/ is not as predicted in (15a) to the degree that /l/ appears to pattern with fortis obstruent C_2s rather than with /g/. In addition, /t/ is relatively short before /g/, and consequently its *voicing ratio* (duration of the voiced interval divided by overall duration of the closed phase) is higher there than before /z/. But note that voicing ratio is only a good indicator of 'degree of assimilation' if both overall duration and voicing duration generally pattern in the fashion predicted by a lexical feature analysis.[1]

In addition to this (limited) evidence for manner-asymmetric RVA in English, N. Thorsen (1971) provides evidence in support of (15b) and especially (15c). No effects of assimilation are discernible in the durations of the closure phases of /t/ and /d/, whilst the length of the vowels preceding the clusters in table 5.1 clearly preserves the contrast between /t/ and /d/ (all differences significant at $p < .01$). Note however, that there is a 13 ms difference between the vowels preceding /t/ + /s/ (32 ms) and /t/ + /z/ (45 ms) in the direction

---

[1] Interestingly, of the means provided in table 5.1 the only pairwise difference(s) that N. Thorsen (1971) lists as statistically significant at the $p < .05$ level is the difference in C_1 voicing duration and/or ratio between /d/ + /s/ and /d/ +/z/ (the type of test is not specified).

predicted by a lexical feature account.

5.3 Methods

Subjects Subjects were 4 native speakers (2 male, 2 female) of British English aged between 24 and 35, and living in or near to London at the time of recording. None of the subjects had a history of speech or hearing impairment. They were not paid for their participation in the experiment. 3 subjects, K6, L7 (both female) and R10 (male) were speakers of a south-eastern variety of British English, while the speech of the remaining subject J11 (male), displayed some characteristics of his native Lincolnshire although it had no strong local features. All 4 subjects were non-rhotic.

Materials The stimuli for this experiment consisted of consonant clusters combining a /k, ɡ, ŋ/ C_1 and a /t, d ,s, z, r/ C_2. Velar stop C_1 were preceded by a long central mid open vowel [ɜː] (V_1)[2], while /ŋ/ followed low back rounded /ɒ/ (V_2). C_2 was always followed by a vowel.

The main reason to use velar rather than alveolar C_1 was that word-final /t/ is often realised as a glottal stop in British English. A different place of articulation was chosen for the C_1 consonants for segmentation purposes; the choice for velar stops over labial ones was determined by the desire to control for the preceding vowel. The choice to use alveolar C_2s was made partially because of the exceptional behaviour of lenis labiodental fricatives with regard to RVA in a number of languages (e.g., Hungarian, Russian), and partially because some claims about the phonetic basis for the nature of laryngeal contrast in fricatives have been made with specific reference to sibilants (Balise & Diehl , 1994).

(17) English sample stimuli

a. How does patchwor*k d*uvet translate?

b. How does headstro*ng z*ealot translate?

c. How does Hambur*g d*airy translate?

The clusters were located at the internal boundary of noun + noun (N_1 + N_2) constructions and further embedded within a carrier phrase (*How does _ translate?*) designed to attract nuclear stress on the second noun. Both N_1s and N_2s were disyllabic with an initial lexical stress. Thus, the rhythmic structure of the stimuli and nuclear accent placement were controlled to maximise the potential effect of RVA, which has been shown to depend on lexical stress in Dutch (see

---

[2] This vowel is transcribed in square brackets in order to side-step questions about the underlying representation of orthographic <V + r + C> sequences, as in, e.g., <work>. Note that all subjects realised such sequences with a long vowel rather than [V + ɹ].

Slis 1986 and chapter 4). Given the sparsity of English words beginning with /z/ no attempt was made to control for the lexical frequency of the target words N_1 and N_2. For each of the 15 different consonant clusters 4 stimuli were prepared. Some sample stimuli are given in 17, with target consonant clusters in a slanted font.[3]

Stimuli containing the sonorant consonants /ŋ, r/ (realised by all subjects as [ŋ, ɹ] in word-final and word-initial contexts respectively) were included to create baseline conditions for the comparison of the relative effects of fortis vs. lenis C_2 on the properties of a preceding obstruent.

Procedure The stimuli were presented to the subjects in a quasi-randomised order to avoid consecutive stimuli with identical consonant clusters. The subjects were asked to repeat each stimulus three times at a comfortable rate and to read a stimulus again if they made a mistake or produced a hesitation. In total, 3 (C_1) * 5 (C_2) * 4 (stimuli) * 3 (repetitions) * 4 (speakers) = 720 utterances were recorded.

Recordings were made onto minidisk in a sound-proofed room using a Brüel and Kjær condenser microphone (Type 4165) and measuring amplifier (Type 2609), and digitised at 22.5 kHz. Segmentation and acoustic measurements were carried out using PRAAT. 23 utterances had to be discarded because they contained small speech errors or (hesitation) pauses between C_1 and C_2 and 37 utterances were excluded because an underlying /k/ was realised as a glottal stop. In addition, all (remaining) clusters starting with a /ŋ/ are excluded from the discussion below because they are largely irrelevant to the hypotheses under consideration, leaving a total of 425 utterances for analysis.

Segmentation and measurements Segment boundaries were determined by visual inspection of waveforms and broadband spectrograms based on Fast Fourier Transforms (*FFT*) on a 5 ms Gaussian window (spectrogram bandwidth 260 Hz). The boundary between a vowel and a following plosive C_1 was placed where there was an abrupt change in the higher frequency energy, as illustrated by figure 5.1. The boundary between a C_1 plosive and a following C_2 was placed at the end of the release burst of the plosive, where *release burst* was defined as the initial transient and any following frication that could be assigned to C_1 rather than to a C_2 fricative.

59% (101 out of 171) of plosive-plosive clusters had a clear C_1 release and could therefore be internally segmented according to these criteria. In the remaining utterances where this was not the case, no boundary was placed and voicing and duration characteristics were measured for the cluster as a whole. In a few cases, mainly involving /g/ followed by /z/, the initial plosive was

---

[3] A full list of stimuli appears in appendix A.

followed by a short period of schwa-like voicing. These intervals were treated as voiced releases (analogously to the 'embryonic vowels' often observed after word-final lenis stops in French), and consequently their duration and voicing were assigned to C_1. In another set of tokens the release was completely obscured by the noise of a following fricative. Here the boundary was set at the onset of the noise signal.

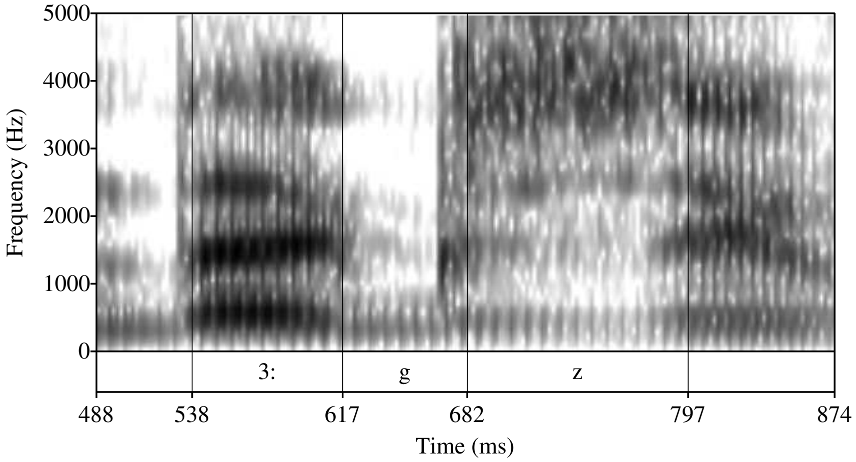

Figure 5.1: Sample spectrogram of an English /gz/ cluster. Speaker: R10 (male).

The offset of C_2 constriction was defined as the offset of frication for /s, z/ and the onset of the release burst for /t, d/. The first measurement point for F_0 was placed at 10 ms from the offset of frication for fricatives and at 10 ms after the onset of post-release voicing for plosives.

Voicing measures were determined on the basis of periodicity in the waveform and the presence of a voice bar in the spectrogram. Note that on the basis of these criteria the /gz/ cluster in figure 5.1 is voiced throughout. VOT was defined in the standard way in terms of the timing of voicing onset relative to the onset of the release burst in plosives.

The measurements that were made on the basis of the hand-segmented speech samples, as well as the relevant derived measures are listed in table 5.2, ordered by speech segment. As described in 2.2.3, preceding vowel (V_1) duration is a crosslinguistically recurrent feature of [±tense], which is generally considered to be particularly salient in English (Chen, 1970; Flege & Hillenbrand, 1987). F_0 and F_1 values were extracted at 10 ms intervals between 50 and 10 ms preceding the onset of C_1 closure, using the autocorrelation and Burg algorithms embedded in PRAAT 4.0. C_1 closure duration and release duration were measured separately for two reasons. The first is that they are not necessarily part of the same cue in articulatory or perceptual terms. In theory it is therefore possible that only one of these two features turns out to cue [tense] in the subjects' speech, and in this case simply measuring overall C_1 duration might lead to a distorted picture of the reflexes of [tense] marking on C_1 and/or C_2 in terms of segmental duration. The second, more practical reason for considering C_1 closure and release duration separately is that when a stop is followed by a fricative, the release noise of the former may be partially or wholly obscured by frication noise of the latter. Again, this might distort the interpretation of C_1 overall duration. On identical grounds, C_1 voicing measures are reported separately for closure and release phases.[4]

Table 5.2: Acoustic measurements and derived measures for Experiment 1.

Segment							
	V_1		C_1		C_2		V_2
(a)	Duration	(d)	Closure duration	(i)	Closure duration (stops)	(n)	F_0 10-50 ms after C_1 offset
(b)	F_0 50-10 ms before C_1 onset	(e)	Release duration	(j)	Overall duration (fricatives)		
(c)	F_1 50-10 ms before C_1 onset	(f)	Overall duration	(k)	Voicing duration (fricatives)		
		(g)	Voicing duration (2 m.)	(l)	Voicing ratio (fricatives)		
		(h)	Voicing ratio (2 m.)	(m)	VOT (stops)		

The nature of the phonetic theory of RVA demands that the phonetic features of C_2 be examined as well. It is particularly important to determine whether the subjects indeed produce tense /t/ with a long lag VOT and lax /d/ without closure voicing and a short lag VOT. Similarly, it is important to establish whether lenis /z/ has any voicing in a postobstruent environment since it would point to

---

[4] Only absolute voicing durations are reported in the main text of this chapter. C_1 voicing *ratios* were also calculated because they are sometimes used as a measure of RVA in other experimental studies. The interested reader can consult them in appendix A. My main motivation for focusing on absolute values of duration and voicing rather than voicing ratios is that the latter type of measure combines two acoustic features of [tense] in way that inflates the distance between two set of obstruents if both its components behave as they do in 'typical' cases of intervocalic fortis-lenis contrast. The relatively short duration and large amount of voicing of lenis obstruents both contribute to a relatively high voicing ratio, whilst the long duration and lack of voicing of fortis obstruents both contribute to a low voicing ratio. However, if either absolute duration or absolute voicing duration behaves contrary to the 'typical' pattern, the effects of underlying [±tense] or RVA on one feature may be (partially) canceled by the other and voicing ratio ceases to be a reliable measure.

the presence of active voicing that is critical to the predictions of the phonetic theory. The phonetic description of C_2 below also includes measurements of segmental duration and F_0 perturbations in the following vowel, but not attempt was made to determine F_1 contours, as V_2 vowel quality was not controlled for.

5.4 Results

5.4.1 Phonetic features of C_2

The data in table 5.3 and figure 5.2 indicates that the subjects use voicing distinctions to signal the distinction between tense and lax stops and fricatives as would be expected of an aspirating language. Thus, stops /t/ and /d/ can be characterised as voiceless aspirated vs. (passively) voiceless whilst the contrast between /s/ and /z/ is realised as voiceless vs. (partially) voiced. This means that prediction (15a) above is indeed applicable to the present corpus.

The mean VOTs for /t/ (70 ms) and /d/ (14 ms) fall into the standard ranges for the long lag and short lag categories, and the difference between them is highly significant according to a t-test: $t(99) = 16.18$, $p < .001$. All tokens of tense /s/ are completely voiceless, whilst /z/ has a substantial amount of voicing (77 ms). The mean voicing ratio for this obstruent is .78 (standard deviation: .22), which is fairly high in comparison with earlier studies such as Haggard (1978) or Smith (1996). Unsurprisingly, the mean difference in absolute voicing duration is statistically highly significant: $t(173) = -23.62$, $p < .001$.[5]

Table 5.3: Experiment 1: duration and voicing of C_2. Closure duration and VOT of /t, d/, and overall duration and duration of the voiced interval for /s, z/. All values in ms, and pooled across preceding contexts (/k, g/). Standard deviations in brackets.

C_2	VOT	Closure duration	N
/t/	70 (15)	56 (16)	44
/d/	14 (19)	71 (15)	57
	Voicing	Duration	N
/s/	0 (0)	132 (18)	92
/z/	77 (31)	99 (17)	83

F_0 microprosody seems to signal the distinction between tense and lax C_2 obstruents, too. Figure 5.2 plots the mean F_0 at five measurement points from 10 to 50 ms into the vowel following C_2 for the two male subjects R10 and J11. It

---

[5] All data on stop + stop clusters in this section pertain to sequences that could be internally segmented, unless indicated otherwise. The result is a fairly large discrepancy in the number of cases for plosive and fricative C_2s.

shows how 10 ms into the vowel, F_0 values for /t, s/ on the one hand and /d, z, r/ on the other are roughly 20-25 Hz apart, and then gradually converge as time progresses. Both the magnitude of the F_0 differences and the grouping of lenis (passively devoiced) /d/ and (actively voiced) /z/ with sonorant (passively voiced) /r/ are in line with earlier observations in the literature. A one-way ANOVA for *C_2 laryngeal specification* (tense obstruents vs. lax obstruents vs. sonorant) confirms that there is a highly significant effect of the phonological status of C_2 on F_0 at the onset of a following vowel: F(1,171) = 24.05, $p < .001$. Tukey and Scheffe post hoc tests indicate that lax /d, z/ and /r/ are both significantly distinct from tense /t, s/ (both $p < .001$) but not from each other.[6]

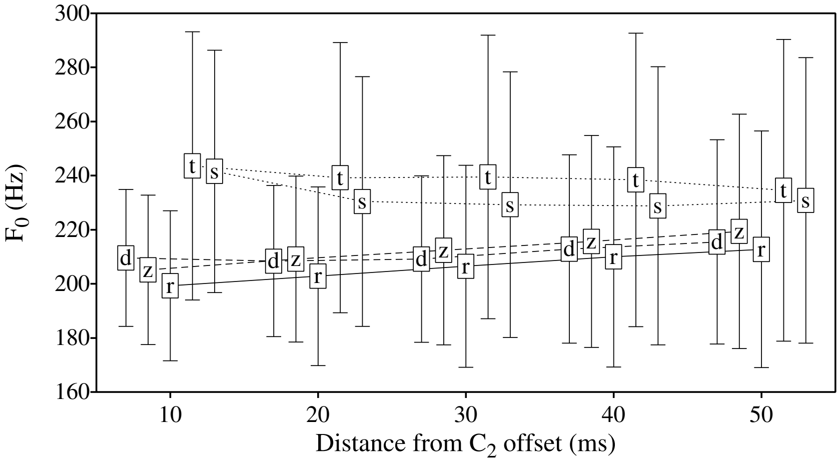

Figure 5.2: Experiment 1: F_0 (Hz) 10-50 ms into the vowel following C_2, for female speakers only. Both internally segmented and unsegmented stop + stop clusters included. Error bars represent the mean ±1 standard deviation.

Finally, fricative (frication) duration but not stop closure duration behave according to the typical [±tense] pattern. On average, /s/ is 33 ms longer than /z/, and this difference is highly significant according to a t-test, t(137) = 12.51, $p < .001$. The 15 ms difference in closure duration between /t/ and /d/ is also statistically significant (t(99) = -4.94, $p < .001$) but patterns in the 'wrong' direction. As closure duration is not known to be a cue to [±tense] in word-initial contexts this finding hardly topples any theories (cf. 2.2.3), and it is hard

---

[6] Utterances from the two female speakers were excluded from figure 5.2 and the ANOVA because of a considerable difference in overall F_0 level between the male and female subjects. However, the behaviour of the female subjects with regard to post-C_2 F_0 perturbations is highly similar to that of the female subjects, with a maximal difference between /d, z, r/ and /t, s/ of approximately 35 Hz.

to say whether any meaningful interpretation can be assigned to it.

5.4.2 Phonetic features of C_1 in the baseline environment

Sonorant /r/ was included as a baseline C_2 environment for the phonetic expression of the contrast between /k/ and /g/. Since /r/ is both passively voiced and phonologically unmarked for [tense] it can be treated as a ‘neutral’ context for the purposes of both the phonetic theory and lexical feature analyses of RVA.

The data summarised in figures 5.3, 5.4, 5.6 and 5.5 further below indicates that the 4 subjects use many of the phonetic features reviewed in chapter 2 to distinguish /k/ from /g/ before sonorants.[7] For example, the bottom two bars of the top panel of figure 5.6 show how /g/ has a shorter release phase than /k/ in this environment (21 vs. 33 ms), and a marginally shorter closure stage, too (44 vs. 50 ms). Similarly, the bottom two bars of figure 5.5 show that the mean duration of vowels preceding /g/ and /k/ pattern as would be expected on the basis of the literature: on average [ɜː] is 27 ms longer before the lax stop than before its tense counterpart (99 vs. 72 ms). Both the release duration and preceding vowel duration differences are highly significant according to t-tests: t(77) = 4.54, $p < .001$ and t(77) = -5.47, $p < .001$. The difference is C_1 closure duration is also statistically significant, but the effect is clearly less strong: t(77) = 2.41, $p < .02$.

The bottom panel of figure 5.4 shows that there is a difference of 21 ms in overall voicing duration between /g/ and /k/ before /r/ (43 vs. 22 ms). The difference in overall voicing ratio is .43 (.27 vs. .70: note that this difference is similar to that obtained by N. Thorsen 1971). All possible measures of C_1 voicing yield statistically highly significant differences between /k/ and /g/. For instance the t-test result for closure voicing duration is t(77) = -5.53, $p < .001$. It is often suggested that the duration of the preceding vowel is the primary cue to [tense] in English word-final obstruents, but these findings suggest that it is, or can be, supported by voicing distinctions.

Of the remaining components of the low frequency feature proposed by Kingston & Diehl (1994, 1995), the test subjects only appear to employ F_1 perturbation (in the present context: see further below). Figure 5.3 plots the first formant contour of the vowel [ɜ] at 10 ms intervals between 50 and 10 ms preceding the onset of C_1. The downward slope of this contour is steeper before /g/ than before /k/, which results in a 26 Hz difference (476 vs. 502 Hz) at 10 ms before the onset of C_1. This pattern agrees with data reported in the literature on the topic (cf. Stevens et al. 1992 and other references mentioned in section 2.2.3 above) and the same applies to differences in F_0 at 10 ms before the onset

---

[7] Exact values for the standard deviations indicated by the error bars in figures 5.4, 5.6 and 5.5 are given in appendix A.

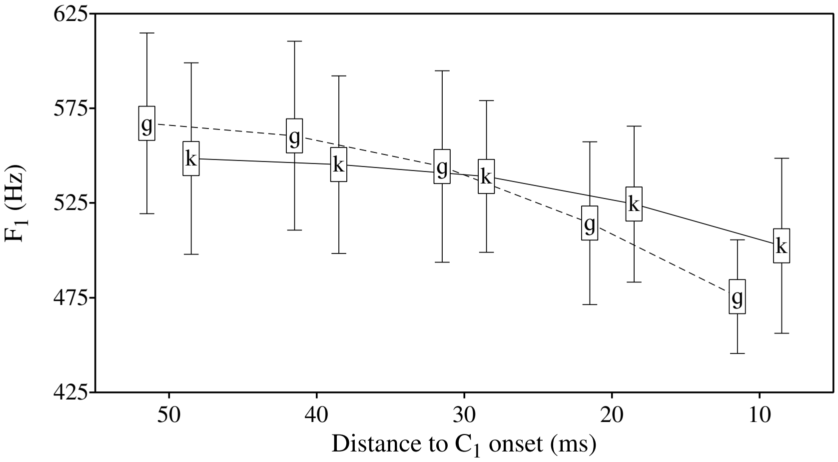

Figure 5.3: Experiment 1: F_1 (Hz) at 5 points from 50-10 ms preceding C_1 onset in /k/ + /r/ and /g/ + /r/ sequences. Error bars represent the mean ±1 standard deviation.

of /k, g/: 192 vs. 183 Hz for the female speakers and 150 vs. 138 Hz for the two male subjects. However, only the difference in F_1 (at 10 ms) is statistically significant according to a t-test: t(77) = 3.13, $p < .005$.

Interestingly, correlations between the individual phonetic features discussed here are generally weak and often not statistically significant. This holds in particular for correlations between V_1 duration and the length of (parts of) C_1. The two strongest correlations are between F_1 value at 10 ms before C_1 and V_1 duration (r = -.35, $p < .005$) and between the former of these and overall C_1 voicing (r = -.32, $p < .005$). The general absence of strong correlations between the values of the individual phonetic features of [tense] is consistent with a view in which they are traded off against each other for perceptual reasons (and manipulated independently). It is inconsistent with models that seek to reduce the cluster of correlates of [tense] to the reflexes of a single or relatively few articulatory gestures.

5.4.3 C_1 voicing in obstruent clusters

Having established the phonetic features of tense and lax velar stops in the baseline environment, it is now possible to assess whether they undergo any form of assimilation in potentially non-neutral environments. The patterning of C_1 voicing in obstruent clusters shows that this is indeed the case, and in a fashion that is entirely consistent with prediction (15a) of the phonetic theory. There is an increase in voicing duration before /z/ but not before /d/ vis-à-vis the __/r/ baseline environment. In addition, the voicing data lends some support to prediction (15c) since voicing duration appears to partially preserve the underlying distinction between /k/ and /g/, even where RVA does apply.

Figure 5.4 provides the means for the duration of the voiced intervals of the C_1 closure and release stages as segments of bars indicating the overall voicing duration of C_1. The most striking generalisation emerging from the data in this figure is the difference in voicing between velar obstruents preceding /r/ and /d/ on the one hand, and /z/ on the other. For example, there is barely any difference in the C_1 closure (1 ms) or C_1 overall (3 ms) voicing of /kd/ and /kr/ sequences, whilst the same measures show a marked increase in voicing in /kz/ clusters. Clusters starting with /g/ behave in exactly the same fashion.

The data in figure 5.4 prompts two additional observations. First, the fortis obstruents /t/ and /s/ appear to have an assimilatory effect on the voicing of a preceding /g/. Relative to the baseline context, the length of the voiced interval of the closure stage of /g/ drops by 10 and 12 ms before /s/ and /t/ respectively. On the other hand, /t/ and /s/ have little effect on the voicing of a preceding /k/. Second, voicing duration preserves the contrast between underlying /k/ and /g/ before /d/, where no assimilation occurs, but there is a hint that even where assimilation does occur, voicing distinctions are incompletely neutralised. Thus, the average overall voicing duration of /g/ is marginally greater than that of /k/ before /t, s, z/.

A number of ANOVAs were carried out to examine whether these impressionistic observations stand up to statistical scrutiny. A first set of three-way ANOVAs for *C_2 laryngeal specification* (/t, s/ vs. /d, z/) * *C_2 manner of articulation* (/t, d/ vs. /s, z/) * *C_1 laryngeal specification* (/k/ vs. /g/) was performed on the voicing data for clusters composed of /k, g/ and /t, s, d, z/. The goal of these ANOVAs was to assess the apparent effects of regressive voicing assimilation and incomplete neutralisation in obstruent sequences. Their results indicate that RVA is indeed manner-asymmetric in the obstruent clusters produced by the test subjects, as predicted by the phonetic theory.

For example, the ANOVA on the C_1 closure voicing data shows highly significant main effects of *C_2 laryngeal specification*, F(1,268) = 73.75, $p < .001$; *C_2 manner of articulation*, F(1,268) = 31.11, $p < .001$; and *C_1 laryngeal specification*, F(1,268) = 12.71, $p < .001$. An ANOVA for C_1 overall voicing (duration + release) duration yields equally significant main effects, and both ANOVAs show a strong interaction between *C_2 laryngeal specification* and *C_2 manner of articulation*: F(1,268) = 23.78, $p < .001$ (closure voicing), and F(1,268) = 28.81, $p < .001$ (overall voicing). The main effects of *C_2 laryngeal specification* and *C_2 manner of articulation* support the idea that the voicing of C_1 is subject to assimilation to a following obstruent. The strong interactions between *C_2 laryn-*

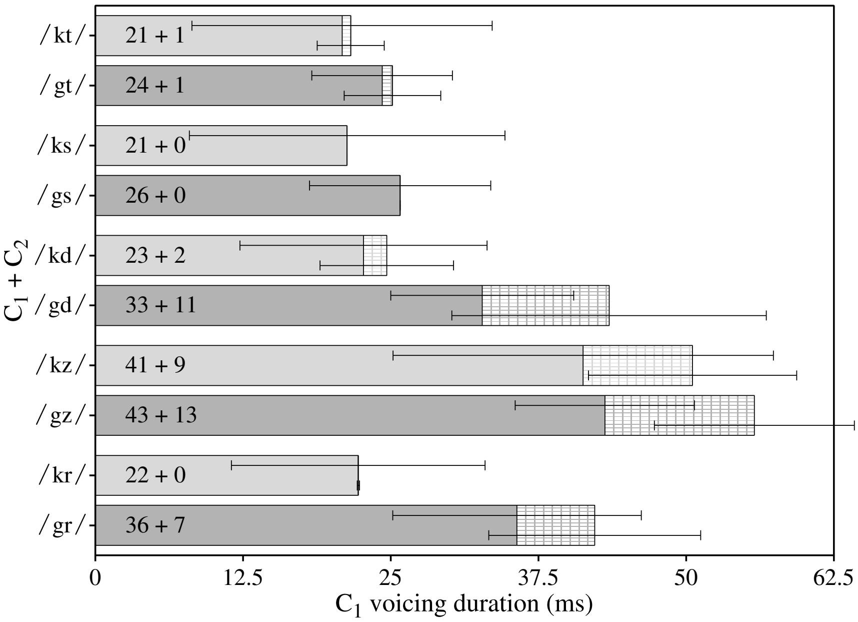

Figure 5.4: C_1 voicing duration across C_2 contexts. All measures in ms; error bars represent the mean ± 1 standard deviation. The top panel represents mean C_1 closure and release voicing durations separately: for each bar the left-hand segment indicates the extent of voicing during the closure phase and the right hand segment represents the duration of voicing during the release stage. Exact values for these means are printed in the leftmost segment for typographical reasons. This diagram shows a marked increase in voicing, relative to the baseline environment, for both /k/ and /g/ before /z/. There is a small decrease in the voicing of /g/ when it precedes /t, s/. These observations suggest that /z/ and, to a lesser extent, the two fortis obstruents, trigger RVA. Before /d/ /k/ and /g/ behave more or less as in the baseline environment, which is an indication that the lenis plosive is unable to trigger voicing assimilation.

geal specification and *C_2 manner of articulation* indicate that both main effects are largely caused by a single laryngeal/manner class, and therefore that only 1 of the 4 obstruents under investigation is a strong trigger of assimilation. Given that it induces the greatest deviations in C_1 voicing (duration) from the baseline condition, this strong trigger is most likely to be /z/.

However, the devoicing of /g/ before /t, s/ probably also contributes something to the main effects of *C_2 laryngeal specification*. A second series of three-way ANOVAs on the closure and overall voicing data summarised in figure 5.4

with /r/ included as a separate C_2 laryngeal category ([0tense]). Tukey and Scheffe post-hoc tests on these ANOVAs show that both tense and lax C_2 environments are distinct from the baseline context provided by /r/ (as well as from each other: $p < .001$ for all pairwise comparisons). Broadly speaking, statistical analysis therefore confirms the idea that RVA is only triggered by obstruents that are actively voiced (/z/) or actively devoiced (textipa/t, s/).

The main effects of *C_1 laryngeal specification* finally, indicate that voicing distinctions between underlying /k/ and /ɡ/ are not entirely neutralised. I suggested above that the principal source of this effect might be the _/d/ context, where assimilation of voicing does not appear to occur. However, the only indication of an asymmetric preservation effect is a weak interaction of *C_1 laryngeal specification* and *C_2 laryngeal specification* revealed by the initial three-way ANOVA on the C_1 overall voicing data (obstruent clusters only): F(1,268) = 5.17, $p < .025$. This suggests that the 5 ms difference in C_1 overall voicing between /kz/ and /ɡz/ reinforces the 18 ms difference between /kd/ /ɡd/, and therefore that C_1 voicing distinctions between underlying /k, ɡ/ are partially preserved before the [-tense] class as a whole.[8]

5.4.4 Segmental duration and obstruent clusters

Whereas C_1 voicing shows reflexes of manner-asymmetric regressive voicing assimilation, the same is not true of the durational measures. These phonetic features generally seem to preserve the contrast between /k/ and /ɡ/. As a result, the segmental duration data contradicts a lexical feature analysis and largely supports prediction (15b of the phonetic theory.

The patterning of V_1 duration offers the most unequivocal evidence for prediction (15b): the data in figure 5.6 shows only small variations of this parameter due to C_2 context. Moreover, the largest difference within a single C_1 laryngeal category in response to a change in C_2 environment occurs between /ɡd/ (89 ms) and /ɡz/ (100 ms) and can therefore not be interpreted in terms of [tense] assimilation to a following C_2 (which would predict that these two environments pattern together). The underlying contrast between /k/ and /ɡ/ on the other hand, induces relatively large differences in V_1 duration in the expected direction (longer vowels before /ɡ/). It seems therefore that preceding vowel duration does not assimilate, and this is confirmed by a three-way ANOVA for *C_2 laryngeal specification* * *C_2 manner of articulation* * *C_1 laryngeal specification*, on V_1 duration before /k, ɡ/ + /t, d, s, z/ sequences. This ANOVA yields a significant main effect of *C_1 laryngeal specification* only: F(1,268) = 72.38, $p < .001$, and no significant interactions.[9]

---

[8] There were no other interactions in any of the ANOVAs reported here.

[9] Adding the unsegmented plosive + plosive clusters only strengthens the effect of *C_1 laryngeal specification*: F(1,338) = 129.90, $p < .001$ and still fails to reveal any other significant effects.

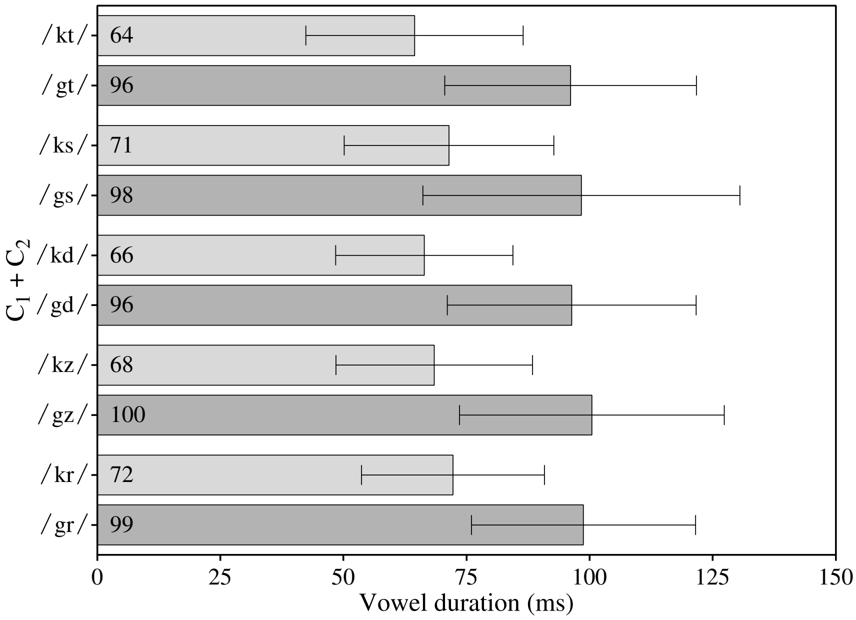

Figure 5.5: V_1 duration across C_2 contexts. All measures in ms; error bars represent the mean ± 1 standard deviation. This diagram shows that vowel duration reflects the status of the immediately following plosive (C_1) but is impervious to the value of [tense] on C_2 consonants: /k/ is preceded by a vowel of relatively short and highly similar duration across C_2 contexts whilst /g/ is preceded by longer vowels of near -identical length.

Figure 5.6 indicates that the behaviour of C_1 duration is a little more complicated. The closure interval of /k/ seems to assimilate to the [tense] value of a following obstruent to the extent that it is somewhat longer before /s/ and especially /t/ than before /d/ and /z/ respectively. However, the difference between /ks/ and /kz/ is marginal, and there does not seem to be any effect on the closure duration of /g/ that can be interpreted as evidence of [tense] assimilation. Given that /g/ does assimilate to a following /t, s/ or /z/ in terms of voicing, this results in a mismatch in the behaviour of closure duration and voicing which clearly contradicts the predictions of a lexical feature account (cf. prediction 10b in 4.1.1 above).

It is not surprising therefore that a three-way ANOVA for *C_2 laryngeal specification* * *C_2 manner of articulation* * *C_1 laryngeal specification* on the C_1 closure duration data (in clusters composed of obstruents only) does not reveal a significant main effect of *C_2 laryngeal specification*. There is a highly significant main effect of C_1 *laryngeal specification*: F(1, 268) = 28.07, $p < .001$. This indicates that C_1 closure duration preserves the distinction between underlying /k/ and /g/, and inspection of the data in figure 5.6 shows that there is indeed a systematic difference in the duration of /k/ and /g/ in all but one context. In addition, the ANOVA yields a significant main effect of C_2 *manner of articulation*, F(1,268) = 25.82, P<.001, and interaction of C_2 *laryngeal specification* * C_1 *laryngeal specification*, F(1,268), p< .01. The first of these effects is likely to stem from the relatively long closure phase of /k/ before /s/ and /z/ whilst the second probably results from the difference in the durations of /k/ closure before /t, s/ and /d, z/. Only the latter effect can be interpreted in assimilatory terms, but as I noted above, this does not vindicate a lexical feature analysis of RVA.

Finally, consider the behaviour of C_1 release duration. The first generalisation concerning this feature that emerges from figure 5.6 is that the release of /k/ and /g/ is relatively short before fricatives. As I hinted in 5.3 this is likely to be a labelling artefact caused by the overlap of release and frication noise in the acoustic signal, and it therefore seems safer to exclude cases involving a fricative C_2 from further analysis.

This leaves the sequences ending in a /d/ or /t/. The data for these clusters may seem to indicate that their internal releases are affected by some form of regressive assimilation, as on average C_1 release duration is somewhat shorter before /d/ than before /t/. However, a two-way ANOVA for C_2 *laryngeal specification* * C_1 *laryngeal specification* on the C_1 release duration data for stop + stop clusters shows that the effect of the first factor is little more than a trend: F(1,97) = 3.29, $p < .075$, which suggests that release duration is subject to little or no regressive assimilation. At the same time there is a weakly significant main effect of C_2 *laryngeal specification*: F(1,97) = 6.82, $p < .015$, which indicates that C_1 release duration at least partially preserves the underlying contrast between /k/ and /g/ (the interaction between the two factors is not significant).

5.4.5 F_0 and F_1 preceding obstruent clusters

No assimilatory patterns can be discerned in the F_0 contours preceding obstruent clusters. F_1 perturbations on the other hand, appear to show a [tense]-symmetric assimilation effect that patterns as would be expected from a rule spreading the lexical laryngeal features of C_2 obstruents: velar stops preceding /t, s/ are marked by a higher F_1 10 ms into the preceding vowel than those preceding /z, d/ (cf. table 5.4). The underlying distinction between /k/ and /g/ is erased before /t/, but in the remaining three obstruent contexts F_1 is lower before /g/ than before /k/ (as it is in the baseline environment), and this suggests that the effect of C_2 is incompletely neutralising.

A three-way ANOVA for C_2 *laryngeal specification* * C_2 *manner of artic-*

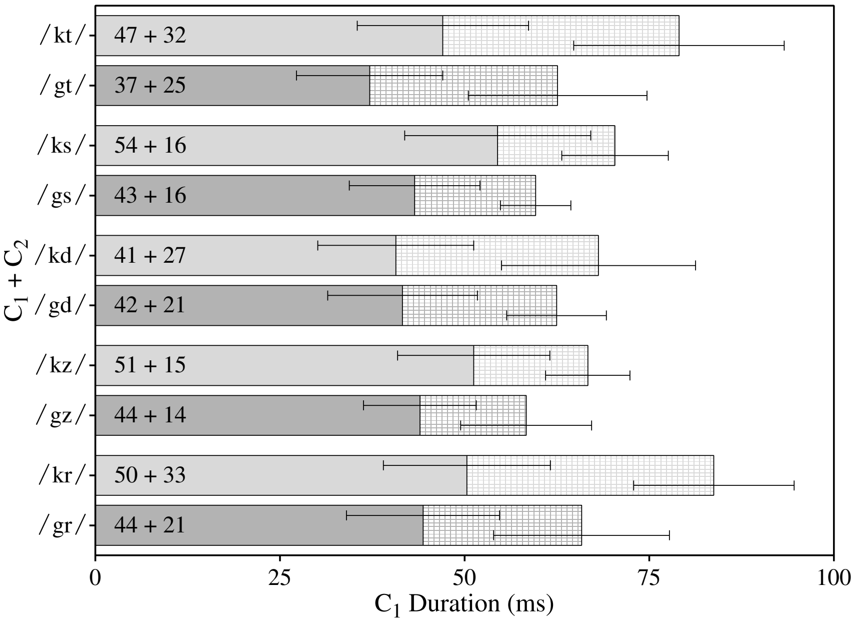

Figure 5.6: C_1 duration across C_2 contexts. All measures in ms; error bars represent the mean ± 1 standard deviation. The diagram represents mean C_1 closure and release durations separately: for each bar the left-hand segment indicates the closure duration and the right hand segment represents the duration of the release phase. Exact values for mean C_1 closure and release duration are given in the leftmost segment for typographical reasons. This diagram provides little or no evidence for an effect of RVA on the closure duration of C_1, because there is little or no systematic shortening (relative to the baseline context) before /t, s/ or shortening before /z/. There is some suggestion of an assimilatory effect on C_1 release duration before plosives, but this is not confirmed by statistical tests.

Table 5.4: Experiment 1: F_1 preceding obstruent clusters. F_1 (Hz) at 10 ms before the onset of C_1. Standard deviations in brackets.

C_1C_2	F_1 at C_1 - 10 ms	N	C_1C_2	F_1 at C_1 - 10 ms	N
/kt/	508 (42)	31	/gt/	508 (34)	26
/kd/	486 (41)	26	/gd/	478 (20)	18
/ks/	519 (43)	47	/gs/	487 (28)	45
/kz/	489 (43)	36	/gz/	479 (34)	47
/kr/	502 (46)	32	/gr/	476 (30)	47

ulation * *C_1 laryngeal specification* on the F_1 data supports both observations. There is a highly significant main effect of *C_2 laryngeal specification*, $F(1,268) = 23.01$, $p < .001$, and a weaker effect of *C_1 laryngeal specification*, $F(1, 268) = 7.36$, $p < .01$, but no significant main effect of *C_2 manner of articulation* and no significant interactions.[10] The absence of any effects related to *C_2 manner of articulation* confirms that unlike C_1 voicing, F_1 assimilates in a [tense]-symmetric fashion. As I noted in the survey of phonetic properties of the tense-lax distinction in chapter 2, lax stops appear to be marked by a lower first formant on flanking vowels regardless of their voicing targets, and so the fact that there is [tense]-symmetric assimilation of F_1 does not necessarily contradict the phonetic theory of RVA. In the light of the C_1 voicing and segmental duration data it certainly does not support a lexical feature analysis, which predicts that assimilation should be [tense]-symmetric or [tense]-asymmetric across phonetic features. However, since the articulatory underpinnings of low frequency spectral cues to [±tense] remain unclear, the data in table 5.4 are difficult to interpret.

5.5 General discussion and conclusions

The aim of this chapter was to test three hypotheses derived from a coarticulation-based view of regressive voicing assimilation at word boundaries. The first of these hypotheses was that only actively (de)voiced obstruents should be able to trigger RVA. In chapter 4 I pointed out how a coarticulatory view of voicing assimilation correctly predicts the distribution of regressive assimilation under word sandhi within the Germanic group of languages. The data from experiment 1 indicates that English obstruents trigger RVA broadly in accordance with this view. Actively voiced /z/ and to a lesser extent, actively devoiced /t, s/ all cause deviations in the phonetic voicing of a preceding obstruent relative to a baseline sonorant context. Crucially, English /d/, which was argued in chapter 2 to be passively voiced, did not trigger any form of voicing assimilation (ignoring for the moment the effect on F_1: see further below).

Regressive voicing assimilation to /s, z/ is qualitatively [tense]-symmetric in the sense that both lenis and fortis obstruents are able to induce changes in the voicing in at least one class of preceding obstruent. However, regressive assimilation is not always observably symmetric with regard to ±tense] target sounds. For example, fortis obstruents do not affect the voicing of a preceding /k/ vis-à-vis the baseline context. The most natural interpretation of these observations is that coarticulation still applies in the relevant clusters but fails to leave a trace

---

[10] Tukey and Scheffe post-hoc tests on a second three-way ANOVA, which included the data from the baseline context shows that the [+tense] C_2 environment is distinct from both the [-tense] and sonorant environments (all $p < .001$), but that the latter two environments are not distinct from each other.

in the speech signal. For example, the devoicing gestures of a /t/ would still be anticipated during the production of a preceding /k/, but because the latter is accompanied by active devoicing measures of its own this has little effect on the voicing of the initial stop. However, this interpretation is in need of further support from articulatory data.

The second hypothesis under hypothesis was that regressive assimilation only affects the voicing of a target obstruent and those features that are mechanically linked to the production of voicing, such as frication duration in fricatives. The data gathered by the present experiment provides an almost perfect match with this hypothesis, because the behaviour of C_1 voicing but not the patterning of C_1 closure and release duration or V_1 duration can be interpreted in terms of RVA. This does not mean that C_1 duration features are not subject to modification when another obstruent follows (there are clear changes relative to the baseline context), but that these modifications cannot be attributed to the same mechanism that controls C_1 voicing. Note that the results of experiment 1 are similar to those obtained by N. Thorsen (1971) who does not find evidence of assimilatory effects on V_1 duration and C_1 duration characteristics either. Furthermore, the lack of regressive assimilation of C_1 duration matches Russian data presented by Burton & Robblee (1997).

The one remaining puzzle with regard to the results of experiment 1 is the observation that the value of F_1 towards the end of V_1 appears to be subject to manner-symmetric but tense-asymmetric assimilation to /d, z/ (assuming that __ is a legitimate baseline condition for this feature). As the articulatory underpinnings of F_1 perturbations by [±tense] obstruents are unclear, any interpretation of this data will be speculative. Note however, that as long as the effect of obstruents on the first formant of flanking vowels can be traced to a definite articulation involved in the expression of the tense-lax contrast in English obstruents, the F_1 data does not contradict the phonetic theory (as the gesture in question would itself be subject to anticipatory coarticulation).

The third and final prediction of a coarticulatory view of regressive voicing assimilation is that the process should not be completely neutralising. This prediction is clearly borne out by the data summarised above. Even C_1 voicing, the primary feature involved in the process, tends to bear residual traces of distinctions between fortis and lenis C_1 obstruents.

Preceding vowel duration is generally regarded as the most important cue to [±tense] distinctions in English postvocalic obstruents (cf. 2). Given that V_1 duration is entirely unaffected by any form of regressive assimilation, it seems hardly surprising that descriptions based on auditory impressions tend to regard (aspirating) English as a language with little or no RVA. However, experiment 1 indicates that this view is not entirely accurate, and that an articulation-driven form of RVA is active at word boundaries even in aspirated varieties of English.

Chapter 6

Experiment 2: regressive voicing assimilation in Hungarian

> „1. A’ Páros Keményeknem szenyvedhetik magok előtt a’ sebes ki mondásban a’ Páros Gyengéket, hanem azokat fel tserélik-az ő Keménypárjaikkal. A’ Liquidákkal pedig széretik.
>
> 2. A’ Páros Gyengék nem szenyvedhetik magok előtt a’ Páros Keményeket, hanem azokát fel tserélik az ő Gyenge Párjaikkal. A’ Liquidákat pedig szeretik.” (Kolmár 1821: 57)

> “1. The paired strong ones [i.e., sounds] cannot bear the paired weak ones to be in front of them in fast speech; rather they trade them up for their strong twin. With liquids however, they get on well.
>
> 2. The paired weak ones cannot bear the paired strong ones to be in front of them in fast speech; rather they transpose them for their weak twin. Liquids however, they love.”[1]

It should not be difficult to motivate Hungarian as a second test case for a coarticulation-based theory of RVA. As the surrounding Slavonic languages and neighbouring Romanian, Hungarian is a voicing language, but in contrast to (most) of the former it lacks a process of across-the-board final laryngeal neutralisation. Assuming for the sake of the argument that *voicing language* and *aspirating language* are coherent notions with regard to the behaviour of word-final stops, it therefore differs from English in terms of just a single variable. Thus, it allows for exactly the same set of hypotheses to be examined as those that were investigated with the previous experiment.

---

[1] Translation by Zoë Toft.

Hungarian possesses a well-documented process of regressive voicing assimilation that is attracting increasing attention in the generative literature. A coarticulation-based theory derives the following set of predictions concerning (the phonetic manifestation of) Hungarian regressive assimilation at word boundaries (cf. 15 above and 11 in chapter 4):

(18) Predictions of a coarticulation-based approach to voicing assimilation regarding obstruent sequences in Hungarian

  a. Hungarian obstruents fall into 2 classes in terms of their influence on (the voicing of) a preceding obstruent. (1) fortis obstruents trigger devoicing (relative to a ‘neutral’ environment); (2) lenis stops and fricatives cause an increase in the voicing of a preceding obstruent, because both classes are actively (pre)voiced (11c)
  b. The assimilatory effects of fortis obstruents and lenis fricatives are limited to voicing and features mechanically dependent on the production of voicing distinctions. Cf. (11a)
  c. Hungarian RVA at word boundaries is a gradient process that is not neutralising in most instances Cf. (11b)

In other words, the phonetic theory predicts that Hungarian RVA behaves in a fundamentally identical fashion to English regressive voicing assimilation except in a single respect: the lax stops /b, d, ɟ, ɡ, d͡ʒ/ should cause an increase in the voicing of a preceding obstruent vis-à-vis a neutral (passively voiced sonorant) environment.

The experiment presented here was designed to test the three hypotheses in (18) against the predictions of a lexical feature analysis. The results of this experiment show that, as in English, regressive voicing assimilation between independent words is a non-neutralising process in Hungarian. However in contrast to the findings on English reported above, the Hungarian data contradict prediction (18b): vowel length distinctions between underlying /k/ and /ɡ/ are near-neutralised when another obstruent follows. The duration of vowels preceding these sequences seems to cue neither the underlying laryngeal specification of the velar stops, nor the laryngeal specification of the obstruents following them, which means that the behaviour of vowel length cannot be regarded as assimilatory in the most straightforward interpretation of the term. Nevertheless, the observations on this point indicate that Hungarian RVA cannot be regarded as a purely coarticulatory process, but may be in part phonologised.

Note that the work reported in this chapter represents part of ongoing collaborative work with Zoë Toft: an earlier report can be found in Jansen & Toft (2002).

6.1 Background

Hungarian is an Uralic (Finno-Ugric, Ugric) language spoken by around 15 million people in Hungary and (as a minority language) in several of the surrounding states. As shown in (19) the obstruent system of Hungarian is bifurcated in the way that is familiar from Germanic and Romance (Kenesei et al., 1998; Siptár & Törkenczy, 2000).[2] According to Kenesei et al. (1998) the fortis stops and affricates of Hungarian are voiceless unaspirated while its lenis stops are prevoiced, and this is corroborated by acoustic data (Meyer & Gombocz, 1909; Gósy, 1999). The same authors characterise the parallel contrast in the fricative inventory as voiceless vs. voiced.

