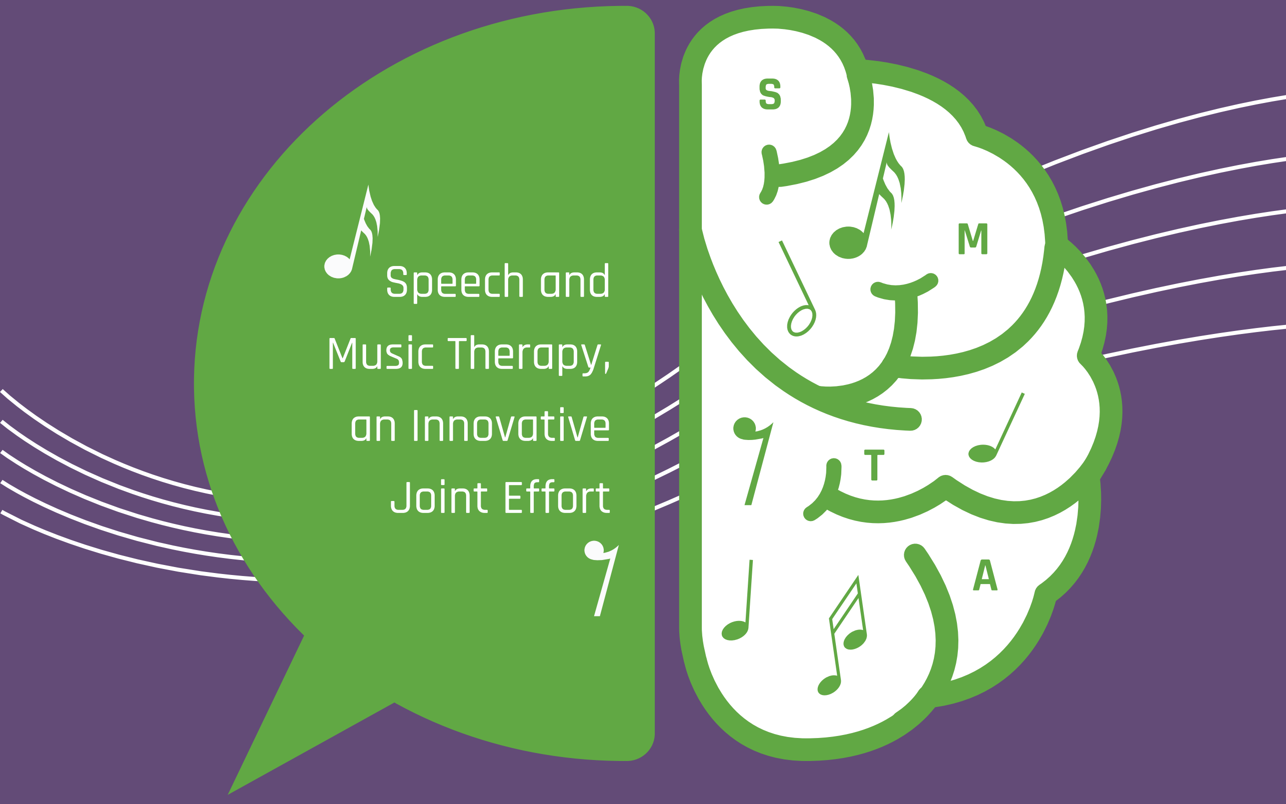
The treatment of apraxia of speech

Hurkmans, Josephus Johannes Stephanus

Publication date:
2016

Chapter 1

Theoretical framework



1.1 | Introduction

The process of speech production has been elaborately outlined in psycholinguistic and neurolinguistic studies. Different models have been proposed based on real-time processing and priming experiments in healthy speakers. One of the most influential models in the literature is the *logogen* model (e.g., Morton, 1969). This model represents various aspects of language processing, such as oral naming, repetition, oral reading and writing. This chapter focuses on one particular part of language processing: spoken word production.

The various levels of spoken word production will be described using linear models. It will become clear that these linear models are not suitable to describe the process of speech motor control. Therefore, non-linear models will be introduced and, in addition, speech motor theories will be discussed in order to outline this process.

1.2 | Linear models of speech production

One of the most influential models of speech production is the model of Levelt, Roelofs and Meyer (1999). Building on the earlier work by Garrett (1975), the original model developed by Levelt in 1989 was adapted in several studies, until its current version.

Levelt et al.'s (1999) model distinguishes various levels of word production (see Figure 1.1). This chapter specifically focuses on phonological and phonetic encoding. First, the preceding processes will be briefly described.

The production of a meaningful word always starts with a speaker's communicative *intention*, which is not, itself, language. The speaker's intention to express information has to be transformed into a verbal message that consists of lexical concepts. The process leading from intention to lexical concepts is called 'conceptual preparation', which is the first process of Level et al.'s (1999) model and provides an interface between thought and language. In this process, there is no one-to-one mapping of notions to be expressed onto messages. This is called the 'verbalisation problem' (Bierwish & Schreuder, 1992), which means that there are multiple ways to refer to the same object, even if a single lexical concept is activated. In picture naming, for example, the same object may be called 'pet', 'bird' or 'canary', depending on the set of alternatives and on the task.

