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: 22-04-2025

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.

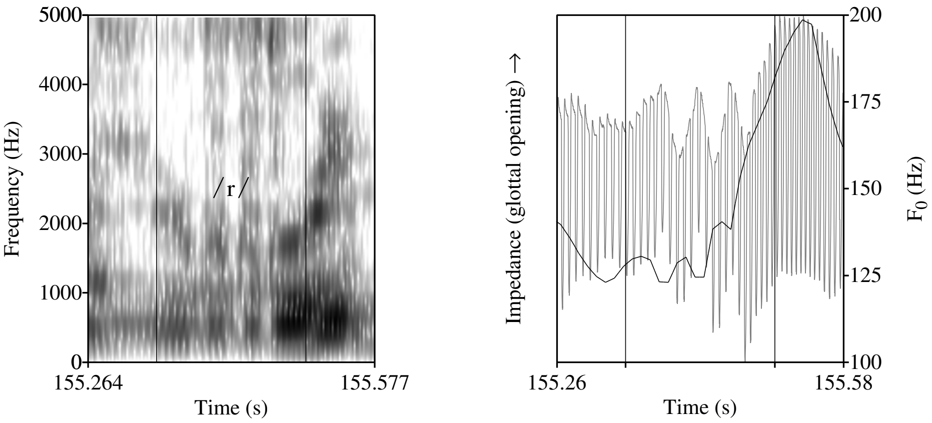

Figure 2.1: Broad band spectrogram (left) and EGG trace (right) with superimposed F_0 contour (in black) of an apical trill in an utterance of Dutch /rɔbən/.

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

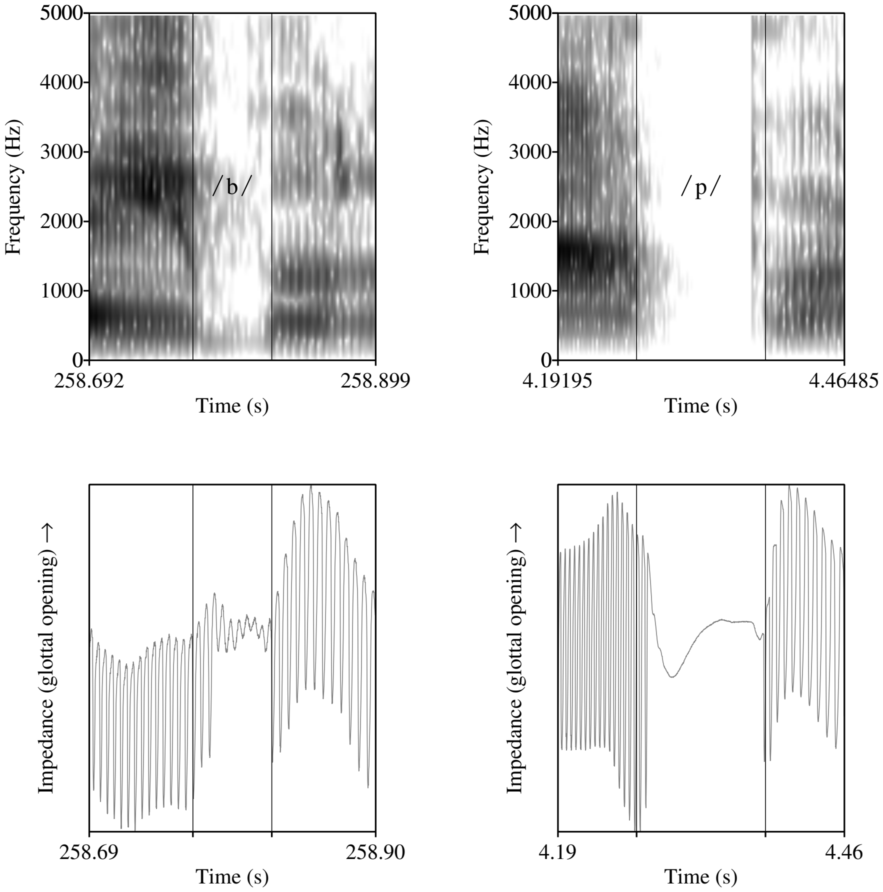

Figure 2.2: Broad band spectrograms (top) and EGG traces (bottom) of the medial labial stops in utterances of English /keibl̩/ (left) and Dutch /kaːpər/ (right).

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//

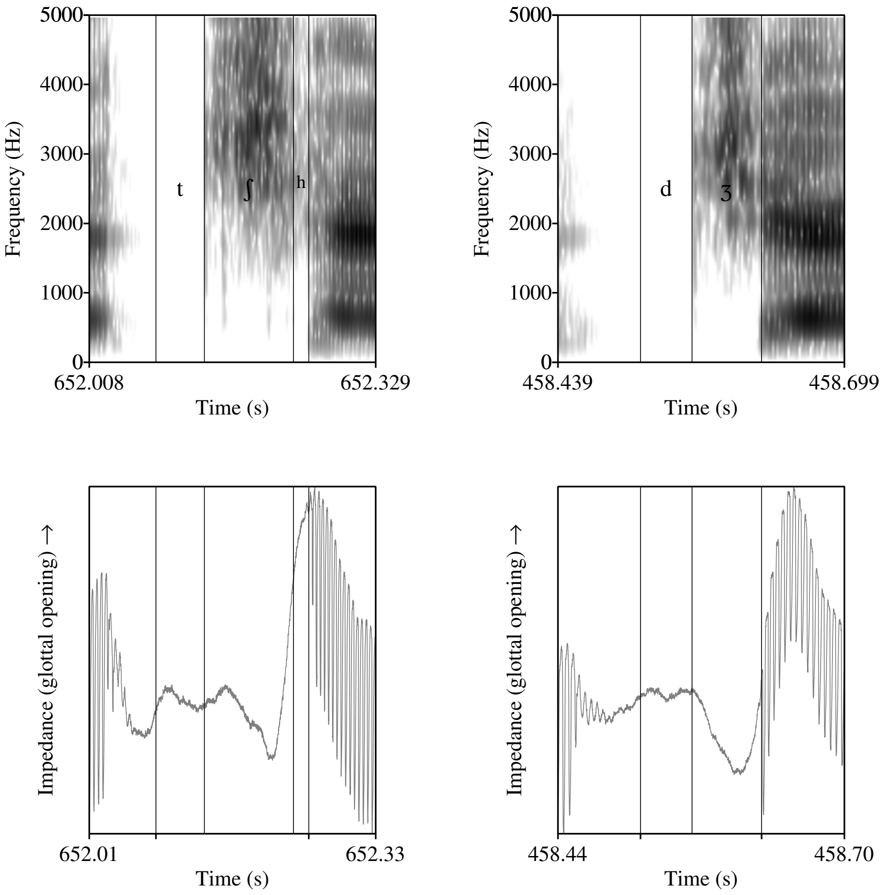

Figure 2.3: Broad band spectrograms and EGG traces of English word-initial affricates

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 /t͡ʃɛrɪ/ 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.