(19) The Hungarian obstruent system

	Labial	Alveolar	Postalveolar	Palatal	Velar
Plosive	p b	t d		c ɟ	k ɡ
Affricate		t͡s (d͡z)	t͡ʃ d͡ʒ		
Fricative	f v	s z	ʃ ʒ		

Just as Yiddish and French, Hungarian preserves the distinction between word final tense and lax obstruents before sonorants and utterance finally. As would be expected under a phonetic theory of the phenomenon, tense and lax obstruents trigger regressive assimilation in obstruent clusters. Hungarian RVA is invariably described as largely symmetric with regard to both [tense] and manner of articulation: with the exception of /v/ all obstruents trigger the process. According to many descriptions it is insensitive to juncture strength and is obligatory in all sandhi obstruent clusters as long as no physical pause intervenes. The examples in (20) are from Kenesei et al. (1998) and Siptár & Törkenczy (2000).[3]

There is a long tradition in Hungarian linguistics that regards regressive voicing assimilation not only as obligatory but also as phonetically neutralising (Hall, 1944; Sauvageot, 1951; Kálmán, 1972; Lotz, 1972, 1988; Siptár, 1991; Olsson, 1992; Kenesei et al., 1998). Kenesei et al. (1998) and Siptár & Törkenczy (2000) emphasise this view by contrasting RVA with a process of regressive place assimilation that affects sibilant fricatives and affricates. Unlike voicing assimilation the latter phenomenon, which is exemplified by several of the forms in (20) (e.g., /brid͡ʒ/ + /sobɔ/ → [brit͡ssobɔ]), is said to be partial

---

[2] The inclusion of [d͡z] in the lexical obstruent inventory of Hungarian remains contentious: Siptár & Törkenczy (2000) argue that on phonological grounds it should be treated as a cluster, but Mária Gósy (p.c.) points out that most Hungarian phoneticians treat it on a par with [t͡s, t͡ʃ, d͡ʒ].

[3] On the basis of a transcription study, Gósy (1999) attempts to demonstrate that Hungarian RVA does apply across certain pauses, but her claims are hard to evaluate as no acoustic definitions to distinguish 'assimilated' from 'unassimilated' obstruents are provided.

and dependent on speaking rate and style. Only a few authors disagree with this assessment, and their objections tend to concentrate on the claim that Hungarian RVA is obligatory: both Kolmár (1821) and Vago (1980) suggest that the process is governed by speech rate, whilst Tompa (1961) claims that it can be suspended in loanwords and when a potential trigger belongs to a contrastively stressed word.

(20) Regressive voicing assimilation in Hungarian (data from Kenesei et al. 1998:445-446 and Siptár & Törkenczy 2000:78)

a. [+tense][-tense] clusters

UR	Phonetic form	Gloss
/kɔlɔp/+ /bɔn/	[kɔlɔbːɔn]	in (a) hat
/kuːt/+ /bɔn/	[kuːdbɔn]	in (a) well
/fyːc/+ /bɔn/	[fyːɟbɛn]	in (a) whistle
/ʒaːk/+ /bɔn/	[ʒaːgbɔn]	in (a) sack
/okoʃ/ + /zɛneːs/	[okozːɛneːs]	smart musician
/kovaːt͡ʃ/ + /zoltaːn/	[kovaːd͡zzoltaːn]	Kovács Zoltán (proper name)
/vɛs/ + /d͡ʒɛmɛt/	[vɛʒd͡ʒɛmɛt]	buy-3.SG.INDEF jam-ACC.
/pɔloːt͡s/ + /d͡ʒidaːʃ/	[pɔloːd͡ʒːidaːʃ]	Northern Hungarian lancer

b. [-tense][+tense] clusters

UR	Phonetic form	Gloss
/rɔb/+ /toːl/	[rɔptoːl]	from (a) prisoner
/kaːd/+ /toːl/	[kaːtːoːl]	from (a) bathtub
/aːɟ/+ /toːl/	[aːctoːl]	from (a) bed
/mɛlɛg/+ /toːl/	[mɛlɛktøːl]	from (the) heat
/montaːʒ/ + /sɛryː/	[montaːsːɛryː]	montage-like
/igɔz/ + /ʃaːg/	[igɔʃːaːg]	truth
/brid͡ʒ/ + /sobɔ/	[brit͡ssobɔ]	bridge room
/vɔraːʒ/ + /t͡sɛruzɔ/	[vɔraːst͡sɛruzɔ]	magic pencil

Whilst Hungarian regressive voicing assimilation has received considerable attention in the recent generative literature, there do not seem to be any quantitative phonetic studies of the process.[4] Gósy (1999) is essentially a transcription-based study, although it is in part based on acoustic rather than impressionistic auditory data. Early work by Meyer & Gombocz (1909) provides some data on segmental duration in (lexical) obstruent clusters, but does not specifically

---

[4] Recent generative work includes Siptár (1996), Szigetvári (1998), Ritter (2000), Siptár & Törkenczy (2000), Siptár & Szentgyörgyi (2002).

investigate assimilation. Consequently, the only material that is available for comparison with the data reported below comes from languages with obstruent systems that are phonologically and phonetically similar to that of Hungarian, such as French (O. Thorsen, 1966) and Syrian Arabic (Barry & Teifour, 1999). Interestingly, these studies show that regressive voicing assimilation is incompletely neutralising in they investigate.

6.2 The Experiment

6.2.1 Methods

Subjects Subjects were 4 native speakers of Hungarian, all female, and aged between 26 and 30 years. All speakers were living in London at the time of recording and had lived in the United Kingdom for up to 4.5 years. None of the subjects reported a history of speech or hearing difficulties but (unavoidably) all of them were proficient to a greater or lesser degree in one or more languages besides Hungarian. Subject K9 grew up in Heves county but describes her speech as ‘standard’ (Budapest) Hungarian. She is fluent in English. Subject M15 also describes her variety of Hungarian as ’standard’, despite having frequently moved around Hungary. This subject describes herself as ‘near-bilingual’ in French and has good English. Subject I16 is a bilingual Hungarian and Slovak speaker from Bratislava. She is fluent in English and has some knowledge of Czech and German. Subject A17 finally, is from Tatabánya, fluent in English, and has a good knowledge of both French and German. She had lived in the United Kingdom for approximately 6 years at the time of recording.

Materials The stimuli for experiment 2 consisted of consonant clusters combining a /k, ɡ, ʃ, ʒ/ C_1 and a /t, d, s, z/ or liquid (/l/ or /r/) C_2. As in experiment 1 stimuli containing a sonorant C_2 were included to create baseline conditions for the comparison of the relative effects of fortis vs. lenis C_2 on the properties of a preceding obstruent. Velar plosive + alveolar obstruent clusters were used for the reasons specified in section 5.3 and also to facilitate comparisons with the English results. The set of obstruent contexts used for the English experiment was expanded somewhat by including the postalveolar fricatives /ʃ, ʒ/ in the C_1 set. Postalveolar fricatives were chosen to minimise the variation in C_1 place of articulation and for segmentation reasons (but see below).

C_1 consonants were preceded by a long vowel or short vowel + glide sequence (phonetic diphthong) from the set /eː, aː, uː, ɔj/, or one of the following short vowels: /i, ɔ, o/. Long vowels and short vowels were evenly distributed across C_1 and C_2 laryngeal specifications and manners of articulation in order to avoid a bias of underlying vowel length in the effects of these factors on vowel duration. Similarly, high and non-high vowels were evenly distributed across C_1 and C_2 laryngeal specifications and manners of articulation in order to control for effects of vowel height on C_1 voicing duration and F0 perturbations. The clusters were located at subject noun + verb boundaries in carrier sentences. As in experiment 1, no attempt was made to control for carrier word frequencies. Some sample stimuli are given in 21 in orthographic and phonological transcription.[5] Target clusters appear slanted.

(21) Hungarian sample stimuli

a. A va*k d*arabolta a húst
/ɔ vɔ*k d*ɔrɔboltɔ ɔ huːʃt/
The blind mince-PAST.3.SG the meat-ACC.
The blind man minced the meat

b. A ké*s d*olgozik a mészáros kezében
/ɔ keː*ʃ d*olgozik ɔ meːsaːros kɛzeːbɛn/
The knife works the butcher hand-3.POSS-in
The knife works in the butcher's hand

c. A ri*zs z*öldül a mezőn
/ɔ ri*ʒ z*øldyl ɔ mɛzøːn/
The rice green-become the field-PL.-LOC.
The rice turns green in the fields

Subject + noun boundaries were chosen over other possible word boundary environments on grounds of the available carriers for C_1, which had to be similar in overall phonological make-up whilst exhibiting a robust contrast between /k, ʃ/ and /g, ʒ/ (and therefore had to be unsuffixed). One potential problem with this choice is that the type of boundary involved usually represents a strong phonological and phonetic juncture, and this is reflected in the number of utterances that had to be excluded because of a physical pause intervening in the target cluster, which was relatively high compared to the English corpus examined above. Strong junctures have a tendency of blocking sandhi processes and it might therefore be argued that the design of the experiment is inherently biased towards non-assimilation or incomplete assimilation (Peter Siptár, p.c.).

However, as can be gleaned from the discussion in chapter 4, it is not an objective of this study to prove that all forms of assimilation or even all forms of regressive assimilation of word boundaries are drive by coarticulation. It's main objective in this area is to investigate the weaker proposition that regressive assimilation at word boundaries either operates as a coarticulatory process or is diachronically grounded in such a process. A first and important step in this argument is to establish that there indeed is a coarticulatory form of regressive voicing assimilation. Since it seems to be typical for sandhi processes that

---

[5] The full stimulus set appears in appendix B.

operate across weak junctures to be subject to phonologisation, the behaviour of obstruent clusters at strong boundaries is therefore not just a legitimate testing ground for the phonetic approach to RVA proposed in chapter 4 but potentially a crucial one.

Note, moreover, that many descriptions of Hungarian RVA, including the one provided by Siptár & Törkenczy 2000, suggest that the process is obligatory (and categorical) regardless of juncture strength as long as no physical pause intervenes. From this perspective, the present design is perfectly valid as long as tokens with a physical pause between C_1 and C_2 are removed from the corpus.

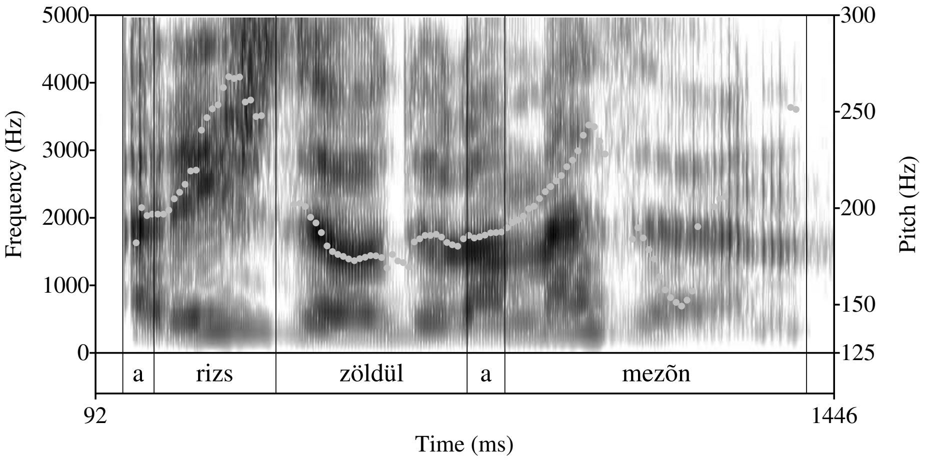

Figure 6.1: Experiment 2: pitch contour of responses. Broad band spectrogram of an utterance of the stimulus sentence in 21c with superimposed F_0 track. The speaker is subject I16.

It was impossible to construct all carrier sentences according to the neutral word order for the propositions they expressed. This raised the possibility that the subjects would assign different prosodic structures to different stimulus sentences. However, the great majority of responses was pronounced with a F_0 peak on the subject noun carrying C_1 followed by a gradual fall across the remainder of the sentence. A variant of this pattern (frequently used by subject I16 and illustrated in figure 6.1) shows what appears to be a secondary pitch accent on the initial syllable of the final word, but under neither of these two contours did the verb acting as the C_2 carrier receive any pitch prominence. Unfortunately, this limits the scope for comparison with the data from experiment 1 somewhat, as all the utterances in the English corpus were produced with a nuclear accent on the syllable containing C_2.

Procedure The stimuli were presented to the subject in a quasi-randomised order to avoid consecutive stimuli with identical consonant clusters. The subjects produced three repetitions of each stimulus and were asked to read a stimulus again if they produced a mistake or hesitation that was clearly audible to the experimenter. In total, 2 (plosive C_1) * 5 (C_2) * 6 (stimuli) * 3 (repetitions) * 3 speakers + 2 (fricative C_1) * 5 (C_2) * 4 (stimuli) * 3 (repetitions) * 4 speakers = 1200 utterances were recorded. Only 4 stimuli each were used for the postalveolar fricative C_1s because of a lack of suitable target words. Recording and acoustic analysis set-ups were the same as for the English experiment. 58 utterances had to be discarded because they contained a pause between C_1 and C_2. In addition, all of the remaining 158 fricative + fricative sequences and 5 plosive + plosive clusters could not be internally segmented in any reliable fashion and had to be discarded too.[6] This left 953 utterances for segmentation and analysis.

Segmentation and measurements Segmentation of the acoustic signals was carried out according to the protocol sketched in 5.3 above, with additional provisions for C_1 fricatives, which were not investigated in experiment 1. The onset of a C_1 fricative was defined as the onset of frication noise, or if present, the appearance of aspiration noise preceding it (cf. Stevens et al. 1992; Stevens 1998). The offset of a fricative C_1 in fricative + stop sequences was defined as the offset of frication noise.

Table 6.1: Acoustic measurements and derived measures for Experiment 2.

Segment							
V_1		C_1		C_2		V_2	
(a)	Duration	(c)	Closure duration (stops)	(h)	Closure duration (stops)	(m)	F_0 10-50 ms after C_1 offset
(b)	F_0 50-10 ms before C_1 onset	(d)	Release duration (stops)	(i)	Overall duration (fricatives)		
		(e)	Overall duration	(j)	Voicing duration (fricatives)		
		(f)	Voicing duration (2 m.)	(k)	Voicing ratio (fricatives)		
		(g)	Voicing ratio (2 m.)	(l)	VOT (stops)		

---

[6] Hungarian sibilant + sibilant clusters are subject to a rule of regressive place assimilation that was mentioned above and illustrated in (20). In most of the fricative + fricative clusters in the present corpus this assimilation is partial (vindicating the description by Siptár & Törkenczy 2000 and others) but it nevertheless proved hard to define sufficiently precise criteria to segment C_1 from C_2 in these clusters.

The phonetic expression of [±tense] and regressive voicing assimilation was quantified in almost exactly the same way as for experiment 1: a summary of the relevant measures appears in table 6.1. Note that no measurements were made of F_1 preceding C_1 onset because the lexical vowel quality of V_1 could not be controlled for.

6.3 Results

The main results of experiment 2 are reported below in roughly the same order as the results of experiment 1 in 5.4 in order to facilitate a comparison of the results. The main focus is on the behaviour of C_1 plosives, in part to highlight similarities and differences with the English data reported above, and in part for practical reasons: the phonetic features of C_1 fricatives could only be examined before C_2 plosives and liquids, which results in a defective paradigm for comparison with C_1 plosives. It is unclear too, whether the segmental duration of fricatives can be meaningfully compared in quantitative terms with the durational features of plosives.

6.3.1 Phonetic features of C_2

The measurements of C_2 voicing are in full agreement with descriptions of Hungarian as a voicing language. /d/ has a negative VOT of 26 ms whilst /t/ has a short lag positive VOT of 23 ms. The amount of prevoicing in the lax stop may seem small in comparison with published data on other voicing languages, but note that the mean duration of the closure stage of /d/ is only 51 ms. The high standard deviation of the VOT for /d/ provides another clue to its mean value: 56 tokens (24.9%) of /d/ are completely voiceless. As 21 of these tokens are preceded by /ʒ/ the assumption seems warranted that this is the result of passive devoicing rather than a rule spreading [+tense].[7] The mean VOT for /t/ is 9 ms longer than the value found for English /d/ in experiment 1, which is consistent with the hypothesis that short lag /d/ and short lag /t/ represent two distinct voicing categories: passively voiced and actively devoiced (2.2.1 above and cf. Raphael et al. 1995).

As shown in table 6.2, there is little unexpected about the behaviour of the fricatives /s/ and /z/. The former is wholly voiceless (there are a few tokens with a minute amount of voicing 'spill' from a preceding voiced obstruent) and relatively long whereas the latter is (partially) voiced and relatively short. The mean voicing ratio of /z/ (.65) is lower than that of English /z/ (.78), but since the latter but not the former was produced in a prosodically strong context it

---

[7] If fully devoiced tokens of /d/ are excluded, the average amount of prevoicing for this category increases to 42 ms.

Table 6.2: Experiment 2: duration and voicing of C_2. Closure duration and VOT of /t, d/, and overall duration and duration of the voiced interval for /s, z/. All values in ms, and pooled across preceding contexts (/k, g, ʃ, ʒ/). Standard deviations in brackets.

C_2	VOT	Closure duration	N
/t/	23 (7)	59 (17)	232
/d/	-26 (30)	51 (14)	225
	Voicing	Duration	N
/s/	0 (2)	123 (22)	136
/z/	56 (31)	92 (18)	135

would be premature to conclude from this that English and Hungarian /z/ have identical voicing targets.

An interesting difference between the English and Hungarian C_2 data is the magnitude of the effect of [±tense] on the F_0 of the following vowel. As illustrated in figure 6.2, the difference between tense and lax obstruents is approximately 10 Hz for the Hungarian subjects as opposed to roughly 35 Hz for the English female speakers and 20 Hz for the male speakers. Note that this discrepancy between the Hungarian and English speakers cannot be attributed to differences in overall F_0 level.

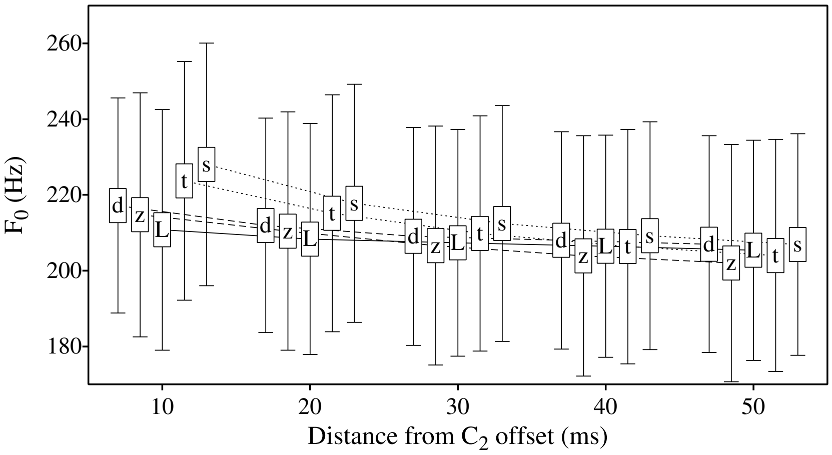

Figure 6.2: Experiment 2: F_0 (Hz) 10-50 ms into the vowel following C_2 (*L* = liquid). Error bars represent the mean ±1 standard deviation.

T-tests show that the differences in duration and VOT/voicing presented in table 6.2 are statistically significant too: t(450) = 24.60, $p < .001$ (plosive VOT); t(450) = 5.48, $p < .001$ (plosive closure duration); t(269) = -21.06, $p < .001$ (fricative voicing); t(269) =20.99 $p < .001$ (fricative duration).

6.3.2 Voicing of C_1

Plosives Figure 6.3 represents the mean voicing of /k, g/ (closure and release) across C_2 contexts. First, the baseline pre-liquid context shows a 33 ms difference in overall voicing between /k/ (32 ms) and /g/ (65 ms), which suggests that phonetic voicing has some role in cueing the [±tense] distinction in Hungarian. The difference in overall voicing is significant according to a t-test: t(135) = -15.97, $p < .001$. Interestingly, the mean voicing *ratio* of Hungarian /g/ is higher than that of its English counterpart, at least judging by the experimental data reported in the previous chapter (.90 vs. .70). This might be interpreted as evidence that the contrast between voicing and aspirating stop systems is maintained in word-final contexts (cf. 2.2.2 above). But in light of prosodic differences between the carrier sentences used for the two experiments, such interpretations remain speculative.

Next, consider the voicing of C_1 plosives before [+tense] C_2 obstruents. Figure 6.3 indicates that C_1 voicing assimilates to a following obstruent, showing a clear reduction in the overall voicing duration of /g/ in this environment. As there is virtually no difference in voicing between /k/ and /g/ before tense /t, s/, it would seem that assimilation neutralises the voicing distinction between the two velar stops.

The patterning of C_1 voicing before the lax obstruents /d, z/ suggests that assimilation occurs in this type of context too, as there is an increase in the overall voicing of /k/ relative to the baseline value of 32 ms (by 21 and 14 ms respectively). In accordance with impressionistic descriptions, and with the predictions of a phonetic theory of RVA, the observed behaviour of Hungarian /d/ contrasts with that of its English counterpart, which patterns with the baseline context rather than with /z/ (cf. figure 5.4 above).

However, in contrast to the [+tense] C_2 contexts, assimilation in the [-tense] environments does not appear to be fully neutralising. There are residual voicing distinctions between /k/ and /g/ both before /d/ (18 ms difference) and /z/ (17 ms). This asymmetry between tense and lax C_2 environments is reminiscent of the voicing patterns of English velar stops preceding /t, s, z/ (see figure 5.4), and therefore suggest that the same mechanism might be at work in both languages.

Statistical tests bolster the impressionistic observations made in the previous paragraphs. First, a two-way ANOVA for *C_1 laryngeal specification* * *C_2 laryngeal specification* was carried out on the overall voicing values of /k/ and /g/ in pre-obstruent contexts (i.e., excluding the baseline environment). This ANOVA shows significant effects of *C_1 laryngeal specification*, F(1,531) = 77.70, $p <$

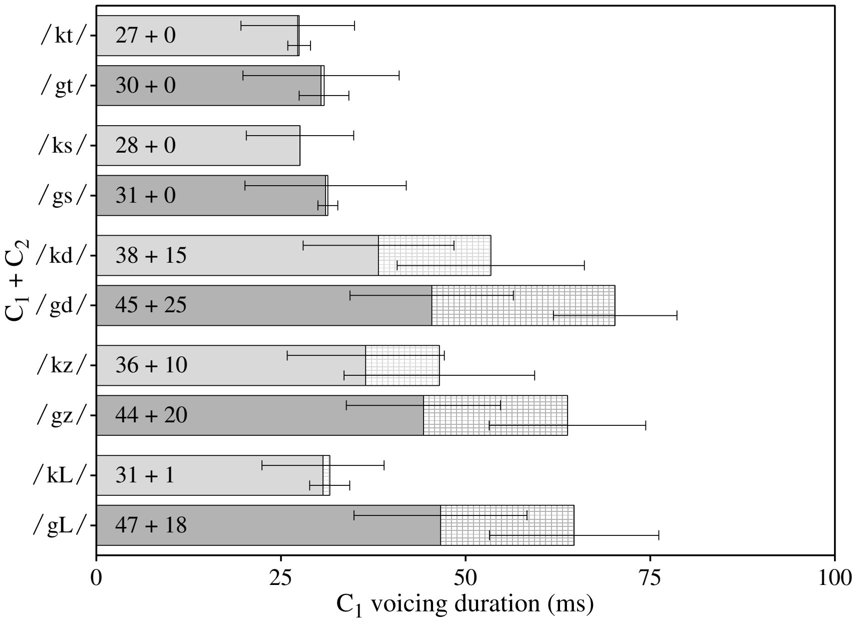

Figure 6.3: Experiment 2: Voicing of /k/ and /g/ across C_2 contexts. All measures are in ms; error bars represent the mean ± 1 standard deviation. The diagram represents the means for voicing duration during C_1 closure and release separately: for each bar the left-hand segment indicates the temporal extent of voicing during the closure stage and the right hand segment represents the voicing duration of the release phase. Exact values for mean C_1 closure and release voicing are given in the leftmost segment for typographical reasons.

.001, *C_2 laryngeal specification*, F(1,531) = 623.04, p $<$ .001, and the interaction between the two main factors *C_1 Laryngeal specification* * *C_2 laryngeal specification*, F(1,531) = 33.16, p ¡ .001. The main effect of *C_2 laryngeal specification* supports the impression that regressive assimilation takes place in Hungarian obstruent clusters whilst the main effect of *C_1 laryngeal specification* indicates that this form of assimilation fails to completely erase underlying voicing distinctions. However, the interaction of the two main factors indicates that the main effects do not apply in equal fashion across contexts, and is most likely caused by the virtual neutralisation of voicing distinctions before tense obstruents vs. the absence of complete neutralisation in lax C_2 environments.

Fricatives Figure 6.4 depicts the mean duration of voicing in /ʃ, ʒ/ before tense /t/, lax /d/ and baseline liquids. In the latter context, there is a marked (45 ms) difference in voicing, which suggests that this feature plays a role in signalling the distinction between tense and lax fricatives word finally. The difference is statistically significant according to a t-test: t(91) = -10.80, $p < .001$.

The voicing pattern that emerges before /t/ and /d/ seems to mirror the pattern observed above for /k/ and /g/ in the same contexts. /ʒ/ (28 ms of voicing) assimilates to the tense alveolar stop to the extent that the difference in voicing with /ʃ/ (26 ms of voicing) is virtually erased. There is evidence of regressive assimilation to /d/ too, as there is a clear (24 ms) increase of voicing in /ʃ/ relative to the baseline environment, but as with /k, g/ before /d, z/, assimilation to [-tense] does not seem capable of completely erasing underlying distinctions.

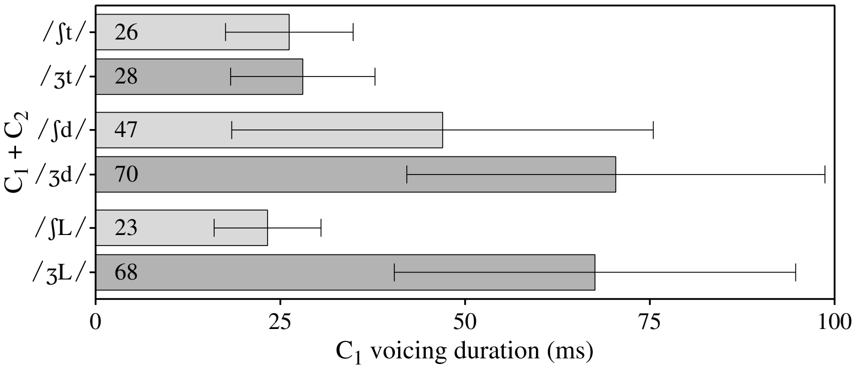

Figure 6.4: Experiment 2: Voicing of /ʃ/ and /ʒ/ across C_2 contexts. All measures are in ms; error bars represent the mean ± 1 standard deviation.

A two-way ANOVA for *C_1 laryngeal specification* * *C_2 laryngeal specification* on the voicing values of /ʃ, ʒ/ in pre-obstruent contexts seems to bear out this apparent parallelism with the assimilatory behaviour of /k, g/ in the same set of environments. Thus, there are highly significant main effects of *C_2 laryngeal specification*, F(1,184) = 107.52, $p < .001$ (evincing regressive assimilation), *C_1 laryngeal specification*, F(1,184) = 17.28, $p < .001$ (an indication of incomplete neutralisation), and a significant interaction of *C_1 laryngeal specification* * *C_2 laryngeal specification*, F(1,184) = 12.63, $p < .001$ (indicating that not all combinations of C_1 and C_2 behave symmetrically). As before, the interaction seems best explained in terms of the asymmetry between [+tense] contexts, where there is virtual neutralisation of C_1 contrast, and [-tense] environments where underlying voicing distinctions between /ʃ/ and ʒ is partially preserved.

6.3.3 Duration of C_1

Plosives Figure 6.5 depicts the duration of /k, g/ across the range of C_2 environments investigated by the present experiment. The bottom two bars of the diagram show how tense /k/ is marked both by a longer closure phase (71 vs. 54 ms) and a longer release burst (35 vs. 23 ms) than /g/. This behaviour is entirely consistent with the phonetic literature on the durational correlates of [±tense] in (medial) plosives in other languages, although this does not in itself constitute evidence that Hungarian listeners make (much) use of either of these features. The observed differences in closure phase and release burst duration between /k/ and /g/ in the baseline pre-liquid context are statistically significant according to t-tests: $t(135) = 10.08$, $p < .001$ and $t(135) = 6.42$, $p < .001$ respectively.

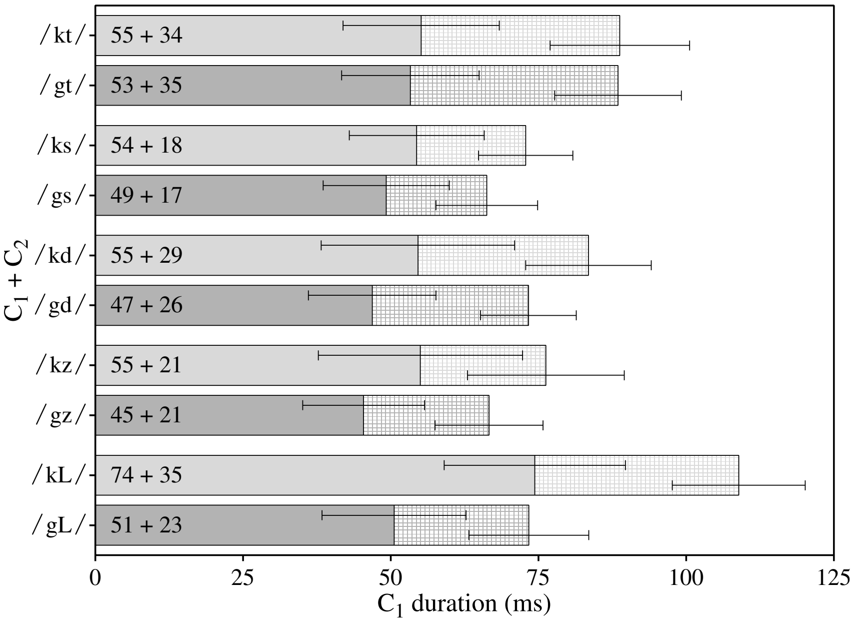

Figure 6.5: Experiment 2: Duration of /k/ and /g/ across C_2 contexts. All measures are in ms; error bars represent the mean ± 1 standard deviation. Means for closure duration and release duration are represented separately: for each bar the left-hand segment indicates the closure duration and the right hand segment represents the duration of the release phase. Exact values for mean C_1 closure and release duration are given in the leftmost segment for typographical reasons.

When followed by another obstruent, the behaviour of the closure phase of Hungarian /k/ and /g/ is strikingly similar to that of their English counterparts, both in terms of pattern and absolute duration values (cf. table 5.6). The closure stage is generally shorter before an obstruent C_2 than in the baseline environment, in particular for /k/. In all four obstruent environments, there is a small positive difference in closure duration between /k/ and /g/: this difference ranges from 2 ms before /t/ to 10 ms before /z/. This positive difference suggests that the duration contrast observed in the baseline environment is incompletely neutralised when an obstruent follows.

Moreover, only the 6 ms lengthening (relative to the baseline) of /g/ before /t/ could be construed as evidence that the partial neutralisation of closure phase duration contrast between /k/ and /g/ constitutes *regressive assimilation* in the conventional sense. However, given that the assimilation of C_1 *voicing* discussed above is triggered by /d, s, z/ as well as by /t/, it would be difficult to attribute the lengthening of /g/ before /t/ to the same underlying mechanism.

Thus, the velar stops of Hungarian appear to exhibit the same mismatch between closure duration and voicing that was observed above for their English counterparts. Consequently, their behaviour poses the same problems to a lexical feature analysis of voicing assimilation, which predicts that voicing and segmental duration maintain their inverse behaviour under assimilation, and therefore that an increase in voicing (as a result of assimilation to a lax stop) should be accompanied by a decrease in duration. The lack of of a systematic relation between C_1 closure duration and C_1 voicing is emphasised by the absence of a (statistically) significant negative correlation between the closure duration and overall voicing of /k/ and /g/ when followed by an obstruent C_2(Pearson's r = -.79, $p < .07$, i.e., significant at trend level only). By contrast, in the baseline environment there is a much stronger negative correlation between closure duration and overall voicing (r = -.45, $p< .001$), which is an indication that in this environment the 'lexical' inverse patterning of voicing and duration does tend to hold. The relation between closure duration and voicing duration in /k, g/ across C_2 contexts is illustrated in figure 6.6.

A three-way ANOVA for *C_2 laryngeal specification* * *C_2 manner of articulation* * *C_1 laryngeal specification* on the C_1 closure duration data (baseline environment excluded) reveals a highly significant main effect of *C_1 laryngeal specification*, F(1,531) = 30.03, $p< .001$, and marginal effects of *C_2 laryngeal specification*, F(1,531) = 5.34, $p < .025$ and *C_2 laryngeal specification* * *C_2 laryngeal specification*, F(1,531) = 5.51, $p < .02$. The first of these effects indicates that on the whole, the distinction between /k/ and /g/ is maintained in terms of C_1 closure duration, at least in speech production. It seems likely that the latter two effects are caused by the 'assimilatory' behaviour of /g/ before /t/, but as argued above, there is little evidence that the mechanism responsible also drives assimilation of C_1 voicing.

Finally, the duration of the *release stage* of /k, g/ does appear to show the

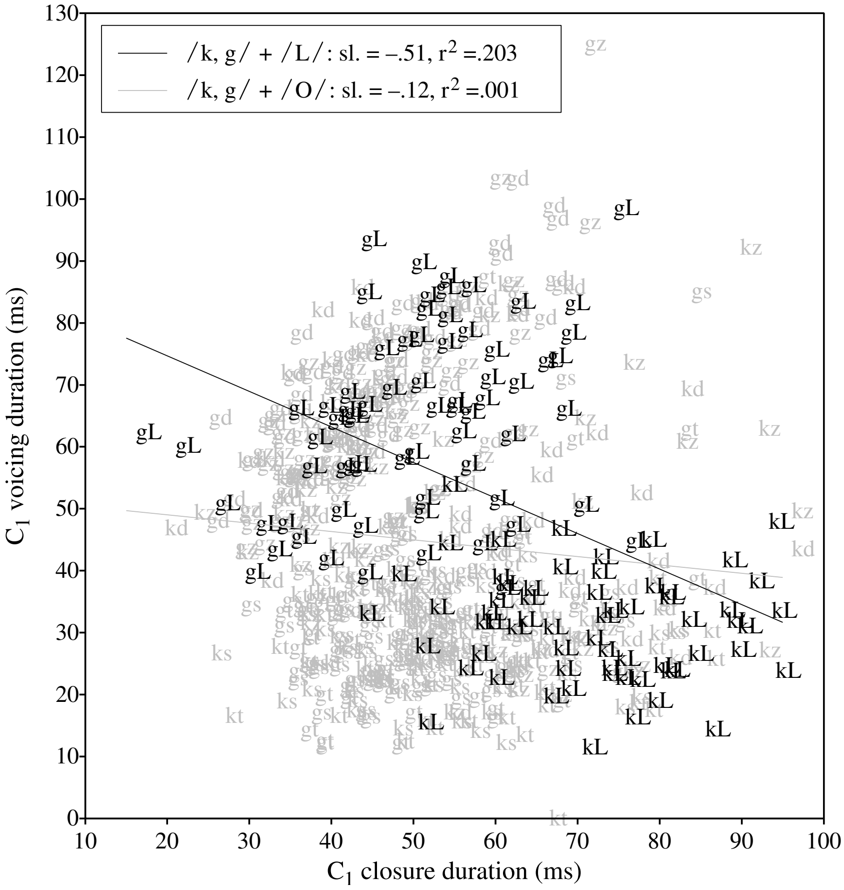

Figure 6.6: Experiment 2: scatter plot of C_1 voicing against C_1 closure duration. C_1 closure duration and C_1 overall voicing values (both in ms) for /k, g/ in obstruent clusters (in grey) and before a liquid (in black) with regression lines.

effects of assimilation to a following stop: relative to the baseline context, the duration of the release of /g/ increases by 12 ms before /t/, and there is a 6 ms decrease in the length of the release of /k/ when it is followed by lax /d/(see figure 6.5; as in the discussion of the English data above I will exclude sequences with a C_2 fricative from the analysis of release duration because it is probable that the relevant values are distorted by the overlap between release and frication noise). A two-way ANOVA for *C_1 laryngeal specification** *C_2 laryngeal specification* on the release duration values in pre-obstruent environments (/t/ and/d/ only) shows that assimilation neutralises the underlying distinction between /k/ and/g/: there is a highly significant effect of *C_2 laryngeal specifica-*

tion , F(1,260) = 27.60, $p < .001$, but not of C_1 *laryngeal specification*, F(1,260) = .123, not significant, or the interaction between C_1 *laryngeal specification* * C_2 *laryngeal specification*, F(1,260) = 2.40, not significant.

Thus, there is an apparent contradiction in the behaviour of the closure and release stages of velar stops preceding obstruents. However, there is a natural account of the release duration pattern that removes this contradiction. I will discuss this account as part of the analysis of fricative C_1 duration immediately below.

Fricatives The durational behaviour of /ʃ, ʒ/ when followed by an alveolar stop parallels that of the release stage of /k, g/. There is an increase (19 ms) in the length of /ʒ/ before tense /t/ and an even clearer decrease (34 ms) in the duration of /ʃ/ before /d/. Neither of these two environments preserves the baseline pattern, which exhibits the expected positive duration difference between the tense and lax postalveolar fricatives (30 ms): preceding /d/ this difference is reduced to a mere 3 ms whilst before /t/ it is reversed (by 9 ms). A t-test shows that the baseline contrast is statistically significant: t(91) = 7.48, $p<$ .001, whilst a two-way ANOVA for C_1 *laryngeal specification** C_2 *laryngeal specification* on the duration values in pre-obstruent contexts reveals a highly significant effect of C_2 *Laryngeal specification*, F(1,184) = 46.74, $p < .001$, but no effect of C_1 *laryngeal specification*, F(1,184) = .938, not significant, and only a very weak interaction of C_1 *laryngeal specification** C_2 *laryngeal specification*, F(1,184) = 4.92, p ¡ .03. The first of these supports the impression that fricative C_1 duration behaves in an assimilatory fashion, whilst the absence of a main effect of C_2 *laryngeal specification* indicates that this assimilation neutralises the distinction between /ʒ/ and /ʃ/ with respect to this feature. The 'reversed' patterning of duration before /t/is likely to be the predominant cause of the interaction between the two main factors, and can therefore not be treated as a sign of incomplete neutralisation.

It appears, therefore, that the duration of the release stage of velar stops and the overall duration of C_1 fricatives pattern identically in assimilating to a following stop in a neutralising fashion. This behaviour might be interpreted in support of a phonological feature analysis of Hungarian RVA. However, this assimilatory behaviour is equally explicable in terms of mechanical linkage between the production of voicing and the generation of turbulence noise in the oral tract. In chapter 2 I pointed out that the production of frication noise (and most of the oral release of the C_1 velar stops is just that) depends on a high transglottal airflow and hence requires some degree of glottal abduction. The production of vocal fold vibration on the other hand requires glottal *adduction*. This means that when voicing gestures are superimposed on a vocal tract configuration suitable for the production of a fricative, the result is a shortening of the interval of frication noise: recall how Stevens et al. (1992) invoke this mechanism to account for the fact that English lenis fricatives have the same duration as their fortis counterparts if measured in terms of F_1 transitions but that they nevertheless have shorter frication intervals.

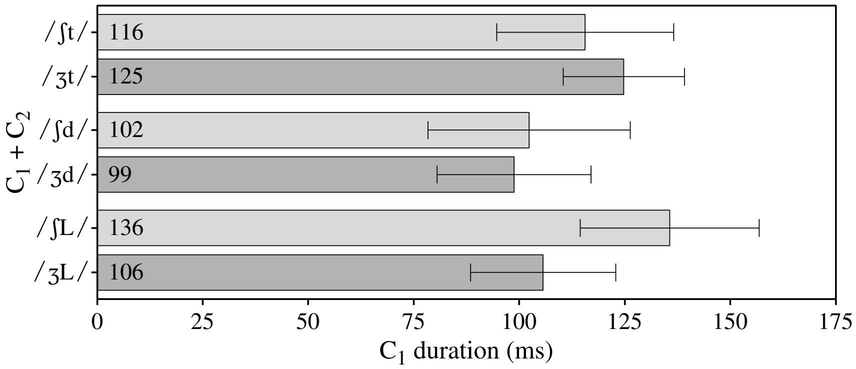

Figure 6.7: Experiment 2: Duration of /ʃ/ and /ʒ/ across C_2 contexts. All measures are in ms; error bars represent the mean ± 1 standard deviation.

The same mechanism can be invoked to capture the duration of C_1 fricatives and release noise: the only difference is that coarticulation rather than an underlying [-tense] specification is the source of the superimposed voicing gestures. I believe that this account should be favoured over a lexical feature analysis because the latter does not resolve the closure duration and voicing observations discussed above.

6.3.4 Duration of preceding vowels

The following paragraphs discuss the behaviour of lexically long vowels only. Short vowels are excluded from the discussion for two reasons. First, factoring in the effects of lexical vowel length would have complicated the presentation and analysis of the results in an unnecessary manner; second, the short vowel data is rather noisy, in particular due to the behaviour of a single stimulus item,/jog/, *law, right*, which tends to have a much shorter vowel than any of the remaining short vowel contexts, and consequently skews the results for __/g/X contexts. Taking on board the effects of this overly short vowel would have further complicated the discussion below.

Plosives Figure 6.8 represents the duration of lexically long vowels across plosive C_1 and C_2 environments. In the baseline pre-liquid environment vowel duration patterns in a way that suggests that, as many other languages, Hungarian utilises preceding vowel duration as a cue to the [±tense] distinction in word final obstruents. The expected negative difference between /k/ and /g/ materialises, is of roughly the same magnitude as the value observed for English in experiment 1 (25 ms), and is statistically significant according to a t-test: t(68) = -3.52, $p < .005$.

But there is a marked difference between English and Hungarian with regard to the behaviour of vowels before obstruent clusters. Recall that in English, the vowel length distinction between /k/ and /g/ is virtually unaffected by the nature of C_2. In Hungarian on the other hand, the vowel length distinction is reduced or erased when C_2 is an obstruent. In the data summarised in figure 6.8 the phenomenon seems most marked before /t/ where for all practical purposes there is complete neutralisation of the contrast. _/z/, which displays a 14 ms difference in vowel length, represents the other end of the scale.

A two-way ANOVA for *C_1 laryngeal specification* * *C_2 laryngeal specification* on the vowel length data summarised in figure 6.8 fails to detect any effects of *C_1 laryngeal specification*, F(1,269) = 3.04, not significant, *C_2 laryngeal specification*, F(1,269) = 2.70, not significant, or *C_1 laryngeal specification* * *C_2 laryngeal specification*, F(1,269) = .784, not significant.

This is an interesting result, since the absence of an effect of *C_1 laryngeal specification* suggests that vowel length distinctions tend to neutralise when velar plosives are followed by another obstruent, whilst the absence of an effect of *C_2 laryngeal specification* indicates that C_2 obstruents are also unable to trigger any consistent length effects, and consequently that the neutralisation process is not *assimilatory* in the conventional sense. In other words, it appears that before obstruent clusters vowel length cues neither the underlying contrast between /k/ and /g/, nor the laryngeal specification of the obstruents that follow them.

Whilst this observation might be problematic for phonological feature-spreading accounts of RVA, it is certainly not predicted by a coarticulatory account of voicing assimilation either. In chapter 4 I argued at length that the coarticulation of articulatory gestures realising [±tense] cannot have an effect on the duration of a preceding vowel, and this means that the only possible conclusion at this point is that Hungarian RVA is not, or at least not solely, based on coarticulation.

Fricatives The patterning of vowel length before clusters starting with a fricative provides an interesting final twist to this argument, as there is evidence that before such clusters the vowel length contrast is retained. Note, first of all that in the baseline environment the vowel length contrast between /ʃ/ and /ʒ/ is more pronounced (at 42 ms) than that between /k/ and /g/ in the same context (25 ms). Unsurprisingly therefore, the difference is statistically significant

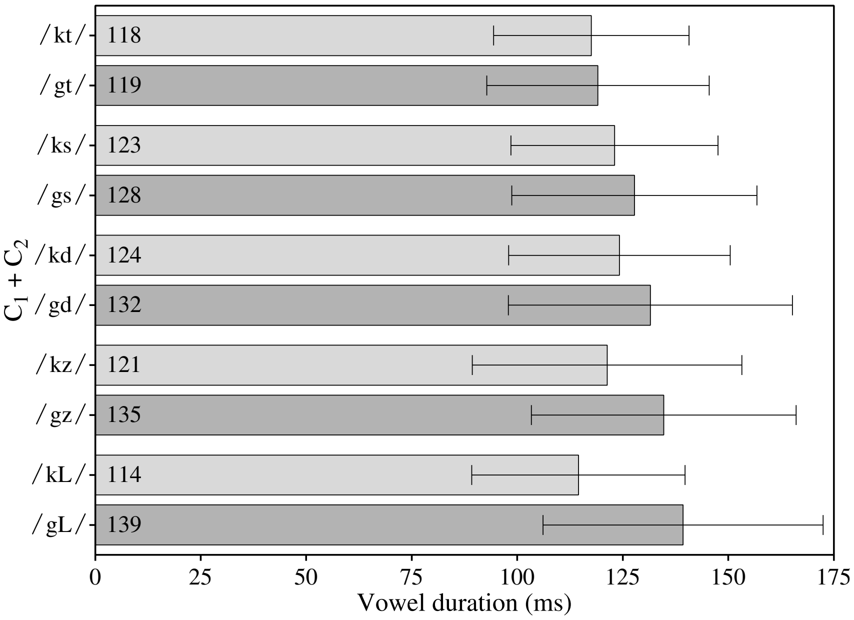

Figure 6.8: Experiment 2: Duration of vowels preceding /k/ and /g/ across C_2 contexts (only lexically long vowels included). All measures are in ms; error bars represent the mean ± 1 standard deviation.

according to a t-test: t(45) = -4.81, $p < .001$.

However, whereas the vowel length contrast between the velar stops is (near-)neutralised before /t, d, s, z/ it is largely retained before /ʃ, ʒ/ + /t, d/. The increased in vowel length before /ʃ/ + /d/ may reflect some degree of assimilation, but if vowel length indeed assimilates to C_2 the effect is far too weak to erase the distinction expressing the lexical contrast between tense and lax postalveolar fricatives.

This impression is borne out by a two-way ANOVA for *C_1 laryngeal specification* * *C_2 laryngeal specification* on the vowel length values found in pre-obstruent contexts. This ANOVA reveals a (highly) significant effect of *C_1 laryngeal specification* only, F(1,90) = 43.42, $p < .001$, whilst the effect of *C_2 laryngeal specification*, F(1,90) = 3.05, can only be regarded as a trend (p ¡ .085). The effect of the interaction of *C_1 laryngeal specification* and *C_2 laryngeal specification*, F(1,90) = 1.31, is not significant.

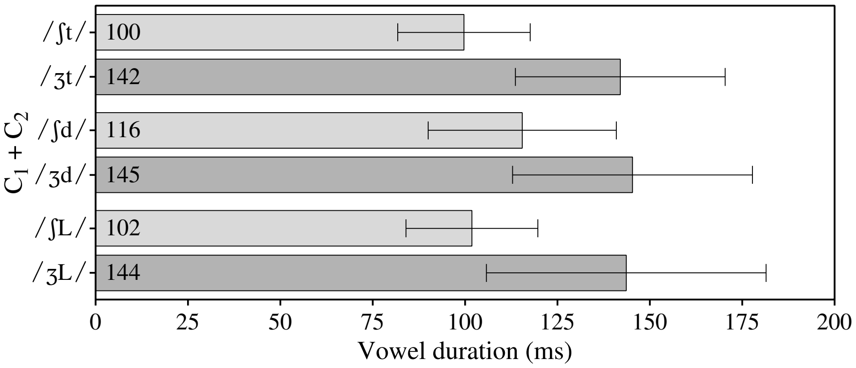

Figure 6.9: Experiment 2: Duration of vowels preceding /k/ and /g/ across C_2 contexts (only lexically long vowels included). All measures are in ms; error bars represent the mean ± 1 standard deviation.

6.4 General discussion and conclusions

Two things stand out in the results of experiment 2. The first is that Hungarian RVA leads to incomplete neutralisation of [tense] distinctions in target sounds. For example, there are residual traces of the the underlying contrasts between /k/ and /g/, and /ʃ/ and /ʒ/ in terms of C_1 voicing. It is interesting that the lack of phonetic voicing neutralisation should occur before the lax obstruents /d, z/, which mirrors the findings with regard to English above. The vowel length contrast between /ʒ/ and /ʃ/ is also preserved in the presence of a following obstruent. In addition, the behaviour of Hungarian /k, g/ is highly similar to that of their English counterparts in that the closure stage of the tense plosive shortens before another obstruent (regardless of its laryngeal specification), whilst there is some indication that the patterning of closure stage duration maintains a faint trace of the underlying [tense] contrast.

Considered in isolation, the similarities of these observations to the English results reported in the previous chapter, suggest a similar conclusion to the one drawn above with regard to the nature of RVA in English. However, this conclusion is contradicted by the second striking fact about the results of experiment 2, viz. that the vowel length distinction between /k, g/ is (near-)neutralised in the presence of a following obstruent. As argued at length in section 4.1.2 above, a purely articulatory form of RVA should leave vowel duration unaffected, and so it would seem that Hungarian RVA is not, or not purely, driven by coarticulation.

This raises a number of questions that must remain largely unanswered here. First and foremost is the question what process is responsible for the neutralisation of vowel length distinctions before velar stop + obstruent sequences. An obvious hypothesis is that Hungarian RVA is a proper phonological process and therefore reflected by all phonetic correlates of [tense]. Under this hypothesis, the incomplete neutralisation effect would be a byproduct of the sort of lexical interference briefly touched on in section 3.2.2 rather than of the coarticulatory nature of the rule. In other words, the Hungarian data would represent “an attempt at neutralisation”, to borrow a phrase from Jim Scobbie (p.c.).

This account might be extended along the lines proposed by Myers (2002) to include the idea that the Hungarian version of RVA is a phonologised version of the English version, caused by the effects of the latter on the perceptibility of plosives in the relevant contexts.

Plausible and attractive as this hypothesis may sound, it raises a number of further issues. One is why Hungarian (part-)phonologised the coarticulatory process found in English, whilst the latter language failed to follow this line of development itself. It is tempting to speculate that the symmetry of (fortis as well as) lenis obstruents (both /d/ and /z/) and/or differences in the role of vowel length in cueing [tense] have played a role here.

A second issue is why the present results show fricatives to be more resistant to neutralisation than plosives (assuming that this is not merely an artefact of the chosen stimulus items). One possible hypothesis in this area is that the cues to [±tense] in fricatives are somehow less vulnerable to the effects of coarticulatory RVA than stops, perhaps because a greater role of vowel length in cueing the contrast (note that there is a greater degree of vowel shortening before ʃ than before /k/ in the baseline context.)

However, the data presented here does not allow for these issues to be resolved, and therefore I will leave them to future research.

Chapter 7

Experiment 3: voicing assimilation in Dutch three-term clusters

The aim of this chapter is to extend the comparative investigation of regressive voicing assimilation that was begun in the previous chapter in a number of ways. The experiment reported here examines the phonetic manifestation of voicing assimilation in a third language, Dutch, and in word-final clusters rather than singleton obstruents. Moreover, it assesses the influence of global variations in speaking register on regressive voicing assimilation.

The results of this experiment broadly support the predictions of a coarticulation-based view of voicing assimilation. They are also consistent with the hypothesis that Dutch dynamic final neutralisation leads to the phonetic underspecification of [tense] features, as proposed by Ernestus (2000). The principal conclusions of the investigation are first, that contrary to assertions in some of the literature on the topic, regressive assimilation does take place in Dutch three-term obstruent clusters with a medial fricative. Second, contrary to the pervasive assumption that Dutch RVA is asymmetrically triggered by lax plosives only, the acoustic data reported below indicate that the process is triggered by tense and lax plosives alike. Third, Dutch RVA affects the voicing of target obstruents, but not the other phonetic correlates of [tense], and therefore operates in a highly similar way to its English counterpart. Somewhat surprisingly however, there is little evidence for either an increase or a decrease in RVA at higher speaking rates.

7.1 Predictions

Dutch word-final neutralisation, regressive voicing assimilation, and postobstruent fricative devoicing were discussed in chapters 2 and 3 respectively. A brief recapitulation of the broadly accepted description of these phenomena should suffice here. According to most phonologists, Dutch has a process of word-final neutralisation that erases all distinctions between word-final tense and lax obstruents. These word-final obstruents are subject to an (optional) rule of regressive voicing assimilation when they precede a lenis plosive. Fortis plosives are generally assumed not to trigger assimilation (or to trigger it vacuously) because word-final obstruents are subject to the more general final neutralisation/devoicing process. Lenis fricatives generally do not trigger regressive assimilation but are devoiced after another obstruent. These three processes were illustrated in (4), (8), and (14) above, and repeated (in part) here as (22) for convenience.

(22) Final neutralisation and regressive voicing assimilation in Dutch

a. Final neutralisation in Dutch

UR	Plural	Citation	diminutive	Gloss
/xrɑp/	[χrɑpən]	[χrɑp]	[χrɑpjə]	joke
/krɑb/	[krɑbən]	[krɑp]	[krɑpjə]	crab
/ɣraːt/	[χraːtən]	[χraːt]	[χraːtjə]	fishbone
/ɣraːd/	[χraːdən]	[χraːt]	[χraːtjə]	degree

b. Regressive voicing assimilation

UR	Phonetic form	Gloss
/ʋeːk/ + /diːr/	[ʋeʲːgdiɹ]	mollusc
/zɑnd/ + /bɑnk/	[zɑndbɑnk]	sand bank
/vɪs/ + /diːfjə/	[vɪzdifjə]	common tern
/rɛiz/ + /duːl/	[rɛizdul]	destination

c. Progressive devoicing of lax fricatives

UR	Phonetic form	Gloss
/drœk/ + /vɑt/	[drʉkfɑt]	pressure vessel
/riːt/ + /zɑŋər/	[ritsɑŋəɹ]	sedge warbler
/lɔf/ +/zɑŋ/	[lɔfz̥ɑŋ]	eulogy
/klɑs/ + /vərklɛɪnɪŋ/	[klɑsv̥ərklɛɪnɪŋ]	class size reduction

With the notable exception of Ernestus (2000), most accounts of Dutch laryngeal phonology subscribe to a lexical feature analysis of RVA and the fortition analysis of final neutralisation. The first part of this section summarises the (shared) predictions of these accounts. I will not delve into the specifics of the wide variety of published and unpublished accounts, but focus on their broad assumptions and their implications for the phonetics of Dutch regressive voicing assimilation, and in particular the behaviour of three-term obstruent clusters. A detailed dissection of recent generative approaches to laryngeal neutralisation and voicing assimilation that is directly applicable to many recent analyses of Dutch is provided in the next chapter, and the reader interested in yet more detail is referred to sources such as Trommelen & Zonneveld (1979), Booij (1981), Berendsen (1983), Booij (1995), and Ernestus (2000). The second part of this section describes what the phonetic manifestations of Dutch RVA should be if it is to be regarded as a purely coarticulatory process, whilst 7.1.3 reviews earlier observations concerning the behaviour of three-term clusters.

7.1.1 Predictions of the 'standard' analysis

The three rules in (23) are the essential components of what I will call the 'standard' analysis of Dutch final neutralisation and regressive voicing assimilation. (23a) expresses the idea that final neutralisation is an asymmetric rule that changes lax obstruents into tense ones word finally. (23b) represents the idea that Dutch regressive voicing assimilation is both manner asymmetric (it is triggered only by plosives) and tense asymmetric (it is triggered only by lax obstruents). Whilst most accounts assume that the manner asymmetry is an idiosyncrasy of Dutch, there has been a tendency in recent generative work to view RVA as typically [tense]-asymmetric (according to this approach languages such as Hungarian and Yiddish belong to a marked or exceptional type: cf. chapter 8). Rule (23c) finally, expresses the idea that despite its similarities to fricative devoicing processes found elsewhere, Dutch postobstruent fricative devoicing reflects true linguistic process that spreads [+tense] rightwards rather than a passive devoicing process (cf. 4.3).

The standard analysis of Dutch final neutralisation and voicing assimilation derives a number of predictions about the phonetic manifestation of these processes. In addition, depending on the relative ordering (or ranking) of the rules in (23) and the precise definition of the RVA rule, it generates a prediction about the (non)-application of RVA in three-term clusters with a medial fricative. First, as any lexical feature analysis, it predicts that Dutch regressive voicing assimilation should apply to all phonetic features involved in the realisation of [±tense] (cf. prediction 10b in section 4.1.1). Second, because it subscribes to a fortition account of final neutralisation, the standard analysis predicts that Dutch RVA is [tense]-asymmetric phonetically as well as phonologically. This means that Dutch word-final obstruents should be phonetically identical before fortis obstruents and sonorants: they are [+tense] in both environments and not subject to regressive assimilation (except in trivial fashion).

(23) The standard analysis of Dutch final neutralisation, RVA, and postobstruent fricative devoicing

a. Final neutralisation as fortition (see chapter 3)

$$\begin{bmatrix}\text{-son}\\ \text{-tense}\end{bmatrix} \longrightarrow [\text{+tense}]\_\#$$

b. Dutch regressive voicing assimilation (e.g., Cohen et al. 1972; Trommelen & Zonneveld 1979; Berendsen 1983)

$$\begin{bmatrix}\text{-son}\\ \text{+tense}\end{bmatrix} \longrightarrow [\text{-tense}]/\_\begin{bmatrix}\text{-son}\\ \text{-cont}\\ \text{-tense}\end{bmatrix}$$

c. Postobstruent fricative devoicing (e.g., Trommelen & Zonneveld 1979; Lombardi 1999)

$$\begin{bmatrix}\text{-son}\\ \text{+cont}\\ \text{-tense}\end{bmatrix} \longrightarrow [\text{+tense}]/[\text{-son}]\_$$

This second prediction of the standard analysis is rarely questioned in the literature on Dutch regressive voicing assimilation and often seems to be regarded as part of the description of the process instead of as part of its analysis. Worse, it has become something of a self-fulfilling prophecy since it is generally left untested by instrumental studies of Dutch. The logic of this state of affairs is made explicit by the following passage from Slis (1986) (emphasis mine):

> Since for syllable-final obstruents a final devoicing rule holds (Trommelen & Zonneveld, 1979; Booij, 1981), the first consonants of the clusters at issue (C_1) will have to be voiceless. This restriction implies that the second consonant in our clusters (C_2) has to be a voiced [i.e., lenis] obstruent; *if it was voiceless the clusters would consist of two voiceless obstruents in which no assimilation of voice could be studied.* (Slis 1986:313)

The predictions of the standard analysis concerning the behaviour of three-term clusters with a medial fricative are less unequivocal. On purely logical grounds, there are 4 possible surface (phonological) forms for the /ts/ sequence in a form such as /fiːts/ + /bɑnd/, *bicycle tyre*: [ts], [tz], [ds], and [dz]. Depending on assumptions about rule ordering (or constraint ranking) and the precise formal definition of the regressive assimilation rule, the standard analysis can be set up to derive all of these sequences except [ds]

Consider first a procedural model that orders (23b) before postobstruent fricative devoicing. This type of model predicts that such forms are realised with a voiceless ([+tense]) medial fricative ([fiːtsbɑnt]) because the former rule feeds the latter. The same is true if the rules are interpreted as violable well-formedness conditions or filters and (23b) dominates (23c). If the order of rule application (or ranking) is reversed it is predicted that /fiːts/ + /bɑnd/ is realised as [fiːtzbɑnt], because in this case (23c) feeds regressive voicing assimilation.

Note that given the definitions in (23), it is impossible to derive a form with a fully voiced three term cluster ([fiːdzbɑnt]) since even if the medial /s/ in the cluster surfaces as [z], further rightward spreading of voicing is prohibited by the manner restriction on (23b). By contrast, obstruent + stop + lenis stop sequences, as in the phrase /kɔlɛkt bɛlən/ *(to) make a collect call* or /ɑxt/ + /baːn/, *rollercoaster*, are predicted to surface with *iterative* RVA, i.e., with voicing throughout, under either ordering or ranking of (23b) and postobstruent fricative devoicing.

A different situation emerges if (23b) is reformulated as an iterative rule that spreads [-tense] leftwards from both plosives and fricatives. Procedural models incorporating such a manner-symmetric regressive assimilation rule predict that all clusters ending in a lenis plosive end up fully voiced (cf. Booij 1981). In such models (23c) has to be ordered before RVA in order to rule out assimilation to lenis fricatives in word-initial position. The derivation of the surface form for /fiːts/ + /bɑnd/ then proceeds as in (24): first, fricative devoicing fails to apply (or applies vacuously) to the cluster-medial /s/ in /fiːts/ + /bɑnd/. This fricative is subsequently voiced by RVA, and because there are no longer any manner restrictions on the process, it can ‘transmit’ voicing to the preceding coronal stop by means of a second iteration of the assimilation rule.[1]

(24) Derivation of surface voicing in obstruent + fricative + lenis plosive clusters using a manner-symmetric RVA rule

Underlying form	/fiːts/ + /bɑnd/
Final neutralisation	/fiːtsbɑnt/
Fricative devoicing	N/A
RVA, iteration 1	/fiːtzbɑnt/
RVA, iteration 2	/fiːdzbɑnt/
Surface form	[fiːdzbɑnt]

A third surface form for the final cluster of /fiː/ in /fiːts/ + /bɑnd/ is pre-

---

[1] Final neutralisation is incorporated for the sake of completeness. The OT analysis of Dutch laryngeal phonology in Grijzenhout & Krämer (1998) also predicts that the medial cluster of /fiːts/ + /bɑnd/, albeit on different grounds. The key to this prediction is that word-initial underlyingly [-tense] fricatives are treated differently from those that acquire this specification through assimilation.

dicted by at least one OT model of Dutch laryngeal phonology. Lombardi (1997) presents a constraint-based analysis of Dutch voicing assimilation built around the interaction of two constraints, *IdentOnset(Laryngeal)* and *AGREE*, which demands manner-symmetric regressive voicing assimilation (cf. section 8.2.5) To avoid assimilation to word-initial lenis fricatives both constraints are dominated by *FricVoice*, which stipulates that postobstruent fricatives be [+tense] (hence voiceless). The result of this ranking is that all-tense sequences emerge as the optimal candidates for underlying obstruent + fricative + lenis stop clusters, because *FricVoice* filters out all candidates with a [-tense] medial fricative and *AGREE* banishes all remaining output forms with mixed voicing. In other words, it predicts that /fiːts/ + /bɑnd/ surfaces as [fiːtspɑnd].

7.1.2 Predictions of the phonetic theory

The two most important predictions of a coarticulation-based theory of RVA with regard to the behaviour of Dutch word-final stop + fricative clusters are set out in (25a) and (25a) (the prediction about the phonetic manifestation of RVA is as before, cf. 25c). First, such clusters are predicted to be subject to regressive assimilation when followed by an actively (de)voiced obstruent. The effect of assimilation on a sequence of obstruents may not be fully proportional to that on a single stop or fricative if anticipatory articulation of the gestures involved in active (de)voicing starts relatively late (and is therefore relatively weak around the time of the onset of the first obstruent in the sequence). Furthermore, if one or both of the obstruents in a word-final clusters are actively devoiced, the combined coarticulatory 'weight' of their voicing targets may temper the effect of a following actively voiced lax obstruent. But coarticulation can not be switched of in the way a phonological assimilation rule can fail to apply, and as a consequence, the phonetic theory of RVA predicts that Dutch word-final obstruent clusters should be subject to some degree of assimilation.

(25) Predictions of a coarticulation-based theory of voicing assimilation concerning the behaviour of word-final plosive + fricative clusters

a. Dutch word-final plosive + fricative clusters are subject to regressive voicing assimilation

b. If these word-final clusters are phonetically underspecified for [tense], regressive assimilation is [tense]-symmetric: a following lax stop should trigger an increase in voicing relative to a baseline sonorant environment whilst a following tense obstruent should cause a decrease (cf. Ernestus 2000)

c. The assimilatory effects of fortis and lenis plosives are limited to voicing and features mechanically dependent on the production of voicing distinctions. Cf. (11a)

According to the model proposed by Ernestus (2000), Dutch neutralised obstruents are phonetically underspecified for [tense], which entails that they lack phonetic targets for voicing, and regressive voicing assimilation is a purely phonetic phenomenon. Ernestus notes that one unexplored prediction of this model is that contrary to the standard view, Dutch regressive assimilation at word boundaries should be (observably) [tense]-symmetric. I briefly sketched the reasoning behind this prediction in 4.1.2 in the context of the assimilatory behaviour of sonorants. The key assumption is that ceteris paribus, [0tense] obstruents have longer voiced intervals when preceded by a vowel than the corresponding [-tense] obstruents. In the former voicing continues from the vowel into the constriction phase of an obstruent until the transglottal pressure difference falls below the critical 200 Pa threshold or the glottis opens for some reason that is unrelated to voicing control. The latter are assumed to be accompanied by active devoicing gestures (glottal abduction, glottal constriction) that force vocal fold vibration to terminate at some time prior to the point of passive devoicing. Now if a [0tense] obstruent is followed by a sonorant, the duration of its voiced interval will be in accordance with the window for passive voicing as dictated by vocal tract aerodynamics and mechanics, because Dutch sonorants lack voicing targets and are therefore unable to exert any coarticulatory pressure on a preceding obstruent. The active devoicing gestures accompanying a [-tense] obstruent on the other hand, should be coarticulated during a preceding [0tense] fricative or stop. As a result, the [0tense] obstruent will itself become actively devoiced to some degree, and at least in principle, this will lead to voicing offset prior to the point determined by passive voicing alone. The same logic predicts that if a [0tense] obstruent precedes an actively voiced [-tense] obstruent, the length of its voiced interval is predicted to increase beyond the length of the passive voicing window.

Whether the predicted three-way pattern of voicing in neutralised obstruents before fortis obstruents (short voicing interval), sonorants (intermediate) and actively voiced lenis obstruents (long) indeed materialises hinges on the actual length of the passive voicing window for particular obstruents with a particular preceding context. Recall from 2.1 that estimates for the amount of passive voicing in a postvocalic obstruent run between 25 and 100 ms. The lower bound of this range is approximately the same as that of the presumably actively devoiced fortis velar stops of English and Hungarian in baseline and __[+tense] environments. This would suggest that if 25 ms was the approximate passive voicing window for (Dutch) neutralised obstruents, the coarticulatory effect of a following [-tense] stop would be impossible to observe in the speech signal (although it would be no less real at the articulatory level). However, if the passive voicing window for a neutralised obstruent is, say, 35 ms or longer, the three-way voicing pattern predicted by the phonetic theory of RVA should emerge from the acoustics.

7.1.3 Observations on assimilation in three-term obstruent clusters

The assimilatory behaviour of clusters composed of more than two obstruents has never played a great role in the modelling of RVA, and that is perhaps part of the reason that observations on this point are relatively rare. Consequently, it is hard to determine whether RVA is typically iterative or not, and whether the manner of articulations of the obstruents involved typically imposes restrictions on iterativity. Nevertheless, Katz (1987) describes Yiddish regressive voicing assimilation as applying to all obstruents preceding the trigger. Thus, he transcribes the realisation of /ɛrʃt/ + /gəʃen/, *just happened*, as [ɛrʒdgəʃen]. Siptár & Törkenczy (2000) claim that Hungarian voicing assimilation is iterative, too. /list/ + /bøːl/ *from flour* for example, is said to surface with a fully voiced medial cluster: [lizdbøːl]. Opinions seem to differ with regard to Frisian: Riemersma (1979) appears to claim that no assimilation occurs in obstruent + fricative + lenis stop sequences, whereas examples provided by van der Meer & de Graaf (1986) indicate that such clusters are subject to RVA like any other sequence of obstruents in Frisian. The data from Yiddish, Hungarian, and Frisian therefore largely supports a phonetic view of regressive voicing assimilation, or, in light of the results presented in the previous chapter, a view of RVA as ultimately grounded in a phonetic process.

However, regressive voicing assimilation in Dutch has traditionally been described as non-iterative in obstruent + /s/ + lenis stop sequences. For example, Brink (1975), citing earlier work on Dutch, states that /fiːts/ + /bɑnd/ can be pronounced with a prevoiced lenis stop as in [fiːtsbɑnd], or with a (partially) devoiced lenis stop as in [fiːtsb̥ɑnt/, but is never realised with any voicing in the obstruents preceding the lenis stop. The status of this description (in part of the research community) as incontrovertible fact is reflected by Camminga & van Reenen (1980) who criticise the model developed by Booij (1981) for predicting RVA in stop + fricative + lenis stop sequences. It ostensibly solves the little rule ordering puzzle sketched above, indicating that the rule in (23c) takes precedence over (23b). At the same time, this description of Dutch regressive voicing assimilation raises doubts about a phonetic account of regressive voicing assimilation.

7.2 Methods

Subjects Subjects were 4 native speakers (MJ1, GBP3, both male, ER2, LB4, female) of Dutch between 21 and 45 at the time of recording. None of the subjects had a history of speech or hearing impairment. They were not paid for their participation in the experiment. Although all speakers were residents of the town of Groningen, where the local dialect is of the aspirating rather than the voicing type, this did not apply to the subject's speech, which can be roughly described as standard with minor (northern and western) local features.

Materials The stimuli for this experiment consisted of clusters combining an initial /p/ C_1, and a medial /s/ C_2 followed by a/p, t, b, d, m, h,/ C_3 or an unreduced vowel (/V/), which is usually preceded by a [ʔ] in Dutch.[2] Although there is evidence that final laryngeal neutralisation is complete in Dutch (Baumann, 1995), C_1 obstruents were consistently /p, k/ and orthographic <p, k>, to avoid a potential bias due spelling pronunciations or other incomplete neutralisation effects (cf. Fourakis & Iverson 1984 and chapter 3 above).

(26) Sample stimuli for experiment 3

a. Het was Jaa*ps t*unnel die onder water stond, niet zijn kelder
/hɛt ʋɑs jaː*ps t*ʏnəl diː ɔndər ʋaːtər stɔnd niːt zən kɛldər/
It was Jaap's tunnel that under-water stood, not his basement
It is Jaap's tunnel that was flooded, not his basement

b. Het was een Kaa*ps m*eisje dat de hoofdprijs won, niet een Kaaps jongetje
/hɛt ʋɑs ən kaː*ps m*ɛisjə dɑt də hoːvdprɛis ʋɔnniːt ən kaːps jɔŋətjə/
It was a Cape-ADJ. girl-DIM. who the head-prize won, not a Cape-ADJ. boy-DIM
It was a little girl from the Cape who won the first prize, not a little boy from the Cape

/m/ functioned as baseline environment in the usual way. /h/ and /V/ were included as C_3 environments to establish the effects of independent glottal articulations on the voicing of a preceding obstruent cluster. Neither /h/ nor/V/ or the glottal stop normally preceding it can be characterised as [±tense] but the articulatory gestures involved in the production of [h] (glottal abduction) and [ʔ] (glottal constriction) are commonly found as active devoicing gestures in the production of plosives. Consequently, the phonetic theory of RVA predicts that the amount of voicing during word-final /ps/ clusters should be highly similar before /h/, /V/, and /p, t/ (Ernestus, 2000).

C_1s were embedded in proper name N_1s consisting of a single syllable and preceded by a long low unrounded vowel /aː/. The medial /s/ always represented a adjectival marker, as in /kaːp/ + /s/, *of, from, pertaining to the Cape*

---

[2] There was an additional set of stimuli combining an initial /k/ and medial /s/ with a /p, t, b, d/ C_3. The responses to these stimuli are excluded from the discussion below because they follow exactly the same pattern as the responses to the /ps/ + C_3 sequences.

or a (related) possessive marker as in /jaːp/ + /s/, *belonging to Jaap*.[3] The carrier words (N_2) for C_3 were disyllabic nouns with an initial lexical stress. C_3 always preceded a long vowel or (phonotactically long) diphthong. The carrier words (N_2) for C_3 were disyllabic nouns with an initial lexical stress. The N_1 + N_2 collocations were further embedded in carrier sentences designed to attract a contrastive nuclear accent on N_2. Some sample stimuli (orthographic and phonological representations) appear in 26. Target clusters are represented in a slanted font.

Procedure The stimuli were presented to the subjects in a quasi-randomised order to avoid consecutive stimuli with identical consonant clusters. The subjects were asked to read the list of stimulus sentences 3 times. For the first, *Normal* reading, the subjects were asked to read the stimulus items at a self-selected comfortable rate. In an attempt to simulate a noisy environment, the subjects were then fitted with sound-treated headphones conveying a 80 dB white noise signal (a noise level roughly comparable to that on a moving city bus) for the second reading, and asked to speak in such a way that they could understand their own speech. The aim of impoverishing the subjects' auditory feedback was to elicit a more hyperarticulated speech variety that is sometimes referred to as the *Lombard reflex* (Lombardi 1991; see Junqua 1996 for an overview). Henceforth the second reading task will therefore be referred to as the *Lombard* condition. For the third, *Fast*, reading, subjects were asked to read the stimulus items as fast as possible in order to create a bias to more hypoarticulated speech.

The three reading tasks or conditions were intended to elicit the same stimulus items on a 3-point hypoarticulation scale, so as to build up a relatively complete 'phonetic map' of the realisation of Dutch three term clusters and to increase the chances of observing any form of regressive voicing assimilation.[4]

With the exception of the use of impoverished auditory feedback to elicit 'clear' speech, this methodology is similar to methods used in a number of experimental studies of speaking rate effects on plosive VOT. Sometimes test subject are simply asked to produce the same set of stimulus items in fast or 'clearly enunciated' speech (e.g., Kessinger & Blumstein 1997), other experiments (Miller et al., 1986; Magloire & Green, 1999) use so-called magnitude production techniques, which essentially consist of instructions to test subjects to speak *n* times faster and slower relative to some self-selected 'normal' baseline.

During each of the three readings subjects were asked to repeat an item if

---

[3] Strictly speaking it is not clear whether these morphemes are [+tense] /s/ or [0tense] /S/, but nothing below crucially hinges on this. For typographical reasons I will represent them as /s/.

[4] Predictions about the behaviour of RVA in different global speaking registers are reviewed in the introduction to 7.2.2.

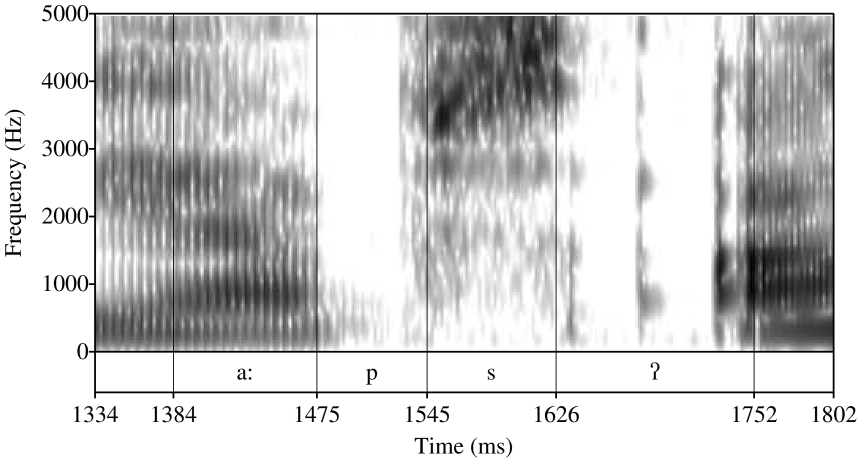

Figure 7.1: Experiment 3: sample segmentation of glottal stop preceded by /ps/. Broad band spectrogram of a /ps/ + [ʔ] cluster: subject = ER2, condition = Normal.

they produced a hesitation or speech error that was clearly audible to the experimenter and that affected the target cluster. In total, 1 (C_1 = /p/, C_2 = /s/) * 7 (C_3) * 10 (stimuli) * 3 (conditions) * 4 (speakers) = 840 utterances were recorded. Recordings were made onto minidisk in a sound-proofed room using a Brüel and Kjær condenser microphone (Type 4165) and measuring amplifier (Type 2609), and digitised at 22.5 kHz. Segmentation and acoustic measurements were carried out using PRAAT. 31 utterances had to be discarded because they contained a pause between C_2 and C_3 or small speech errors, leaving 809 utterances for segmentation and analysis.

The segmentation protocol was as for experiments 1 and 2 with added provisions for /m, h, V/. The boundary between /m/ a following vowel were determined on the basis of the offset of the nasal formant, whilst the offset of /h/ was defined as the onset of high frequency energy carried by the periodic source. [ʔ] was only marked as such if there was evidence of irregular glottal pulsing in the signal (this was virtually always the case): for an example, see figure 7.1. For reasons of time, V_1 was initially segmented and measured only for clusters beginning in /ps/ and ending in /p, t, b, d/. The results of the measurements was such that is was deemed unnecessary to investigate V_1 duration for clusters beginning in /ks/, or those ending in /m, h, V/.

The measurements that were made on the basis of the hand-segmented speech samples, as well as the relevant derived measures are listed in table 7.1, ordered by speech segment. In the light of the previous chapter the ra-

Table 7.1: Acoustic measurements and derived measures for Experiment 3.

Segment							
V_1		C_1C_2		C_2		V_2	
(a)	Duration	(d)	Duration	(g)	Closure duration (stops)	(i)	F_0 10-50 ms after C_1 offset
(b)	F_0 50-10 ms before C_1 onset	(e)	Voicing duration	(h)	VOT (stops)		
(c)	F_1 50-10 ms before C_1 onset	(f)	Voicing ratio				

tionale for the measures summarised in table 7.1 should require little further comment. Note that all measurements in the column for C_1C_2 were performed twice, for /p/ and /s/ individually. Because there was no variation in C_2 context, no attempt was made to segment the release of /p/ from the preceding closure phase. As stated above, V_1 duration was only measured for clusters ending in a plosive C_3.

7.2.1 Main results

Phonetic features of C_3 plosives As can be gleaned from table 7.2 and figure 7.2, the contrast between the tense and lax C_3 plosives is as would be expected from a voicing language. /p, t/ have a small positive VOT that is comparable to the value for English /d/ found in experiment 1. The average VOT for /b, d/ is -54 ms, which is somewhat larger than the value reported for Hungarian reported in the previous chapter, but note that the average duration of oral closure is considerably longer too for the Dutch stops, so that the mean proportion of oral closure that is voiced (.55) is lower than the value found for Hungarian (.74). The difference in F_0 between tense and lax stops is comparatively large for both the female and male subjects: at 10 ms into the vowel the gap is 44 Hz (274 vs. 230 Hz) for the former and 31 Hz (214 vs. 183 Hz) for the latter. As in English and Hungarian, the sonorant baseline environment patterns with the lax rather than with the tense stops.[5]

These differences in VOT and F_0 stand up to statistical scrutiny. A t-test on the VOT data reveals a highly significant effect, t(465) =23.24, p $<$ .001 and the same applies to a three-way ANOVA for *C_3 laryngeal specification* on the F_0 values at 10 ms into the following vowel (female speakers only, clusters ending in /h, V/ excluded): F(1,298) = 31.55, p$<$ .001. Tukey and Scheffe post-hoc tests show that as in English and Hungarian, the tense stops are distinct from

---

[5] Clusters ending in /h, V/ are excluded from figure 7.2 for clarity, and because segmentation of the offset of [h, ʔ] was felt to be relatively unreliable.

Table 7.2: Experiment 3: closure duration and VOT for C_3 plosives. All values in ms, and pooled across places of articulation (labial and alveolar) and reading tasks. Standard deviations in brackets.

C_3	VOT	Closure duration	N
Tense (/p, t/)	16 (9)	89 (39)	234
Lax (/b, d/)	-54 (45)	85 (33)	233

both the lax stops and sonorant /m/ ($p < .001$ for all pairwise comparisons), whilst the means for the latter two groups are not significantly different.

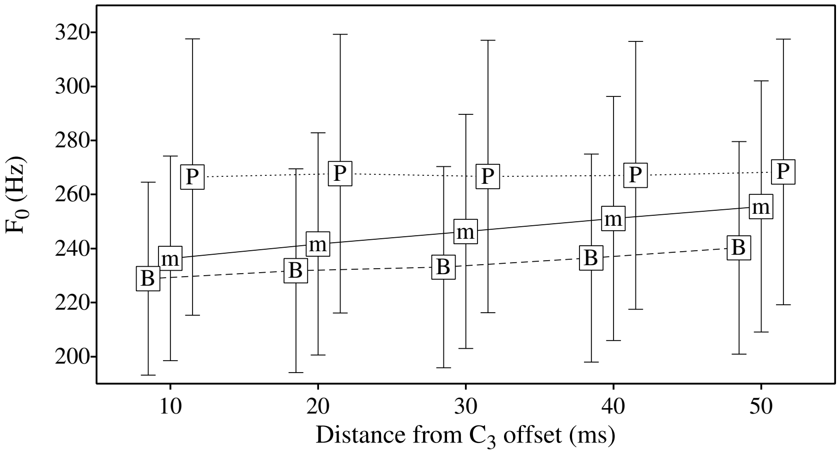

Figure 7.2: Experiment 3: F_0 (Hz) 10-50 ms into the vowel following C_3, for female speakers only. Clusters ending in /h/ and /V/ excluded, and data pooled over reading tasks. Error bars represent the mean ± 1 standard deviation.

C_1 + C_2 voicing and duration Voicing and duration data for /ps/ clusters, which are summarised in figure 7.3, lend virtually unequivocal support to an articulation-based view of RVA on all counts. First, there is a clear increase in the duration of voicing before lax stops vis-à-vis all other contexts, which means that word-final /p/ + /s/ clusters are subject to regressive voicing assimilation under any viable phonetic definition of the term. Second, the duration of the voiced interval before /m/ is roughly intermediate between the durations of voicing before tense and lax stops, which indicates that, contrary to the received view, Dutch RVA is tense-symmetric. Third, /h/ and (preglottalised) /V/ pattern with tense stops as far as the voicing of a preceding obstruent cluster is concerned, which supports the idea that RVA is largely a matter of the coarticulation of gestures involved in the realisation of voicing targets. Fourth, voicing and segmental duration do not maintain the inverse pattern that is typical of the phonetic expression of [tense] outside assimilation contexts. Variations in segmental duration rather seem to reflect mechanical interactions between glottal articulations involved in the expression of laryngeal (segmental) contrast and those involved in the production of fricatives.

Recall that according to the standard view of Dutch regressive voicing assimilation, there should be no difference in voicing between obstruents preceding a sonorant consonant and those followed by a tense obstruent. The (overall) voicing durations depicted in 7.3 plainly contradict this view and support prediction (25b) of the phonetic theory because there is a 13 ms difference between the two contexts. Note that the increase of voicing before lax stops is itself at odds with the assertion by Brink (1975) and others that no assimilation takes place in three-term clusters with a medial fricative. Interestingly, the effect of a fortis stop on the voicing of a preceding /ps/ cluster is highly similar to that of a preceding /h/ or preglottalised vowel. Given that glottal abduction and glottal compression are known active devoicing strategies this observation is consistent with the idea that the tense stops of voicing languages (in contrast to the lax stops of aspirating languages) are actively devoiced.

At first sight, the overall duration of /ps/ before lenis plosives, fortis plosives, /m/ and /h/ looks consistent with a representation of RVA as phonological feature spreading. $C_1 + C_2$ segmental duration is relatively long before /p, t/ (142 ms), short before /b, d/ (120 ms), whilst /m, h/ represent a more or less intermediate class. However, this classification of C_3 does not match the grouping implied by the voicing data, which classifies /h/ with the fortis stops rather than with /m/. Moreover, the standard theory places /V/ in the set of ‘neutral’ contexts (along with tense stops, /m/, and /h/), but it shortens the duration of a preceding /ps/ sequence even more than the lenis plosives. In other words, the inverse patterning of segmental duration and voicing that is predicted by a lexical feature analysis of RVA does not hold for the data reported here. This ‘mismatch’ between C_1C_2 voicing duration and segmental duration, is reminiscent of the behaviour of English and Hungarian C_1 as reported above.

This argument is bolstered if the segmental durations of /p/ and /s/ are considered separately. The right panel of figure 7.3 shows that differences in overall C_1C_2 segmental length are mainly due to differences in the duration of /s/ rather than the initial /p/: the means for the latter all cluster within a 8 ms band, whereas the maximal difference for the former is 20 ms. This pattern is familiar from the behaviour of C_1 plosive release duration and C_1 fricative duration observed in the results of experiment 2 (cf. table 7.3) and suggests an explanation along similar lines. In other words, it appears likely that mechanical

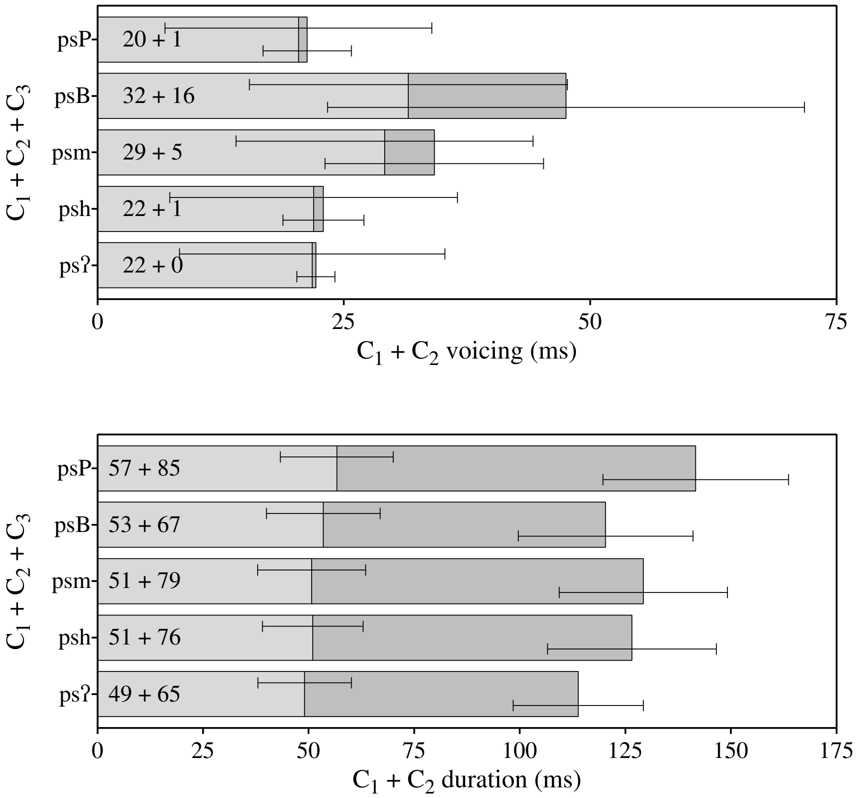

Figure 7.3: Experiment 3: voicing and duration of C_1 + C_2. Voicing (left) and segmental duration (right) of /p/ (bottom segments, dark grey fill) and /s/ (top segments, light grey) before fortis stops, lenis stops, /m/, /h/, and /V/ (all values in ms). Data pooled across registers.

interactions between the (glottal) articulations of C_3 and the medial /s/ form at least part of the explanation for the segmental duration facts.

At several previous points I have invoked the purely mechanical account of the linkage of duration and voicing in fricatives as proposed in Stevens et al. (1992): the vocal fold adduction required for the production of voicing inhibits the high transglottal airflow required for the production of frication noise, and consequently the frication phase of a fricative shortens if active voicing measures are imposed on it (e.g., by coarticulation). This idea extends to the shortening of a fricative before [ʔ] because the latter is also produced with an adduction gesture that inhibits airflow across the glottis. As noted above, it is the same gesture that impedes voicing (in a preceding sound) because it is stronger than the adduction involved in modal voicing and leads to glottal compression. Thus, glottal coarticulation offers a natural account for the assimilatory behaviour of /s/ followed by [ʔ], which is rather Janus-faced from the perspective of a lexical feature analysis.

To test whether the above impressionistic observations stand up to statistical analysis, a number of tests were performed. First, clusters followed by a plosive C_3 were directly compared by means of a t-test on the voicing duration data, in order to demonstrate that RVA takes place even from the perspective of the standard view. This t-tests indicates that the differences in C_1C_2 voicing duration is statistically highly significant, t(465) = -10.73, p $<$ .001, and thus leaves little doubt about the inaccuracy of descriptions of Dutch three-term clusters with a medial fricative as impervious to RVA.

Table 7.3: Experiment 3: results of Tukey and Scheffe post hoc tests on the ANOVAS for C_1 + C_2 voicing (top) and segmental duration (bottom). $\star$: significant difference (p $<$ .05) according to both tests; $\diamond$: significant difference (p$<$ .05) according to Tukey only; n.s.: difference not significant on either test.

	/p, t/	/b, d/	/m/	/h/	/V/
/p, t/		$\star$	$\star$	n.s.	n.s.
/b, d/	$\star$		$\star$	$\star$	$\star$
/m/	$\star$	$\star$		$\star$	$\star$
/h/	n.s.	$\star$	$\star$		n.s.
/V/	n.s.	$\star$	$\star$	n.s.	

	/p, t/	/b, d/	/m/	/h/	/V/
/p, t/		$\star$	$\star$	$\star$	$\star$
/b, d/	$\star$		$\diamond$	n.s.	n.s.
/m/	$\star$	$\diamond$		n.s.	$\star$
/h/	$\star$	n.s.	n.s.		$\star$
/V/	$\star$	n.s.	$\star$	$\star$	

Next, 3 one-way ANOVAs for *C_3 laryngeal specification* were performed on the C_1C_2 voicing and duration data with the baseline environment /m/ and /h, V/ included as three separate laryngeal specifications in addition to [±tense]. The ANOVAs on the voicing duration data, F(4,804) = 48.92, p $<$ .001, and segmental duration data F(4,804) = 25.77, p $<$ .001, both yield highly significant effects. Results of Tukey and Scheffe Post Hoc tests are summarised in table 7.3. The top panel shows the pairwise comparisons for C_1C_2 voicing duration, which clearly supports the impressionistic grouping of C_3 contexts for this parameter as /p,t/, /h/, /V/ vs. /b,d/ vs. /m/: for example, mean C_1C_2 voicing durations before the [+tense] plosives, /h/, and /V/ are all significantly different from those before both /b, d/ and /m/ but there are no such differences within the first group. The fact that the mean C_1C_2 voicing duration before /m/ is significantly different from the mean voicing durations before /p, t/ as well as /b, d/ indicates that Dutch RVA is indeed [tense]-symmetric.

The Tukey and Scheffe results for C_1C_2 segmental duration do not support the classification indicated by the left panel, however. As suggested by the right panel of figure 7.3, the clearest split here appears to be between /p, t/ and all other contexts. But no clear grouping emerges within the set of remaining C_3 environments: for example, whereas /b,d/ is distinct from /V/ in terms of voicing duration the difference in C_1C_2 segmental duration is not statistically significant. Thus, the statistical tests confirm the mismatch between voicing duration and segmental duration identified above (as well as in the English and Hungarian data), thereby casting yet more doubt on a lexical feature analysis of RVA which predicts that such mismatches should not occur.

Classifying target clusters as (perceptibly) 'assimilated' vs. 'unassimilated'

In an experimental production study of regressive voicing assimilation in Dutch two-way obstruent clusters, Slis (1986) employs a technique of quantifying RVA that is different from the methods used so far in this study. Slis classifies all obstruents preceding a lenis plosive C_2 as 'unassimilated' if they have a VTT < 50 ms and as 'showing regressive assimilation' if their VTT exceeds 50 ms. The cut-off point is based on the VTT of singleton intervocalic stops in Dutch as measured by Slis (1970), which indicates that there is a probability $< .0025$ that fortis stops have a VTT equal or greater to the mean of 25 ms + 2 standard deviations (10 ms) + 5 ms = 50 ms.[6] The definition of regressive voicing assimilation used by Slis (1986) assigns strict acoustic criteria to a method that is used in transcription studies and consequently his methodology exposes the size of the effect of [±tense] on C_1 voicing duration relative to the inherent variance within a laryngeal category in an intuitively transparent way. The relative magnitude of [±tense] effects vis-à-vis the noise caused by within-category variation has been used as a rough indicator of perceptual salience: O'Shaughnessy (1981) suggests that effects smaller than or equal to a single standard deviation from a baseline mean should be treated as below the threshold of perception.

There are several ways in which this technique can be applied to the C_1C_2 voicing data from the present experiment. Given that assimilation of C_1C_2 voicing appears to be [tense]-symmetric a natural procedure is to define three classes of /ps/ sequences using the *overall* mean C_1C_2 voicing duration of 31 ms and its standard deviation of 26 ms: a *'voiceless'* category or 'band' with relatively short intervals, a *'neutral voicing'* class centered around the overall mean, and a *'voiced'* category with relatively long voiced intervals.

---

[6] I.e., provided that the sample size is sufficiently large and VTT is normally distributed.

Regardless of the precise settings of the boundary values delimiting the three categories, the picture of Dutch regressive voicing assimilation that emerges from this method is not substantially different from the one drawn above. Figure 7.4 depicts the classification of /ps/ clusters if the neutral band is defined as the overall mean of 31 ms ± 1 standard deviation of 26 ms. There are very few 'voiced' clusters preceding [+tense] plosives, /h/, or /V/, whereas 30 % of the cases before /b,d/ belong in this class, with the _/m/ context almost exactly halfway in between (15%). The frequencies of 'devoiced' clusters hint at the same natural classes of assimilation environments, with similar frequencies before /p,t/ and /h/ and /V/ (17, 18, and 15% respectively), and lower figures for /m/ (8%) and /b,d/ (3%).

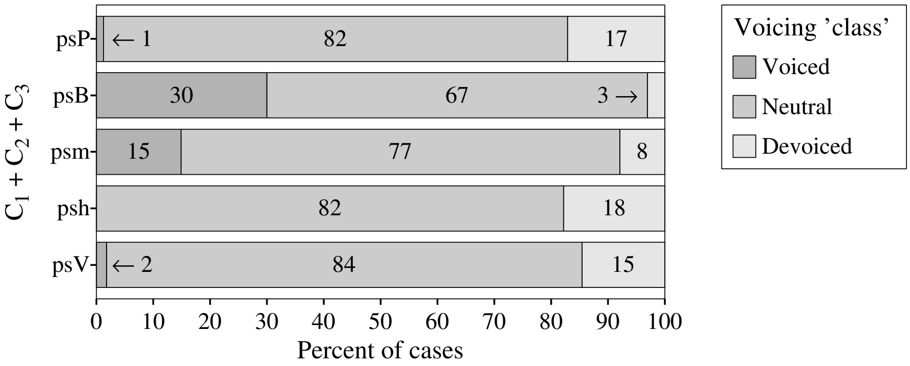

Figure 7.4: Experiment 3: frequencies of 'devoiced', 'neutral', and 'voiced' /ps/ clusters before fortis stops, lenis stops, /m, h, V/ if the neutral category is defined as 31 ± 26 ms (the overall mean ± 1 standard deviation). Numbers inside columns represent numbers of cases.

In his study of RVA in Dutch, Slis (1986) reports that lenis stops trigger 86% 'regressive assimilation' in preceding singleton plosives across a word boundary and before stress. This figure is considerably higher than the proportion of (equivalent) 'voiced' cases before [-tense] plosives in the present study. Note, however, that in the classification illustrated in figure 7.4 the cut-off point between the 'neutral' and 'voiced' bands is 57 ms of voicing as opposed to Slis's 50 ms. If the cut-off point is lowered to 50 ms the proportion of 'voiced' cases before [-tense] plosives rises to 36%, which is identical to the frequency of 'assimilated' singleton fricatives found by Slis (in the relevant environment). This implies that, however real, the effect of RVA on plosive + fricative clusters are in some sense weaker than the effect on singleton plosives and therefore perhaps less audible. This may in turn account for the claims in the descriptive literature that regressive voicing assimilation does not apply to plosive + fricative clusters in Dutch.

V_1 duration The mismatch between voicing duration and segmental duration in /ps/ sequences is consistent with prediction 25c of the phonetic theory. The same applies to absence of an assimilation reflex in V_1 duration. Vowel duration is known to pattern with [±tense] in the familiar way before word-medial obstruents in Dutch (e.g., Slis & Cohen 1969a), and can therefore not be dismissed as irrelevant by proponents of a lexical feature analysis. Figure 7.5 represents the mean duration of V_1 preceding /ps/ + [+tense] plosive and /ps/ + [-tense] plosive sequences (note again that V_1 was not segmented in any of the remaining cluster types for reasons of time). V_1 is slightly (3 ms) longer when C_3 is a lenis plosive, but a t-test, t(463)= -.82, indicates that this difference is far from statistically significant. It appears therefore, that regressive assimilation in Dutch behaves much as its English counterpart in affecting obstruent voicing but not any of the other correlates of [tense].

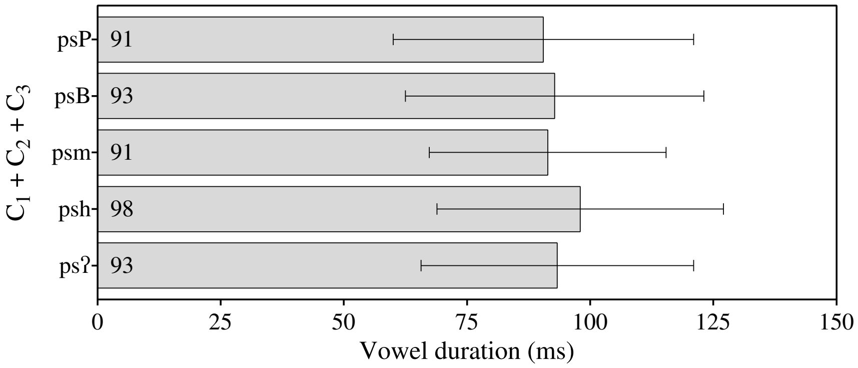

Figure 7.5: Experiment 3: mean duration (ms) of V_1 before /ps/ preceding tense and lax plosives. Error bars represent the mean + 1 standard deviation

Low-frequency spectral features F_0 and F_1 preceding /ps/ clusters mimic the behaviour observed for these features in the data from experiment 1. Despite the relatively large differences in F_0 following C_3 there is no regressive assimilation of fundamental frequency. F_1 values preceding C_1 on the other hand, do show some effect of C_3. Table 7.4 gives values for the first formant of the vowel /aː/ at 10 ms before the onset of C_1. Just as English /d/, Dutch /b, d/ appear to cause a decrease in the F_1 of the vowel preceding C_1, and the magnitude of the difference between tense and lax stops is similar in the two languages, too (approximately 20 Hz). Likewise, the behaviour of Dutch /m/ mirrors that of English /r/: both sonorants pattern with tense rather than with lax obstruents. Note that the values for /h, V/ appear to be intermediate between the extremes defined by the lax stops on the one hand and sonorants and tense stops on the other. A one-way ANOVA for C_3 *laryngeal specification* on the F_1 values at 10 ms before the onset of C_1 (clusters ending in /h, V/ excluded) reveals an effect that is significant, $F(2,598) = 4.13$, $p < .02$, but considerably weaker than that obtained from experiment 1.

Table 7.4: Experiment 3: F_1 values (Hz) of the vowel /aː/ at 10 ms before the onset of C_1. Standard deviations in brackets, and data pooled across reading tasks.

C_3	F_1 at C_1 - 10 ms	N
/p, t/	698 (94)	242
/b, d/	678 (87)	242
/m/	700 (77)	114
/h/	689 (81)	118
/V/	690 (95)	110

As noted in 5.4, the articulatory source of low-frequency spectral cues to [±tense] remains unclear, and consequently it is difficult to gauge the significance of assimilatory effects on F_1.

7.2.2 Reading task effects

Descriptions often suggest that the degree of voicing assimilation between words somehow increases in ‘casual’ or fast speech or becomes applicable in a wider range of contexts, and there is a range of possible explanations for such observations, including a decrease in the number and length of physical pauses. However, it is difficult to state any precise predictions about the interaction of regressive voicing assimilation and global speech register without a highly formalised speech production model. Generative frameworks such as those of Kaisse (1985) and Nespor & Vogel (1986) tend to model rules that vary with speech register as ‘late’ and/or optional parts of a phonological derivation, or in terms of prosodic reanalysis. But such models are inherently committed to a phonological, all-or-nothing view of rule application and therefore shed no light on the question of how a (potentially) gradient assimilation process might be shaped by changes in global register.

In its present prose formulation, the model described in chapter 1 does not derive any precise predictions about the interaction between global register variation and regressive voicing assimilation either. Because hypoarticulation (conceived as a relaxation of auditory targets) plays a key role in this model, there is a further complication in the interpretation of the present results if any variation induced by the three reading tasks cannot be characterised in terms of hypoarticulation. To the extent that the subjects' responses to the three reading tasks does show systematic variations in hypoarticulation, it might be expected that increased hypoarticulation would lead to three-term obstruent clusters converging on a completely passively voiced configuration (i.e., with an initial voicing tail and if the constrictions remain in place long enough, subsequent devoicing) regardless of their underlying specification. Note that this prediction runs counter to the common claim that assimilation increases in fast and 'casual' speech. However, a number of other factors including prosodic phrasing can be expected to distort this convergence. Moreover, even in the absence of such possibly confounding factors it might be difficult to identify any convergence on a passively voiced configuration, since global hypoarticulation is likely to affect segmental duration as well as the implementation of voicing contrasts.

In the following I will therefore simply describe the effect of the different reading tasks on the voicing of Dutch three-term obstruent clusters in the context of more global properties of the subjects' speech. Perhaps the most surprising part of this description is the observation that there is little evidence for an increase in regressive assimilation in fast speech. This finding contradicts earlier work on Dutch by Menert (1994).

Global effects of reading task Table 7.5 summarises the effects of reading task on a number of utterance-level variables: overall utterance duration, utterance mean F_1, and utterance F_0 range. Utterance duration was determined on the basis of segmentation by hand, and can be interpreted as an indicator of overall speaking rate. Utterance mean F_1 was defined as the raw mean of the F_1 values of all voiced samples of an utterance. It was extracted automatically from the formant tracks produced by the Burg algorithm embedded in PRAAT 4.0. F_1 has been observed to increase in 'loud' and shouted speech, presumably because of a greater average jaw opening (hence a lower tongue height) during vowels under such conditions (Rostolland, 1982; Bladon, 1986; Junqua, 1996). Consequently, utterance mean F_1 was expected to be highest in the Lombard and lowest in the Fast readings. F_0 range was defined as the distance in Hz between utterance F_0 maxima and minima which were again extracted automatically, from the pitch tracks produced by the autocorrelation routine of PRAAT 4.0. F_0 range may be seen as a rough indicator of (perceptual) pitch range. Note that these are fairly noisy measures and the main motivation to use them was that they involved a minimum of additional hand labelling of speech samples.

All three measures could be said to exhibit register variation because they show different values for the 3 reading tasks, in a fashion that is consistent with the literature on the topic. Broadly speaking, utterance duration, mean F_1, and F_0 range all decrease between the Lombard and Fast conditions. The one exception is represented by the F_0 range for male speakers, which is 5 ms greater in the Normal than in the Lombard condition. However, perhaps the most conspicuous fact about the data summarised in table 7.5 is that the differences between the Lombard and Normal conditions are (considerably) smaller than those between the Normal and Fast conditions: there is a marginal difference (8 ms) in utterance duration between the Lombard and Normal conditions, as opposed to a 65 difference between the Normal and Fast conditions, a 20 (vs. 83) Hz difference in mean F_1, and a 14 Hz (vs. 24) difference in F_0 range for the female speakers.

A series of ANOVAs lends support to the impression that the speech produced in response to the different reading tasks has different global characteristics, with the strongest contrast emerging between the Fast condition on the one hand and the Lombard and Normal readings on the other. A one way ANOVA for *reading task* (Lombard vs. Normal vs. Fast) on the utterance duration data shows that the effect is highly significant, F(1,1264) = 422.66 , $p < .001$, but Tukey and Scheffe post hoc tests indicate that whereas the mean utterance duration for the Fast condition is significantly different from both the Normal and Lombard conditions (all $p < .001$), the difference between the means for the Lombard and Normal conditions is not significant. A one way ANOVA for *reading task* on the utterance mean F_1 data also yields a highly significant effect, F(1,1264) = 358.18, $p < .001$, but in this case Tukey and Scheffe post hoc tests show that all pairwise comparisons yield highly significant results, despite the relatively small difference between the Lombard and Normal conditions.

Table 7.5: Global phonetic effects of reading task: utterance duration (ms), utterance mean F_1, and F_0 range (Hz) for female and male speakers (all combinations of /ps/ and C_3 included. Standard deviations appear in brackets.

Variable	Lombard	N	Normal	N	Fast	N
Duration	3553 (711)	426	3545 (651)	420	2480 (457)	421
Mean F_1	799 (58)	426	779 (57)	420	696 (62)	421
F_0 range (female)	210 (75)	214	196 (72)	211	172 (66)	210
F_0 range (male)	130 (26)	212	135 (32)	209	114 (53)	211

A two way ANOVA for *reading task* * *gender* (female vs. male) yields highly significant effects of *reading task*, F(1,1261) = 27.86, $p < .001$, as well as of *gender* F(1,1261) = 422.59, $p < .001$, and *reading task* * *gender*, F(2,1261) = 4.72 , $p < .01$. The main effect of *reading task* evinces register variation even, in the F_0 range of male and female speakers pooled together. However, Tukey and Scheffe post hoc tests for *reading task* show that as for utterance duration, the means for the Lombard and Fast readings are not significantly different, although both are significantly different from the mean utterance duration in the Fast condition (both $p< .001$). It seems likely that the behaviour of the male speakers, which produced a smaller F_0 range in the Normal than in the Lombard condition contributed to the lack of statistically significant contrast between these two conditions. The same phenomenon (and the reverse patterning of F_0 range in the female subjects' speech) may well underlie the significant interaction between *reading task* and *gender*. Finally, the main effect of *gender* attests to the greater pitch range, across registers, for the female speakers.[7].

Reading task effects on VOT in C_3 plosives The effects of reading task on the VOT of C_3 plosives was examined because the phonetic literature offers some good ground for (cross-linguistic) comparison.

It is a recurring observation in experimental studies on the topic that the VOT of short lag ([p, t, k] and [b̥, d̥, g̊]) plosives is impervious or only slightly sensitive to variations in speaking rate, whereas the VOT of long lag and prevoiced stops is sensitive to such variations (Miller et al., 1986; Pind, 1995; Kessinger & Blumstein, 1997; Magloire & Green, 1999). As visible in figure 7.6 the same applies to the present experiment: the mean VOT of the fortis plosives varies within a 2 ms window whereas there is considerable register-based variation in the prevoicing of the lenis class. The only result that is somewhat surprising is that lenis plosive VOT in the Lombard condition is smaller than in the Normal condition. Note however, that Magloire & Green (1999) report a similar reverse register effect on the VOT of Spanish fortis stops, which some of their test subject was produced with a longer VOT at faster rates. More generally, it appears that the effects on VOT of (elicitation techniques aimed at) slow speech are far less pronounced than those of fast speech (when compared with a 'normal' condition: cf. Miller et al. 1986; Magloire & Green 1999). Thus the results of the present experiment are more or less consistent with earlier studies.[8]

A two way ANOVA for *reading task* * C_3 *laryngeal specification* ([+tense] vs. [-tense]) reveals highly significant effects of *reading task*, $F(2,461) = 13.98$, $p < .001$, C_3 *laryngeal specification*, $F(1,461) = 607.49$, $p < .001$, and *reading task* * C_3 *laryngeal specification*, $F(2,461) = 16.98$, $p < .001$. The effect of C_3 *laryngeal specification* can hardly be surprising given the distinctly different VOT values for fortis and lenis plosives, while the effect of *reading task* indi-

---

[7] Recalculating F_0 on the basis of F_0 minima and maxima expressed in (base 100 Hz) *Semitones* has little effect on this outcome: the mean difference in F_0 range between male and female speakers is 2.77, 1.84, and 2.18 ST under the Lombard, Normal, and Fast conditions respectively.

[8] VOT was calculated as 0 - C_3 closure voicing for lenis plosives with closure voicing > 0 ms, and the interval between the onset of the release burst and voicing onset in ('devoiced') lenis plosives with no closure voicing and fortis plosives (all of which were fully voiceless). Data for the /ks/ + plosive clusters exhibits a VOT patterning that is nearly identical to that illustrated in figure 7.6.

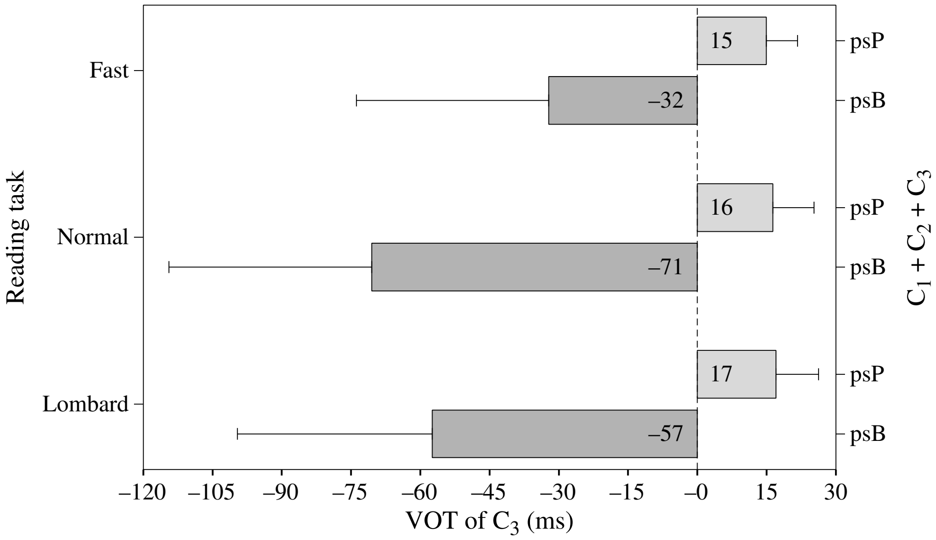

Figure 7.6: Experiment 3: effects of reading task on the VOT of [±tense] plosives preceded by /ps/. Error bars represent the mean ± 1 standard deviation.

cates that there is a statistically significant amount of register variation in the realisation of VOT. However, the interaction between *reading task* and *C_3 laryngeal specification* implies that the main effect of *reading task* may be almost fully due to the variation of the lenis stops. Moreover, Tukey and Scheffe post hoc stops for *reading task* show that the mean Fast VOT is significantly different from those in the Lombard and Normal two conditions, but that there is no significant difference between the latter two. This reinforces the conclusion that the change from Lombard to Normal speech affects the subject's speech (and VOT in particular) far less than the change from Normal to Fast speech.

The interaction between register and RVA In the light of the foregoing it is somewhat surprising that the present experiment reveals little evidence for the idea that register or even plain rate have an observable effect on regressive voicing assimilation.

Figure 7.7 plots $C_1 + C_2$ voicing duration (as an index of RVA) against the combined duration of $C_1 + C_2 + C_3$ closure for clusters composed of /ps/ + /b, d, m, p, t/. sequences ending in /h, V/ were excluded for the sake of simplicity but nothing crucial hinges on this as they behave in a near-identical fashion to the clusters ending in a fortis plosive. $C_1 + C_2 + C_3$ closure duration was chosen as an index of hypoarticulation (assuming that speech rate translates into hypoarticulation) because it reflects the impact of the reading tasks at the same point in the signal where RVA occurs (but note that using any of the global measures discussed above leads to the same conclusions).

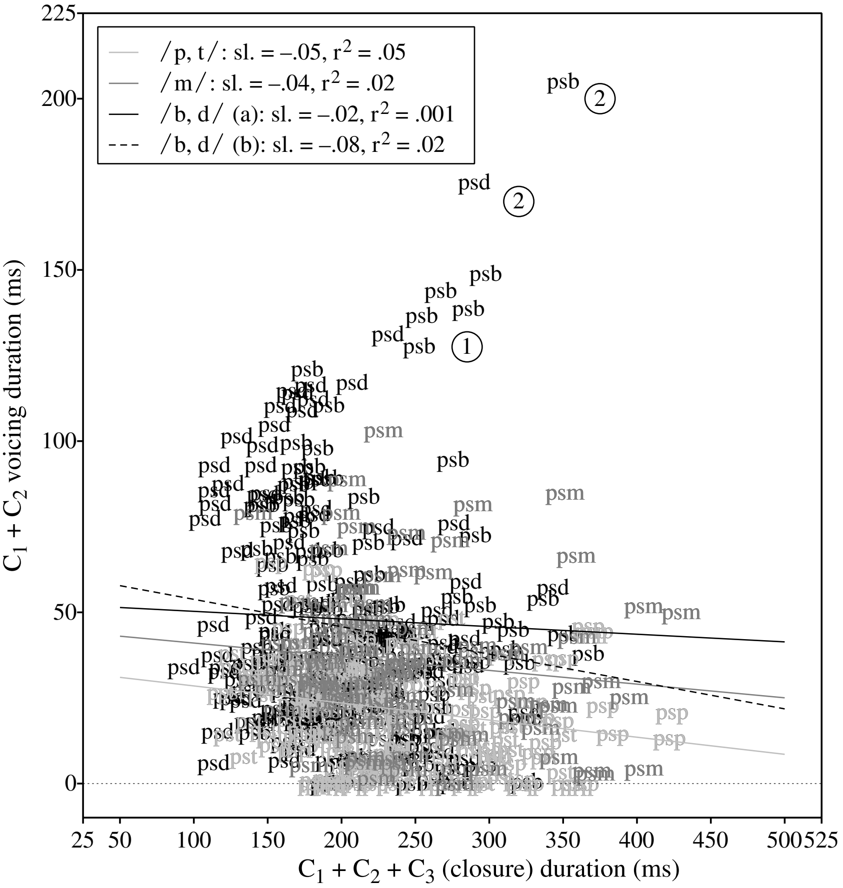

Figure 7.7: Experiment 3: scatter plot of $C_1 + C_2$ voicing duration (ms) against $C_1 + C_2 + C_3$ duration (ms) and results of linear regression analyses. The solid black line (a) represents the regression line for clusters ending in /b, d/ if all cases are included whilst the dashed black line represents the regression line for these clusters if the outliers labelled (2) are excluded.

Given that Dutch RVA is tense-symmetric, the most obvious measure of *degree of assimilation* is the difference in voicing between /ps/ clusters preceding lax stops and those preceding tense stops. An ‘increase’ in assimilation then means a divergence of C_1C_2 voicing values before tense and lax stops, and an ‘increase’ in assimilation as a result of higher speaking rate should appear as a convergence of voicing values towards the left-hand side of the scatter plot in figure 7.7. In the introduction to this section I suggested a hypothesis which cast increased hypoarticulation as a 'leveller' of voicing distinctions in obstruent clusters and hence of distinctions transmitted backwards by means of regressive assimilation. This hypothesis predicts a *convergence* towards the left-hand side of the scatter plot (assuming again that speech rate may be used as an index of hypoarticulation).

However, no clear trends in either direction are visible in figure 7.7. A linear regression analysis indicates that across C_3 contexts, the amount of $C_1 + C_2$ voicing decreases with increasing $C_1 + C_2 + C_3$ closure duration (the slope of the regression line is -.08), and thus increases with degree of hypoarticulation, but although the effect is significant ($p < .001$), the reduction in variance, r_2, is only .03. One plausible source for this increase in voicing in more hypoarticulated speech is increased coarticulation with the preceding vowel (see chapter 4) Separate regression analyses indicate a similar (significant) decrease in $C_1 + C_2$ voicing with increasing $C_1 + C_2 + C_3$ closure duration in clusters ending in [+tense] plosives, slope = -.05, $r_2 = .05$, $p < .005$, but fail to reveal effects for clusters with C_3 = /m/, slope = -.04, $r_2 = .02$, not significant, or with C_3 = /b, d/, slope = -.02, $r_2 = .001$, not significant. Thus, the slight convergence of the regression lines for these classes at lower values of $C_1 + C_2 + C_3$ closure duration, which is illustrated by the solid lines in figure 7.7, does not constitute evidence for the levelling hypothesis.

Moreover, the slope of the regression for clusters ending in /b, d/ is heavily influenced by the two outlier values labelled (2) in figure 7.7, and to a lesser extent by the small cloud of 6 tokens labelled (1). Impressionistic inspections of the recordings of these cases, which were produced by a single speaker (MJ1) reveal nothing remarkable to the ear of the author that would form independent grounds for excluding them from the database. Thus, there seems little ground for excluding the cases labelled (1): their voicing values are equal to or smaller than the /ps/ + lenis plosive mean + 3 standard deviations (149), and are more or less contiguous with the upper $C_1 + C_2$ voicing ranges of the main cloud. But there does seem to be a case for treating the two tokens labelled (2) as genuine outliers: their $C_1 + C_2$ voicing durations represent the /ps/ + /b ,d/ mean (47 ms) + 4.64 and 3.79 standard deviations (34 ms) respectively, and there are large gaps between the category (2) voicing values and the values for neighbouring tokens.

A regression analysis on the clusters ending in a lenis plosive with these two outliers excluded yields a weakly significant effect, $r_2 = .02$, $p = .05$ and a slope that is greater (-.08) than that for /p, t/, which in turn results in a small divergence between the regression lines for [+tense] and [-tense] in more hypoarticulated speech (cf. the dashed black line (b) in figure 7.7), as is predicted

Table 7.6: Experiment 3: effects of reading task on C_1C_2 voicing ratio (standard deviations in brackets).

C_3	C_1C_2 voicing ratio					
	Lombard	N	Normal	N	Fast	N
/p, t/	.12 (.09)	79	.12 (.10)	78	.25 (.15)	77
/b, d/	.36 (.27)	77	.29 (.23)	78	.61 (.33)	78
/m/	.23 (.15)	39	.24 (.20)	39	.39 (.20)	36
/h/	.16 (.10)	39	.17 (.11)	39	.23 (.16)	40
/V/	.19 (.09)	35	.16 (.13)	37	.25 (.15)	38

by optional rules models of regressive voicing assimilation. But note that the effect is a best a weak one: the proportion of variance that is accounted for by the regression lines for /ps/ + /p, t/ and /ps/ + /b, d/ does not exceed .08, even if the outliers labelled (1) are excluded from the analysis.[9]

In fact, the best evidence for an increase in regressive voicing assimilation comes from the marked increase in C_1C_2 voicing *ratio* before lenis plosives in the Fast condition, which is brought about by the concomitant increase in voicing and compression of segmental duration in the Fast condition. As table 7.6 illustrates, C_1C_2 voicing ratio increases across C_3 contexts in the Fast condition, but not to the same extent as before /b/ and /d/. A two-way ANOVA for *reading task* * *C_3 laryngeal specification* on the voicing ratio data (clusters ending in /h, V/ excluded) shows highly significant main effects of *reading task*, F(2,572) = 47.65, $p < .001$, and *C_3 laryngeal specification*, F(2,572) = 87.47, $p < .001$, as well as a significant interaction of *reading task* and *C_3 laryngeal specification*, (4,572) = 4.94, $p< .005$. Whilst the main effects indicate that C_1C_2 voicing ratio is subject to effects of reading task and regressive assimilation, the interaction show that not all conditions contribute evenly to these effects. The sharp increase in voicing ratio before lax stops in the Fast condition is the most likely cause of this interaction. Provided that voicing ratio is perceptually relevant, an increased voicing ratio before /b, d/ may explain why RVA is perceived to increase with speaking rate.

7.3 Conclusions

With regard to the question of whether regressive voicing assimilation applies in Dutch obstruent + fricative + lax stop sequences, the findings of the experiment reported in this chapter seem to me fairly unequivocal. The voiced interval of

---

[9] Excluding the tokens labelled (1) and/or (2) does not affect the conclusions reached in the previous sections in any (statistically) significant way.

word-final /ps/ clusters is affected by the status of a following obstruent even if baseline environments are set aside, and this finding is at odds with traditional accounts of the behaviour of such sequences. From a typological perspective this result bring Dutch RVA in line with the processes found in Hungarian and Yiddish, which have long been described as iterative.

The observation that assimilation does occur in three-term clusters with a medial fricative is conform prediction (25a) but will in itself not rock major theories of laryngeal phonology. However, some of the other results reported above have more far-reaching ramifications. First and foremost, Dutch regressive voicing assimilation appears to be [tense]-symmetric rather than asymmetrically triggered by lax plosives, as held by the standard theory. This finding is consistent with a coarticulation-based view of RVA (25b), which predicts that both the tense and lax obstruents of voicing languages are able to trigger voicing assimilation. Perhaps more importantly, it is entirely consistent with the hypothesis in Ernestus (2000) that final obstruents in Dutch are phonetically underspecified for [tense]. Even if this hypothesis was not directly put to the test (this would involve comparing neutralised with non-neutralised obstruents in phonetically similar contexts) the observation that regressive assimilation to tense stops and [h, ʔ] is responsible for what is heard (by linguists) as final devoicing, suggests that the standard view of final neutralisation as fortition may be flawed.

Experiment 3 supplied further evidence for hypothesis (25c). V_1 duration nor C_1C_2 segmental duration nor F0 microprosody pattern with C_1C_2 voicing, which is the only feature that displays clear traces of regressive voicing assimilation. This finding is entirely consistent with the results of experiment 1. As in Hungarian and English C_1C_2 segmental duration varies under the influence of the following context, but as in the case of Hungarian C_1 fricative duration and C_1 release burst duration it seems that the variations in question can to a large extent be explained in terms of glottal coarticulation between gestures involved in the production of voicing distinctions or [ʔ] and the glottal abduction required for the generation of turbulence noise in the oral tract. The data reported above also indicate that the somewhat puzzling [tense]-asymmetric effect of C_2 on the first formant of a vowel preceding C_1 is not limited to the British English of the subjects of experiment 1. The mechanism responsible for this effect remains unclear however, as does the answer to the question why F_0 microprosody should not also 'spread' backwards.

Finally, this chapter attempted to examine the effects of global register variation on the phonetic manifestation of RVA. It is commonly observed that voicing assimilation phenomena increase with increases in speaking rate, and Menert (1994) offers experimental evidence for this view. However, it is only supported by the results of the present experiment to the extent that there is an increase in the voicing ratio of /ps/ preceding a lax stop in the Fast reading condition. An alternative hypothesis about the relation between speaking rate (as a bias towards a certain degree of hypoarticulation) and the realisation of voicing distinctions in obstruents holds that rate increases act to level underlying distinctions in voicing and hence to diminish the effects of RVA. However, no support was found for this hypothesis.

Chapter 8

Formalist models of laryngeal neutralisation and voicing assimilation[1]

In the previous chapters I have pursued a distinctly functionalist approach to laryngeal neutralisation and voicing assimilation phenomena. A key feature of this approach is that it posits two separate mechanisms that operate in separate modalities to account for neutralisation rules and regressive voicing assimilation. Following Steriade (1997), I argued that neutralisation of fortis-lenis distinctions should be attributed to their perceptibility in context. 3 represents an attempt to extend Steriade's original proposal, which mainly examines the effects of flanking sounds on the perceptibility and neutralisation of laryngeal contrast in obstruents, to positional neutralisation, and neutralisation asymmetries between stops and fricatives. An articulatory mechanism on the other hand forms the core of my analysis of regressive voicing assimilation. As Slis (1985) and Ernestus (2000) I have argued that RVA is rooted in the coarticulation of voicing targets of obstruents. This coarticulation process may feed neutralisation in a diachronic sense, by diminishing the perceptibility of laryngeal contrast in pre-obstruent contexts and biasing the reanalysis of neutralised obstruents, but it remains a distinct mechanism.

Some phonologists, even those who in principle accept the evidence and arguments of the preceding chapters, may maintain that formalist phonological theory still has a role to play in the analysis of laryngeal neutralisation and voicing assimilation phenomena. In 1.4.1 I sketched the position of Hale & Reiss (2000a,b), who see phonology as the study of the set of sound systems that can be encoded by human mental representation rather than as the study

---

[1] An earlier and much condensed version of this chapter can be found in Jansen (2001a).

of the proper subset that speakers and listeners are able to acquire and use in practice. According to this view, formalist models should act as a source of metaconstraints on phonological rules that are (in part) motivated by perceptual, articulatory, or acquisition considerations. It is important to note that such metaconstraints are strictly formal and should therefore hold across rules originating from different modalities, and (ideally) account for recurrent features of those rules that are otherwise inexplicable. A slightly different view of the relation between formalist phonological theory and functionalist accounts of assimilation and neutralisation is suggested by Mascaró & Wetzels (2001) (fn. 20) who speculate that their essentially formalist approach may be complimentary to the functionalist model proposed by Steriade (1997).

The aim of this chapter is not to search for evidence for purely formal metaconstraints on laryngeal neutralisation and voicing assimilation processes: it should be abundantly clear from the preceding chapters that I do not believe that such constraints exist. The aim of this chapter is rather to demonstrate that existing formalist models of laryngeal phonology do not offer any useful predictions regarding the data examined in this study that can act as metaconstraints on a functionalist account or complement it.

This critique focuses on the sets of possible rules that are defined by formalist models rather than on the devices responsible for specifying the environments in which rules apply. The reason for this focus is that contexts for rule application are usually simply stated in terms of syllable structure or higher-order prosodic domains or morpheme boundaries. Many proposals in this vein fail to justify the prosodic structure invoked to account for neutralisation rules on independent grounds (Brockhaus 1995 is an exception) and can therefore hardly be counted as explanations. Moreover, I have already discussed the inadequacies of a syllable-based approach to laryngeal neutralisation in 3.

Models that derive specific and testable predictions about possible rule inventories tend to be implemented in terms of procedural rules or 'hard' (inviolable) declarative constraints. Consequently, little (specific) attention will be paid to optimality-theoretic treatments of laryngeal neutralisation and voicing assimilation. It is true that OT models produce 'factorial typologies' of phonological rules of the sort extensively discussed by for instance Lombardi (1999), but generally speaking they fail to put limits on what constitutes a possible phonological constraint. In the absence of such limits, any logically conceivable statement about the combination of representational primitives (or the mapping between underlying and surface levels) is a possible constraint, and factorial typologies do not generate any real predictions.[2]

---

[2] Note that this is an observation about specific optimality-theoretic accounts of laryngeal neutralisation and voicing assimilation, not a complaint about the overarching framework itself, which can in principle be combined with any theory of phonological or phonetic representations

The bulk of this chapter is reserved for an assessment of what are probably the two most influential types of model forms, both of which can be classified as monovalent lexical feature models. These models can be classified as formalist to the extent that their representational primitives and rule templates are stipulated or derived within the grammar rather than based on external (functional) considerations. Section 8.2 examines *[tense]-based models*, which encode the fortis-lenis distinction identically across voicing and aspirating languages by designating (voiceless) fortis obstruents as universally unmarked. The principal proponent of this approach is Lombardi (1994, 1995a,b). Its main advantage lies in the prediction that (final) laryngeal neutralisation is equally likely in voicing and aspirating languages. However, it fails to predict manner asymmetries in laryngeal neutralisation, the phonetic conditioning of regressive voicing assimilation, and on a more general level, the phonologically 'active' status of unaspirated voiceless fortis obstruents. This section closes with an brief excursus on the OT model proposed in Lombardi (1995c, 1997, 1999), which purports to maintain monovalency whilst solving some of the problems faced by her earlier model, but which is formally equivalent to any traditional account based binary [tense] (or some equivalent).

The second type of model is examined in 8.3. *VOT-based models*, independently proposed by Harris (1994) and Iverson & Salmons (1995, 1999) effectively build lexical representations on the VOTs of fortis and lenis plosives in word-initial position. Whilst this approach yields an adequate account of the assimilatory capacities of actively and passively voiced lenis plosives, it also fails to predict the behaviour of unaspirated fortis obstruents. In addition, under the assumption (made by Iverson & Salmons 1999) that final neutralisation represents fortition, VOT-based models are unable to predict that the process is equally likely in voicing and aspirating languages. As in [tense]-based models, attempts to make the model fit the data tend to undermine its basic premise rather than strengthen its predictions.

Furthermore, the phonetic interpretation of both types of model is beset by problems, in particular because they provide exactly the sort of lexical feature analysis of RVA that is contradicted by the acoustic data reported above. The final section of this chapter investigates how phonetically more fine-grained representations can alleviate these and other problems by improving the fit to the

---

and rules (cf. Boersma 1998; Polgárdi 1998; Flemming 2001). Optimality Theory is a theory of constraint interaction (like Lexical Phonology is a theory of the phonology-morphosyntax interface), and not a theory of phonological representations, which (within the generative paradigm) is an essential prerequisite for a theory of possible phonological rules. It is therefore natural for OT to inspire work on constraint interaction (typologies) rather than on what constitutes a possible constraint. However, as illustrated in a brief discussion of the OT model defended by Lombardi (1999) in section 8.2 this does not dispense with the need for a theory of possible constraints (or a theory of phonetic interpretation).

(phonetic) data under consideration. However, whereas monovalent lexical feature approaches typically undergenerate rules and (phonetic) inventories, refined autosegmental accounts overgenerate, and their output can only be constrained on grammar-external grounds. Thus, rather than solving the problems posed by laryngeal neutralisation and voicing assimilation they merely underline the shortcomings of the formalist enterprise.

8.1 Binary and unary features

As noted at various points in chapters 1-3 it is common practice to represent the contrast between fortis and lenis obstruents in terms of a single phonological feature, and this practice extends to the models examined in the next two sections. Because only a single feature is involved, the predictive power of these models largely derives from the idea that only a single value of [±tense] (or its equivalent) plays a significant role in the phonology and the way this *marked* value is assigned to specific phonetic classes of obstruent. The stipulation that only a single value of [tense] is 'visible' or 'active' in the phonology reduces the number of possible rules that can affect the feature by half (relative to the number allowed by a binary feature) and the same applies to the number of rules that can refer to it as a context. For example, if values of binary [tense] are visible to the phonology, it is possible to derive 4 'grammars' of regressive assimilation: (1) neither tense nor lax obstruents trigger assimilation; (2) [+tense] but not [-tense] obstruents trigger assimilation; (3) [-tense] but not [+tense] obstruents trigger assimilation; (4) both [+tense] and [-tense] obstruents trigger assimilation. If only [-tense] is phonologically active on the other hand, only grammars (1) and (3) can be derived.

Features with a single phonologically active value are usually called *unary*, *privative*, or *monovalent*. Dependency Phonology (Anderson & Ewen 1987) and Government Phonology (Kaye et al., 1985, 1990) are two frameworks that (almost) exclusively employ monovalent features, but they have also been used in feature-geometric models, and in particular models of laryngeal phonology. A universal notational convention of these models is to omit phonologically inert feature values from representations altogether. Although in principle there is nothing wrong with this convention, it sometimes disguises analyses that rely in part or wholly on binary features as strictly monovalent. The next two sections of this chapter identify several instances in the generative literature where this is the case. The remainder of this section spells out what constitutes a true monovalent analysis an what counts as binarity in disguise.

It is not often spelt out explicitly whether the inert value of a feature is invisible for both autosegmental spreading and delinking operations and for the description of the contexts in which such operations apply, or only for the former.

Yet this choice has critical consequences for the restrictiveness of monovalent models. Consider for example the two rules in figure 8.1. The left-hand side of this figure represents a rule that spreads a monovalent feature [A] to a docking site (not containing [A]) preceding it. This sort of rule, which is perhaps the most common of all in autosegmental models, is equivalent to the more traditional 'agreement' rule [-A] → [+A]/__[+A]. Both rules are equally asymmetric in that they propagate a single value of the feature [A] backwards in sequences ending in that same value, whilst they leave sequences ending in other values of [A], e.g., [+A][-A] unaffected. The fundamental claim which, in a sense, makes a monovalent feature model is that the mirror images of these rules, i.e., an autosegmental rule spreading [-A] or [+A] → [-A]/__[-A] are not available. This leads to the hard prediction that languages will only exhibit harmony and assimilation processes involving single feature values such as [+nasal] (nasal harmony) or (as in the example above) [+voice] (assimilation to voiced sounds), but not their opposites, [-nasal] (oral harmony) or [-voice] (assimilation to voiceless sounds).

However, if the description of rule contexts is allowed to make reference to unmarked feature values this prediction is lost and monovalent models become all but indistinguishable from otherwise equivalent binary models, even if spreading and delinking rules are still restricted to single feature values. It is therefore surprising that some authors (e.g., Harris & Lindsey 1995) allow for reference to unmarked feature values in the statement of rule contexts, whilst at the same time expounding the restrictiveness of monovalent models. This illustrated by the rule at the right-hand side in figure 8.1, which delinks the marked value of [A] before a docking site specified as unmarked for A. Note that a rule of precisely this type is proposed by Harris & Lindsey (1995) to represent facts about height harmony in the Pasiego dialect of Spanish. The notational conventions of autosegmental phonology may be deceptive here in obscuring the fact that the context for the delinking rule really is [-A], but the rule at the right-hand side in figure 8.1 is nevertheless fully equivalent to [+A] → [-A]/__[-A], the mirror image or the rule to its left. If reference to unmarked features is allowed in the statement of rule contexts, originally monovalent models are in principle endowed with the same power as binary feature frameworks.

For example, the combination in a single grammar of the two rules in figure 8.1 is equivalent to the inclusion of the symmetric agreement rule [αA] → [-αA]/__[-αA]. So rather than proving that monovalent models can handle phenomena that seem to call for a binary feature framework (Harris & Lindsey 1995), rules of the type represented at the right of figure 8.1 are inconsistent with the main premiss of the approach and relax its constraints on the set of possible phonological rules.[3]

---

[3] Harris & Lindsey (1995) only incorporate the delinking rule in their analysis of Pasiego

$[-\mathrm{A}] \rightarrow [+\mathrm{A}] / \_\_ [+\mathrm{A}]$ $\quad$ $[+\mathrm{A}] \rightarrow [-\mathrm{A}] / \_\_ [-\mathrm{A}]$

Figure 8.1: Binary rules in monovalent models.

There is one way in which a model that allows access to unmarked feature values in the statement of rule contexts can remain less expressive than full-blown binary models. If the set of autosegmental rule templates is restricted to spreading and delinking, neutralisation triggered by prosodic structure or a context defined in extra-phonological (e.g., morphosyntactic) terms can only produce unmarked feature values. In other words, a model of this kind is able to encode [A] → [-A]/__# but not [-A] → [A]/__#. For example, if [+voice] (as a feature representing lenis obstruents) is marked as opposed to unmarked [-voice] (fortis) the former can be delinked to derive the latter, but if insertion of [+voice] is not allowed the reverse mapping is impossible. It follows that word-final [voice] neutralisation is always fortition (and devoicing). By contrast, fully binary feature frameworks allow both mappings and therefore predict that final ‘lenition’ and voicing are equally possible as final fortition. This distinction may seem too fine to be ever important, but it is clearly relevant to the performance of the model pursued by Iverson & Salmons (1995, 1999) (cf. 8.3).

8.2 [tense]-based models

In the early work of Lombardi (1994, 1995a,b) the representation of [tense] oppositions is conceived as part of a comprehensive approach to laryngeal contrast in obstruents and sonorants.[4] Lombardi defines 3 privative features, i.e., [voice], [asp(iration)], and [gl(ottalised)], which are suspended from a LAR(yngeal) articulator node in a feature-geometric representation. Phonetically, [asp] is interpreted as aspiration noise, whilst [gl] represents the increased glottal constriction found in for example ejectives ([gl]) and implosives ([gl, voice]). On the interpretation of the [voice] feature, see below. Provided that [asp] and [gl] cannot combine under the same LAR node (in the same way that vocalic [+high] and

---

Spanish, but they do not offer any principled grounds on which the coocurrence of the rules in figure 8.1 in a single grammar can be ruled out.

[4] Lombardi (1994, 1995a,b) are all based on the author’s PhD dissertation, Lombardi (1991).

[+low] are traditionally said to be incompatible for ‘articulatory’ reasons) the maximal number of laryngeal distinctions (per language) that can be encoded by this feature inventory is 6, which corresponds to the attested maximal number of laryngeal contrasts in obstruent inventories (Ladefoged & Maddieson, 1996). This maximally 6 term contrast includes the fully inert (i.e., [-voice, -asp, -gl]) category, which is represented as lacking a LAR node altogether.

It is this last structure that is consistently used by Lombardi and others following a similar approach to represent fortis obstruents (defined as in the introduction to 2) regardless of their phonetic interpretation as plain voiceless or voiceless unaspirated. As shown in figure 8.2, lenis obstruents possess a LAR node with a [voice] feature suspended from it, again irrespective of their VOT/voicing. No distinction is drawn between the representation of laryngeal contrast in plosives and fricatives. Consequently, the structures in figure 8.2 classify the obstruents of aspirating and voicing languages in exactly the same way as the descriptive feature [tense] as I have used it in this study.

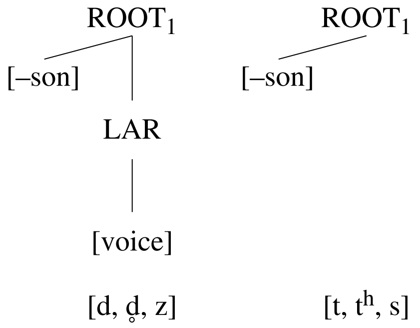

Figure 8.2: Representation of laryngeal contrast in [tense]-based models

Although Lombardi has in some ways erected the most elaborate [tense]-based framework available, the philosophy, if not the technical devices, of her approach are shared by a number of other accounts of the laryngeal phonology and phonetics of languages with a fortis-lenis opposition in their obstruent systems. Among the more notable of these studies are Mester & Îto (1989), Cho (1990a/1999,b), and to a slightly lesser degree by Mascaró (1987/1995). Broadly speaking, the merit of these and other [tense] based models lies in their ability to account for properties and behaviour shared by fortis obstruents and by lenis obstruents across the voicing-aspirating distinction, including some aspects of laryngeal neutralisation and the realisation of fricatives. However, because they are built on [tense] rather than voicing distinctions, their weaknesses reside in their inability to relate the occurrence of voicing assimilation to the voicing (targets) of the triggers. To the extent that they subscribe to the Laryngeal Constraint, they also fail to offer an account of the Germanic past tense paradigm and progressive devoicing.

8.2.1 Phonetic interpretation of fortis and lenis obstruents

Since the models discussed in this section capture [tense] rather than phonetic voicing distinctions, they offer a transparent representation of the phonetic properties that pattern with [±tense] in both aspirating and voicing languages. For instance, the unmarked structure in figure 8.2 maps consistently into raised F_0, raised F_1, shorter duration of preceding vowels, relatively long and high amplitude release bursts, and other cues that correlate with [+tense] across languages (cf. 2). Similarly, the representational format in figure 8.2 captures the fact that both in voicing and aspirating languages the fortis-lenis distinction in fricatives is realised as voiceless vs. voiced. To the extent that voicing and aspirating languages employ voicing (distinctions) in similar fashions in word or morpheme-final environments, [tense]-based models again offer a transparent account of phonetic interpretation.

On the other hand, as already hinted above, the interpretation of [tense] in terms of VOT and voicing does not follow naturally from the structures in figure 8.2. The interpretations of [voice] and the unmarked class have to be stipulated separately for aspirating and voicing languages as zero to short lag vs. long lag, and prevoiced vs. zero to short lag respectively. An additional difficulty in the description of aspirating languages is that the mapping between phonological structure and phonetic voicing is not consistent across manners of articulation, since [voice] has to be interpreted as zero to short lag VOT for (word-initial) lenis plosives but as voicing for lenis fricatives, whilst the unmarked structure represents a long lag VOT for [+tense] plosives but plain voicelessness for fricatives. It is true that Kingston & Diehl (1994, 1995) provide evidence that zero to short lag and prevoiced plosives are *perceptually* similar, but this evidence falls short of rescuing [tense]-based models, as there seems little doubt that they are systematically different in the *production* of speech by monolingual speakers of voicing and aspirating languages. More importantly, bilingual speakers who have both a voicing and an aspirating language also seem to be able to distinguish actively and passively voiced lenis plosives (e.g., Magloire & Green 1999).

A problem of a somewhat different order is the voiced realisation of some neutralised obstruents, e.g., those flanked by sonorants in Catalan and Old English (fricatives), and worse, the neutral but nevertheless voiced initial fricatives of (standard) German and a number of eastern dialects of Dutch (cf. German <See>, [zeː]). If such obstruents are represented by the unmarked laryngeal structure in figure 8.2, its phonetic interpretation is rendered ambivalent between voiced and voiceless, that is, unless some sort of subsequent sonorant-to-obstruent assimilation rule is invoked to account for the voicing (cf. Lombardi 1995b).

8.2.2 Modelling regressive voicing assimilation in [tense]-based models

Whether or not a lack of phonetic transparency in phonological representation counts in itself as a drawback for a model of course depends on one's beliefs about the nature of the relation between phonology and phonetics. According to the extreme formalist position briefly referred to in 1, it matters little if a model of phonetic interpretation for Swedish should have to stipulate that [voice] corresponds to short lag VOT in lenis plosives, whilst a model for Dutch interprets it as prevoicing. This position is ultimately based on the belief that shape of phonological rules stands in an arbitrary relation to the phonetic substance of the features they operate on. But as shown in the previous chapters, regressive voicing assimilation under word sandhi is clearly conditioned by the phonetics of the trigger obstruents, and therefore the fact that lenis obstruents are invariably marked [voice] constitutes a problem for Lombardi's model. An additional and equally serious problem is that under the most strict interpretation of monovalency, [tense]-based models predict that only lenis obstruents trigger RVA.

Following slightly earlier proposals by Mester & Îto (1989) and Cho (1990a/1999), Lombardi (1994, 1995a,b) analyses regressive voicing assimilation to [-tense] obstruents in terms of two mechanisms, the first of which postulates an intrinsic link between assimilation and final laryngeal neutralisation. RVA to [+tense] sounds does not really exist in this account, at least not in the intuitive sense conveyed by older models employing agreement rules for [αvoice] or autosegmental feature ([±voice]) spreading. Instead it is captured by delinking the LAR node of a lenis obstruent when another obstruent follows: if the following obstruent is a fortis one, the result is a sequence of unmarked sounds (which are interpreted as phonetically voiceless; cf. the left panel of figure 8.3). Thus, the model conceives of assimilation to fortis obstruents as neutralisation, and without having to refer to the inert feature value [-voice], at least technically speaking. It is possible to refer to [(+)voice], and therefore RVA to [-tense] obstruents is modelled in the more intuitive way as the backward spreading of a LAR node dominating a [voice] feature (figure 8.3: right panel). In a cluster consisting of two lenis obstruents both operations apply: delinking first derives an unmarked form before [voice] spreading restores the original marked value.[5]

---

[5] The account proposed by Mascaró (1987/1995), although similar in many other respects, differs from the approach discussed in this paragraph, in allowing spreading of [-voice] as well as [+voice].

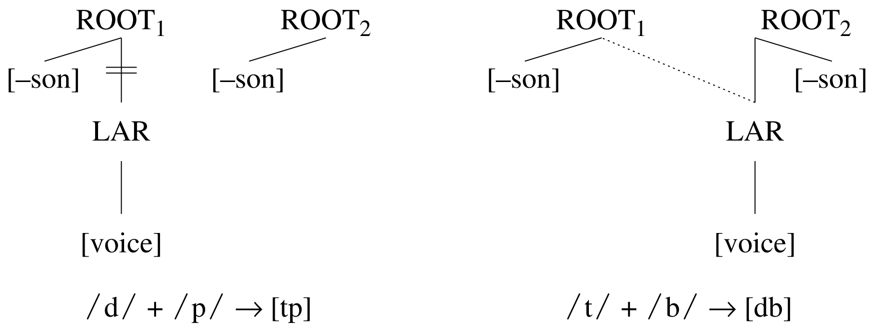

Figure 8.3: Regressive voicing assimilation according to Lombardi (1994, 1995a). Left-hand panel: ‘assimilation’ to fortis obstruents by delinking; right: RVA to lenis obstruents by (delinking and) spreading.

Languages with across-the-board final neutralisation such as Dutch and German are represented in this account with the Laryngeal Constraint switched on, which results in delinking of all word-final LAR nodes, and hence derives the first step in the assimilation process independently for all obstruent clusters straddling a word boundary. This leaves a two way taxonomy of regressive voicing assimilation in these languages: if [voice] spreading is switched on there is RVA of the trivially [tense]-symmetric form that is usually described for Dutch and Polish (although the results of experiment 3 indicate that this description is inaccurate); if it is switched off, assimilation does not occur in any intuitive sense, although clusters ending in a fortis obstruent of course end up as homogeneous sequences of unmarked obstruents.

Lombardi (1994, 1995a), Cho (1990a/1999) as well as Siptár & Törkenczy (2000) agree in the idea that in languages such as Yiddish and Hungarian, which have [tense]-symmetric RVA but not across-the-board final neutralisation, the same two operations illustrated in figure 8.3, delinking and [voice] (i.e., LAR) spreading, are responsible for RVA, although they differ on the technical point of how the delinking of [voice] from lenis plosives and fricatives is restricted to (certain) preobstruent contexts. This means that realisations such as [aːctoːl] in (27c) are derived by delinking [voice] from the palatal plosive. Note that delinking and subsequent [voice] spreading apply to the initial obstruent in [-tense][-tense] sequences too if the delinking mechanism is not allowed to refer to the unmarked status of fortis obstruents such as the alveolar plosive in /aːɟ/+ /toːl/.

Because preobstruent delinking and [voice] spread constitute two independent parameters, which are switched on or off with independent possibilities and/or have independent UG default settings, a four way typology of regressive voicing assimilation is derived for the languages in question. RVA is predicted to be either [tense]-symmetric (preobstruent delinking and [voice] spreading are both active), or asymmetrically triggered by [-tense] (only [voice] spreading is active), or not to occur at all (both mechanisms are switched off). In the fourth type of language where only preobstruent delinking is active, a single series of voiceless obstruents emerges before both fortis and lenis plosives and fricatives. In other words, underlying /d/ + /p/ is predicted to surface as [tp] (or [tp^{h}] in an aspirating language) whilst /d/ + /b/ as well as /t + b/ should be realised as [tb] ([tb̥]) rather than [db] ([db̥]). Mascaró & Wetzels (2001) point out that languages exhibiting this fourth pattern do not seem to have been documented, although I argued in 2 that there may be several independent explanations for this (apparent) typological gap. Of course the latter type of language (as well as the voice-only spreading type) may be ruled out if the model somehow states that delinking and spreading always apply in tandem, but since there is no way of deriving this state of affairs this could hardly be said to be an improvement.

(27) [tense]-symmetric RVA in Hungarian and Yiddish (cf. 6.1 and Katz (1987))

a. Hungarian [+tense][-tense] clusters

/kuːt/+ /bɔn/	[kuːdbɔn]	in (a) well
/fuːc/+ /bɔn/	[fuːɟbɛn]	in (a) whistle

b. Yiddish [+tense][-tense] clusters

/bak/ + /bejn/	[bagbejn]	cheekbone
/bux/ + /gəʃɛft/	[buɣgəʃɛft]	bookstore

c. Hungarian [-tense][+tense] clusters

/rɔb/+ /toːl/	[rɔptoːl]	from (a) pri- soner
/aːɟ/+ /toːl/	[aːctoːl]	from (a) bed

d. Yiddish [-tense][+tense] clusters

/klug/ + /kind/	[klukːind/	clever child
/briv/ + /trɛgər/	[briftrɛgər]	mailman

Meanwhile, the four-way typology for languages without across-the-board laryngeal neutralisation excludes a pattern that has been attested. Asymmetric RVA to [+tense] obstruents is predicted not to occur, because preobstruent delinking is insensitive to the laryngeal class of the trigger. Yet this is the pattern displayed to some degree by the English plosives examined in chapter 5, and as Mascaró & Wetzels (2001) note (perhaps to a greater extent) by Yorkshire English, and Yaːthê.

The latter issue is part of the much more serious difficulty identified at the outset of this section (but completely overlooked by Mascaró and Wetzels), namely that [tense]-based models are not able to capture the relation between the voicing of obstruents and their capacity for triggering regressive voicing assimilation. This problem extends to all [tense]-based models, irrespective of the valency of the [voice] feature, and therefore including Mascaró (1987/1995) and the OT analyses of Grijzenhout & Krämer (1998) and Lombardi (1999). For example, in Lombardi's and Cho's models, the lenis plosives of voicing as well as aspirating languages have a LAR node carrying a [voice] available for spreading, but there is no way of predicting that [voice] spreading should be activated for all and only the first group of languages. Instead the value of the [voice] parameter has to be stipulated in the model for every individual voicing language ('on' for Dutch, Yiddish, Hungarian) and every single aspirating language ('off' for English, German, Swedish). Similarly, [tense]-based models are unable to predict voicing-based manner asymmetries of the type established for English by experiment 1. Again, the fact that English /z/ but not /d/ triggers RVA can only be stipulated in the model, because both segments are specified as in figure 8.2.

Moreover, the observation that fortis obstruents have to capacity to trigger RVA as well as lenis obstruents is incongruent with the basic hypothesis of monovalent models that only marked feature values should be phonologically active. In the case at hand this entails that only lenis ([+voice]) obstruents should trigger regressive voicing assimilation. Although the use of pre-obstruent [voice] delinking to represent RVA to fortis obstruents avoids reference to inert [-voice] (as at the right-hand side in figure 8.1) by capitalising on the fact that [voice] is only marked on obstruents, it still equips the grammar of Hungarian, Yiddish and similar languages with the formal equivalent of $[\alpha\text{voice}] \rightarrow [-\alpha\text{voice}]/\_\_[-\alpha\text{voice}]$, as of course it has to. Given the asymmetry hypothesis underlying monovalent feature models this is a patch to cover an inaccurate prediction rather than a real solution.

In sum, it seems that quite apart from their ability to represent the phonetic manifestation of regressive voicing assimilation, [tense]-based models fail to predict the phonetic conditioning of RVA and (as a result) fail to predict too, that the process can pattern asymmetrically with respect to [tense] and manner of articulation where there are mismatches between [tense] and phonetic voicing. Since the models reviewed in this section invariably employ a fortition analysis of final neutralisation as well as a lexical feature account of RVA they are also unable to represent the symmetric nature of the latter process in Dutch that was established by experiment 3. Representation of the voicing of Dutch final obstruents before fortis obstruents (and /h/, [ʔ]), sonorants and lenis plosives requires a 3 point scale rather than the two way contrast available in lexical feature accounts. However, assumptions about the structure preserving (‘lexical’) nature of final neutralisation and RVA are deep-seated, even in some of the experimental literature, and it is therefore unsurprising that they have gone unquestioned in accounts that pay little heed to phonetic data more generally speaking.[6]

8.2.3 Laryngeal neutralisation

One of the stronger points of [tense]-based frameworks, is their modelling of laryngeal neutralisation. If the fortis-lenis distinction is encoded without regard for its phonetic interpretation, as in figure 8.2, it follows that laryngeal neutralisation (i.e., delinking) also applies without regard to phonetic interpretation. LAR delinking is a single parameter that has the same probability of being switched on or off and/or the same UG default settings, irrespective of the phonetic type of the contrast involved. In 3 I tried to demonstrate that there is little evidence for the claim that the occurrence of laryngeal neutralisation depends on the phonetic interpretation of [tense]: it occurs in aspirating as well as in voicing languages, both within (German, Dutch, Frisian) and outside (e.g., in various Slavonic and Turkic languages) the Germanic group. The predictions of a tense-based account are therefore in accordance with the data.

However, as noted before, neither Lombardi’s work, nor Cho (1990a/1999,b), nor Mascaró (1987/1995), draw any distinction between plosives and fricatives in the representation of laryngeal contrast. This implies

---

[6] This section glosses over a number of additional problems with the model developed in Lombardi (1994, 1995a,b). Several of these problems involve unorthodox claims about assimilation, neutralisation, and syllabification that are not always defended in a satisfactory manner. One issue that deserves some attention here are the hints throughout Lombardi’s work that regressive assimilation in languages such as Yiddish and Hungarian is somehow a *lexical* process and/or restricted to word internal contexts (cf. Mascaró & Wetzels (2001)). For example, the constraint that Lombardi (1994, 1995a) employs to block across-the-board final neutralisation in the languages in question (*Final Exceptionality*), takes scope over all word-final obstruents and therefore also blocks RVA at word boundaries, at least assimilation to [+tense] obstruents. Furthermore, whilst Lombardi (1996) retracts the stronger position taken up by the earlier work by stating that [-voice] may be phonologically active *postlexically*, [tense]-symmetric RVA is not used as an argument for this revised model. Thirdly, whilst claiming that the analysis does not extend to the postlexical level, Lombardi (1999) discusses Yiddish RVA data taken from Katz (1987), whose description of the phenomenon indicates that it applies across as well as within (morphologically complex) words.

As Mascaró & Wetzels (2001) observe, Lombardi never offers any argument for the (post)lexical or word internal status of RVA in Yiddish, Hungarian, or any similar language, although the relevant diagnostics (e.g., the presence of lexical exceptions, cyclic effects) are fairly straightforward. Furthermore, there may be hints in the descriptive literature that regressive voicing assimilation is ‘optional’ across word boundaries or otherwise less pervasive than word internally, but I am not aware of any claims that the process is restricted to word internal contexts. Moreover, the data from experiment 2 unambiguously supports the claims in, e.g., Kenesei et al. (1998) and Siptár & Törkenczy (2000) that Hungarian RVA applies between words (as well as word internally).

that laryngeal neutralisation affects both classes of obstruent in similar degrees. This may be true for word-final neutralisation in, e.g., the Germanic and Slavonic languages, but in 3 I showed that it does not apply to (lexical) neutralisation in other contexts, nor does it hold universally for final neutralisation phenomena. For example, in the Germanic languages laryngeal contrast is generally less stable in (sibilant) fricatives than in plosives, even if the contrast is not suspended across the board.

8.2.4 Postobstruent devoicing and the Germanic past tense paradigm

I argued in 4 that progressive devoicing of lenis obstruents across word boundaries and the progressive assimilation of the past tense suffix in a number of Germanic languages are two distinct processes. Nevertheless, I will discuss these phenomena together here, because they are treated on a par by most formalist models, witness papers such as Lombardi (1997) on the 'special' status of progressive voicing assimilation.

One pertinent observation about the 'Germanic' past tense paradigm (data from 4 repeated here as examples 28 and 29) is that its occurrence is not conditioned by the voicing-aspirating distinction: the regular past tense suffix patterns in essentially the same way in aspirating languages such as English and Swedish, and voicing languages like Dutch. Perhaps the most convincing illustration of the lack of a correlation between the use of VOT in initial plosives and the behaviour of the regular past tense morpheme is that there is no dialectal variation in past tense marking coinciding with variation in VOT use: for example, Scottish English, a voicing dialect, has the same past tense allomorphy as the aspirating dialects of English.

(28) Dutch past tense paradigm

/zɑk/ + /də/	[zɑktə]	came down, descended
/ʋɑs/ + /də/	[ʋɑstə]	washed
/dœb/ + /də/	[dʉbdə]	doubted, wavered
/xraːz/ + /də/	[xraːzdə]	grazed
/krœl/ + /də/	[krʉldə]	curled

(29) English past tense paradigm

/lɒp/ + /d/	[lɒpt]	lopped
/lɒk/ + /d/	[lɒkt]	locked
/lɒb/ + /d/	[lɒbd]	lobbed
/lɒg/ + /d/	[lɒgd]	logged
/loːn/ + /d/	[loʊnd]	loaned

In the light of the discussion of laryngeal neutralisation above, the insensitivity of regular past tense allomorphy to the use of VOT categories might seem to constitute an argument for [tense]-based privative representations of the fortis-lenis contrast. After all, such models are able to capture the class of obstruents that cause the (initial) obstruent of the past tense suffix to devoice in a crosslinguistically uniform way: both the /k/of Dutch /zɑk/ (cf. /zak/ + /ən/, *to come down, descend*) and the /k/ of English /lɒk/ belong to the unmarked category in figure 8.2.

However, the fact that they are unmarked also mean that they have no feature available that can spread onto the underlying /d/s of the English and Dutch past tense suffixes, or otherwise force feature agreement. The laryngeal feature configuration of the plosive clusters in Dutch /zɑk/ + /də/ and English /lɒk/ + /d/ is illustrated in figure 8.4. This is of course the same configuration that emerges where a word-initial lenis obstruent is preceded by a fortis one (cf. figure 8.3), as (at least under a fortition analysis of final neutralisation) in Dutch /riːt/ + /zɑŋər/, *sedge warbler (acrocephalus schoenobaenus)*, which is typically realised with a fully voiceless alveolar fricative, [ritz̥ɑŋəɹ] (or [ritsɑŋəɹ] according to conventional transcriptions). Recall from 4.3 that Dutch postobstruent fricative devoicing is generally considered to be within the scope of generative models, and therefore presents the same problem to [tense]-based privative frameworks as the Germanic past tense rule.

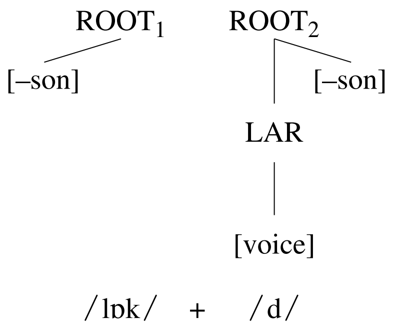

Figure 8.4: Laryngeal feature configuration for fortis + lenis obstruent clusters in privative tense-based models

Because there is no LAR node that can spread from ROOT$_1$ to ROOT$_2$ in figure 8.4, such frameworks have to resort to devices that somehow removes the license for the LAR node or the [voice] feature of ROOT$_2$ to account for the behaviour of the past tense rule and postobstruent devoicing. However, given that the processes in question occur outside ‘typical’ (i.e., constituent-final, preobstruent) laryngeal neutralisation contexts, such devices are difficult to formulate in a principled or general fashion within purely phonological (rather than morphology or phonetics driven) models. This difficulty is illustrated well by the various wellformedness principles that Lombardi has employed or proposed to handle instances of delinking outside the scope of the Laryngeal Constraint, two of which appear in example (30).[7]

The first principle harks back to typological observations, primarily about morphologically simplex words, made by Harms (1973) and Greenberg (1978) but is implemented in various ways as a formal constraint by Mester & Îto (1989), by Cho (1990a/1999) and by Lombardi (1994, 1995a, 1997, 1999), to deal with the English regular past tense and regular plural rules. The constraint in (30a) forces the final [voice] feature (and presumably its dominating LAR node) in sequences such as /lɒk/ + /d/ to delink, at least under the (questionable) assumption that they are parsed into a single syllable. As a formal device, *Harms' generalisation* raises two serious objections, both of which hold within the methodological confines of a generative approach as well as on more general grounds. First and foremost, the filter in (30a) is a statement of fact rather than anything else. The way it is formulated implies that the reverse (voiceless obstruents must be closer than voiced to the syllable nucleus) is equally possible (and likely) as a formal grammatical constraint, whether it is attested by typological evidence or not.

The second objection is that using (30a) to derive the allomorphy of the English regular past tense and plural morphemes fails to draw any connection with the similar patterning of the regular past tense in Dutch or Swedish, or more generally with other instances in which affixes adapt to the phonological properties of their hosts, as for instance in the formation of the Dutch diminutive. Harms' generalisation is insufficient to account for the behaviour the regular past tense morpheme of Dutch or Swedish because its underlying /d/ is parsed into a separate syllable (or onset rhyme sequence) with the following vowel under every (re)syllabification algorithm available in the literature. As a result very similar (looking) patterns of allomorphy have to be accounted for in terms of completely disjunct formal mechanisms, which implies that they somehow possess different properties. However, evidence to this effect is not supplied by any of the accounts cited at the beginning of this paragraph.[8]

---

[7] The wellformedness condition in (30b) is slightly rephrased from the version in Lombardi (1994), but is formally equivalent.

[8] The idea that /lɒk/ + /d/ is parsed into a single syllable is questionable on the grounds that it requires a word-based definition of the English syllable that fails to capture root-level phonotactic constraints (e.g., Harris 1994). Note furthermore that Borowsky (2000) presents a morphology-driven analysis of the Dutch past tense and English regular plural rules that draws on *Correspondence Theory* (McCarthy & Prince, 1995; Benua, 1995). This account nominally maintains privative [voice] but as the OT model of Lombardi (1995c, 1997, 1999) effectively allows reference to both [+voice] and [-voice], and therefore merely highlights the problematic nature of the Germanic past tense paradigm for privative models. See further below.

(30) Constraints on laryngeal structure in postobstruent contexts

a. *Harms' Generalisation* (see Harms 1973; Mester & Îto 1989; Cho 1990a/1999; Lombardi 1994, 1999): Voiced obstruents must be closer than voiceless to the syllable nucleus

b. *Postobstruent Fricative Devoicing* (cf. Lombardi 1994, 1997): [voice] cannot be licensed in the following configuration:

$$[\text{-son}]\begin{bmatrix}\text{-son}\\ \text{+cont}\\ \_\_\end{bmatrix}$$

Lombardi (1997, 1999) asserts that progressive assimilation to [-voice] is relatively rare and only occurs under special circumstances. The special circumstances notwithstanding, this constitutes an admission that the laryngeal specification of fortis obstruents can be phonologically active, and thereby undermines a basic premiss of Lombardi's framework. The prose definition of Harms' Generalisation may suggest otherwise, but it is a mechanism along the lines of the delinking rule in figure 8.1: it delinks marked [voice] after an obstruent that bears the unmarked value for the same feature. Consequently, the constraint is equivalent to the traditional agreement rule [+voice] → [-voice]/[-voice]__, or an autosegmental rule spreading [-voice], which (given that [+voice] spreading is also allowed in [tense]-based frameworks) is precisely the sort of rule monovalent feature models purport to be ruling out.

A final problem within the specific context of Lombardi's model that affects both the rules in (30a) and (30b), is that they refer to [voice] rather than to the dominating LAR node, as the other rules and filters in the model. Despite the suggestion in Lombardi (1995b) that all rules of laryngeal phonology refer to LAR instead of to the specific features it dominates, this is tantamount to the admission that there are [voice]-specific rules, and thereby underlines the failure of privative [tense]-based accounts to yield any solid predictions about the behaviour of the Germanic past tense paradigm and progressive devoicing.

8.2.5 Excursus: Lombardi's (1999) OT model and privativity in Optimality Theory

Up to this point I have described Lombardi's work as consistently adopting a [tense]-based and privative approach to laryngeal contrast and voicing assimilation. However, as far as the latter phenomenon is concerned, the OT implementations of Lombardi (1995c, 1997, 1999) represent a break with earlier work, in that voicing assimilation is effectively captured in terms of a binary [voice] feature. The same applies to the analyses of the Dutch past tense and English regular plural rules in Borowsky (2000), which are also nominally based on monovalent [voice]. I will briefly discuss the relevant properties of Lombardi (1999), the latest published version of the OT-based model, because it emphasises the problematic nature of the mechanisms employed to capture [tense]-symmetric RVA, progressive devoicing, and the Germanic past tense paradigm in monovalent models. In my opinion they also underline the shift of attention in much optimality-driven work away from constraining rule inventories and towards the modelling of rule interaction.

(31) Constraints relating to (symmetric) RVA in Lombardi (1999)

a. *IdentOnset(Laryngeal)* (Lombardi 1999:270): Consonants in the position stated in the Laryngeal Constraint (cf. figure 3.1 in section 3.1) should be faithful to underlying laryngeal specification
b. *Ident(Laryngeal)* (Lombardi 1999:270): Consonants should be faithful to underlying laryngeal specification
c. *AGREE* (Lombardi 1999:272): Obstruent clusters should agree in voicing

Lombardi (1999) employs the wellformedness conditions in (31) to model regressive voicing assimilation. The first (abbreviated *IDOnsLar*) is a *positional faithfulness* constraint of the type first proposed by J. Beckman (1997). Both this constraint and the second (abbreviated *IDLar*) belong to the *Identity(Feature)* family (McCarthy & Prince, 1995), but the former is relativised to apply only in a certain context. Consequently, IDOnsLar is violated by any output obstruent that (a) precedes a tautosyllabic sonorant and (b) has a laryngeal specification that is in any way different from its correspondent in the input. Thus, given an underlying form /bæd/ this constraint is violated by surface /pæd/ but not by [bæt]. IDLar on the other hand is violated by any obstruent in the output that meets condition (b), which means it penalises both /pæd/ and [bæt] as surface forms of /bæd/. Technically speaking, AGREE is also an Identity constraint, but one demanding identical laryngeal structure in adjacent segments in the same output string, much like the B(ase)-R(eduplicant) faithfulness of McCarthy & Prince (1995), instead of between sounds in different output strings or between correspondent segments in input and output strings.

The tableau in (32) demonstrates how AGREE and IDOns interact to select the optimal candidates for underlying forms with laryngeally heterogeneous obstruent clusters in languages with [tense]-symmetric RVA, such as Yiddish and Hungarian. The examples here (/fyːc/ + /bɔn/, *in (a) whistle*; /aːɟ/ + /toːl/, *from a bed*) are from Hungarian, but the top and bottom parts of table (32) correspond directly to constraint tableaux (22) and (23) in Lombardi (1999). The interaction between AGREE and IDOnsLar derives [tense]-symmetric regressive voicing assimilation in a straightforward way. AGREE filters out any surface forms with a heterogeneous cluster, including the perfectly faithful ones in (b) and (g). IDOnsLar then determines the choice between the candidates showing progressive voicing assimilation (a and h) and those with regressive assimilation (d and e), all of which incur a single violation for IDLar, in favour of the latter two forms: (a) and (h) have an obstruent with a different laryngeal specification in the context defined by the Laryngeal Constraint and therefore violate IDOnsLar as well as IDLar. Thus, the OT account of Lombardi (1999) captures voicing assimilation to both lenis *and* fortis obstruents in terms of the same mechanism: the AGREE constraint, with IDOnsLar determining the direction of assimilation.

(32) Constraint tableau for [tense]-symmetric RVA according to Lombardi (1999). Examples from (27a) and (27c) above

	/fyːc/ + /bɔn/	AGREE	IDOnsLar	IDLar
(a)	[fyːcpɛn]		*!	*
(b)	[fyːcbɛn]	*!		
(c)	[fyːɟpɛn]	*!	*	**
(d)	[fyːɟbɛn] √			*
	/aːɟ/ + /toːl/			
(e)	[aːctoːl] √			*
(f)	[aːcdoːl]	*!	*	**
(g)	[aːɟtoːl]	*!		
(h)	[aːɟdoːl]		*!	*

This uncomplicated account of [tense]-symmetric RVA seems surprising in the light of the discussion of this issue above, which attempted to show that the phenomenon is something of an embarrassment for privative [tense]-based models, which are forced to capture RVA to fortis and lenis in terms of two separate mechanisms, i.e., [voice] delinking and [voice] spreading. However, any surprise disappears on closer inspection of the correspondence constraints in (31), which reveals that they equip the model with the full power of binary rather than privative [voice]. First, note that the representations for all the output forms in (32) can be derived from their respective underlying representations by operations on privative [voice]. For example, [fyːɟpɛn] can be derived from /fyːc/ + /bɔn/ by delinking [voice] from the underlying /b/ and inserting a LAR node carrying voice to the preceding /c/. In other words, it would be correct to say that privativity is maintained in GEN.

But this is not true of H-EVAL, because given the constraint definitions in (31), the algorithm scans and uses the laryngeal specifications of lenis (marked, [+voice]) and fortis (unmarked, [-voice]) in exactly the same way. Take for example the surface forms [fyːcbɛn] (b) and [aːɟtoːl] (g). To establish the performance of the first with respect to AGREE, H-EVAL has to check the laryngeal specification of the labial (marked, i.e., [+voice]) and palatal plosives (unmarked, i.e., [-voice]). This information leads to a violation of AGREE, which assigns the unassimilated form (b) suboptimal status. In many ways, this procedure resembles the application of the [voice] spreading rule discussed above to the same input, i.e., /fyːc/ + /bɔn/. The latter also has to check the laryngeal specification of the palatal, which establishes that it is a legitimate target for spreading, and the specification of the labial, which triggers the decision to spread [voice].

Crucially, the OT model processes the performance of the second form, [aːɟtoːl], in much the same way as that of the first. Here the information that [t] is [-voice] (unmarked) whilst [ɟ] is [+voice] (marked) again leads to a fatal violation of AGREE. It is this violation that critically distinguishes it from the optimal form with regressive assimilation (d). But in this instance, the OT model and the earlier framework developed by Lombardi (1994, 1995a,b) behave differently. Whereas in the former the unmarked ([-voice]) specification of the alveolar plosive 'triggers' RVA by rendering the candidate without assimilation suboptimal, an autosegmental rule spreading privative [voice] does nothing on encountering the unmarked /t/ in /aːɟ/ + /toːl/. This means that AGREE is equivalent to a rule spreading *binary* [voice] rather than to its privative counterpart, and it is therefore hardly surprising that the OT model is able to deal with [tense]-symmetric RVA in terms of a single constraint.[9]

So although Lombardi claims that her OT analysis of laryngeal neutralisation and voicing assimilation:

> "will get the result that both [assimilation to lenis and assimilation

---

[9] IDOnsLar and IDLar have the same binary power as AGREE. For example, both the realisation of an underlying lenis (marked) obstruent as fortis (unmarked; a and c in 32) and the realisation of a fortis obstruent as lenis (f and h) lead to violations of the first two constraints. This does not mean that privativity cannot be, or has not been, implemented or simulated in OT. For instance, **Struc(ture)*- type constraints (cf. Prince & Smolensky 1993) generally seem to be limited to single values of traditionally binary features, so that **nas(al)* and **voice* (which for present purposes is equivalent to the **Lar* of Lombardi 1995c, 1997, 1999) are possible constraints but not **-nas* and **-voice*. Furthermore, IDOnsLar and IDLar could be redefined as *DEP*-type correspondence constraints (McCarthy & Prince, 1995) that are sensitive to marked but not unmarked structure in output forms. These constraints would still be violated by (f) and (h) in (32), which 'add' marked [voice] to the underlying form, but no longer by (a) and (c), which 'remove' marked structure, and thus act in a [tense]-asymmetric fashion. AGREE could be redefined along similar lines so as to penalise fortis + lenis (b, f) but not lenis + fortis constraints (c, g). Needless to say, restoring privativity in Lombardi's model in this way would reintroduce the problems caused by [tense]-symmetric RVA as well.

A different OT-based approach to the sort of asymmetric phenomena that have been accounted for in terms of monovalent features, is to (universally) rank faithfulness constraints for marked feature values above those for unmarked values, as in *Ident(+voice) >> Ident(-voice)* or *AGREE(+voice) >> AGREE(-voice)*. Depending on how such fixed feature hierarchies are interleaved with other constraints, they supply unary or binary feature power and can therefore describe both symmetric and asymmetric RVA. This approach, which in some ways is closest to the original spirit of OT, is exemplified by the work of Gnanadesikan (1997).

> to fortis obstruents] are equally natural, while still allowing us to maintain the result that voice is privative. (Lombardi 1999:280)”

this analysis in fact merely underlines the fundamental problem that [tense]-symmetric RVA poses for truly privative feature accounts.

8.3 VOT-based models

The models of Lombardi (1994, 1995a,b), Cho (1990a/1999,b) and others I discussed in the previous section are both [tense]-based and privative. These are strictly speaking independent properties, but since the hard predictions of these models flow from their privative representation of [tense], I mostly referred to them simply as [tense]-based models. However, most binary feature accounts of the fortis-lenis distinction and voicing assimilation are [tense]-based in that they consistently represent fortis obstruents as [-voice] or [+tense] and lenis obstruents as [+voice] or [-tense], irrespective of their phonetic voicing.

The models reviewed in this section on the other hand, encode the lexical contrast between fortis and lenis obstruents in a way that is transparently (and universally) related to the use of VOT to signal distinctions between word-initial and prestress plosives. Just as Lombardi’s model they employ monovalent features, but they assign the unmarked structure to the series of plosives that has a zero-to-short lag VOT utterance initially. Consequently, laryngeal contrast is represented differently for aspirating and voicing languages, and this leads to the prediction that laryngeal neutralisation and voicing assimilation behave differently in the two types of languages: in this sense VOT-based models are diametrically opposed to [tense]-based accounts.

Perhaps the most comprehensive defense of a privative VOT-based approach to the languages in the Germanic group appears in the work of Iverson & Salmons (1995, 1999), but the idea to encode lexical laryngeal contrast in obstruents in terms of their voicing is present in Kaye et al. (1990) and is fleshed out for English by Harris (1994). Moreover, both of the latter draw on earlier work by Halle & Stevens (1971) to justify their proposals to link the representation of laryngeal contrast in obstruents and F_0/lexical tone on sonorants. Jessen (1998) also proposes different phonological representations for the fortis-lenis distinction in aspirating and voicing languages but employs binary rather than monovalent features.

Iverson & Salmons (1995, 1999) adopt the articulation-based terminology of recent versions of Feature Geometry (cf. Clements & Hume 1995) whereas the interpretation of the phonological elements used by Harris (1994) is couched in auditory/acoustic terms (see also Harris & Lindsey 1995). Nevertheless, both models recognise the set of VOT classes for utterance-initial stops that was identified by Lisker & Abramson (1964), and both models assign formally identical structures to these classes. Figure 8.5 displays the representations of the 3 VOT classes in the notation of Harris (1994): *H(igh tone)* represents the long lag VOT element, *L(ow tone)* the negative VOT element, whilst *h* represents the *noise* element which (roughly) defines the class of obstruents and can for present purposes be treated as equivalent to [-son]. These names for the long lag and negative VOT elements betray the influence of Halle & Stevens (1971) and their attempt to unify the representation of lexical tone and laryngeal contrast in obstruents using the features [±stiff vocal folds] and [±slack vocal folds]. H and L correspond to the features [spread glottis] and [voice] in the work of Iverson & Salmons (1999, 1995).

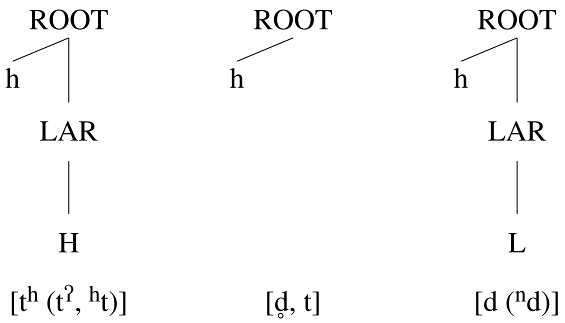

Figure 8.5: Representation of the fortis-lenis contrast according to Harris (1994) and Iverson & Salmons (1995, 1999)

On of the principal advantages of VOT-based models is that they capture the connection between prevoicing in lenis plosives and RVA in a natural way. However, reducing (the representation of) the fortis-lenis contrast to (utterance-)initial VOT distinctions has several serious drawbacks as well. Most of these drawbacks are related to the fact that VOT-based models cannot easily express the various similarities that justify the use of *fortis* and *lenis* as phonetic and phonological classes across voicing and aspirating languages. For example, they are unable to capture the similar behaviour of word-final laryngeal neutralisation and past tense paradigms in voicing and aspirating languages. The unification of passively voiced lax stops and actively devoiced unaspirated tense stops into a single phonological category is another source of problems for VOT-based frameworks. It becomes impossible to predict, for example, that unaspirated fortis stops, but not passively voiced lenis stops are RVA triggers. Finally, the monovalent feature representations of the VOT-based accounts discussed below are unable to capture symmetric RVA in a satisfactory way, much as [tense]-based privative accounts.

8.3.1 Phonetic interpretation of fortis and lenis obstruents

The phonetic interpretation of the structures in figure 8.5 in terms of initial and prestress VOT is obviously unproblematic. However, the broader picture of the phonetic interpretation of the fortis-lenis contrast in VOT-based models is fraught with difficulties, because they posit a single category unifying the tense stops of voicing languages with the lax stops of aspirating languages. Yet it is evident from the work of Kingston & Diehl (1994) and numerous earlier authors that the tense stops of voicing languages and the lax stops of aspirating languages are phonetically distinct in terms of duration, preceding vowel duration, release burst characteristics, low-frequency spectral features. In fact, as argued in sections 2.2.1 and 2.2.2, the two types of stop are most probably distinct even in terms of phonetic voicing. The tense stops of voicing languages behave as if they are actively devoiced whereas the lax stops of aspirating languages are most likely to be passively (de)voiced in context: in a sense, their similar VOTs in utterance-initial and postobstruent environments can be regarded as accidental.[10] As a consequence, VOT-based models have to specify the phonetic interpretation of the unmarked structure on a language-specific basis, whilst the interpretation of marked stops is universally fixed

Moreover, since both Harris (1994) and Iverson & Salmons (1995, 1999) align the laryngeal representation of fricatives with that of stops, there is a many-to-many mapping between the structures in figure 8.5 and [±tense] fricatives. The phonological structure and phonetic interpretation of laryngeal contrast in fricative inventories VOT-based frameworks is schematised in figure 8.6. The voiced lenis fricatives of voicing languages are assigned the laryngeal element L, which is consistent with the interpretation of this element as prevoicing (and low F_0) for plosives. However, because both Harris and Iverson and Salmons adopt a manner-symmetric approach to laryngeal contrast in obstruents, the lenis fricatives of aspirating languages are assigned the unmarked structure. Since H, L and the unmarked structure correspond universally to the same 3 VOT categories for (prestress and word-initial) plosives under VOT-based models, this implies that the lenis fricatives of voicing and aspirating languages belong to 2 separate voicing categories. Yet experiments 1 and 2 fail to support this prediction, and in an ironic twist, the only language that supplies evidence for

---

[10] This observation also undermines the suggestion by Harris & Lindsey (1995) that phonologically unmarked structure is phonetically underspecified in the sense of Keating (1988) or Pierrehumbert & Beckman (1988) (cf. 1.3.5). Note that even if other adherents of Government Phonology do not sign up to this idea as a general principle, they show a strong tendency to interpret the element H in terms of various forms of active devoicing, L as active voicing, and the unmarked structure as passively voiced. Thus, H is sometimes interpreted as the type of stiff voice/laryngealisation associated with Korean ‘tense’ unaspirated stops, or as preaspiration. Likewise, the element L has been used to represent stop (pre)nasalisation. See, e.g., Kaye et al. (1990), Heo (1994), Gussmann (1999), Ploch (1999), and section 8.4 below.

the existence of phonologically inert (passively voiced) lenis fricatives is Dutch, a voicing language which should have L-marked /v, z/ according to the VOT-based approach (see section 4.3).

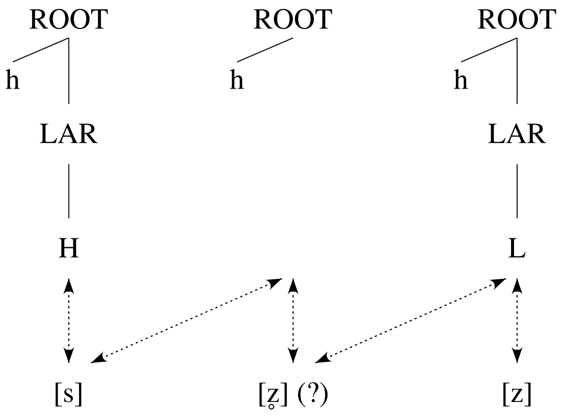

Figure 8.6: Representation of fortis and lenis fricatives in VOT-based models.

The transparency of the relationship between voicing and laryngeal structure in the models of Harris (1994) and Iverson & Salmons is further compromised by the representation of fortis fricatives, again as a result of maintaining manner symmetry in laryngeal representation. Fortis fricatives are marked H in aspirating languages because the corresponding plosives are, whereas they are unmarked in voicing languages. VOT-based models therefore do not only have to fit 4 phonetic stop categories to the Procrustean bed of 3 laryngeal classes, they are also forced to adopt a many-to-many mapping for fricatives: there are 2 possible structures each for voiceless fortis and voiced lenis fricatives, and conversely, the unmarked laryngeal structure can represent at least two categories of fricative.

As before, it might be objected that the relative complexity of phonetic interpretation models is of no concern to the phonologist, and therefore that the intransparencies identified here cannot be held against VOT-based models, but as with [tense]-based models this autonomist position backfires in the analysis of regressive voicing assimilation.

8.3.2 Regressive voicing assimilation in VOT-based models

Regressive voicing assimilation is ostensibly the mainstay of VOT-based models, because they are able to capture the observation that in voicing languages but not in aspirating languages, lenis stops trigger RVA: both Harris (1994) and

Iverson & Salmons (1995, 1999) make this point explicitly. It appears rather ironic in this light that their models perform hardly better in the analysis of RVA than the [tense]-based models of Lombardi (1994, 1995a,b) and others. The behaviour of the sounds encoded by the unmarked structure in figure 8.5 is (again) a major problem, because contrary to the predictions of the approach in voicing languages they trigger regressive voicing assimilation. Furthermore, as all other accounts that subscribe to manner symmetry in laryngeal representation, the VOT-based models reviewed in this section are unable to capture the observation that English /z/ but not /d/ triggers RVA, and as under all privative approaches their account of symmetric RVA leads to some inaccurate predictions.

The single advantage of using the VOT-based structures in figure 8.5 over the [tense]-based representations in figure 8.2 should be easy to appreciate in the context of the present study: prevoiced stops of the type found in Dutch, Yiddish, Hungarian, and similar languages bear (a LAR node dominating) the element L, and can therefore spread their laryngeal specification backwards. Zero to short lag VOT lenis stops of the type found in English, German and the North Germanic languages on the other hand have no such element available for spreading and are consequently predicted not to trigger RVA. Thus, VOT-based models, like the phonetic theory of RVA pursued in this study, capture an important observation about the assimilatory behaviour of lenis stops.

However, whereas the phonetic theory attributes the different assimilatory properties of actively vs. passively voiced /b, d, ɟ, g, ɢ/ and other obstruents to the articulatory gestures involved in their production, VOT-based theories seek an account in terms of the *acoustic results* (in a specific phonetic context) of these gestures. The latter approach is problematic because it predicts that [b̥, d̥, ɟ̊, g̊, ɢ̥] and [p, t, c, k q] should behave alike with respect to regressive voicing assimilation: both categories are represented by the unmarked laryngeal structure and should therefore be phonologically inert. Yet it is evident from the literature and in particular from the experiments reported in the previous two chapters the that fortis zero to short lag plosives of Yiddish, Hungarian (cf. the examples in 27), Dutch, and other voicing languages are capable of triggering assimilation whereas their lenis counterparts in e.g. English do not.

Although neither Harris (1994) not Iverson & Salmons (1995, 1999) discuss this issue, VOT-based models have to resort to L delinking to describe RVA to unmarked [p, t, c, k] just as [tense]-based models have to resort to [voice] delinking to represent this phenomenon. If preobstruent delinking is incorporated as a separate parameter, VOT-based models derive the 4 way taxonomy of assimilation for voicing languages without across-the-board final neutralisation that was described for [tense]-based frameworks above, including the ‘anomalous’ delinking-without-spreading type that derives [tb] for underlying /t/ + /b/ and

/d/ + /b/. As pointed out in the previous section, preobstruent delinking works around the problem of assimilation to voiceless unaspirated fortis plosives in a technical sense, but it introduces binary power into the model, and consequently weakens the basic hypothesis of monovalent feature models.

Moreover, introducing preobstruent delinking to VOT-based models has an undesirable side-effect in the analysis of H-based (aspirating) languages. Note that laryngeal spreading needs to be available as a parametric option for such languages because it allows for the representation of [tense]-asymmetric assimilation to fortis obstruents as found in Yorkshire English and to some extent for English /t/ in experiment 1 (cf. the right panel of figure 8.7). This form of assimilation cannot be derived for L-based languages for the same reason it cannot be modelled in [tense]-based models (see above), which amounts to the prediction that [tense]-asymmetric RVA to fortis obstruents only occurs in aspirating languages. However, if delinking of preobstruent laryngeal structure is also available for H-based languages (and this can be ruled out only by stipulation if it is employed in L-based systems) a 4 way typology of regressive voicing assimilation is derived for this group which contains two systems with RVA to lenis plosives, i.e., those with H delinking switched on. This is illustrated in the left panel of figure 8.7: if the lexical H of a fortis obstruent is delinked before a lenis one, the result is an ‘assimilated’ lenis + lenis cluster, which (assuming spontaneous voicing for the unmarked category) would be phonetically realised as [db̥] or [d̥̹b̥], where [̥̹] represents partial final devoicing of the alveolar stop. The fact that zero to short lag VOT lenis plosives never trigger RVA is one of the key observations behind VOT-based models and it is therefore unfortunate that to maintain this generalisation preobstruent delinking has to be ruled out for H-based languages on entirely arbitrary grounds.

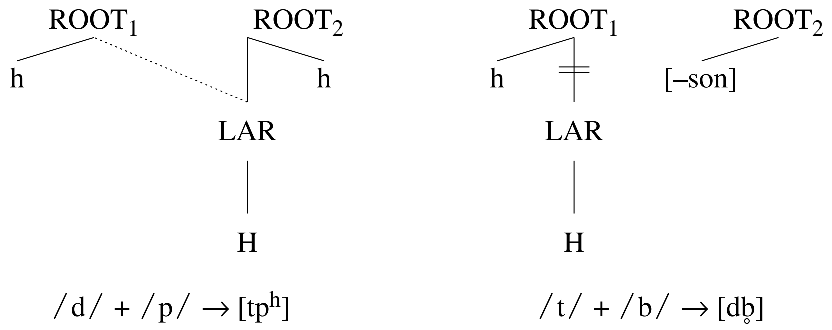

Figure 8.7: RVA in aspirating systems according to VOT-based models. Left: assimilation to fortis obstruents; right: assimilation to lenis obstruents.

Finally, ignoring the complications introduced by preobstruent delinking as a device for representing RVA, VOT-based models reviewed in this section are unable to predict the assimilatory behaviour of fricatives. Both Harris (1994) and Iverson & Salmons (1995, 1999) (tacitly) assume that laryngeal representation is manner-symmetric, and thus [v, z, ʒ, ɣ] are marked L in voicing languages but left unmarked in aspirating languages. This implies that only the lenis fricatives of the former are able to trigger regressive voicing assimilation in a preceding obstruent. Conversely, only the fortis fricatives of aspirating languages have a H element available for spreading, which predicts that voicing assimilation to fortis fricatives occurs only in this group. These suggestions are plainly contradicted by the results of experiments 1 and 2 which show that (aspirating) English [z] and (voicing) Hungarian [s] are both able to trigger RVA.

8.3.3 Laryngeal neutralisation

Iverson & Salmons (1999) maintain a fortition analysis of final neutralisation in German and Dutch, which entails that the phenomenon is represented in different ways for these two languages: for an aspirating language such as German the change of a lenis obstruent into a fortis one involves the insertion of H, whereas for voicing languages such as Frisian and Dutch it involves L delinking (cf. figure 8.8). Under a uniform analysis of neutralisation as delinking such as that of Lombardi (1994, 1995a,b), final neutralisation would manifest itself as H delinking in aspirating languages and hence as lenition rather than fortition and phonetic ‘deaspiration’ instead of ‘devoicing’. Given that the final neutralisation rules of German, Dutch and Frisian are descriptively very similar, invoking two distinct devices to represent these rules raises the suspicion that their principal motivation is to make the fortition analysis work.[11]

Iverson & Salmons (1999) appear to be aware of this issue and defend their position with the claim that in German the phonetic result of neutralising laryngeal contrast between plosives is an aspirated plosive whereas in Dutch it is a plain voiceless plosive. Since H is interpreted as long lag VOT (produced by means of aspiration) prevocalically this would be a sound argument in favour of an insertion analysis if Iverson & Salmons’ description of German final plosives were accurate, but unfortunately it is not. It is true that German plosives have been described as being postaspirated, but in many cases this description seems to refer to the presence of an audible oral release. Note that according to this definition of the term *aspiration*, which extends well beyond the literature on German, the final neutralised obstruents of Dutch are aspirated too. But as shown by Jessen (1998) word-final plosives in German are not normally accompanied by a wide glottal abduction gesture that peaks around the time of oral

---

[11] Although Harris (1994) does not propose an analysis of final neutralisation along these lines (see below) I have maintained his notation throughout this section for reasons of clarity.

release (see also Moulton 1962; Knetschke & Sperlbaum 1987). Apart from the evidence reviewed in 3 that final neutralisation in German is phonetically incomplete, descriptions suggest that the phonetic manifestation of the phenomenon is highly similar in German, Dutch and Frisian. Thus, there seems little phonetic justification for the account of Iverson & Salmons (1995, 1999).

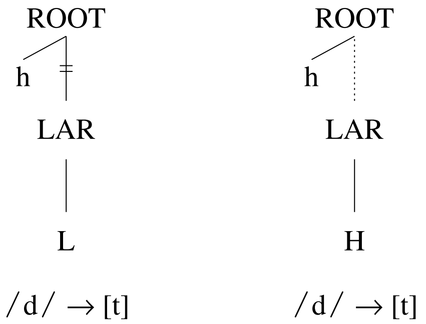

Figure 8.8: Final neutralisation in aspirating and voicing languages, according to a VOT-based model. Left: final neutralisation in voicing languages; right: neutralisation in aspirating languages.

Harris (1994) does not offer an analysis of final laryngeal neutralisation, but his general approach, in which phonetic reduction as well as phonological neutralisation are represented in terms of delinking, is hardly consistent with Iverson and Salmons' account of neutralisation in aspirating languages. Rejecting H insertion leaves only the H delinking, 'deaspiration' route open as a way of capturing final neutralisation in aspirating languages using a VOT-based framework. Although it runs counter to the long tradition of viewing final neutralisation as fortition or *Auslautverhärtung* (sharpening), this solution does not complicate the phonetic interpretation of the model to any greater degree than the H insertion analysis, whilst it eliminates the question (left unanswered by Iverson & Salmons 1999) why H insertion and L delinking should be available but L insertion and H delinking (both resulting in final obstruent 'voicing') not. It is also congruent with the spirit of the VOT-based approach, which acknowledges the observation that phonetic voicelessness is not invariably tied to phonological [+tense].[12]

---

[12] Jessen (1998) represents German final neutralisation as lenition and deaspiration rather than as fortition using a binary feature [tense]. Brockhaus (1995) is a detailed account of German final laryngeal neutralisation couched in Government Phonology, but it is crucially different from the models discussed in this section and section 8.2 above in that it abandons monovalent representation of the fortis-lenis distinction and a structure preserving analysis of final neutralisation. See section 8.4 below.

A VOT-based delinking account of final neutralisation (cf. figure 8.9) regains an important generalisation captured by [tense]-based frameworks (as well as the cue-driven account of Steriade 1997), namely that final neutralisation is equally likely to occur in aspirating and voicing languages. The presence or absence of final neutralisation is again controlled by a single parameter (LAR delinking in an independently specified set of contexts) with a single probability of being switched on (and/or a single UG default setting) whereas the model of Iverson & Salmons (1999) requires two parameters (delinking and insertion), which can only be assigned identical probabilities of being switched on or identical default settings by brute force stipulation.

Note that a VOT-based delinking analysis is still structure preserving in the sense that the output of laryngeal neutralisation is a feature structure that is present at the underlying level, i.e., the laryngeally unmarked one. As pointed out above the phonetic interpretation of this structure is problematic, and employing it to represent the class of neutralised obstruents hardly improves the situation. For German the unmarked structure represents underlying /b, d, g/ (as well as /v, z/), which are realised as [b̥, d̥, g̊] in word (utterance-)initial and postobstruent environments but as voiced or partially voiced plosives postvocalically, where they are also preceded by longer vowels than their fortis counterparts (e.g., Jessen 1998). For Dutch and Frisian on the other hand, the unmarked structure represents underlying /p, t, k, f, s, x/, which trigger vowel shortening postvocalically (Slis & Cohen, 1969a) and are generally speaking mostly voiceless across the environments in which these languages maintain laryngeal contrast. Thus, representing laryngeal neutralisation in terms of delinking implies that all else being equal, neutralised obstruents are phonetically identical to the unmarked series outside neutralisation contexts, and since the unmarked series is not implemented uniformly across voicing and aspirating languages, that there are phonetic differences between the neutralised obstruents of aspirating and voicing languages. In other words, the prediction is that the final obstruent of (the citation form of) Dutch /rɑd/, *wheel* is phonetically similar to the medial obstruent of /rɑt/ + /ən/, *rats* whereas the final plosive of German /raːd/, *bicycle*, is phonetically more similar to its medial counterpart in /reːdər/, *bicycles*.

Although a great deal of research remains to be done on the phonetics of laryngeally neutralised obstruents, this prediction appears to be inaccurate, as impressionistic phonetic descriptions suggest that the neutralised obstruents of Dutch, Frisian, and German (as well as Polish and other languages outside Germanic) are phonetically highly similar (again ignoring the incomplete neutralisation debate). Such descriptions may be (partially) biased by assumptions about the nature of final neutralisation, but as pointed out in 3 there is experimental evidence from Dutch (Ernestus, 2000) and Taiwanese (Hsu, 1996) that neutralised obstruents are phonetically distinct from both fortis and lenis obstruents (in contexts where an opposition is maintained) and similar crosslinguistically in lacking phonetic targets for [±tense]. Nevertheless, since the model of Iverson & Salmons (1999) faces similar problems with regard to phonetic interpretation, a delinking model remains preferable over a VOT-based fortition account.

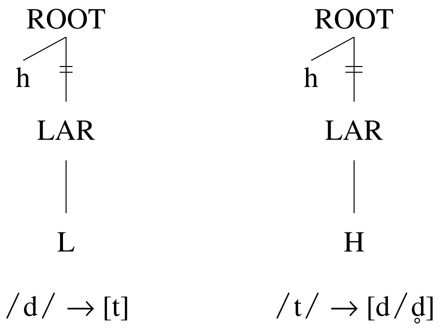

Figure 8.9: Final neutralisation in aspirating and voicing languages, according to a VOT-based delinking model. Left: final neutralisation in voicing languages; right: neutralisation in aspirating languages.

8.3.4 Progressive devoicing and the Germanic past tense paradigm

One of the motivations offered by Harris (1994) for employing the high tone/long lag VOT element to represent the fortis-lenis distinction in English is that it allows the regular past tense and plural paradigms of this language to be captured as H spreading. The underlying representation for the past tense of stems ending in a fortis obstruent such as /lɒk/, *lock*, is depicted in the left-hand panel of figure 8.10. Forward spreading of the H onto the laryngeally unmarked /d/ of the past tense suffix derives the surface form [lɒkt] with a voiceless final obstruent cluster. Stems ending in a sonorant or lenis obstruent have no laryngeal element available for this spreading operation, and consequently the past tense suffix surfaces in its unmarked, (partially) voiced form. Note that under this account it is essential that despite their difference in word-initial voicing both [+tense] fricatives and [+tense] plosives are marked for H, because the English regular past tense and plural paradigms behave symmetrically with respect to obstruent manner of articulation

However, spreading of H cannot be responsible for the allomorphy of the past tense suffix in (voicing varieties) of Dutch or in Scottish English, because fortis obstruents are unmarked in these varieties of Germanic. The right-hand side of figure 8.10 illustrates the underlying representation for past tense forms of stems ending in a fortis obstruent in voicing languages, which is of course

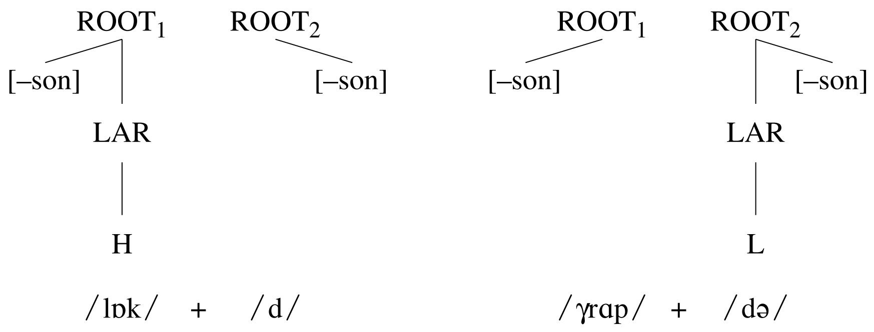

Figure 8.10: Progressive devoicing and the Germanic past tense paradigm in VOT-based models.

identical to the underlying representation assigned to such forms in the [tense]-based frameworks discussed above (cf. figure 8.4). Neither Harris (1994) not Iverson & Salmons (1995, 1999) discuss this issue at any length, but the only way to arrive at the correct surface forms for stems such as Dutch /ɣrɑp/ (cf. /ɣrɑp/ + /ən/, [χrɑpən] *to (make a) joke*) is to delink the L element of the past tense suffix. As reference to the unmarked laryngeal feature value of stem-final [+tense] obstruents is disallowed, this delinking rule would have to apply after all obstruent-final stems. In order to restore the original L of the past tense suffix after lenis obstruent-final stems and derive appropriate surface forms such as Dutch [krɑbdə], *scratched*, with voiced obstruent clusters, an additional rule would have to be posited that spreads L forward from stem-final obstruents. The two-step derivation of past tense forms of stems ending in a lenis obstruent is illustrated in figure 8.11 below.

A VOT-based account of the Germanic past tense paradigm and similar phenomena begs the same sorts of questions as Iverson & Salmons's approach to final neutralisation in German and Dutch. First and foremost, it implies that the past tense rule of (voicing varieties of) Dutch is somehow a very different species from its (aspirating) English and Swedish (and aspirating Dutch) counterparts, whilst there is little data to suggest that this is indeed the case. It is true that the Dutch past tense rule is lacking in phonetic motivation to the extent that it preserves a contrast between stem-final /x/ and /ɣ/ which is neutralised in all other contexts for many speakers. But this 'abstractness' primarily concerns stem-final lenis obstruents: the underlying form /də/ and, crucially, its progressive assimilation after fortis stem-final obstruents are used productively, as testified by past tense forms of relatively recent borrowings from English such as [blo^{ʊ}ːdə], *smoked dope* (cf. /blo^{ʊ}ː/ + /ən/, *to smoke dope*), and [sʏɹftə] *surfed*. Moreover, there is no evidence that the regular past tense and plural rules of Scottish English behaves in any way differently to its counterpart in aspirating dialects. Consequently, there seems little empirical support for the distinct analyses they (necessarily) receive under VOT-based models.

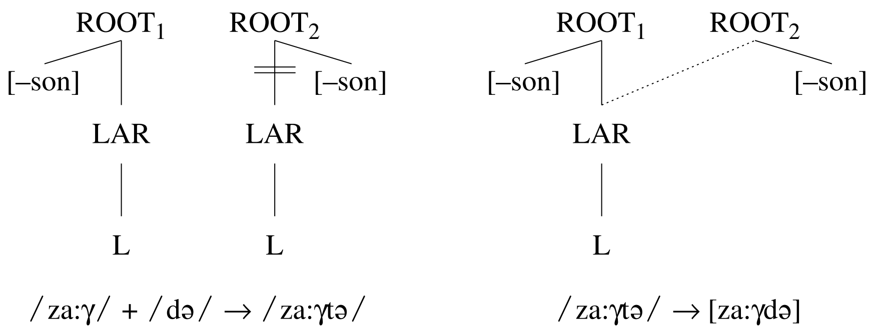

Figure 8.11: Two step derivation of past tense forms of stems ending in a lenis obstruent in voicing languages, according to VOT-based models.

In addition, the analysis of the Germanic past tense paradigm again prompts the more technical question of why L and H are typically subject to different types of rules, whilst it is expected under the general logic of monovalent models that they are subject to the same range of operations. For example, L and H must both be available for backward spreading (to model RVA) but for reasons outlined above only the former can be subject to preobstruent delinking (to capture [tense]-symmetric RVA in voicing languages). In the fortition-based account of Iverson and Salmons, final neutralisation, word final L is delinked whereas H is spread, but the authors fail to explain why word-final L insertion and H delinking are unavailable (or rare). Furthermore, on the one hand, L has to be delinked in postobstruent position to derive the correct allomorphs of the regular past tense suffix in (voicing varieties of) Dutch as well as to capture the devoicing of word-initial lenis fricatives in postobstruent environments. On the other hand, there seems to be little or no data to motivate a parallel rule of postobstruent H delinking, which (among other things) would predict a variety of English in which fortis fricatives are voiced after another obstruent: the pronunciation of <quicksand> as [kwɪkzænd] (with H delinking) in this unlikely dialect would be phonologically parallel to the realisation of Dutch /riːt/ + /zɑŋər/, *sedge warbler*, as [ritz̥ɑŋəɹ] (with L delinking).

As I argued in 8.1 above, a serious theory-internal objection to allowing both the spreading of a marked feature value x to a particular context y and delinking of the same marked feature value in the same context, is that it compromises monovalency. By also allowing *insertion* of x in y (i.e., H in word-final contexts)

Iverson & Salmons (1999) equip their model with full binary feature power. Needless to say, any additional constraints to protect monovalency, such as a ban on inserting L word finally in voicing languages or deleting H in aspirating languages have to be stipulated separately in the model.

8.4 Between discrete and continuous representations

In the previous two sections I have attempted to demonstrate that monovalent lexical feature models have little to offer that in any way complements the predictions of the functionalist approach(es) to neutralisation and RVA pursued in chapters 2 through 5, or places useful metaconstraints on these predictions. It is true that both [tense]-based and VOT-models ‘connect’ predictions about domains (neutralisation and assimilation) that are treated as separate according to the functionalist approach, but such predictions are often inaccurate, e.g., the connection between the ability of lax stops and lax fricatives to trigger regressive voicing assimilation.

Three particularly notable problems for the models I reviewed are: (1) the behaviour of the plain voiceless fortis obstruents [p, t, c, k, f, s, ʃ, x] of voicing languages, which are predicted to be phonologically (and phonetically) inert, but which are nevertheless capable of triggering RVA and the allomorphy of the Dutch and Frisian regular past tense suffixes; (2) the phonetic interpretation and assimilatory inertia of the (passively voiced) lenis plosives [b̥, d̥ ɟ̊, g̊] of aspirating languages, which are either classified with the actively (pre)voiced plosives of voicing languages (in [tense]-based models) or with the actively devoiced fortis obstruents of the same group of languages (by VOT-based models), but which are clearly distinct from both; (3) the behaviour of plain voiceless fortis and voiced lenis fricatives, which are predicted to always pattern with the corresponding plosives in the same language, which yields the wrong results for aspirating languages.

It would be unfair to dismiss all (contemporary) generative work on the laryngeal phonology of voicing and aspirating languages on the basis of the two previous sections alone. There are models that take a less parsimonious line on the representation (and phonological visibility) of the fortis-lenis distinction, and therefore appear to make the problems identified here more tractable. However, the goal of this section is to demonstrate that the solutions offered by these models are little more than apparent, since they undermine the basis of the formalist enterprise in phonological theory. First and foremost, the proliferation of structural categories in frameworks with enriched representations leads to an explosion of the number of possible rules that are available on formal grounds. It seems highly unlikely that it is possible to contain this explosion without invoking grammar-external (functional) principles. Moreover, enriched models blur the distinction between discrete phonological and continuous phonetic levels of representation (or modules), and given that human phonetic knowledge extends to the continuous domain, this prompts the question whether a level of discrete representation might be entirely superfluous.

Rather than discussing a selection of the models offered by the recent literature on an individual basis, I have opted here to combine what I consider to be the key aspects of these models in an outline of a single refined autosegmental framework for laryngeal representation, and examine its performance against the data surveyed in the previous chapters. The principal reasons for pursuing this approach rather than embarking on a model-by-model survey are to keep the size of this section manageable, and to avoid repetition.

8.4.1 A refined autosegmental model

Although this is not necessarily the case for the individual models it is based on, the refined model developed here is a surface underspecification model in the sense of Pierrehumbert & Beckman (1988): if a particular feature is underspecified it is not only inert in the phonology, but also with regard to phonetic interpretation (i.e., it is not interpreted at all). As far as the phonological invisibility of underspecified features is concerned I assume the strong interpretation described in section 8.1, which means that no reference can be made to an underspecified feature in the statement of a rule environment. In addition, I have made the following basic assumptions. Terminal nodes in the subsegmental feature tree have universal phonetic interpretations, which means that even if two segments are phonologically identical their representations contain different terminal features to the extent that they have distinct phonetic targets. Non-terminal (class) nodes are devoid of any phonetic content (cf. Harris & Lindsey 1995) and merely serve to facilitate the expression of phonological rules. Finally, only two autosegmental operations are available for either type of node: spreading and delinking.

The refined model assigns the structures in figure 8.12 to aspirated and plain voiceless fortis plosives, passively voiced (zero to short lag VOT) and actively (pre)voiced lenis plosives, and voiceless fortis and voiced lenis fricatives. The *X* at the top of each structure represents a single timing slot. Nothing important hinges on whether this slot is interpreted as a slot on the skeletal (*X* or *CV*) tier adopted in many autosegmental models, or in moraic terms. Timing slots dominate *oral aperture nodes* of the type introduced by Steriade (1992, 1993) (see also Clements & Hume 1995). Aperture nodes are similar to the more familiar root nodes in that they mediate between subsegmental and prosodic (timing) structure, but bear information about the degree of oral aperture involved in the production of a segment. A_0 represents full closure, A_f the critical constriction during the friction phase of a fricative, and A_{Max} a degree of aperture that appropriate for a vowel or glide, i.e., for laminar airflow. In the account of Steriade (1992, 1993), released plosives are represented as contour segments consisting of a closure (A_0) node and a release node, i.e., A_f for affricates and A_{Max} for stops (including nasal stops) with a plain release. Fricatives, approximants and vowels on the other hand, consist of single A_f and A_{Max} nodes respectively.[13]

Aperture nodes serve as anchors for the LAR node and the remaining elements of subsegmental structure, which, as elsewhere in this chapter, is omitted from segmental representations. By tying together information about the laryngeal phonology and phonetics of segments, the LAR node fulfills the same role as its namesake in other recent frameworks, including the ones reviewed above. Its presence encodes the availability of a lexical laryngeal contrast and with respect to the languages that form the focus of this study, it therefore has the same purpose as the descriptive feature [tense].

LAR dominates two features: [voice] and [L/H]. The former is a ternary feature intended to capture active obstruent voicing ([+voice]) and devoicing ([-voice]), and passive voicing ([0voice]: i.e., lack of a voicing target). Because [0voice] is assumed to be phonologically as well as phonetically inactive it is graphically represented as absent, according to standard practice in monovalent autosegmental models. L encodes the phonetic features shared by lenis obstruents across the voicing-aspirating divide such as shorter closure duration and F_0 lowering, whereas H represents the phonetic features shared by fortis obstruents. L and H may be seen as two marked values of two binary features, each with one active value, or a single ternary feature, with a third ‘0’ value representing the absence of targets for the phonetic features in question. In addition, both [voice] and [L/H] could be conceived as nonterminal nodes dominating the separate articulatory ([spread glottis], [stiff vocal folds], etc.) or acoustic features involved in their production, much as the *dimensions* of Avery & Idsardi (2001). They also bear some resemblance to the *Intermediate Perceptual Properties* (IPPs) of Kingston & Diehl (1994, 1995). Note that the split representation of phonetic voicing and the other phonetic properties involved in the signalling of the fortis-lenis distinction has precedents in early generative work on Dutch RVA such as Hubers & Kooij (1973) and Brink (1975), as well as the ‘syncretic’ representation of lenis fricatives in German as [+voice, -tense] proposed by Jessen (1996, 1998).

Both long lag VOT and plain voiceless fortis obstruents are represented in the refined model as containing the feature H and as actively devoiced ([-voice]). The difference in VOT between these two phonetic categories is encoded in terms of the docking site for the LAR node: the release (A_{Max}) for the former,

---

[13] Since the structures in figure 8.12 are intended to facilitate discussion of certain proposals in the literature rather than as a fully-fledged framework for laryngeal representation, I have ignored a number of issues that would have been relevant otherwise, such as the representation of (contrastively) voiced aspirates, aspirated fricatives, or voiceless sonorants.

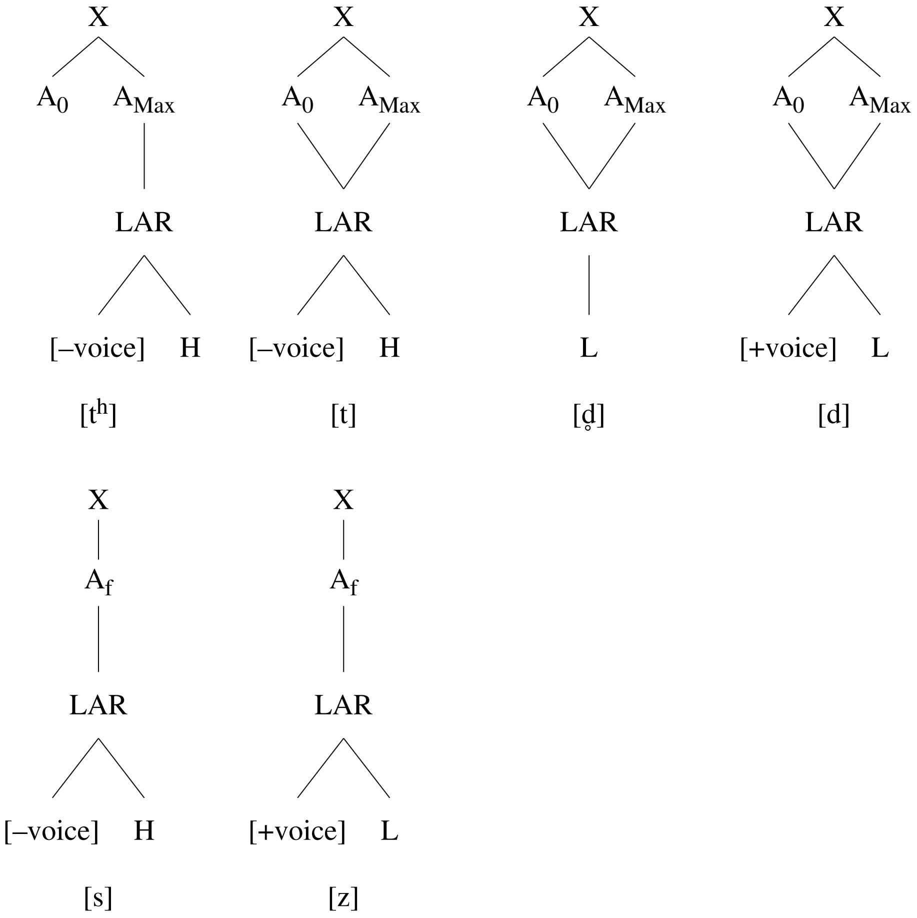

Figure 8.12: Refined autosegmental representations for plosives and fricatives. Place and manner features omitted.

and both A_0 and A_{Max} for the latter. This encoding is more or less iconic in terms of the timing of the glottal abduction gestures involved in the production of long lag and zero to short lag VOT in fortis plosives, and parallels the representation of postnasalised vs. nasal and glottalised vs. postglottalised stops in the work of Steriade (1992, 1993) (see also Keating 1990a). Furthermore, it is reminiscent of the encoding of the glottalised and aspirated plosives of Korean in Heo (1994) and Ploch (1999). According to these accounts, both glottalised and aspirated obstruents contain the element H, which is 'fused' into a single phonological expression with the other relevant elements in the former, and 'extraposed' as a separate phonological expression in the latter instance.

The structures for lenis plosives similarly capture the phonetic similarities and differences between the passively voiced (zero to short lag VOT) and actively (pre)voiced classes: both contain the feature L, which maps into, e.g., relatively long preceding vowels and relatively low F_0 on a following vowel, but only the latter includes the active voicing ([+voice]) feature. In the fricative structures, LAR can only be associated with the single aperture node A_f, which captures the observation that there appear to be only 2 distinctive voicing targets for fricatives across (most of) the languages examined in this study.[14] Both fortis and lenis fricatives are assigned an active value for [voice] as well as for [L/H], whilst voiced sonorants are assumed to be phonologically and phonetically fully inert (underspecified) in terms of laryngeal structure. The representation of fricatives in the refined model therefore entails abandoning the assumption that laryngeal representation is symmetric with respect to obstruent manner of articulation, which is central to many lexical feature models.

It is important to note that, as a surface underspecification model, the framework for laryngeal representation presented here is intended to be phonetically transparent only with respect to the phonetic *targets* associated with [±tense], and the coarticulation of those targets. It is not designed to derive the effects of hypoarticulation or supralaryngeal articulatory settings on the manifestation of the fortis-lenis distinction in the speech signal: such effects are assumed to arise in the course of phonetic implementation and as byproducts of the mechanics and aerodynamics of the vocal tract respectively. Thus, the deaspiration of fortis plosives in poststress prevocalic positions in aspirating languages is not represented as the association of the LAR node of aspirated stops to A_0, but is attributed to local articulatory reduction, which affects all articulatory gestures in a gradient fashion. Similarly, word-initial lenis stops in aspirating languages are represented as lacking an active value for [voice], in spite of the fact that these stops may be (partially) voiced, especially after sonorants, since this voicing is most likely passive (cf. chapters 4, 5).

Conversely, where there are grounds to assume that positional variation in the realisation of [±tense] reflects genuinely different targets, this variation has to be incorporated in the model, even if the phonetic differences involved are not lexically contrastive. Since there is good phonetic evidence for treating the unaspirated realisation of final fortis plosives in many aspirating languages as stemming from a distinct positional target, this phenomenon provides the clearest case for inclusion in the model. The fortis plosives of English and similar aspirating languages are therefore represented by the leftmost structure in figure 8.12 if they occur before a sonorant within the same word, but with a doubly linked LAR node word finally. In addition, a [-voice, H] LAR node with a single association to A_0 can serve to represent preaspirated or preglottalised fortis stops (Steriade, 1992, 1993).

---

[14] On the representation of lenis fricatives in Dutch, see below.

8.4.2 Advantages of the refined model

One of the advantages of the model sketched in figure 8.12 is that it allows for a crosslinguistically uniform and phonetically falsifiable analysis of laryngeal neutralisation. Since both fortis and lenis obstruents in aspirating as well as voicing languages are marked with a LAR node, a single operation of LAR delinking suffices to eliminate the contrast between the two lexical classes. As in [tense]-based frameworks and VOT-based delinking models, LAR delinking can be stated as a single (UG) parameter and consequently the refined model predicts correctly that (final) neutralisation does not depend on the phonetic manifestation of a laryngeal contrast. It is different from the models discussed in the previous two sections but similar to the analyses of Gussman (1992), Brockhaus (1995), Steriade (1997), and Ernestus (2000) in that laryngeal neutralisation is neither fortition nor lenition but a symmetric operation that affects both fortis and lenis obstruents to produce a separate, fully inert, laryngeal category. Under the assumption that phonologically unmarked always equals phonetically underspecified, this account predicts that laryngeal neutralisation results in obstruents that are phonetically distinct from both their lenis and fortis counterparts in acquiring their voicing, duration and other 'laryngeal' properties from the phonetic context (and their own remaining phonological features). 3 discussed how phonetic evidence from Dutch, Taiwanese, and English /s/ + stop clusters is consistent with this prediction.[15]

Furthermore, since the refined model outlined in figure 8.12 encodes active obstruent (de)voicing in a transparent fashion, it is able to capture the set of obstruents that trigger regressive voicing assimilation in terms of a simple [±voice] spreading-cum-delinking rule. The two classes of fortis plosives, fortis as well as lenis fricatives, and prevoiced lenis plosives all have a [voice] feature with an active value available for this spreading operation, which is illustrated in figure 8.13, and so the model correctly predicts that these obstruents are able to trigger voicing assimilation in a preceding obstruent. Passively voiced lenis plosives are unable to trigger RVA because they are assigned the inert value for [voice]. Consequently, it appears that the [voice] spreading rule does not have to be parametrised, as for example in Lombardi's model: if it is universally switched on the assimilatory behaviour of fortis and lenis obstruents follow from

---

[15] Gussman (1992) retains the notion that final laryngeal neutralisation ultimately results in fortition/devoicing by using a 'fill in ' rule late in the derivation that specifies all underspecified obstruents as [-voice]. Brockhaus (1995) on the other hand maintains the traditional idea that laryngeal neutralisation in German is something that *happens to lenis obstruents only*, but nevertheless treats the output of the process as laryngeally unmarked and therefore distinct from lexically fortis obstruents, which are marked H in her model. Data showing that German final laryngeal neutralisation is incomplete provides the main argument for this 'hybrid' position, which is independently suggested by Charles-Luce (1985).

the representations in figure 8.12.[16]

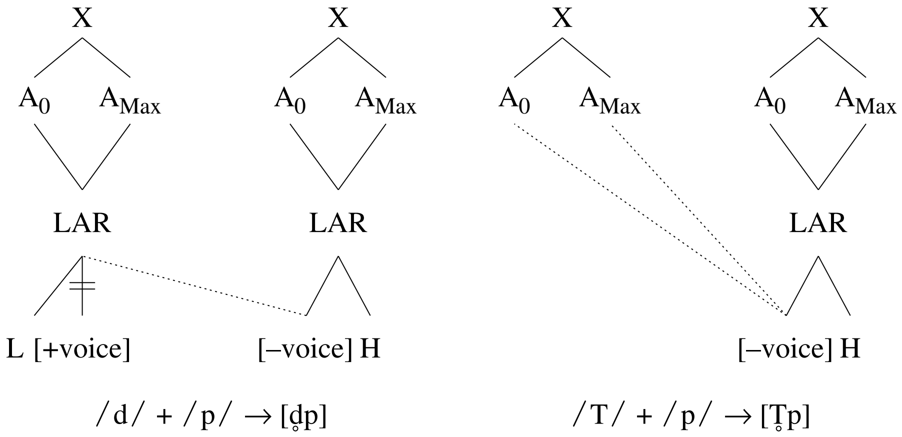

Figure 8.13: Regressive voicing assimilation in the refined autosegmental model. Left-hand side: non-neutralising RVA in languages that maintain a fortis-lenis distinction in word-final contexts. Right: RVA in languages with final laryngeal neutralisation.

Modelling RVA in terms of [voice] delinking and spreading has two additional advantages over the models described in the previous sections of this chapter, and over lexical feature analyses more in general. First, it accounts for the non-neutralising nature of the process as it was established for English and Hungarian in chapters 5 and 6, because it operates on a ‘sublexical’ feature. For example, if [-voice] spreads from an underlying /p/ to a preceding /d/ with concomitant [+voice] delinking in the latter, the alveolar obstruent still retains its lexical [L] specification and thus remains distinct from underlying /t/ (cf. the left-hand side of figure 8.13). In IPA notation this corresponds to, e.g., /d/ + /p/ → [d̥p] rather than the /d/ + /p/ → [tp] which is implied by lexical feature analyses. Second, in conjunction with an analysis of final neutralisation in terms of LAR delinking, sublexical [voice] spreading captures the [tense]-symmetric RVA effect reported for Dutch in 7. Because neutralised obstruents are [0voice], spreading of [-voice] from a following fortis plosive is no longer vacuous or inapplicable (as it is in lexical feature models), and neutralised are predicted to show a 3 way voicing distinction before lenis plosives ([+voice] through spread-

---

[16] To the extent that underlying voicing distinctions are maintained in obstruents targeted by RVA the delinking part in the left-hand panel of figure 8.13 could be omitted. The output of the spreading rule would then be a segment display contour or doubly articulated behaviour with respect to [voice]. See Hayes (1992) for an approach in roughly these terms to incomplete neutralisation in English coronal place assimilation.

ing), sonorants (inert [0voice]) and fortis obstruents ([-voice]: cf. figure 8.13)[17]

The refined model is also able to handle (most versions of) the Germanic past tense paradigm and similar rules of allomorphy as well as traditional binary feature accounts, at least under the assumption that these processes involve complete neutralisation. Spreading the LAR node of stem-final fortis obstruents forward and delinking the underlying LAR node from the initial /d/ of the past tense suffix derives the right pattern for both aspirating and voicing languages, since fortis obstruents have a node available for spreading in both types of language. Note that the spreading rule can but does not have to be defined as targeting stem-final fortis obstruents only: given that stem-final lenis obstruents have the same laryngeal specification as the initial /d/ of the suffix, and given that voiced sonorants are inert in terms of laryngeal structure, the output is the same if all available stem-final LAR nodes are spread rightward. The analysis is illustrated in figure 8.14 for Dutch /ɣrɑp/ + /də/, [χrɑptə], *joked*, and /krɑb/ + /də/, [krɑbdə], *scratched*.

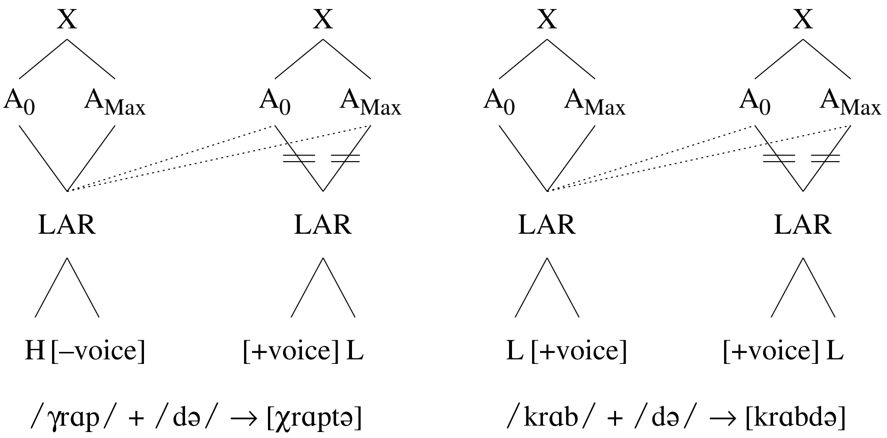

Figure 8.14: The ‘Germanic’ past tense paradigm in the refined autosegmental model. Left-hand side: LAR spreading from stem-final fortis obstruents. Right: (vacuous) spreading from lenis obstruents.

Finally, in chapters 2 and 4 I argued that Dutch postobstruent lenis frica-

---

[17] For the sake of the argument I have assumed that [±voice] attaches directly to the aperture nodes in neutralised obstruents. A more detailed version of the refined model would have to include a precise statement of the possible docking sites for [±voice], but this is not relevant to the point made here. Also note that the assimilatory behaviour of Dutch word-initial /h/, [ʔ] demands that both segments be specified [-voice]. There do not seem to be compelling arguments against a [-voice] representation of these sounds, given that they can be distinguished in terms of the lower level articulatory features [(+)spread glottis] for /h/ and [(+)constricted glottis] for [ʔ].

tive devoicing is not a form of voicing assimilation in the same sense as the regressive assimilation to lenis and fortis plosives found in the same language. The main argument for this position is that in languages that maintain a fortis-lenis contrast word finally, lenis fricatives are devoiced after both classes of obstruent, even to a smaller degree than in Dutch. Dutch, and as shown in 6, Hungarian prevoiced plosives are also subject to (partial) devoicing following another obstruent, especially in in unstressed environments. It seems therefore plausible that the particularly strong devoicing of Dutch lenis fricatives arises through the interaction of normal obstruent aerodynamics with a voicing target that is somehow weaker than that of the corresponding fricatives in English and Hungarian, even though the clinching evidence (A substantial intraspeaker correlation between the degrees of lenis fricative devoicing in utterance-initial and postobstruent environments) is not yet available.

In contrast to lexical feature models, the refined model goes some way in modelling this analysis. Weakly voiced lenis fricatives can be represented analogously to passively voiced lenis plosives as possessing a LAR node that bears active L but inactive [0voice]. As [0voice] cannot be spread, this would mean that Dutch /v, z/ do not have the capacity to trigger RVA in a preceding obstruent, and since [0voice] is phonetically interpreted as the absence of a voicing target, fricative devoicing would follow (passively) from the aerodynamics of obstruent sequences. On the grounds of the same aerodynamics it would be predicted that utterance initially, Dutch lenis fricatives are voiceless, but shorter than their lenis counterparts and followed by a lower F_0. Between sonorants however, they would be subject to a greater degree of passive voicing than /f, s, x/, which are represented as actively devoiced ([-voice]).

8.4.3 Problems with the refined model

To return to the three especially problematic observations highlighted in the introduction to this section, it seems that a marked improvement can be achieved with the refined model on all three counts. Given the nature of the data it is not surprising that the model's comparative success is largely due to its phonetic realism, and to its symmetric encoding of the fortis-lenis distinction (the latter could be regarded as a corollary of the former). Thus, in contrast to monovalent lexical feature analyses, the refined model represents plain voiceless fortis obstruents as phonologically marked and therefore active rather than inert. Phonetic realism entails that actively voiced lenis stops, passively voiced lenis stops and actively devoiced fortis stops receive different structures, and consequently different assimilatory properties. It also entails abandoning manner symmetry as an a priori principle in laryngeal representation, and this in turn leads to more accurate predictions about the (crosslinguistic) assimilatory behaviour of fricatives.

However, as it stands, the refined model still leaves a number of empirical issues unresolved, and raises a number of more serious theoretical problems. The different propensities of stops and fricatives for (lexical) laryngeal neutralisation represent one prominent empirical issue. Another is the observation that voicing, but generally not durational correlates of [±tense] are involved in regressive assimilation: the model in 8.12 accounts for the fact that voicing can spread separately, but offers no explanation why L/H cannot spread independently. The former problem could be tackled by exploiting differences between fricatives and plosives at the aperture node level. For instance, the LAR class node could always be linked to both aperture nodes in plosives (this would involve re-encoding of the difference between aspirated and unaspirated fortis plosives). Just as double linking has been used to represent geminate integrity (e.g., Hayes 1986), this 'double bond' could then be employed to encode the relative resistance of plosives to laryngeal neutralisation.

As far as the durational correlates of [±tense] are concerned, their failure to spread backwards would follow if they were incorporated into prosodic rather than melodic (sub)segmental structure. I discussed this idea early on in 4.1.2 to illuminate the predictions of a coarticulation-driven account of RVA. It involves representing the durational contrast between fortis and lenis obstruents by assigning the former a larger number of slots on a timing tier (cf. figure 8.15), much as the contrast between singleton and geminate consonants is usually encoded in terms of positions on a *X*, *CV* or moraic tier (see Perlmutter 1995 for a survey and references). Under this analysis, the lack of durational assimilation in obstruent sequences would follow from the general prohibition in autosegmental models against the spreading of prosodic structure (i.e., anything above the root level). At least in principle, it would also allow for the durational tradeoff between obstruent duration and preceding vowel duration to be treated as compensatory lengthening, again along the same lines as familiar analyses of fixed quantity syllable rhymes (e.g., Hayes 1989).[18]

Note that a serious proposal in this vein would be beset by all sorts of complications, including the awkward specification of the fixed quantities involved (given that many languages allow lexically contrastive length regardless of the laryngeal specification of the following obstruent, e.g., English or allow contrastive length to cross classify with [±tense]), e.g., Italian, and the difficulty of making durational contrast dependent on the presence of a LAR node. Moreover, prosodic representation of the durational correlates would still leave the behaviour of F_0 and F_1 cues unaccounted for.

Unfortunately, given a formalist perspective, these complications are part of

---

[18] In the light of the grid-based representation of phrase-final lengthening developed by Selkirk (1984), the use of a phonological timing tier to represent this sort of phonetic detail is less odd than it may seem.

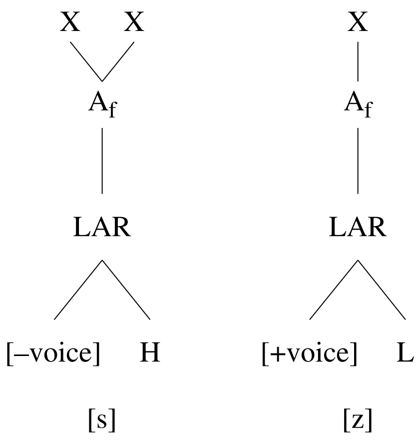

Figure 8.15: Prosodic representation of durational cues to the fortis-lenis distinction.

a much heavier price of overgeneration that comes with the structures in figure 8.15. According to a strictly formalist logic, the inventory of possible phonological rules is derived from the available structural primitives and a small number of principles governing their combination. Consequently, an increase in the number of structural primitives leads to an increase in the number of possible rules, and to the extent that the resulting rules have to be ruled out on arbitrary grounds they attenuate the predictive power of the model in question. It is certainly no accident therefore that those who take the formalist enterprise most seriously (e.g., Jensen 1994; Ploch 1999) generally seek to reduce rather than expand the number of structural primitives. Recall too that one of the arguments used by Lombardi (1994, 1995b) in favour of her three term ([voice], [asp], [gl]) monovalent feature inventory is that it maximally generates a 6 term system, which is exactly the maximum number of contrasts that has been established for a single language.

Because the refined model encodes laryngeal contrast on three tiers (nodes) and allows for two autosegmental rules (delinking and spreading) it generates 6 types of rule with phonetically distinct outputs. For instance, the model predicts a three way taxonomy of final neutralisation, comprising a language with complete neutralisation of phonetic contrast between fortis and lenis obstruents (LAR delinking), a language that erases only voicing distinctions ([voice] delinking) whilst leaving durational, burst-related and F_0/F_1 cues intact, and a type of language in which final laryngeal contrast is solely marked in terms of voicing (L/H delinking).[19] Needless to say, the latter two types have not been

---

[19] None of the combinations of these rules can be distinguished from LAR delinking in terms

attested. Likewise, the refined model generates a number of unlikely underlying forms and lexical inventories in addition to the ones illustrated in figure8.12. For example, there is no model-internal reason why lexical laryngeal contrast between obstruents could not be solely based on [voice] or L/H features, or worse, why there should not be a system that cross-classifies these features to derive a wholly unattested 12 term lexical contrast.

It seems difficult if not impossible to rule out this hypothetical lexical inventory or the related rule taxonomies on formal grounds, other than by brute force stipulation. Simply stating, e.g., that LAR only spreads within words whereas [voice] spreads (presumably) both within and across word boundaries, whilst L/H does not spread at all represents the relevant observations fairly accurately but does nothing to predict them. Dispersing the cues related to the L/H feature across other layers of structure does not provide a solution either, but rather compounds the problem. For example, prosodic representation of the durational cues to [±tense] implies the possibility of a four-term lexical length contrast without concomitant distinctions in terms of voicing or other correlates of [±tense] (i.e., through crossclassification of the durational distinctions between fortis and lenis obstruents with the regular singleton-geminate contrast).

It is possible, on the other hand, to rule out 12 term lexical contrasts involving voicing, F_0/F_1 perturbation and duration on external, functional, grounds. The absence of this type of lexical inventory is plausibly attributed to the small amount of perceptual contrast between the individual terms, and by the same token the coocurrence of certain values of [voice] and L/H can be seen as auditory and/or articulatory enhancement in the sense of Stevens & Keyser (1989). But the introduction of such external constraints undermines not only the formalism of the model but also the need for discrete phonological representations as part of a modular model of the phonology-phonetics interface. Recall from 1.3.2 that in principle, external mechanisms such as auditory enhancement, articulatory effort reduction (mediated by various forms of feedback), misperception, and phonological (re)analysis by listeners are able to generate, maintain, and change ‘implicit’ structure in continuous phonetic space. This means that the need for a level of categorical phonological representations hinges on arguments for purely formal constraints on phonological patterns. The previous chapters of this study have built a case for an analysis of voicing assimilation and laryngeal neutralisation phenomena wholly in terms of grammar-external principles. The previous sections of this chapter show that at least the predictions of monovalent lexical feature models do not reveal any tenable metatheory of assimilation and neutralisation. Consequently, the observation that the refined autosegmental model, whilst improving on the predictions of lexical feature frameworks, overgenerates rules and (phonetic) inventories cannot be dismissed as a minor drawback,

---

of its output.

but undermines the most basic premises of formalist approaches to laryngeal neutralisation and voicing assimilation.

8.5 Conclusions

The aim of this chapter was to demonstrate that current generative models of laryngeal phonology do not provide a theory or metatheory of laryngeal neutralisation and/or voicing assimilation that adds to, or reaches beyond, the predictions of a functionalist approach. This investigation was inspired by the thinking of citethare00a,hare00b and others who see a role for formalist phonology next to, or as a prerequisite to, functionalist models. Its principal method was an assessment of the predictions of two influential generative frameworks of laryngeal phonology: the [tense]-based approach of Lombardi and others, and the VOT-based approach espoused by Harris (1994), Iverson & Salmons (1995, 1999), and in a modified form by Avery & Idsardi (2001).

Both types of model were found to be fundamentally inadequate in a number of respects, even if the phonetic detail of regressive assimilation at word boundaries is ignored. [tense]-based models fail to predict the critical difference in assimilatory capacity between actively voiced and passively voiced lax stops, and derive the grossly incorrect prediction that all fortis obstruents are phonologically inactive. Both types of model fail to predict the phonological and phonetic activity of fortis obstruents in voicing languages, and are unable to handle asymmetries between plosives and fricatives. Neither of the two models has a phonetic interpretation that is transparent across contexts, obstruent types, and/or languages. Worst of all perhaps, the patches that are introduced (or have to be introduced) to deal with some of the more glaring mispredictions effectively involve the use of binary [±tense]. This seriously reduces the predictive power of the models in question, which is mostly founded on unary feature marking.

Using only devices that have been proposed in the published literature, I then constructed an alternative autosegmental model of fortis-lenis laryngeal phonology and phonetics that improves on feature-based models by explicitly incorporating ('phonemic') information about lexical contrast and ('subphonemic') information about phonetic detail. The former allows for a uniform statement of (final) neutralisation rules across voicing and aspirating languages (as in [tense-based lexical feature frameworks) whilst the inclusion of phonetic detail allows for more accurate predictions about the triggers and phonetic manifestation of regressive voicing assimilation, and potentially about fricative-stop asymmetries in laryngeal marking.

However, this refined autosegmental model is untenable on formalist grounds because its proliferation of structural categories results in overgeneration. Without evidence for purely formal constraints on the laryngeal phonology of fortis-lenis systems it is equally untenable from a functionalist perspective since the external principles that it requires to avoid overgeneration render explicit categorical phonological structure superfluous.

It is impossible to show that purely formal constraints on laryngeal neutralisation and voicing assimilation do not exist; it is also true that I made no attempt in this chapter to unearth any such constraints. But I have demonstrated that current generative models of laryngeal phonology are fundamentally flawed even on their own terms, and bring no insights that complement or enable the explanations provided by a functionalist approach. I believe that this puts the burden of evidence in this matter firmly back on proponents of formalist models.

Chapter 9

Summary, conclusions, remaining issues

> “Het is uiterst bezwaarlijk eenigszins betrouwbare gegevens omtrent de assimilatie bij de samenstelling en afleiding van woorden te verzamelen. Zoodra men iemand toch verzoekt een woord of zin te zeggen, zet hij zich schrap om het zo goed mogelijk te doen en het resultaat is gekunsteld. …Het materiaal voor op te stellen regels kan derhalve slechts te hooi en te gras verzameld worden en moet noodzakelijk zeer onvolledig en van verschillende waarnemers zeer uiteenloopend zijn.” (Zwaardemaker & Eijkman 1928: 223-224)

> “It is most troublesome to collect any reliable data with regard to assimilation under the compounding and derivation of words. As soon as one requests someone to say a word or sentence after all, he will strain to pronounce it as good as he can, and the result will be artificial. …Material to base rules on can therefore be collected only in a haphazard fashion and will necessarily be very incomplete and very divergent for different observers.”

One of the principal aims of this chapter was to develop a functionalist perspective on the phonetics and phonology of fortis-lenis systems, i.e., obstruent systems that use voicing as a cue to a two-term lexical contrast. This enterprise was organised into three broad parts, comprising chapters 1 to 3, chapters 5 through to 7, and chapter 8 respectively. The first of these was concerned with the theoretical underpinnings of a functionalist model of (laryngeal) phonology and phonetics, its basic architecture and predictions, as well as with a survey of the relevant phonetic and phonological data in the literature. The second part described three experiments designed to test the predictions of the model concerning the phonetics of regressive voicing assimilation. The third part was devoted to a critique of formalist analyses of fortis-lenis systems.

Chapter 1 described the analytical framework for this study, which was inspired by Ohala's (1981, 1993) theory of language change and more recent work on what I have referred to as *diachronic* or *evolutionary* functionalism (de Boer 1999, 2001; Blevins to appear). This brand of functionalism is distinct from the *synchronic* functionalism of Boersma (1998), Kirchner (1998), Flemming (2001) and others in its hypothesis that speakers' grammars have no direct access to functional or 'ecological' principles such as articulatory effort minimisation or perceptual optimisation. It is distinct from formalist frameworks in its assumption that all phonological and phonetic constraints are ultimately derived from such principles. An additional difference with formalist models and some (early) models of the phonetics-phonology interface associated with work on laboratory phonology is that all constraints are stated in terms of continuously-valued auditory and articulatory features

The fundamental components of the model are rote learning, transmission noise, and various forms of feedback. The first of these embodies the assumption that language learners strive to approximate the (ambient) language produced by older generations as closely as they are able to. However, because the speech transmission chain is noisy in both directions, some errors are introduced in the copying process. These errors are likely to be non-random in being approximations of the categories of the ambient language, and can develop into linguistic innovations that are retained and transmitted to subsequent generations if they receive a sufficient amount of positive feedback. Because positive feedback to a phonetic form is a function of its utility (to speaker and addressees alike) innovations and will conform to functional constraints at the time they are adopted into the phonetic grammar. One of the advantages of diachronic functionalist models is that new forms become exempt from functional pressures afterwards: this means that such models can accommodate so-called crazy rules.

Chapter 2 motivated the terms fortis/tense and lenis/lax as convenient descriptive labels for the phonetic categories found in obstruent inventories bifurcated by a two-term contrast that is phonetically supported in terms of voicing distinctions, and attempted a review of the vast literature on the phonetics of such systems.

Chapter 3 provided the phonological counterpart to the phonetic investigations of chapter 2. It attempted to identify a number of generalisations about laryngeal neutralisation in fortis-lenis systems including the type of dynamic word-final neutralisation that can be found in Dutch and German. Its theoretical point of departure was the work of Steriade (1997) which tries to derive generalisations about the effects of flanking contexts on laryngeal neutralisation from the effects of those contexts on the perceptibility of distinctions between fortis and lenis obstruents (and thus on the likelihood of such obstruents to be subject to copying errors in acquisition).

As it stands, Steriade's theory deals only with the effects of flanking contexts and not with neutralisation asymmetries between fricatives and plosives or the asymmetry between word-initial and word-final environments. Following suggestions by Balise & Diehl (1994) and the work of J. Beckman (1996, 1997) I argued that a perceptibility-driven account of laryngeal neutralisation can at least in principle be extended to incorporate positional and manner-based asymmetries. First, the well-documented phenomena of articulatory weakening and strengthening are likely to have an asymmetric effect on the perceptibility of word-initial and word-final contrasts which is consistent across flanking contexts. Second there is evidence that voicing distinctions inhibit the expression of place cues in fricative systems, which biases any functionalist model towards fricative inventories composed of only voiceless fricatives.

Chapter 3 is in many ways the most speculative of this study because several of its assumptions about perceptibility remain to be confirmed. However, I think it is important to emphasise yet again that perceptibility hierarchies represent propositions about the relative salience of specific phonetic features to speakers with specific native languages at particular times in history that can be tested in perception experiments. Thus a perceptibility-driven account of laryngeal neutralisation is empirically accountable.

Chapter 4 developed a preliminary typology of voicing assimilation phenomena, showing that there are important differences between assimilation in restricted morphological contexts and regressive assimilation across word boundaries. Whereas the former occurs regardless of the voicing categories employed by a language to cue the distinction between tense and lax obstruents, the latter is clearly dependent on the active (de)voicing of trigger obstruents. In addition, experimental studies indicate that regressive assimilation at word boundaries tends to be phonetically gradient whereas morphologically restricted assimilation (at least) seems to operate in a neutralising fashion.

These observations suggest that voicing assimilation occurs in two forms: as a phonological rule that operates on the feature [tense] or its formal equivalent(s), and as a purely articulation-based process driven by the mechanisms underlying the production of voicing contrast. I hypothesised that the former type of process is the one typically found in morphological paradigms and that the latter is responsible for regressive assimilation across word boundaries. Coarticulation-based approaches to voicing assimilation rules have been proposed before, e.g., by Slis (1985) and Ernestus (2000), but such proposals rarely spell out the phonetic typology of articulation-driven assimilation rules. Three principal features of this typology are (1) that only actively (de)voiced obstruents are able to trigger coarticulatory voicing assimilation; (2) that the only correlates of [tense] affected by assimilation obstruent voicing and phonetic features mechanically dependent on voicing; (3) that assimilation is always gradient.

The experiments reported in chapters 5 and 6 were designed to test whether regressive assimilation at word boundaries is always of the coarticulation-driven type, as suggested by chapter 4. Experiment 1 investigated patterns of assimilation in British English obstruent clusters whilst experiment 2 was an attempt to apply the same design to regressive voicing assimilation in Hungarian. In many respects the results of these experiments are in accordance with the predictions of the phonetic theory, and in some respects surprisingly so in the light of descriptions in the literature. The hypothesis that receives almost completely unequivocal support from these experiments as well as from experiment 3 is the one that states that only actively (de)voiced obstruents can trigger regressive voicing assimilation.

However, whereas the results of experiment 1 match the predictions of the phonetic theory more generally, the behaviour of vowel duration before Hungarian velar stop + obstruent sequences represents the most notable problem since it cannot be attributed to the coarticulation of voicing targets. In 6.4 I suggested that this might be interpreted as evidence for the idea that Hungarian RVA is a part-phonologised process, and a process that was perhaps sparked by the effects of phonetic RVA on the perceptibility of [±tense] in word-final plosives.

It is perhaps important to emphasise that whilst the data reported in chapter 6 contradict a purely phonetic analysis of Hungarian RVA, it does not vindicate recent generative analyses of the phenomenon. Such analyses describe Hungarian RVA as categorical, non-manner specific, and imply that the length of vowels preceding obstruent clusters should cue the laryngeal specification of the final obstruent in such clusters. All these claims are contradicted by the results of experiment 2

Chapter 7 investigated regressive assimilation of voicing in Dutch three-term clusters with a medial fricative. Part of the descriptive literature has it that assimilation does not apply in such clusters. Given that Dutch devoices word initial lenis fricatives that are preceded by an obstruents it is difficult to see this description as completely unconnected to phonological analysis. It is an inaccurate description in any case, because regressive assimilation clearly does apply in three-term clusters with a medial fricative, exactly as predicted by the phonetic theory. However, there is some evidence that the effect of assimilation is weaker in the clusters investigated in chapter 7 than in the corresponding singleton obstruents examined by Slis (1986), and this may well have given rise to the perception that assimilation does not apply at all.

The observation that Dutch RVA is [tense]-symmetric is probably the more exciting conclusion of the work reported in chapter 7: the observation that /ps/ clusters have less voicing before a fortis plosive than before a sonorant /m/ undermines one of the most pervasive and unquestioned assumptions about RVA in Dutch. It is entirely consistent with the phonetic theory of RVA (because Dutch fortis obstruents are arguably actively devoiced) and with Ernestus' (2000) hypothesis that Dutch word-final neutralisation leads to the phonetic underspecification of [tense].

Chapter 8 finally, tried to dispel the notion that formalist phonological theory has a role to play as a source of metaconstraints on functionalist analyses or at least as a source of complimentary constraints that cannot be derived otherwise. This chapter went into considerable detail in fleshing out the predictions of current generative models of laryngeal phonology. I believe this detail was essential for pinning down the predictions of the models in question and exposing the inconsistencies introduced by patches designed to make these predictions to fit the data. The final section of this chapter brought the overall argument of this study full circle by showing how autosegmental models that improve on lexical feature analyses by incorporating phonetic detail need to be constrained by external principles and thus dissolve into the type of framework set out in chapter 1.

Appendix A

Stimuli and supplementary data for experiment 1

A.1 Stimuli

/k/ + /t/

How does Falkirk tonic translate?
How does Selkirk topping translate?
How does brickwork tunnel translate?
How does patchwork tartan translate?

/k/ + /d/

How does patchwork duvet translate?
How does Selkirk devil translate?
How does Falkirk dagger translate?
How does brickwork depot translate?

/k/ + /s/

How does Falkirk singer translate?
How does Selkirk saga translate?
How does patchwork surface translate?
How does brickwork siphon translate?

/k/ + /z/

How does Falkirk zipper translate?
How does brickwork zester translate?
How does patchwork zebra translate?
How does Selkirk zygote translate?

/k/ + /s/

How does patchwork rigging translate?
How does Falkirk river translate?
How does Selkirk raven translate?
How does brickwork rafter translate?

/g/ + /t/

How does Hamburg tenant translate?
How does Limburg timber translate?
How does Lindberg tactic translate?
How does Strindbergh temper translate?

/g/ + /d/

How does Limburg daisies translate?
How does Hamburg dairy translate?
How does Lindberg diary translate?
How does Strindbergh Danish translate?

/g/ + /s/

How does Hamburg satin translate?
How does Limburg singer translate?
How does Lindberg summon translate?
How does Strindbergh sermon translate?

/g/ + /z/

How does Limburg zombie translate?
How does Hamburg Zulu translate?
How does Strindbergh zenith translate?
How does Lindberg zephyr translate?

/g/ + /r/

How does Limburg relish translate?
How does Strindbergh rigour translate?
How does Hamburg rifle translate?
How does Lindberg rumour translate?

A.2 Supplementary data

Table A.1: Experiment 1: C_1 voicing. Absolute duration of the voiced interval during C_1 closure, release, overall (all in ms) and C_1 overall voicing ratio. Standard deviations in brackets.

C_1C_2	C_1 voicing				N
	Closure	Release	Overall	Ratio overall	
/kt/	21 (13)	1 (3)	22 (14)	.29 (.20)	31
/kd/	23 (10)	2 (6)	25 (15)	.38 (.24)	26
/ks/	21 (13)	0 (0)	21 (13)	.31 (.19)	47
/kz/	41 (16)	9 (9)	51 (21)	.76 (.30)	36
/kr/	22 (11)	0 (0)	22 (11)	.27 (.13)	32
/gt/	24 (6)	1 (4)	25 (7)	.42 (.13)	26
/gd/	33 (8)	11 (13)	43 (19)	.70 (.27)	18
/gs/	26 (8)	0 (0)	26 (8)	.44 (.13)	45
/gz/	43 (8)	13 (8)	56 (11)	.97 (.12)	47
/gr/	36 (11)	7 (9)	42 (15)	.70 (.30)	47

Table A.2: Experiment 1: segmental duration. Durations of V_{1}, C_1 closure, C_1 release, and C_1 overall duration (all in ms). Standard deviations in brackets.

C_1C_2	Duration				N
	V_{1} overall	C_1 closure	C_1 release	C_1 overall	
/kt/	69 (23)	47 (12)	32 (14)	79 (18)	31
/kd/	68 (17)	41 (11)	27 (13)	68 (16)	26
/ks/	71 (21)	54 (13)	16 (7)	70 (14)	47
/kz/	68 (20)	51 (10)	15 (6)	67 (10)	36
/kr/	72 (19)	50 (11)	33 (11)	84 (18)	32
/gt/	93 (22)	37 (10)	25 (12)	63 (14)	26
/gd/	89 (24)	42 (10)	21 (7)	62 (12)	18
/gs/	98 (32)	43 (9)	16 (5)	60 (10)	45
/gz/	100 (27)	44 (8)	14 (9)	58 (12)	47
/gr/	99 (23)	44 (10)	21 (12)	66 (16)	47

Appendix B

Stimuli and supplementary data for experiment 2

B.1 Stimuli

/**k**/ + /**t**/

A bak találkozott az egérrel
A szék támaszkodik a falnak
A vak találkozott az orvossal
A lak tönkrement a földrengés miatt
A rák találkozott a hallal
A pék találkozott a tanárral

/**k**/ + /**d**/

A rák dolgozott a tenger alatt
A vak darabolta a húst
A szék datálódik a harmincas évekből
A vak darabolta a húst
A bak dalolt az erdőben
A pék darabolta a tésztát

/**k**/ + /**s**/

A pék szagolta a levegőt
A szék szabályozza a testtartásomat
A bak szagolta a levegőt
A vak szagolta a levegőt
A lak szaglik mint a disznóól
A rák szedi a kis köveket

/**k**/ + /**z**/

A rák zabálta a kis halacskát
A vak zabálta a tortát
A bak zabálta a virágokat
A lak zárva van
A pék zabálta a kenyeret
Egy szék zuhant ki az erkélyről

/**k**/ + /**L**/

A lak lángban állt a falu közepén
A vak lakott a nagymamájával
A szék lángolt a tűzben
A bak lakott az erdőben
A rák lakik a tenger alatt
Egy pék lakott az ötödik emeleten

/**g**/ + /**t**/

A tag találkozott a főnökkel
Az ág támaszkodott a falnak
A jog tagadja a lehetőséget
A vég távolodik, nem közeledik
A szag távolodik, szerencsére
A cég tagadta a bűnösségét

/**g**/ + /**d**/

A vég dominál minket
A jog datálódik a második háborúból
A tag dalolta az új slágert
Az ág dobálta a leveleket a viharban
A szag dagad a szobában
A cég datálódik a húszas évekből

/**g**/ + /**s**/

A jog szabadította ki a vádlottat
A tag szagolta a levegőt
A vég széttépi múltadat
A szag szétterjed a szobákon át
Az ág szakadt el a fától

/g/ + /z/

A vég zavarba hozta a közönséget
A szag zúdult a szobába
Az ág zavarta a kilátást
A jog zavarba hozta a jogászokat
A tag zabálta a kenyeret
A cég zárta az épületet

/g/ + /L/

A jog lapult egy középkori traktátusban
A tag lakott a negyedik emeleten
A vég láthatóvá válik a szemhatáron
A cég lassult a gyenge piac miatt
A szag lágyul idővel
Az ág lángolt az éjszakában

/ʃ/ + /t/

A sas találkozott a sirállyal
A hús túlterheli a mérleget
A kos találkozott a bakkal
A kés tompult, minél többet használta a mészáros

/ʃ/ + /d/

A kés dolgozik a mészáros kezében
A hús duzzad a melegben
A sas darabolta a nyúlnak a húsát
A kos dolgozott a mezőn

/ʃ/ + /s/

A hús szaglik mintha rothadt lenne
A sas szárnyalt a fák fölött
A kos származik Egyiptomból
A kés szakítja a húst a csontról

/ʃ/ + /z/

A kés zengett, mikor leesett a kövekre az asztalról
A hús zuhant rá a mészáros táblájára
A kos zabálta a szénát
A sas zuhant le a földre

/ʃ/ + /L/

A kés lakik az evőeszközökkel
A sas rágta a húst
A hús lehűl a tányéron
A kos lakik a mezőn

/ʒ/ + /t/

A rozs terem a mezőn
A rúzs tart egész nap
A pajzs tágult a fémmunkás ütései alatt
Rizs terem a mezőkön

/ʒ/ + /d/

A rozs dohosodik a magtárban
A rúzs díszíti a szájat
A rizs dohosodik a zsákban
A pajzs datálódik a vikingkorból

/ʒ/ + /s/

A rizs szárad az aszály miatt
A pajzs széttörött a kezében
A rúzs szépíti az arcot
A rozs szárad az aszály miatt

/ʒ/ + /z/

A rizs zöldül a mezőn
A rúzs zöldíti az arcot
A pajzs zörgött, mikor leesett a kövekre
A rozs zöldül a mezőn

/ʒ/ + /L/

A rizs lelapult a monszunban
A rozs lelapult a viharban
A pajzs lángolt a harcban
A rúzs loccsant a vízben, mikor kiesett a retikülömből

B.2 Supplementary data

Table B.1: Experiment 2: duration of /k, ɡ/ across C_2 contexts. Absolute duration of the voiced interval during C_1 closure, release, overall (all in ms) and C_1 overall voicing ratio. Standard deviations in brackets.

C_1C_2	C_1 voicing				N
	Closure	Release	Overall	Ratio overall	
/kt/	27 (8)	0 (2)	27 (8)	.32 (.12)	64
/kd/	38 (10)	15 (13)	53 (18)	.69 (.29)	62
/ks/	28 (7)	0 (0)	28 (7)	.39 (.13)	66
/kz/	36 (11)	10 (13)	46 (17)	.68 (.31)	63
/kL/	31 (8)	1 (3)	32 (9)	.30 (.11)	66
/ɡt/	30 (11)	0 (3)	31 (12)	.35 (.13)	71
/ɡd/	45 (11)	25 (8)	70 (14)	.96 (.12)	67
/ɡs/	31 (11)	0 (1)	31 (11)	.48 (.15)	70
/ɡz/	44 (10)	20 (11)	64 (15)	.96 (.13)	72
/ɡL/	47 (12)	18 (11)	65 (15)	.90 (.17)	70

Table B.2: Experiment 2: duration of /k, ɡ/ across C_2 contexts. Duration of C_1 closure, release, and overall duration (all in ms). Standard deviations in brackets.

C_1C_2	C_1 duration			N
	Closure	Release	Overall	
/kt/	55 (13)	34 (12)	89 (18)	64
/kd/	55 (16)	29 (11)	83 (20)	62
/ks/	54 (11)	18 (8)	73 (14)	66
/kz/	55 (17)	21 (13)	76 (27)	63
/kL/	74 (15)	35 (11)	109 (21)	67
/ɡt/	53 (12)	35 (11)	88 (12)	71
/ɡd/	47 (11)	26 (8)	73 (13)	67
/ɡs/	49 (11)	17 (9)	66 (13)	70
/ɡz/	45 (10)	21 (9)	67 (14)	72
/ɡL/	51 (12)	23 (10)	73 (14)	70

Table B.3: Experiment 2: voicing duration and overall duration (both in ms) and voicing ratio of /ʃ, ʒ/ across C_2 contexts. Standard deviations in brackets.

C_1C_2	C_1 property			N
	Voicing duration	Duration	Voicing ratio	
/ʃt/	26 (9)	116 (21)	.23 (.07)	50
/ʃd/	47 (29)	102 (24)	.51 (.35)	45
/ʃL/	23 (7)	136 (21)	.18 (.07)	47
/ʒt/	28 (10)	125 (14)	.23 (.08)	47
/ʒd/	70 (28)	99 (18)	.73 (.29)	46
/ʒL/	23 (7)	106 (17)	.65 (.26)	46

Table B.4: Experiment 2: duration of V_1 (ms) across C_1 * C_2 contexts. Standard deviations in brackets. Note that only cases with a lexically long V_1 are represented here.

C_1C_2	V_1 duration	N	C_1C_2	V_1 duration	N
/kt/	118 (23)	32	/ʃt/	100 (18)	
/kd/	125 (27)	29	/ʃd/	116 (25)	
/ks/	123 (25)	35			
/kz/	121 (32)	33			
/kL/	114 (25)	35	/ʃL/	102 (18)	
/gt/	119 (26)	36	/ʒt/	142 (28)	
/gd/	129 (31)	36	/ʒd/	145 (32)	
/gs/	128 (29)	35			
/gz/	135 (31)	37			
/gL/	139 (33)	35	/ʒL/	144 (38)	

Appendix C

Stimuli and supplementary data for experiment 3

C.1 Stimuli

/ps/ + /p/

Het was een Kaaps pandje dat ze aangeschaft hebben, niet een Kaapse wijngaard
Het was een Kaaps paadje waarover wij liepen, geen Kaapse autoweg
Het was een Kaaps pondje dat zij daarvoor betaalde , geen Kaaps tientje
Dat is een Kaaps paaltje waar we net tegenaan reden, niet een Kaapse wegwijzer
Dat is een Kaaps peultje dat op je bord ligt, Geen Kaapse spercieboon
Dat waren Jaaps peren die ik weggooide, niet zijn appels
Het was Jaaps pelgrim die in haar huis logeerde, niet z’n opa
Dat was Jaaps parel die in die ring gezet is, niet zijn diamant
Het was Jaaps paling die de kat opslokte, niet zijn parkiet
Het is Jaaps penning die op de bodem ligt, niet zijn medaille

/ps/ + /t/

Dat is een Kaaps tentje dat daar in de struiken hangt, geen Kaapse dweil
Het is een Kaaps tientje dat daar wegwaait, geen Kaaps geeltje
Het was een Kaaps tintje in dat schilderij, geen Kaaps kleurenschema
Het was een Kaaps tempo waarmee dat allemaal gebeurde, niet een Kaapse mentaliteit
Het is een Kaaps toontje dat ze gebruikt, geen Kaaps wijsje
Het zijn Jaaps termen die in die afspraak vastliggen, niet zijn wensen
Het is Jaaps tafel waaraan hij zo gehecht is, niet zijn schommelstoel
Dat was Jaaps tijger die daar tussen de bomen sloop, niet zijn luipaard
Het was Jaaps tunnel die onder water stond, niet zijn kelder
Het zijn Jaaps taarten die daar zijn blijven staan, niet z’n tassen

/ps/ + /b/

Dat is een Kaaps boompje dat daar groeit, niet een Kaapse struik
Het is een Kaaps bankje waarop hij zit, niet een Kaapse stoel
Dat was een Kaaps baantje dat hij misliep, niet een Kaaps huisje
Dat is een Kaaps beestje wat dat woord betekent, niet een Kaapse wijn
Het is een Kaaps beekje waaruit dat water komt, niet een Kaapse bron
Dat was Jaaps berging die hij verkocht heeft, niet zijn woning
Dat was Jaaps bende die de brand veroorzaakte, niet zijn groepje
Het zijn Jaaps benen die hij gebroken heeft, niet zijn armen
Dat is Jaaps balie die bij de vuilnis staat, niet zijn werkbank
Het zijn Jaaps bullen die daar hangen, niet zijn diploma’s

/ps/ + /d/

Dat is een Kaaps deeltje dat ze ondekt hebben, geen Kaaps beeldje
Het was een Kaaps dansje dat ze uitvoerden, geen Kaaps toneelstuk
Het was een Kaaps dijkje dat doorbrak, geen Kaapse stormvloedkering
Het is een Kaaps deuntje dat hij fluit, niet het Kaapse volkslied
Het is een Kaaps duintje waarop wij daar zitten, geen Kaapse heuvel
Het was Jaaps dame die het spel besliste, niet zijn koning
Het is Jaaps daalder die zij verloren heeft, niet zijn dukaat
Dat is Jaaps deken daar op het bed, niet zijn sprei
Dat is Jaaps dadel die overgebleven is, niet zijn vijg
Het was Jaaps demo die de mist inging, niet zijn lezing

/ps/ + /m/

Het was een Kaaps meertje waarin het monster woonde, niet een Kaapse binnenzee
Het is Kaaps marmer waarin dat gebouw opgetrokken is, geen Kaapse zandsteen
Het was een Kaaps mandje waarin het fruit was uitgestald, geen Kaaps bakje
Het was een Kaaps meisje dat de hoofdprijs won, niet een Kaaps jongetje
Het was een Kaaps monster dat zich in het meer verschool, niet een Kaapse witte haai
Het waren Jaaps maanden die hij nog moest doorbrengen, niet z’n weken
Het was Jaaps mummie die langzaam tot stof verging, het waren niet z’n fossielen
Het is Jaaps molen waar dat meel vandaan komt, niet z’n bakkerij
Het was Jaaps manie die hem nachten wakker hield, niet z’n verslaving
Het is Jaaps marmer in de muren van dat gebouw, het is niet z’n kalksteen

/ps/ + /h/

Het is een Kaaps hulpje dat daar nu werkt, geen Kaapse dienstbode
Het was een Kaaps heertje dat gisteren op bezoek kwam, geen Kaapse dame
Het is een Kaaps huidje dat zo gevoelig is, niet de Kaapse haren
Het is een Kaaps huisje dat al weken te koop staat, niet een Kaapse bungalow
Het was een Kaaps hulsje waarin dat verpakt zat, geen Kaaps papiertje
Het was Jaaps hamer waarmee ze het werk klaarden, niet z’n zaag
Het zijn Jaaps haaien die de visstand bedreigen, niet z’n zeehonden
Het is Jaaps hanger die daar in de kast hangt, niet z’n schoenlepel
Het is Jaaps harnas dat tentoongesteld wordt, het zijn niet z’n zwaarden
Het zijn Jaaps handen waarop hij zuinig is, niet z’n schouders

/ps/ + /V/

Het was een Kaaps euvel dat het mechanisme kwelde, niet een Kaapse monteur
Het was een Kaaps orgel dat daar stond te jengelen, niet een Kaapse fluitspeler
Het is een Kaaps ijsje dat zo zoet smaakt, niet de Kaapse bonbons
Het was een Kaaps epos dat voorgedragen werd, niet een Kaaps sprookje
Het is een Kaaps anker waaraan dat schip ligt afgemeerd, geen Kaapse kade
Het was Jaaps ijzer dat wij omsmolten, niet z’n koper
Het was Jaaps angel die verwijderd moest worden, niet z’n steenpuist
Het is Jaaps emmer die we nodig hebben, niet z’n tuinslang
Het is Jaaps enkel die in het verband zit, niet z’n rechterpols
Het waren Jaaps armen die het zware werk deden, niet z’n benen

C.2 Supplementary data

Table C.1: Experiment 3: voicing duration of C_1 and C_2 separately, voicing duration of C_1 and C_2 combined (all in ms), and voicing ratio of C_1 and C_2 combined across C_3 environments. Standard deviations in brackets. *P* = [+tense] plosive (/p/ or /t/); *B* = [-tense] plosive (/b/ or /d/).

$C_1C_2C_3$	C_1 voicing	C_2 voicing	C_1C_2 voicing	C_1C_2 voicing ratio	N
[psP]	20 (14)	1 (4)	21 (15)	.16 (.13)	234
[psB]	32 (16)	16 (24)	47 (34)	.42 (.31)	233
[psm]	29 (15)	5 (11)	34 (21)	.28 (.19)	114
[psh]	22 (15)	1 (4)	23 (15)	.19 (.13)	118
[psʔ]	22 (13)	0 (2)	22 (14)	.20 (.13)	110

Table C.2: Experiment 3: V_1 duration, and duration of C_1 and C_2 separately and combined (all in ms). Standard deviations in brackets. *P* = [+tense] plosive (/p/ or /t/); *B* = [-tense] plosive (/b/ or /d/).

$C_1C_2C_3$	V_1 duration	C_1 duration	C_2 duration	C_1C_2 duration	N
[psP]	91 (31)	57 (13)	85 (22)	142 (30)	234
[psB]	93 (30)	53 (13)	67 (21)	120 (29)	233
[psm]	91 (24)	51 (13)	79 (20)	129 (26)	114
[psh]	98 (29)	51 (12)	76 (20)	127 (27)	118
[psʔ]	93 (28)	49 (11)	65 (15)	114 (21)	110

Bibliography

Abramson, A. & L. Lisker (1972) Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics* 1: 1-8.

Andersen, H. (1972) Diphthongization. *Language* 48: 11-50.

Anderson, J. & C. Ewen (1987) *Principles of Dependency Phonology.* Cambridge: CUP.

Anderson, S. (1981) Why Phonology isn't ”natural”. *Linguistic Inquiry* 12: 493-539.

Archangeli, D. & D. Pulleyblank (1994) *Grounded Phonology* Cambridge MA: MIT Press.

Laryngeal dimensions, completion, and enhancement. In T. Hall (ed.) *Distinctive Feature Theory.* Berlin: Mouton de Gruyter.

Aylett, M. (2000) *Stochastic Suprasegmentals: Relationships between Redundancy, Prosodic Structure and Care of Articulation in Spontaneous Speech.* PhD dissertation, University of Edinburgh.

Bach, E. & R. Harms (1969) How do languages get crazy rules? In R. Stockwell & K. Macaulay (eds.) *Linguistic change and generative theory.* Bloomington: Indiana University Press.

Baer, T. (1975) *Investigation of phonation using excised larynxes.* PhD dissertation, MIT.

Baitchura, U. (1975) The sound structure of the Turkic languages in connection with that of the Finno-Ugric ones. *Central Asiatic Journal* 19:241-263.

Balise, R. & R. Diehl (1994) Some distributional facts about fricatives and a perceptual explanation. *Phonetica* 51: 99-110.

Barry, M. (1992) Palatalisation, assimilation and gestural weakening in connected speech. *Speech Communication* 11:393-400.

Barry, M. & R. Teifour (1999) Temporal patterns in Arabic voicing assimilation. *Proceedings of the XIVth International Congress of Phonetic Sciences* (San Francisco) Vol 3: 2429-2432.

Baumann, M. (1995) *The production of syllables in connected speech.* PhD dissertation, University of Nijmegen.

Beckman, J. (1996) *Positional Faithfulness.* PhD dissertation, University of Massachusetts, Amherst.

Beckman, J. (1997) positional faithfulness, positional neutralisation and Shona height harmony. *Phonology* 14: 1-46.

Belasco, S. (1953) The influence of articulation of consonants on vowel duration. *Journal of the Acoustical Society of America* 25: 1015-1016.

Benua, L. (1995) Identity effects in morphological truncation. In J. Beckman, L. Walsh Dickey & S. Urbanczyk (eds.) *University of Massachusetts Occasional Papers 18: Papers in Optimality Theory.* Amherst MA: GSLA.

Berendsen, E. (1983) Final devoicing, assimilation, and subject clitics in Dutch. In H. Bennis & W. van Lessen Kloeke (eds.) *Linguistics in the Netherlands 1993.* Dordrecht: Foris.

Bhat, D. & M. Ningomba (1997) *Manipuri Grammar.* München: Lincom.

Birnbaum, S. (1979) *Yiddish. A Survey and a Grammar.* Manchester: Manchester University Press.

Bladon, A. (1986) Phonetics for hearers. In G. McGregor (ed.) *Language for hearers.* Oxford: Pergamon.

Blevins, J. (1995) The syllable in phonological theory. In J. Goldsmith (ed.) *The Handbook of Phonological Theory*. London: Blackwell.

Blevins, J. (to appear) *Evolutionary Phonology. The Emergence of Sound Patterns.* Cambridge: CUP.

Boersma, P. (1998) *Functional Phonology* PhD dissertation, University of Amsterdam.

Bolognesi, R. (1998) *The Phonology of Campidanian Sardinian: A Unitary Account of Self-Organizing Structure.* PhD dissertation, University of Amsterdam.

Booij, G. (1981) *Generatieve Fonologie van het Nederlands.* Utrecht: Spectrum.

Booij, G. (1995) *The Phonology of Dutch.* Oxford: Clarendon.

Borowsky, T. (2000) Word faithfulness and the direction of assimilations. *The Linguistic Review* 17: 1-28.

Bradlow, A. (1995) A comparative study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97: 1916-1924.

Brink, D. (1975) Voice assimilation in Dutch: some refinements. *Acta Linguistica Hafniensa* 16: 11-19.

Briscoe, E. (2000) Grammatical acquisition: inductive bias and coevolution of language and the language acquisition device. *Language* 76: 245-296.

Brockhaus, W. (1995) *Final devoicing in the phonology of German* (Linguistische Arbeiten 336). Tübingen: Niemeyer.

Bromberger, S. & M. Halle (1989) Why phonology is different. *Linguistic Inquiry* 20: 51-70.

Browman, C. & L. Goldstein (1986) Towards an articulatory phonology. *Phonology Yearbook* 3: 219-252.

Browman, C. & L. Goldstein (1990) Tiers in Articulatory Phonology, with some implications for casual speech. In J. Kingston & M. Beckman (eds.) *Papers in Laboratory Phonology 1. Between the Grammar and the Physics of Speech.* Cambridge: CUP.

Browman, C. & L. Goldstein (1992) Articulatory Phonology: an overview. *Phonetica* 49: 155-180.

Burton, M & K. Robblee (1997) A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics* 25: 97-114.

Busá, M. & J. Ohala (1999) In search of the perceptual correlates of vowel harmony. *Proceedings of the XIVth International Congress of Phonetic Sciences* (San Francisco) Vol 1: 357-360.

Butskhridze, M. (1998) *Typology of Fixed Accent.* English summary of a Tbilisi State University dissertation in Georgian.

Bybee, J. (2001) *Phonology and Language Use.* Cambridge: CUP.

Byrd, D. (1996a) A phase window framework for articulatory timing. *Phonology* 13: 139-169.

Byrd, D. & E. Saltzman (1998) Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26: 173-199.

Byrd, D. & C. Tan (1996) Saying consonant clusters quickly. *Journal of Phonetics* 24: 263-282.

Caisse, M. (1982) *Cross-linguistic differences in fundamental frequency perturbations induced by voiceless unaspirated stops.* MA dissertation, UC Berkeley.

Camminga, J. & P. van Reenen(1980) Review of Trommelen, M. & W. Zonneveld (1980) *Inleiding in de Generatieve Fonologie. Glot* 3: 183-189.

Chang, K. & B. Shefts (1964) *A Manual of Spoken Tibetan (Lhasa Dialect).* Seattle: University of Washington Press.

Charles-Luce, J. (1985) Word final devoicing in German: effects of phonetic and sentential contexts. *Journal of Phonetics* 13: 309-324.

Charles-Luce, J. (1993) The effects of semantic context on voicing neutralisation. *Phonetica* 50: 28-43.

Chen, M. (1970) Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22: 129-159.

Chistovich, L. & V. Lublinskaya (1979) The center of gravity effect in vowel spectra and critical distance between formants. *Hearing Research* 1: 185-195.

Cho, T. & P. Ladefoged (1997) Variations and universals in VOT: evidence from 17 endangered languages. *UCLA Working Papers in Phonetics* 94: 18-40

Cho, T. & P. Ladefoged (1999) Variations and universals in VOT: evidence from 18 endangered languages. *Journal of Phonetics* 27: 207-222

Cho, Y-M. (1990a/1999) *Parameters of Consonantal Assimilation.* PhD dissertation, Stanford University, and Munich: Lincom.

Cho, Y-M. (1990b) A typology of voicing assimilation. *Proceedings of the ninth West Coast Conference on Formal Linguistics.* Stanford: CSLI.

Chomsky, N. & M. Halle (1968) *The Sound Pattern of English.* New York: Harper.

Choo, W. & M. Huckvale (1998) Spatial relationships in fricative perception. *UCL Working Papers in Linguistics* 10.

Cohen, A., K. Ebeling & A. van Holk (1972) *Fonologie van het Nederlands en het Fries.* Den Haag: Nijhoff.

Clark, J. & C. Yallop (1995) *An Introduction to Phonetics and Phonology.* 2nd ed. London: Blackwell.

Clark, L. (1998) *Turkmen reference grammar* (Turcologica 34). Wiesbaden: Harrassowitz.

Clements, N. & E. Hume (1995) The internal organization of speech sounds. In J. Goldsmith (ed.) *The Handbook of Phonological Theory*. London: Blackwell.

Coleman, J. (1992) The phonetic interpretation of headed phonological structures containing overlapping constituents. *Phonology* 9: 1-44.

Crystal, T. & A. House (1982) Segmental durations in connected speech: preliminary results. *Journal of the Acoustical Society of America* 72: 705-716.

Crystal, T. & A. House (1988a) Segmental durations in connected speech: current results. *Journal of the Acoustical Society of America* 83: 1553-1573.

Crystal, T. & A. House (1988b) Segmental durations in connected-speech signals: syllabic stress. *Journal of the Acoustical Society of America* 83: 1574-1585.

Crystal, T. & A. House (1988c) The duration of English stop consonants: an overview. *Journal of Phonetics* 16: 285-294.

Dauer, R. (1983) Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11: 51-62.

Dawkins, R. (1988) *The Blind Watchmaker.* London: Penguin (first published by Longman, 1986).

de Boer, B. (1999) *Self-organization in vowel systems.* PhD dissertation, Free University of Brussels.

de Boer, B. (2001) *The Origins of Vowel Systems.* Oxford: Oxford University Press.

De Jong, K., M. Beckman & J. Edwards (1992) The interplay between prosodic structure and coarticulation. *Language and Speech* 36: 127-212.

De Jong, K. (1995) The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97: 491-504.

Denes, P. (1955) Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27: 761-764.

Dinnsen, D. & J. Charles-Luce (1984) Phonological neutralization, phonetic implementation, and individual differences. *Journal of Phonetics* 12: 49-60.

Disner, S. (1983) Vowel quality: The relation between universal and language-specific factors. *UCLA Working Papers in Phonetics* 58: .

Docherty, G. (1992) *The timing of voicing in British English obstruents.* Dordrecht: Foris.

Docherty, G. & P. Foulkes (1999) Derby and Newcastle: instrumental phonetics and variationist studies. In P. Foulkes & G. Docherty (eds.) *Urban Voices.* London: Arnold.

Downing, L. & B. Gick (2001) Voiceless tone depressors in Nambya and Botswana Kalang'a. In *Proceedings of the 27th meeting of the Berkeley Linguistics Society.*

Drosdowski, G. & P. Eisenberg (1995) *Duden: Grammatik der Deutschen Gegenwartssprache.* 5th ed. Mannheim: Duden.

Elman, J. & J. McClelland (1986) Exploiting lawful variability in the speech wave. In J. Perkell & D. Klatt (eds.) *Invariance and Variability in Speech Processes.* Hillsdale, NJ: Erlbaum.

Engstrand, O. (1998) Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *Journal of the Acoustical Society of America* 83: 1863-1875.

Ernestus, M. (2000) *Voice assimilation and Segment Reduction in casual Dutch.* PhD dissertation, Free University of Amsterdam/Holland Institute of Generative Linguistics.

Ernestus, M. & H. Baayen (2003) Intraparadigmatic effects on the perception of voice. Ms., Max Planck Institute for Psycholinguistics, Nijmegen.

Farnetani, E. (1997) Coarticulation and connected speech processes. In W. Hardcastle & J. Laver (eds.) *The Handbook of Phonetic Sciences.* London: Blackwell.

Fischer-Jørgensen E. (1954) Acoustic analysis of stop consonants. *Miscellanea Phonetica II.* London.

Fischer-Jørgensen E. (1968) Voicing, tenseness and aspiration in stop consonants, with special reference to French and Danish. *Annual Report of the Institute of Phonetics, University of Copenhagen* 3:63-114.

Flege, J. (1982) Laryngeal timing and phonation onset in utterance-initial stops. *Journal of Phonetics* 10: 177-192.

Flege, J. & J. Hillenbrand (1987) A differential effect of release bursts on the stop voicing judgments of native French and English listeners. *Journal of Phonetics* 15: 203-208.

Flemming, E. (1995) *Auditory features in phonology*. PhD dissertation, UCLA.

Flemming, E. (2001) Scalar and Categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7-44.

Forrez, G. (1966) Relevante parameters van de stemhebbende fricatief /z/. *Instituut voor Perceptie Onderzoek, Verslag 55*

Foley, J. (1977) *Foundations of Theoretical Phonology.* Cambridge: CUP.

Fougeron, C. (1999) Articulatory properties of initial segments in several constituents in French. *UCLA Working Papers in Phonetics* 97: 74-99.

Fourakis, M. & G. Iverson (1984) On the 'incomplete neutralization' of German final obstruents *Phonetica* 41: 140-149.

Fowler, C. (1980) Coarticulation and theories of extrinsic timing *Journal of Phonetics* 8: 113-133.

Fowler, K. (1981) Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 46: 127-139.

Fox, R. & D. Terbeek (1977) Dental flaps, vowel duration, and rule ordering in English. *Journal of Phonetics* 5: 27-34.

Gaskell, M. & W. & Marslen-Wilson (1996) Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance.* 22: 144-158.

Gaskell, M. & W. & Marslen-Wilson (1998) Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance.* 24: 380-396.

Giegerich, H. (1994) The limits of phonological derivation. Spelling pronunciation and schwa in English. *Linguistische Berichte* 142: 413-436.

Gimson, A. (1994) *Gimson's Pronunciation of English.* 5th ed., revised by A. Cruttenden. London: Arnold.

Gnanadesikan, A. (1997) *Phonology with ternary scales.* PhD dissertation, University of Massachusetts, Amherst, and Ms., Rutgers Optimality Archive.

Gordon, M. (1996) The effects of stress and prosodic phrasing on duration, acoustic amplitude and airflow of nasals in Estonian. *UCLA Working Papers in Phonetics* 92: 151-159.

Gósy, M. (1999) Predictability of voicing assimilation in speech production. *Proceedings of the XIVth International Congress of Phonetic Sciences* (San Francisco) Vol 3: 1905-1908

Gósy, M. (2000) A [p, t, k] mássalhangzók zöngekezdési ideje. *Magyar Nyelvor* 124: 195-203

Greenberg, J. (1978) Some generalizations concerning initial and final consonant clusters. In J. Greenberg (ed.) *Universals of Human Language.* Vol. 2: *Phonology.* Stanford: Stanford UP.

Grijzenhout, J., & M. Krämer (1998) *Final devoicing and voicing assimilation in Dutch derivation and cliticization.* Ms., Rutgers Optimality Archive.

Gussenhoven, C. (1986) Over de fonologie van de Nederlandse clitica. *Spektator* 15: 180-200.

Gussenhoven, C. (1996) *Mooie Bedoelingen. Over de rol van de fonetiek in de fonologie.* Inaugural, University of Nijmegen.

Gussmann, E. (1992) Resyllabification and delinking: the case of Polish voicing. *Linguistic Inquiry* 23: 29-56.

Gussmann, E. (1999) Preaspiration in Icelandic: unity in diversity. *Anglica Wratislaviensa* 35 (*Acta Universitatis Wratislaviensis* 2127): 163-181.

Haggard, M. (1978) The devoicing of voiced fricatives. *Journal of Phonetics* 6: 95-102.

Haggard, M., S. Ambler & M. Callow (1970) Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47: 613-617.

Hahn, R. (1998) Yellow Uyghur and Salar. In Johanson, L. & Éva Csató (eds.) *The Turkic Languages.* London: Routledge.

Hale, M. & C. Reiss (2000a) Phonology as cognition. In N. Burton-Roberts, P. Carr & G. Docherty (eds.) *Phonological Knowledge. Conceptual and Empirical Issues.* Oxford: OUP.

Hale, M. & C. Reiss (2000b) Substance abuse and dysfunctionalism: current trends in phonology. *Linguistic Inquiry* 31: 157-169.

Hall, R. (1944) *Hungarian Grammar.* Language Monograph 21. Baltimore: Linguistic Society of America.

Halle, M. (1975) Confessio Grammatici. *Language* 51: 525-535.

Halle, M., G. Hughes & J. Radley (1957) Acoustic properties of stop consonants. *Journal of the Acoustical Society of America* 29: 107-116.

Halle, M. & K. Stevens (1971) A note on laryngeal features. *Quarterly Progress Report, Research Laboratory of Electronics, MIT* 101: 198-211.

Halliday, M. (1960) Categories in the theory of grammar. *Word* 17: 241-292.

Harms, R. (1973) *Some nonrules of English.* Distributed 1973, Indiana University Linguistics Club.

Harris, J. (1994) *English Sound Structure.* London: Blackwell.

Harris, J. & G. Lindsey (1995) The elements of phonological representation. In J. Durand & F. Katamba (eds.) *Frontiers of Phonology. Atoms, structures, derivations.* London: Longman.

Harris, J. & G. Lindsey (2000) Vowel patterns in mind and sound. In N. Burton-Roberts, P. Carr & G. Docherty (eds.) *Phonological Knowledge. Conceptual and Empirical Issues.* Oxford: OUP.

Hayes, B. (1986) Inalterability in CV phonology. *Language* 62: 321-351.

Hayes, B. (1989) Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20: 253-306.

Hayes, B. (1992) Comment on F. Nolan, the descriptive role of segments. In D. Docherty & D. Ladd (eds.) *Papers in Laboratory Phonology 2. Segment, gesture, prosody.* Cambridge: CUP.

Hayes, B. & T. Stivers (1996) *The phonetics of post-nasal voicing.* Ms., UCLA.

Helgason, P. (1998) On-line preaspiration in Swedish: implications for historical sound change. In *Proceedings of Sound Patterns of Spontaneous Speech 98 (September 24-26, 1998).*

Helgason, P. (1999) Phonetic preconditions for the development of normative preaspiration. In *Proceedings of the 13th International Congress of the Phonetic Sciences..*

Helgason, P. (2001) Convergence in historical phonetics: stop contrasts in Åland and Iceland. *Lund Working Papers in Linguistics* 49: 54-57.

Hellberg, S. (1974) *Graphonomic Rules in Phonology: Studies in the Expression Component of Swedish.* Göteborg: University of Göteborg.

Hellberg, S. (1980) Apparent naturalness in Faroese phonology. *Nordic Journal of Linguistics* 3: 1-24.

Henderson, E. (1952) The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies* 14: 149-174.

Henderson, J. & B. Repp (1982) Is a stop consonant released when followed by another consonant? *Phonetica* 39: 71-82.

Heo, Y. (1994) *Empty categories and Korean phonology.* PhD dissertation, School of Oriental and African Studies, University of London.

Hombert, J.-M., J. Ohala & W. Ewan. (1979) Phonetic explanations for the development of tones. *Language* 55: 37-55.

House, A. & Fairbanks, G. (1953) The influence of consonantal environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25: 105-113.

Hsu, C.-S. (1996) *Laryngeal features in Taiwanese.* Ms., UCLA.

Hsu, C.-S., & S.-A. Jun (1998) Prosodic strengthening in Taiwanese: syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics* 97: 152-156.

Hualde, I. (1991) *Basque Phonology.* London: Routledge.

Hubers, G. & J. Kooij (1973) Voice assimilation in Dutch. *Acta Linguistica Hafniensa* 14: 25-33.

Huffman, M. (1993) Phonetic patterns of nasalization and implications for feature specification. In Huffman, M. & R. Krakow *Nasals, Nasalization, and the Velum* (Phonetics and Phonology 5). San Diego: Academic Press.

Hutters, B. (1985) Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica* 42: 1-24.

Ishizaka, K., & J. Flanagan (1972) Synthesis of voiced sounds from a two-mass model. *Bell System Technical Journal* 51: 1233-1268.

Iverson, G. & J. Salmons (1995) Aspiration and laryngeal representation in Germanic. *Phonology* 12: 369-396.

Iverson, G. & J. Salmons (1999) Glottal spreading bias in Germanic. *Linguistische Berichte* 178: 135-151.

Jacobs, N., E. Prince & J. van der Auwera (1994) Yiddish. In E. König & J. van der Auwera (eds.) *The Germanic Languages.* London: Routledge.

Jansen, W. (2001a) Phonetic voice and phonological assimilation in the Germanic languages. In M. Andronis, C. Ball, H. Elston & S. Neuvel (eds.) *CLS 37: The Panels. Papers from the 37th Meeting of the Chicago Linguistic Society.* Vol. 2. Chicago: Chicago Linguistic Society, 43-59.

Jansen, W. (2001b) Jansen, W. (2001) Review of B. de Boer (2001) *The Origins of Vowel Systems.* Oxford: OUP. Published as *LinguistList* issue 13.541, http://www.linguistlist.org/issues/13/13-541.html

Jansen, W. & Z. Toft (2002) On sounds that like to be paired (after all): an acoustic investigation of Hungarian voicing assimilation. *SOAS Working Papers in Linguistics* 12: 19-52.

Jassem. W., & Richter, L. (1989) Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17: 317-325.

Javkin, H. (1976) The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Phonology Laboratory, UC Berkeley* 1: 78-92.

Jensen, S. (1994) Is ʔ an element? Towards a non-segmental phonology. *SOAS Working Papers in Linguistics* 4: 71-78.

Jessen, M. (1996) The relevance of phonetic reality for underlying phonological representation: the case of tense vs. lax obstruents in German. In U. Kleinhenz (ed.) *Interfaces in Phonology. (Studia Grammatica 41).* Berlin: Akademie Verlag.

Jessen, M. (1998) *Phonetics and Phonology of Tense and Lax Obstruents in German.* Amsterdam: John Benjamins.

Johanson, L. & Éva Csató (eds.) (1998) *The Turkic Languages.* London: Routledge.

Johnson, K., E. Flemming & R. Wright (1993) The hyperspace effect: phonetic targets are hyperarticulated. *Language* 69: 505-528.

Jones, D. (1956) *An Outline of English Phonetics.* 8th ed. Cambridge: Heffer.

Jones, D. (1977) *Everyman's English Pronouncing Dictionary.* 14th ed., revised by A. Gimson. London: Dent

Jones, M. (2001) The historical development of a phonological rarity: postnasalised stops in Lancashire dialects. Paper presented at the *9th Manchester Phonology Meeting*, University of Manchester, May 2001.

Jongman, A., J . Sereno, M. Raaimakers & A. Lahiri (1992) The phonological representation of [voice] in speech perception. *Language and Speech* 35: 137-152.

Jun, S.-A. (1995) Asymmetrical prosodic effects on the laryngeal gesture in Korean. In B. Connell & A. Arvaniti (eds.) *Papers in Laboratory Phonology 4. Phonology and Phonetic Evidence.* Cambridge: CUP.

Junqua, J.-C. (1996) The influence of acoustics on speech production: a noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication* 20: 13-22.

Kahlen-Halstenbach, B. (1990) Zur psychologischen Realität der Auslautverhärtung im Deutschen. *Zeitschrift für phonetische Sprachwissenschaft und Kommunikationsforschung* 43: 645-655.

Kahn, D. (1976) *Syllable-based Generalizations in English Phonology.* Distributed by the Indiana University Linguistics Club, Bloomington, Indiana.

Kaisse, E. (1985) *Connected speech: the interaction between syntax and phonology.* Orlando, Florida: Academic Press.

Kaisse, E. & P. Shaw (1985) On the theory of lexical phonology. *Phonology Yearbook* 2: 1-30.

Kálmán, B. (1972) Hungarian historical phonology. In L. Benkő & S. Imre (eds.) *The Hungarian Language.* The Hague: Mouton

Katz, D. (1987) *Grammar of the Yiddish Language.* London: Duckworth.

Kaye, J. (1989) *Phonology: A Cognitive View.* Hillsdale, NJ: Erlbaum.

Kaye, J. (1992) Do you believe in magic? the story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics.* 2: 293-313.

Kaye, J., J. Lowenstamm & J.-R. Vergnaud (1985) The internal structure of phonological elements: a theory of charm and government. *Phonology Yearbook* 2:305-208

Kaye, J., J. Lowenstamm & J.-R. Vergnaud (1990) Constituent structure and government in phonology. *Phonology* 7: 193-232.

Keating, P. (1984) Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286-319.

Keating, P. (1988) Underspecification in phonetics. *Phonology* 5: 275-292.

Keating, P. (1990a) Phonetic representations in a generative grammar. *Journal of Phonetics* 18: 321-334.

Keating, P. (1990b) The window model of coarticulation: articulatory evidence. In J. Kingston & M. Beckman (eds.) *Papers in Laboratory Phonology 1. Between the Grammar and the Physics of Speech.* Cambridge: CUP.

Keating, P., T. Cho, C. Fougeron & C.-S. Hsu (1998) *Domain-initial articulatory strengthening in four languages.* Ms., (draft of LabPhon 6 paper), UCLA.

Keller, R. (1961) *German dialects :phonology and morphology, with selected texts.* Manchester, Manchester University Press.

Kenesei, I., R. Vago & A. Fenyvesi (1998) *Hungarian.* (Routledge Descriptive Grammar.) London: Routledge.

Kenstowicz, M. (1995) Base-Identity and Uniform Exponence: Alternatives to cyclicity. Ms., Rutgers Optimality Archive.

Kessinger, R. & S. Blumstein (1997) Effects of speaking rate on voice onset time in Thai, French, and English. *Journal of Phonetics* 25: 143-168.

Kingston, J. & R. Diehl (1994) Phonetic knowledge. *Language* 70: 419-454.

Kingston, J. & R. Diehl (1995) Intermediate properties in the perception of distinctive feature values. In B. Connell & A. Arvaniti (eds.) *Papers in Laboratory Phonology IV: Phonology and Phonetic evidence.* Cambridge: CUP.

Kirby, S. (1999) *Function, Selection, and Innateness: The Emergence of Language Universals.* Oxford: OUP.

Kirchner, R. (1998) *An Effort-Based Approach to Consonant Lenition.* PhD dissertation, UCLA, and Rutgers Optimality Archive.

Kiss, J. (ed.) (2001) *Magyar dialektológia.* Budapest: Osiris.

Klatt, D. (1975) Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3: 129-140.

Kluender, K., R. Diehl, & B. Wright (1988) Vowel length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics* 16: 153-169.

Knetschke, E. & M. Sperlbaum (1987) Zur Orthoepie der Plosiva in der deutschen Hochsprache. Eine Auditiv-komparative Untersuchung. Tübingen: Niemeyer.

Kochetov, A. (2003) Positional Markedness as a by-product of the learning situation: a simulation. Paper presented at the *29th meeting of the Berkeley Linguistics Society*, February 2003.

Kohler, K. (1979) Parameters in the production an the perception of plosives in German and French *Arbeitsberichte Institut für Phonetik Kiel* 12: 261-291.

Kohler, K. (1984) Phonetic features in phonology: the feature fortis-lenis. *Phonetica* 41: 150-174.

Kohler, K. (1985) The perception of lenis and fortis fricatives in French. A critical re-evaluation. *Phonetica* 42: 116-123.

Kolmár, J. (1821) *Próbatétel a Magyar Helyesirás Philosophiájára.* In *Jutalom feleletek a' Magyar Nyelvrol, a Magyar Nemzeti Muzeum 1815, 1816, 1817 esztendei kérdéseire.* Pest: Horvát István.

König, E. & J. van der Auwera (eds.) (1994) *The Germanic Languages.* London: Routledge.

Kopkalli, H. (1993) *A Phonetic and Phonological Analysis of Final Devoicing in Turkish.* PhD dissertation, University of Michigan.

Kostic, D. & R. Das (1972) *A Short Outline of Bengali Phonetics.* Calcutta: Indian Statistical Institute.

Kress, B. (1982) *Isländische Grammatik.* München: Huber.

Kristoffersen, G. (2000) *(*The Phonology of Norwegian.) Oxford: Clarendon.

Ladd, D. (1996) *Intonational Phonology.* Cambridge: CUP.

Ladd, D. & Scobbie, J. (forthcoming) External sandhi as gestural overlap? Counter-evidence from Sardinian. In J. Local, R. Ogden, & R. Temple (eds.) *Papers in Laboratory Phonology VI* Cambridge: CUP.

Ladefoged, P. (1973) Features of the larynx. *Journal of Phonetics* 1: 73-83.

Ladefoged, P. & T. Cho (2000) Linking linguistic contrasts to reality: the case of VOT. *UCLA Working Papers in Phonetics* 98: 1-9.

Ladefoged, P. & I. Maddieson (1996) *Sounds of the World's Languages.* Oxford: Blackwell.

Lakoff, G. (1972) The arbitrary basis of transformational grammar. *Language* 48: 76-87.

Lavoie, L. (2001) *Consonant Strength. Phonological Patterns and Phonetic Manifestations.* New York: Garland.

Lehiste, I. (1970) *Suprasegmentals.* Cambridge MA: MIT Press.

Levelt, W. (1989) *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Liberman, A., K. Harris, P. Eimas, L. Lisker & J. Bastian (1961) An effect of learning on speech perception: The discrimination of durations of silence with and without linguistic significance. *Language and Speech* 4: 175-195.

Lieberman, P. (1970) Towards a unified phonetic theory. *Linguistic Inquiry* 1:307-322.

Lieberman, P. (1977) *Speech physiology and acoustic phonetics.* New York: MacMillan.

Lindblom, B. (1963) Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 1773-1781.

Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H & H theory. In W. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modelling.* Dordrecht: Kluwer.

Lisker, L. (1986) ”Voicing” in English: A catalogue of acoustic features signalling /b/ vs. /p/ in trochees. *Language and Speech* 29: 3-11.

Lisker, L. & A. Abramson (1964) A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20: 384-422.

Lisker, L. & A. Abramson (1970) The voicing dimension: some experiments in comparative phonetics. *Proceedings of the 6th International Congress of the Phonetic Sciences.*

Löfqvist, A. (1980) Interarticulator programming in stop production. *Journal of Phonetics.* 8: 475-490.

Löfqvist, A. (1981) Laryngeal activity in Icelandic obstruent production. *Nordic Journal of Linguistics* 4: 1-18.

Löfqvist, A. & H. Yoshioka (1980) Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America* 63: 792-801.

Löfqvist, A. & H. Yoshioka (1984) Intrasegmental timing: laryngeal-oral coordination in voiceless consonant production. *Speech Communication* 3: 279-289.

Lombard, E. (1910) Le signe d de l'élévation de la voix. *Annales Maladies Oreille, Larynx, Nez, Pharynx* 37: 101-119.

Lombardi, L. (1991) *Laryngeal features and laryngeal neutralisation.* PhD dissertation, University of Maryland.

Lombardi, L. (1994) *Laryngeal features and laryngeal neutralisation.* New York: Garland.

Lombardi, L. (1995a) Laryngeal neutralisation and syllable wellformedness. *Natural Language and Linguistic Theory* 13: 39-74.

Lombardi, L. (1995b) Laryngeal features and privativity. *The Linguistic Review* 12: 35-59.

Lombardi, L. (1995c) *Why place and voice are different: constraint interactions and featural faithfulness in Optimality Theory*. Ms., University of Maryland, and Rutgers Optimality Archive.

Lombardi, L. (1996) Postlexical rules and the status of privative features. *Phonology* 13: 1-38.

Lombardi, L. (1997) Restrictions on direction of voice assimilation: an OT account. In *Maryland Working Papers in Linguistics* 4.

Lombardi, L. (1999) Positional faithfulness and voicing assimilation. *Natural Language and Linguistic Theory* 1: 267-302.

Lotz, J. (1972) *Two papers on English-Hungarian Contrastive phonology.* Budapest: Linguistic Institute of the Hungarian Academy of Sciences and Center for Applied Linguistics.

Lotz, J. (1988) *Das Ungarische Sprachsystem.* Bloomington, Indiana: Eurolingua.

Luce, P. & J. Charles-Luce (1985) Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America* 78: 1949-1957.

Lyons, R. (1982) A computational model of filtering, detection, and compression in the cochlea. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing.*

Mack, M. (1982) Voicing-dependent vowel duration in English and French: monolingual and bilingual production. *Journal of the Acoustical Society of America* 71: 171-178.

Maddieson, I. (1984) *Patterns of Sounds.* Cambridge: CUP.

Magloire, J. & K. Green (1999) A cross-language comparison of speaking rate effects on the production of Voice Onset Time in English and Spanish. *Phonetica* 56: 158-185.

Marr, D. (1982) *Vision.* San Francisco: Freeman.

Mascaró, J. (1987/1995) A reduction and spreading theory of voicing and other sound effects. Circulated as ms., later published as *Catalan Working Papers in Linguistics* 4: 267-328

Mascaró, J. & L. Wetzels (2001) The typology of voicing and devoicing. *Language* 77: 207-244.

Massaro, D. (1998) *Perceiving Talking Faces. From Speech Perception to a Behavioral Principle.* Cambridge, MA: MIT Press

Massaro, D. & M. Cohen (1983) Consonant/vowel ratio: an improbable cue in speech perception. *Perception and Psychophysics* 33: 502-505.

McCarthy, J. & A. Prince (1995) Faithfulness and reduplicative identity. In J. Beckman, L. Walsh Dickey & S. Urbanczyk (eds.) *University of Massachusetts Occasional Papers 18: Papers in Optimality Theory.* Amherst MA: GSLA.

Menert, L. (1994) *Experiments on voice assimilation in Dutch: Prosodic structures and tempo.* PhD Thesis, University of Utrecht.

A. Mester & Îto, J. (1989) Feature predictability and underspecification: palatal prosody in Japanese mimetics. *Language* 65: 258-293.

Meyer, E. & Gombocz, Z. (1909) Zur Phonetik der ungarischen Sprache. In *Le Monde Oriental 1907-1908* (published 1909), 122-187.

Mielke, J. (2001) Explaining directional asymmetry in Turkish [h] deletion: a crosslinguistic study of perceptibility. *OSU Working Papers in Linguistics* 55: 117-171.

Miller, G. & Nicely, P. (1955) Analysis of perceptual confusions among some English consonants.. *Journal of the Acoustical Society of America* 27: 338-353.

Miller, J., K. Green & A. Reeves (1986) Speaking rate and segments: a look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43: 106-115.

Mitleb, F. (1981) *Segmental and non-segmental structure in phonetics: evidence from foreign accent.* PhD dissertation, University of Indiana.

Mohanan, K. (1986) *The Theory of Lexical Phonology.* Dordrecht: Reidel.

Moon, S.-J., & B. Lindblom (1994). Interaction between duration, context, and specking style in English stressed vowels. *Journal of the Acoustical Society of America* 96: 40-55.

Moulton, W. (1962) *The Sounds of English and German.* Chicago: University of Chicago Press.

Myers, S. (2000) Boundary disputes: the distinction between phonetic and phonological sound patterns. In N. Burton-Roberts, P. Carr & G. Docherty (eds.) *Phonological Knowledge. Conceptual and Empirical Issues.* Oxford: OUP, 245-272.

Myers, S. (2002) Gaps in factorial typology: The case of voicing in consonant clusters. Ms., Rutgers Optimality Archive.

Nagajara, K. (1990) *Khasi Phonetic Reader.* CIIL phonetic readers 29. Mysore: Central Institute of Indian Languages.

Nespor, M. & I. Vogel (1986) *Prosodic Phonology.* Dordrecht: Foris.

Nolan, F. (1992) The descriptive role of segments: evidence from assimilation.

In D. Docherty & D. Ladd (eds.) *Papers in Laboratory Phonology II. Segment, gesture, prosody.* Cambridge: CUP.

Noss, R. (1964) *Thai Reference Grammar.* Washington, DC: Foreign Service Institute.

O'Dell, M. & R. Port (1983) Discrimination of word-final voicing in German. *Journal of the Acoustical Society of America* 73 supplement 1.

Ohala, J. (1981) The listener as a source of sound change. In C. Masek, R. Hendrik & M Miller (eds.) *Papers from the Parasession on Language and Behaviour. (CLS 17)* Chicago: Chicago Linguistics Society.

Ohala, J. (1983) The origin of sound patterns in vocal tract constraints. In P. MacNeilage (ed.)(1983) *The Production of Speech.* New York: Springer.

Ohala, J. (1993) The phonetics of sound change. In C. Jones (ed.) *Historical Linguistics: Problems and Perspectives.* London: Longman.

Ohde, R. (1984) Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustic Society of America* 75: 224-230.

Öhman, S. (1962) *On the perception of Swedish consonants in intervocalic position.* Report No. 25, Royal Institute of Technology, Stockholm.

Olsson, M. (1992) *Hungarian phonology and morphology.* Lund: Lund University Press.

Orton, H. & W. Halliday (eds.) (1963) *Survey of English Dialects.* Vol. 1: *The Six Northern Counties and the Isle of Man.* Leeds: Arnold.

O'Shaughnessy, D. (1981) A study of French vowel and consonant durations. *Journal of Phonetics* 9: 385-406.

Panzer, C. (1981) *A Generative Phonology of Danish.* PhD dissertation, City University of New York.

Parker, E., R. Diehl & K. Kluender (1986) Trading relations in speech and nonspeech. *Perception and Psychophysics* 39: 129-142.

Parker, F. (1981) A functional-perceptual account of final devoicing. *Journal of Phonetics* 9: 129-137.

Pater, J. (1996/1999) Austronesian nasal substitution and other NC effects. Ms., Rutgers Optimality archive. Published in R. Kager, H. van der Hulst & W. Zonneveld (eds.) (1999) *The Prosody-Morphology Interface.* Cambridge: CUP.

Peterson, G. & H. Barney (1952) Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24: 175-184.

Perkell, J. (1969) *Physiology of speech production: results and implications of a quantitative cineradiographic study.* Cambridge, MA: MIT Press.

Perkell, J., M. Matthies, M. Svirsky & M. Jordan (1995) Goal-based speech motor control: a theoretical framework and some preliminary data. *Journal of Phonetics* 23: 23-35.

Perlmutter, D. (1995) Phonological quantity and multiple association. In J. Goldsmith (ed.) *The Handbook of Phonological Theory.* Oxford: Blackwell.

Pierrehumbert, J. & M. Beckman (1988) *Japanese Tone Structure.* Cambridge MA: MIT Press.

Pierrehumbert, J., M. Beckman & D. Ladd (2000) Conceptual foundations of phonology as a laboratory science. In N. Burton-Roberts, P. Carr & G. Docherty (eds.) *Phonological Knowledge. Conceptual and Empirical Issues.* Oxford: OUP, 273-304.

Pierrehumbert, J. & D. Talkin (1992) Lenition of /h/ and glottal stop. In D. Docherty & D. Ladd (eds.) *Papers in Laboratory Phonology 2. Segment, gesture, prosody.* Cambridge: CUP.

Pind, J. (1995) Speaking rate, voice onset time and quantity: the search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics* 57: 291-304.

Piroth, H., L. Schiefer, P. Janker & B. Johne (1991) Evidence for final devoicing in German? An experimental investigation. In *Proceedings of the 12th International Congress of the Phonetic Sciences.*

Ploch, S. (1999) *Nasals on My Mind. The Phonetic and the Cognitive Approach to the Phonology of Nasality.* PhD dissertation, School of Oriental and African Studies, University of London.

Polgárdi, K. (1998) *Vowel Harmony. An Account in Terms of Government and Optimality.* PhD dissertation, University of Leiden.

Port, R. (1996) The discreteness of phonetic elements: response to A. Manaster-Ramer. *Journal of Phonetics* 24: 491-511.

Port, R. & Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics*32: 141-152.

Port, R. & Crawford, P. (1989) Pragmatic effects on neutralization rules. *Journal of Phonetics* 16: 257-282.

Port, R. & M. O'Dell (1985) Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13:455-471.

Port, R. F. Mitleb, M. O'Dell (1981) Neutralization of obstruent voicing in German is incomplete. *Journal of the Acoustical Society of America* 70 supplement 1.

Prince, A. & P. Smolensky (1993) *Optimality Theory: constraint interaction in generative grammar.* Ms., Rutgers Universty and Johns Hopkins University.

Quené, H. & J. Krull (1999) Recognition of assimilated words in normal and fast speech. *Proceedings of the XIVth International Congress of Phonetic Sciences* (San Francisco) Vol 3: 1831-1834.

Raphael, L (1981) Durations and contexts as cues to word-final cognate oppositions in English. *Phonetica* 38: 126-147.

Raphael, L., Y. Tobin, A. Faber, T. Most, H. Kollia, D. Milstein (1995) Intermediate values of voice onset time. In F. Bell-Berti & L. Raphael (eds.)

Producing speech: contemporary issues. New York: American Institute for Physics.

Riemersma, T. (1979) *Sylabysjerring, nazzeljerring, assymljerring.* Ljouwert: Koperative Utjowerij.

Ritter, N. (2000) Hungarian voicing assimilation revisited in Head-Driven Phonology. In G. Alberti & I. Kenesei (eds.) *Papers from the Pécs Conference. Approaches to Hungarian 7.* Szeged: JATE Press.

Rostolland, D. (1982) Acoustic features of shouted voice. *Acustica* 50: 118-125.

Rubach, J. (1993) *The Lexical Phonology of Slovak.* Oxford: Clarendon.

Ruhlen, M. (1975) *A Guide to the Languages of the World.* Palo Alto: Stanford University Press.

Saltzman, E. & K. Munhall (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1: 333-382.

Sauvageot, A. (1951) *Esquisse de la langue hongroise.* Paris: Klincksieck.

Schachter, P. (1969) Natural assimilation rules in Akan. *International Journal of American Linguistics* 4: 342-355.

Schiefer, L. (1992) Trading relations in the perception of stops and their implications for phonological theory. In D. Docherty & D. Ladd (eds.) *Papers in Laboratory Phonology 2. Segment, gesture, prosody.* Cambridge: CUP.

Schwartz, J.-L., L.-J. Boe, N. Vallée and C. Abry (1997a) Major trends in vowel system inventories. *Journal of Phonetics* 25: 233-253.

Schwartz, J.-L., L.-J. Boe, N. Vallée and C. Abry (1997b) The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics* 25: 255-286.

Selkirk, E. (1980) The phrase phonology of English and French. New York: Garland.

Selkirk, E. (1984) *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge, MA: MIT Press.

Selkirk, E. (1986) On derived domains in sentence phonology. *Phonology Yearbook* 3: 371-405.

Siptár, P. (1991) Fast speech processes in Hungarian. In M. Gósy (ed.) *Temporal factors in speech: a collection of papers.* Budapest: Magyar Tudományos Akadémia, Nyelvtudományi Intézete.

Siptár, P. (1996) A Janus-faced Hungarian consonant. *ELTE SEAS Working Papers in Linguistics. The Even Yearbook* 2: 83-96.

Siptár, P., & M. Törkenczy (2000) *The Phonology of Hungarian.* Oxford: Clarendon.

Siptár, P. & S. Szentgyörgyi (2002) H as in Hungarian. *Acta Linguistica Hungarica* 49: 427-456.

Slis, I. (1970) Articulatory measurements on voiced, voiceless, and nasal consonants: a test of a model. *Phonetica* 2: 193-210.

Slis, I. (1985) *The voiced-voiceless distinction and voice assimilation in Dutch.*

PhD dissertation, University of Nijmegen.

Slis, I. (1986) Assimilation of voice in Dutch as a function of stress, word boundaries and sex of speaker and listener. *Journal of Phonetics* 14: 311-426.

Slis, I. (1987) Rules for assimilation of voice in Dutch. In R. Channon and L. Shockey (eds.) *In Honor of Ilse Lehiste.* Dordrecht: Foris.

Slis, I. & A. Cohen (1969a) On the complex regulating the voiced-voiceless distinction I. *Language and Speech* 12: 80-102.

Slis, I. & A. Cohen (1969b) On the complex regulating the voiced-voiceless distinction II. *Language and Speech* 12: 173-155.

Slowiaczek, L., & D. Dinnsen (1985) On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13: 325-341.

Slowiaczek, L., & H. Szymanska (1989) Perception of word-final devoicing in Polish. *Journal of Phonetics* 17: 205-212.

Smith, C. (1996) The devoicing of /z/ in American English: effects of local and prosodic context. *UCLA Working Papers in Phonetics* 92: 125-150.

Steever, S. (1990) Tamil and the Dravidian languages. In B. Comrie (ed.) *The Major Languages of South Asia, The Middle East and Africa.* London: Routledge.

Steriade, D. (1992) Mazateco onsets. In J. Cole & C. Kisseberth (eds.) *Proceedings of the University of Illinois Conference in Phonology.* Stanford, CA: CSLI.

Steriade, D. (1993) Closure, release and nasal contours. In M. Huffman & R. Krakow (eds.) *Phonetics and Phonology 5: Nasals, Nasalization and the Velum.* New York: Academic Press.

Steriade, D. (1997) *Phonetics in phonology: the case of laryngeal neutralization.* Ms., UCLA.

Steriade, Donca, & J. Zhang (2001) *Context dependent similarity evaluations: the phonetics and phonology of half rhymes.* Paper presented at *CLS 37*, University of Chicago, April 2001.

Stevens, K. (1975) The potential role of property detectors in the perception of consonants. In G. Fant & M. Tatham (eds.) *Auditory Analysis and Perception of Speech.* New York: Academic Press.

Stevens, K. (1998) *Acoustic Phonetics* Cambridge, MA: MIT Press.

Stevens, K., S. Blumstein, L. Glicksman, M. Burton & K. Kurowski (1992) Acoustic and Perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America* 91: 2979-3000.

Stevens, K. & S.-J. Keyser (1989) Primary features and their enhancement in consonants. *Language* 65: 81-106.

Strevens, P. (1960) Spectra of fricative noise in human speech. *Language and Speech* 3:32.

Svirsky, M., K. Stevens, M. Matthies, J. Manzella, J. Perkell& R. Wilhelms-

Tricarico (1997) Tongue surface displacement during obstruent stop consonants. *Journal of the Acoustical Society of America* 102: 562-571.

Szigetvári, P. (1998) Voice assimilation in Hungarian: the hitches. In László Varga (ed.) *The Even Yearbook 3. ELTE SEAS Working Papers in Linguistics.* Budapest: Department of English Linguistics, Eötv"os Loránd University.

Ternes, E. & T. Vladimirova-Buhtz (1999) Bulgarian. In *Handbook of the International Phonetic Association.* Cambridge: CUP.

Tiersma, P. (1985) *Frisian Reference Grammar.* Dordrecht: Foris.

Thorsen, N. (1971) Voicing in British English t and d in contact with other consonants. *Annual Report of the Institute of Phonetics, University of Copenhagen* 5:1-39.

Thorsen, O. (1966) Voice assimilation of stop consonants and fricatives in French and its relation to sound duration and intra-oral air pressure. *Annual Report of the Institute of Phonetics, University of Copenhagen* 1: 67-76.

Thráinsson, H. (1978) The phonology of Icelandic preaspiration. *Nordic Journal of Linguistics* 1: 3-54.

Tompa, J. (ed.) (1961) *A mai magyar nyelv rendszere: Leíró nyelvtan.* Kötét 1. Budapest: Akadémiai kiadó

Trommelen, M., & W. Zonneveld (1979) *Inleiding in de generatieve fonologie.* Muiderberg: Coutinho.

Turk, A. (1992) The American English flapping rule and the effect of stress on stop consonant duration. *Working Papers of the Cornell Phonetics Laboratory* 7: 103-133.

Vago, R. (1980) *The Sound Pattern of Hungarian.* Washington, DC: Georgetown University Press.

Vallée, N., L.-J. Boë, J.-L. Schwartz, P. Badin & C. Abry (2002) The weight of phonetic substance in the structure of sound inventories. *ZAS Papers in Linguistics* 28: 145-168

van den Berg, J. (1958) Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research* 1: 227-244.

van der Meer, G. & T. de Graaf (1986) Sandhi Phenomena in Frisian. In H. Andersen (ed.) *Sandhi Phenomena in the Languages of Europe.* Berlin: Mouton.

van Dommelen, W. (1983a) Parameter interactions in the perception of French plosives. *Phonetica* 40: 32-62.

van Dommelen, W. (1985) Further experiments in the perception of French plosives. *Phonetica* 42: 100-115.

Vanvik, A. (1972) A phonetic-phonemic analysis of Standard Eastern Norwegian. *Norwegian Journal of Linguistics* 26: 119-164.

Vaux, B. (1998) The laryngeal specifications of fricatives. *Linguistic Inquiry* 29:

497-511.

Wajskop, M. & J. Sweerts (1973) Voicing cues in oral stop consonants. *Journal of Phonetics* 1: 121-130.

Wells, J. (1982) *Accents of English.* Vol 2: *The British Isles.* Cambridge: CUP.

Wells, J. (1982) *Accents of English.* Vol 3: *Beyond the British Isles.* Cambridge: CUP.

Wells, J. (2000) *Longman Pronunciation Dictionary.* 2nd ed. Harlow: Pearson Education.

Westbury, J. & P. Keating (1986) On the naturalness of stop consonant voicing. *Journal of Linguistics* 22: 145-166.

Wiese, R. (1996) *The Phonology of German.* Oxford: Clarendon.

Windsor Lewis, J. (1972) *A Concise Pronouncing Dictionary of British and American English.* London: Oxford University Press.

Wissing, D. (1991) Regressiewe stemassimilasie in Afrikaans. *South African Journal of Linguistics* Supplement 11: 132-156.

Wright, P. (1952) Parasitic nasals at Marshside, Lancashire. *Leeds studies in English* 7-8: 92-96.

Yoshioka, H., A. Löfqvist, and H. Hirose (1981) Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America* 70: 1615-1623.

Yoshioka, H., A. Löfqvist, and R. Collier (1982) Laryngeal adjustments in Dutch voiceless obstruent production. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 16: 27-35.

Zonneveld, W. (1983) Lexical and phonological properties of Dutch devoicing assimilation. In M. van den Broecke, V. van Heuven & W. Zonneveld (eds.) *Sound Structures: Studies for Antonie Cohen.* Dordrecht: Foris.

Zsiga, E. (1997) Features, gestures, and Igbo vowels: an approach to the phonetics-phonology interface. *Language* 73: 227-274.

Zue, V. (1976) *Acoustic Characteristics of Stop Consonants.* Sc.D. thesis, MIT.

Zwaardemaker, H. & L. Eijkman (1928) *Leerboek der Phonetiek.* Haarlem: Bohn.

Nederlandse samenvatting

Deze dissertatie behandelt de fonologie en fonetiek van obstruentsystemen waarin het stemgeluid wordt aangewend om een binair lexicaal contrast tussen *gespannen* en *ongespannen* te realiseren. Dit type systeem is zeer wijdverbreid in de in Europa vertegenwoordigde taalfamilies en wordt ook daarbuiten veel gevonden.

Het stemgeluid, het quasi-periodieke geluid dat wordt voortgebracht door het trillen van de stembanden, wordt universeel gebruikt als brongeluid voor het spraaksignaal. In een groot aantal talen wordt het stemgeluid ook voor een specifiek talig doeleinde aangewend, namelijk om het onderscheid tussen bepaalde (contrastieve) spraakklanken in het spraaksignaal tot uitdrukking te brengen. Zo gebruikt het Nederlands stembandtrilling om de beginklanken van bijvoorbeeld *paling* en *polsen* (fonetisch [paːlɪŋ], [pɔlsən] met stemloze beginklanken) te onderscheiden van de beginklanken van bijvoorbeeld *baken* en *bolder* (fonetisch [baːkən], [bɔldər], met een stemhebbende beginklank). Talen die de stem op deze wijze benutten om plosieven te onderscheiden worden hier omschreven als *stemtalen*.

Een tweede type taal maakt eveneens gebruik van stembandtrilling om plosieven van elkaar te onderscheiden, zij het op een ietwat andere wijze. Zo onderscheiden (de meeste dialecten van) het Engels de beginklanken van woorden als *pollen* en *parsley* van de beginklanken van *ballot* en *banjo* door de eerste groep zowel stemloos als geaspireerd en (daardoor) met een verlate stemaanzet uit te spreken (fonetisch [pʰɒlən], [pʰɑːzlɪ]. Wanneer een sonorante klank onmiddellijk voorafgaat, worden de klanken in de tweede groep min of meer stemhebbend uitgesproken, maar voorafgegaan door een pauze of een andere obstruent zijn zij veelal volledig stemloos (fonetisch [b̥ælət], [b̥ænd͡ʒoʊ]). Aan dit tweede taaltype wordt hier gerefereerd als *aspiratietaal*.

Er bestaat dus een fonetisch onderscheid tussen de stemloze plosieven van stemtalen (als in *paling, polsen* in het Nederlands) en de stemloze geaspireerde plosieven van aspiratietalen als het Engels (*pollen, parsley*). Tegelijkertijd moeten deze klanken tot op zekere hoogte als een groep beschouwd worden, daar zij een aantal fonetische en fonologische overeenkomsten vertonen. Dit is de groep van gespannen klanken. Gespannen explosieven in zowel stem- als aspiratietalen duren bijvoorbeeld relatief lang, worden voorafgegaan door relatief korte vokalen, gaan gepaard met relatief luide explosies, en verhogen de toonhoogte van naburige vokalen enigzins. Een voorbeeld van de fonologische overeenkomsten tussen gespannen obstruenten met verschillende stemaanzeteigenschappen is het gedrag van de (initiële) alveolaire explosief van de verleden tijdssuffix in het Nederlands en het Engels. In beide talen wordt deze klank als [t] gerealiseerd wanneer een gespannen klank voorafgaat.

De over het algemeen stemhebbende plosieven van stemtalen (als in *baken, bolder* in het Nederlands) en de dikwijls stemloze plosieven van aspiratietalen (*banjo, ballot* in het Engels) vertonen vergelijkbare overeenkomsten, en kunnen derhalve tezamen als ongespannen klanken worden bestempeld. Ongespannen explosieven in zowel stem- als aspiratietalen zijn bijvoorbeeld relatief kort, worden voorafgegaan door relatief lange vocalen, gaan gepaard met relatief zacht klinkende explosies, en verlagen de toonhoogte van naburige vocalen enigszins. Merk op dat deze overeenkomsten en de hierboven omschreven overeenkomsten tussen de twee klassen gespannen klanken in zekere zin in fonemische transcripties en ook in de orthografie tot uitdrukking komen.

Dit proefschrift richt zich met name op de fonetische en fonologische regels die het gedrag bepalen van gespannen en ongespannen obstruenten in de Germaanse taalfamilie en het Hongaars, waarbij een bijzondere nadruk wordt gelegd op stemassimilatieverschijnselen. Op grond van een literatuuronderzoek en drie spraakproduktie-experimenten betoogt het dat de typologie van deze regels van een grotere complexiteit is dan vaak wordt voorgesteld in traditionele beschrijvingen en generatieve modellen. Deze complexiteit lijkt zich in hoge mate te laten verklaren binnen een functionalistisch kader, mits de productie en waarneming van de individuele fonetische correlaten van gespannen en ongespannen obstruenten in beschouwing worden genomen. Het ten dele formalistische en volledig categorische karakter van de vigerende generatieve modellen schiet daarentegen tekort in zowel de beschrijving als de verklaring van het gedrag van gespannen en ongespannen obstruenten.

Het betoog is als volgt opgebouwd. Hoofdstuk 1 schetst het beschrijvingskader en de theoretische beginselen die aan het proefschrift ten grondslag liggen. hier wordt uitgebreid stilgestaan bij recente modellen van de interface tussen fonologie en fonetiek, en bij de voor- en nadelen van formalistische en functionalistische verklaringsmodellen.

Hoofdstuk 2 begint met een beschrijving van de mechanismen die ten grondslag liggen aan de productie van stemcontrasten in obstruenten. Het begrip *passieve stemvorming* (ook wel: spontane stemvorming) wordt hier op de min of meer bekende manier gedefinieerd gedefinieerd als een situatie waarin de supraglottale configuratie van de spraakbuis, de aanzet of voortzetting van stembandtrilling niet in de weg staat. *passieve verstemlozing* wordt gebruikt om te refereren aan situaties waarin de stand van de articulatoren stemvorming in de weg staat, en geen articulatorische compensatiestrategieën worden aangewend om stemvorming alsnog mogelijk te maken. Van *actieve stemvorming* wordt hier gesproken als dergelijke strategieën wel worden benut. Tenslotte doelt de term *actieve verstemlozing* op gevallen waarin articulatiebewegingen gericht lijken te zijn op het tijdelijk blokkeren van passieve stemvorming tijdelijk wordt geblokkeerd.

Hoofdstuk 3 richt zich op de typologie en analyse van regels die het contrast tussen gespannen en ongespannen obstruenten volledig (lijken) te neutraliseren, zonder dat daarbij van assimilatie sprake is. Zulke neutralisatieregels treden op als ‘statische’ fonotactische generalisaties op het lexicale niveau, en ook als ‘dynamische’ processen die door de morfologie worden aangedreven: de vorm van ‘finale verstemlozing’ die in onder meer het Nederlands en het Duits wordt gevonden is een voorbeeld van de tweede groep.

Twee hoofdthema’s komen in dit hoofdstuk aan de orde. Het eerste is de fundamentele aard van neutralisatieprocessen. Neutralisatie van gespannenheids-opppositites wordt vaak gezien als *fortitie* of *verharding* waar het resultaat een stemloze klank is, en als *lenitie* of *verzachting* indien een proces leidt tot een stemhebbende obstruent. Dit betekent dat neutralisatie als een asymmetrisch verschijnsel gezien wordt, dat hetzij een ongespannen obstruent in de corresponderende gespannen klank omzet, hetzij een gespannen klank verandert in zijn ongespannen tegenhanger. Zo wordt het proces van ‘finale verstemlozing’ in het Nederlands veelal beschouwd als een proces dat de slotklanken van *hand* (onderliggend /hɑnd/) of *reis* (onderliggend /rɛiz/) verhardt tot respectievelijk [t] en [s], maar de slotklanken van *kant* (/kɑnt/) en *eis* (/ɛis/)ongemoeid laat.

Een tweede opvatting over de aard van neutralisatieprocessen stelt dat deze juist fundamenteel symmetrisch zijn. Volgens deze opvatting treft het Nederlandse finale verstemlozingsproces zowel de slotklanken van *kant* en *reis* als die van *hand* en *reis*, en produceert het een serie ‘neutrale’ slotklanken die noch als gespannen noch als ongespannen te karakteriseren zijn. Dit idee is op zichzelf al in de pregeneratieve structuralistische fonologie voorgesteld, maar meer recent is geopperd dat de reeks neutrale klanken geen fonetische specificaties ontvangt voor het geneutraliseerde kenmerk.

De in de voorafgaande alinea’s omschreven concepties van neutralisatie nemen aan dat elke uitspraak van iedere allomorf van een gegeven stam wordt afgeleid van een enkele fonologische vorm. Zo wordt vaak verondersteld dat de allomorfen van *hand* alle worden afgeleid van de onderliggende fonologische vorm /hɑnd/. In de meervoudsvorm handen blijft dan het ongespannen karakter van de alveolaire explosief bewaard, terwijl hij in de enkelvoudsvorm wordt omgezet in een gespannen of neutrale tegenhanger. Een derde visie op fonologische neutralisatie stelt echter dat de allomorfen van een stam of suffix normaliter ge-

genereerd worden op basis van onafhankelijk gerepresenteerde lexicale vormen, en dat de uitspraak van een bepaalde allomorf onderhevig is aan de invloed van meer dan één lexicale vorm. Volgens deze opvatting wordt de uitspraak van de enkelvoudsvorm *hand* bepaald door tenminste twee vormen: een ‘eigen’ vorm [hɑnt] met finale verstemlozing, en de paradigmatisch verwante vorm, [hɑnd], zonder verstemlozing, die vooral actief is in de productie van de meervoudsvorm *handen*. Een belangrijke voorspelling die deze derde benadering onderscheidt van de eerste twee is dat neutralisatieprocessen fonetisch onvolledig zijn indien er paradigmatische interferentie mogelijk is.

Uit het in hoofdstuk 3 opgemaakte inventaris van regels die het onderscheid tussen gespannen en ongespannen obstruenten opheffen blijkt dat er evidentie bestaat voor alle hierboven geschetste theorieën over de aard van neutralisatieverschijnselen. Het is onduidelijk in hoeverre deze stand van zaken valt te wijten aan de verschillende experimentele methodes die in het gepubliceerde fonetisch onderzoek worden gebruikt, maar vooralsnog lijkt het raadzaam elk geval op zijn eigen merites te beschouwen

Niet alle fonetische en morfosyntactische omgevingen zijn hebben hetzelfde potentieel voor neutralisatie van het contrast tussen gespannen en ongespannen obstruenten, en eenzelfde observatie kan worden gemaakt voor de verschillende obstruenttypes (explosieven, affricaten en fricatieven). De beschrijving en analyse van deze neutralisatie-assymetrieën vormt het tweede hoofdthema van hoofdstuk 3. Een aantal factoren waarvan het bekend is dat zij van invloed zijn bij het optreden van de neutralisatie van gespannenheid worden nader onderzocht op grond van een literatuuronderzoek, en ook de mogelijke effecten van een aantal andere fonologische en fonetische parameters worden bij dit onderzoek betrokken. De belangrijkste nieuwe observatie die dit deel van het proefschrift oplevert is de generalisatie dat stem- en aspiratietalen hetzelfde repertoire aan (niet-assimilatorische) neutralisatieverschijnselen bezitten en dat zij even gevoelig zijn voor het optreden van zulke verschijnselen.

Voorts wordt hier de houdbaarheid onderzocht van de neutralisatietheorie die is voorgesteld door Steriade (1997) en de mogelijkheid deze uit te breiden. Deze theorie stelt dat neutralisatie van een gegeven contrast eerder optreedt in omgevingen waar dit contrast relatief slecht waarneembaar is dan elders. Een van de grote voordelen van deze theorie is dat zij in staat is de zowel de neutralisatie van de slotklanken van woorden of lettergrepen, als de neutralisatiepatronen die optreden aan het begin van deze domeinen, te verklaren op grond van hetzelfde mechanisme. In dit opzicht is de theorie van Steriade superieur aan modellen die de neutralisatie van het contrast tussen gespannen en ongespannen obstruenten in verband brengen met de (gepostuleerde) speciale status van de het codadomein van de syllabe.

Tenslotte laat het tweede deel van hoofdstuk 3 zien hoe de theorie van Steriade, die voornamelijk is gebaseerd op neutralisatie-effecten die worden veroorzaakt door het contact tussen naburige klanken, in principe kan worden gegeneraliseerd naar andere neutralisatieverschijnselen, zoals bijvoorbeeld regels die betrekking hebben op verschillen in articulatiewijze. Alhoewel de hier geformuleerde hypothesen nog experimenteel getoetst moeten worden, levert de gegeneraliseerde versie van Steriade's theorie mogelijkerwijs een model dat alle bekende neutralisatiepatronen verklaart op grond van hetzelfde mechanisme. Een vergelijkbaar algemeen en eenvoudig model gaat alle gangbare generatieve kaders ver te boven.

Hoofdstuk 4 beschrijft de fonetische kenmerken van twee types stemassimilatie die mogen worden verwacht op grond van de literatuur over sandhiverschijnselen. Het eerste type is een fonologisch proces dat alle fonetisch correlaten van gespannenheid beïvloedt en dat kan optreden ongeacht de rol van stemgeving in de realisatie van het contrast tussen gespannen en ongespannen klanken. Het tweede type is een coarticulatieproces dat wordt aangedreven door de articulatie van actieve stemvorming en verstemlozing. De voornaamste fonetische eigenschappen van dit tweede type zijn de volgende: (1) assimilatie van stem wordt alleen uitgedrukt in de vorm van veranderingen in stemhebbendheid, en in fonetische kenmerken die mechanisch afhankelijk zijn van de productie van stemcontrasten; (2) het optreden van assimilatie is afhankelijk van de aanwezigheid van actieve stemvormings- of verstemlozingsprocessen in de betrokken klanken; (3) assimilatie leidt niet tot de neutralisatie van het onderscheid tussen gespannen en ongespannen obstruenten.

De voornaamste hypothese die in dit hoofdstuk wordt geopperd is dat assimilatieprocessen die zich binnen de grenzen van (morfologisch complexe) woorden afspelen normaliter tot het eerste, fonologische, type behoren, maar dat regressieve assimilatie van stem over woordgrenzen heen normaal gesproken een coarticulatieproces is. De volgende drie hoofdstukken beschrijven de resultaten van een drietal spraakproduktie-experimenten die werden ontworpen om dezeen een aantal ander hypothesen te toetsen.

Hoofdstukken 5 en 6 doen verslag van een vergelijkend onderzoek naar stemassimilatie in het Zuidelijk Brits Engels en het Hongaars. De onderzochte variant van het Engels is een aspiratietaal en het Hongaars een stemtaal en daar geen van beide talen het contrast tussen gespannen en ongespannen obstruenten neutraliseren aan het woordeinde, vormen zij ideaal testmateriaal voor de in hoofdstuk 2 ontwikkelde hypotheses. Experiment 1 onderzoekt het fonetisch gedrag van clusters die bestaan uit een velaire plosief /k/ of /g/, gevolg door een alveolaire plosief /t/ of /d/, een alveolaire sibilante fricatief /s/ of /z/ of een vloeiklank /r/. Experiment 2 onderzoekt het gedrag van vergelijkbare medeklinkerclusters in het Hongaars.

De resultaten van deze twee experimenten zijn in hoge mate consistent met de in hoofdstuk 2 geopperde hypothese dat regressieve stemassimilatie geconditioneerd is door actieve stemvorming in de klank die het proces aandrijft. De Engelse obstruenten /t, s, z/ en de Hongaarse obstruenten /t, d, s, z/, die alle met reden als klanken met actieve stemvorming of verstemlozing beschouwd kunnen worden, wekken ook zonder uitzondering stemassimilatie op in een voorafgaande velaire plosief. De Engelse ongespannen /d/ daarentegen, kan worden gezien als een klank waarvan de stemvorming en verstemlozing op volledig passieve gronden geschied, en wekt ook geen assimilatie van stem op.

De resultaten van experiment 1 ondersteunen eveneens de hypothese dat stemassimilatie alleen de stemvorming aanpast van de klanken die het proces ondergaan, maar niet de andere fonetische correlaten van gespannenheid. In de resultaten van experiment 2 tekent zich echter een ingewikkelder patroon af. Hier ondergaan bepaalde andere correlaten van gespannenheid, zoals de lengte van de vocaal die aan de velaire plosief voorafgaat, wel veranderingen onder invloed van een obstruent.

De voor het Hongaars gevonden effecten kunnen niet worden afgeleid uit een puur articulatorisch model van stemassimilatie, en impliceren daarmee dat regressieve assimilatie in het Hongaars door een ander type proces wordt aangedreven dan de voor het Engels gevonden assimilatie. een tweede mogelijkheid is dat beide talen hetzelfde assimilatieprocess bezitten, maar dat dit proces alleen in het Hongaars gepaard gaat met een onafhankelijke neutralisatieregel.

Hoofdstuk 7 rapporteert de resultaten van experiment 3, dat de fonetische manifestatie van stemassimilatie in het Nederlands onderzoekt. Dit experiment vergelijkt de effecten van gespannnen explosieven (/p, t/), ongespannen explosieven (/b, d/), de klanken /m/, /h/ en vokalen op de eigenschappen van een voorafgaande /p/ + /s/ cluster.

De resultaten van dit experiment zijn consistent met de in hoofdstuk 2 geformuleerde coarticulatie-theorie van stemassimilatie en impliceren bovendien dat de traditionele beschrijving van stemassimilatie in het Nederlands dien te worden herzien. De eerste conclusie die aan de uitkomsten van experiment 3 verbonden kan worden is dat in clusters die zijn samengesteld uit een explosief, een fricatief en een tweede explosief (als in *fietsbel* of *rijksdaalder*) regressieve stemassimilatie plaatsvindt. Deze conclusie is strijdig met beweringen elders in de literatuur (bijvoorbeeld in Brink 1975; Camminga & van Reenen 1980) dat dit type cluster is uitgesloten van RVA. Ten tweede blijkt dat het assimilatieproces *symmetrisch* is: zowel /p, t/ als /b, d/ hebben invloed op de stemvorming in een voorafgaande /ps/ cluster. Dit beeld is strijdig met de wijdverbreide opvatting dat in het Nederlands alleen /b/en /d/ in staat zijn RVA op te wekken. Ten derde zijn de resultaten van experiment 3 consistent met het idee dat RVA ontstaat door de coarticulatie van stemvorming, omdat alleen de stemhebbendheid van een /ps/ cluster op de verwachte wijze wordt beïvloed door een volgende explosief.

Het werk in de eerste vijf hoofdstukken van dit proefschrift is in hoge mate bepaald door twee algemene hypotheses. De eerste stelt dat het mogelijk is om het gedrag van spraakklanken te beschrijven aan de hand van een continue fonetische representatie en dat dit in een aantal gevallen zelfs noodzakelijk is. Deze hypothese ligt ten grondslag aan de coarticulatie-theorie van stemassimilatie die is beschreven in hoofdstuk 2 en wordt getoetst in hoofdstukken 5, 6 en 7. De tweede algemene hypothese is dat alle fonetische en fonologische regels zich uiteindelijk laten verklaren op grond van de eigenschappen van het menselijke spraakproductiesysteem, het spraakwaarnemingssysteem, en de rol van taal als een communicatiemedium. Deze functionalistische benadering vormt de basis voor de analyse van neutralisatieregels in hoofdstuk 3.

Op verschillende plaatsen in de eerste vijf hoofdstukken worden de (consequenties van) deze ‘fonetische’ en functionalistische uitgangspunten vergeleken met het perspectief van de generatieve fonologie, waarin fonetisch gesproken relatief abstracte en discrete structuren de basis vormen voor de analyse van spraakklanken, en die in elk geval ten dele een formalistische inslag heeft. Dat laatste houdt in dat taalverschijnselen worden verklaard op grond van formele principes die zijn beperkt tot het taalvermogen en waarvan de vorm niet wordt beïnvloed door (taal-)externe factoren. Nergens gaan de eerste vijf hoofdstukken echter in op de specifieke voorspellingen van de relevante generatieve modellen.

Het voornaamste doel van Hoofdstuk 8 is het ontrafelen van zulke specifieke voorspellingen, en dit hoofdstuk vormt daarmee het sluitstuk in de these dat een ‘fonetisch’ en functionalistisch model betere beschrijvingen en verklaringen biedt voor de onderzochte fonologische en fonetische verschijnselen.

Twee soorten generatieve modellen staan hier centraal, en voor beide modellen wordt aangetoond dat zij niet in staat zijn de fonetiek en fonologie van gespannenheid afdoende te behandelen. Het eerste soort model is vertegenwoordigd in het werk van bijvoorbeeld Mascaró (1987/1995); Lombardi (1994, 1995a,b, 1996); Cho (1990a/1999), en representeert de gespannen en ongespannen obstruenten van stem- en aspiratietalen op exact dezelfde wijze. Een van de voornaamste problemen van deze benadering is dat de fonetische conditionering van regressieve stemassimilatie niet kan worden voorspeld. Dit probleem wordt ondervangen door het tweede modeltype, dat is vertegenwoordigd in het werk van bijvoorbeeld Harris (1994); Iverson & Salmons (1995, 1999). In deze benadering is de representatie van gespannenheid gebaseerd op de stemhebbendheid van plosieven aan het woordbegin, en de verschillende assimilatorische eigenschappen van de twee typen ongespannen plosieven worden daarmee voorspelbaar. Onder bepaalde aannames gaat als gevolg van dit representatieschema echter de (correcte) voorspelling verloren dat stem- en aspiratietalen even gevoelig zijn voor het optreden van neutralisatie van gespannenheid. De keuze voor stemhebbendheid of stemloosheid in plaats voor het verschil tussen actieve en passieve stemvorming en verstemlozing als basis voor de representatie van gespannenheid leidt voorts tot de voorspelling dat de gespannen obstruenten van stemtalen en de ongespannen obstruenten van aspiratietalen exact hetzelfde gedrag vertonen, hetgeen duidelijk niet het geval is.

In meer algemene zin wreken zich in beide modellen het gebruik van zogenoemde monovalente fonologische kenmerken en de keuze voor een atomistische representatie van gespannenheid, dat will zeggen, het gebruik van één enkel kenmerk om het verschil tussen gespannen en ongespannen obstruenten weer te geven. Dit leidt tot problemen in de analyse van het gedrag van gespannen obstruenten in stemtalen, die onterecht als fonologisch ‘inert’ worden bestempeld. Het betekent eveneens dat de beperking van stemassimilatie tot stemhebbendheid (tenminste aan woordgrenzen in het Nederlands en Engels) niet adequaat beschreven kan worden.

Op het eerste gezicht zouden deze en andere problemen wellicht kunnen worden verholpen door een complexere representatie van gespannenheid te gebruiken. Het laatste deel van hoofdstuk 8 toont aan dat een generatief model dat de in dit proefschrift onderzochte verschijnselen min of meer afdoende representeert een veel te groot aantal processen voorspelt en daarmee onbruikbaar wordt. De enige manier om de kracht van dit model in goede banen te leiden en zijn voorspellend vermogen te herstellen is de introductie van functionalistische principes. Door zowel de complexiteit van fonologische structuren te vergroten (en meer fonetisch realistisch te maken) en externe factoren in de vormgeving van de fonologische grammatica toe te laten, benadert dit herziene generatieve model een ‘fonetisch’ en functionalistisch model echter dermate, dat het karakter van het generatieve kader in wezen ondermijnd wordt.

Hoofdstuk 9 tenslotte, biedt een overzicht van de belangrijkste uitkomsten van de voorafgaande hoofdstukken en stipt een aantal onderwerpen aan die verder onderzoek verdienen.

Groningen dissertations in linguistics (GRODIL)

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach.*
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure.*
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation.*
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation.*
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar.*
6. Peter I. Blok (1993). *The Interpretation of Focus.*
7. Roelien Bastiaanse (1993). *Studies in Aphasia.*
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist.*
9. Wim Kosmeijer (1993). *Barriers and Licensing.*
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach.*
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity.*
12. Ton van der Wouden (1994). *Negative Contexts.*
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar.*
14. Petra Hendriks (1995). *Comparatives and Categorial Grammar.*

15. Maarten de Wind (1995). *Inversion in French.*
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance.*
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition.*
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items.*
19. Karen Lattewitz (1997). *Adjacency in Dutch and German.*
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch.*
21. Henny Klein (1997). *Adverbs of Degree in Dutch.*
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du'-NPs.*
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure.*
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program.*
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers.*
26. Erik Tjong Kim Sang (1998). *Machine Learning of Phonotactics.*
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses.*
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure.*
29. H. Wee (1999). *Definite Focus.*
30. E-H. Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse.*
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing.*
32. Klarien van der Linde (2001). *Sonority substitutions.*
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information.*
34. Shalom Zuckerman (2001). *The Acquisition of "Optimal"Movement.*

35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding*
36. Esther Ruigendijk(2002). *Case assignment in agrammatism: a cross-linguistic study*
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures.*
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.*