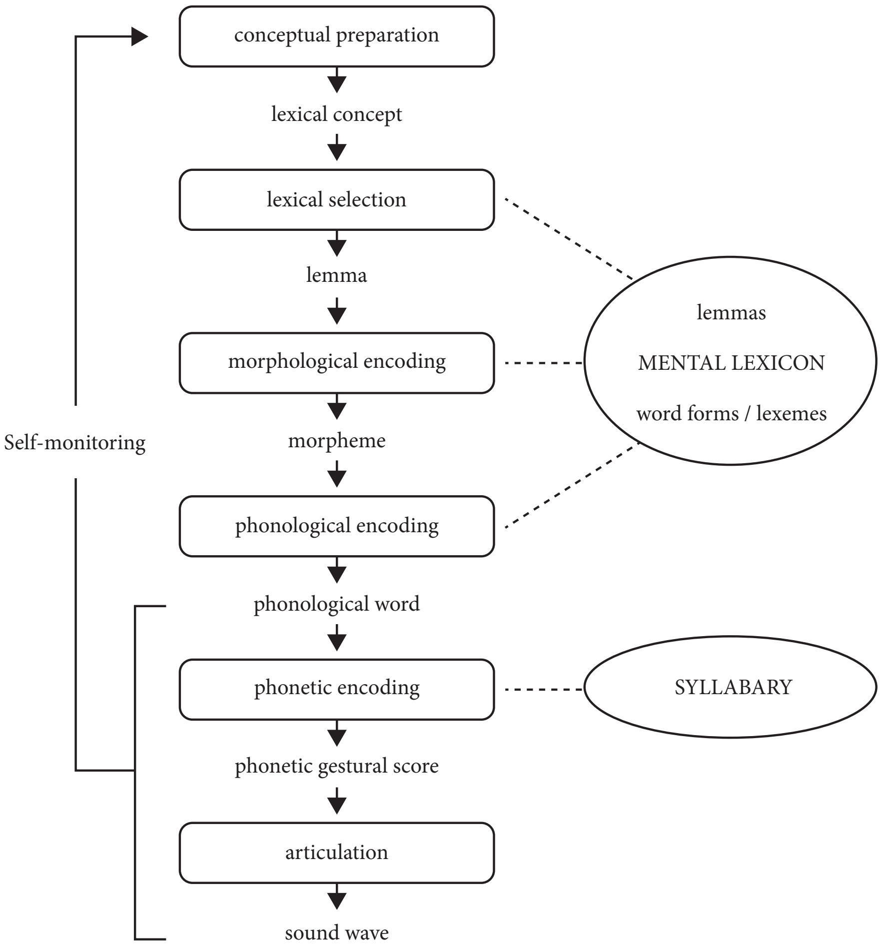


Figure 1.1 | Speech production model of Levelt et al. (1999; 3)

Apart from the verbalisation problem, additional semantic-related activations are triggered during lexical concept activation (Levelt et al.,

1999) resulting in a conceptual network. In this conceptual network, many links to other semantically related concept nodes are activated. For example, the concept node 'bike' links to other concept nodes, such as 'cycle', 'wheel' and 'car'.

The next level in Levelt et al.'s (1999) model is lexical selection. The lexical selection theory has been proposed by Roelofs (1992) and is modelled in terms of a feed-forward activation-spreading network. Lexical selection implies the retrieval of a lemma from the mental lexicon. As a result of the co-activation of lexical concepts, there is a spread of activation to the lemma nodes. Therefore, various semantically related lemmas in the mental lexicon become active. This process is similar to the previous level of spread activation for concept nodes. However, this level concerns *lexical* nodes. For example, in naming a picture of a dog, the concept for dog will activate the lemma 'dog' in the mental lexicon. Also, related lemmas will be co-activated, such as 'cat'. Within an efficiently running system, inappropriate candidates are rejected. The state of activation of non-target words follows a mathematical rule. This is called 'competition'. The target lemma will be the strongest: the lemma of 'dog' in picture naming of a dog.

Once a lemma is retrieved from the mental lexicon, morphemes are encoded at the level of morphological encoding. At this level, features for number, person, tense and mood are added to the lemma. The verb lemma 'walk', for example, can be phonologically realised as *walk, walks, walked, walking,* depending on the values of its features.

1.2.1 | Phonological encoding

The level after morphological encoding is phonological encoding. This level begins with the activation of the word form. Encoding the word form is divided into two separate processes: (1) retrieval of the phonological content of the word (segments), and (2) syllabification, which is the retrieval of the word structure (e.g., number of syllables; cf.

Hartsuiker, Bastiaanse, Postma, & Wijnen, 2005). Figure 1.2 illustrates the process of phonological encoding of the word ‘table’.

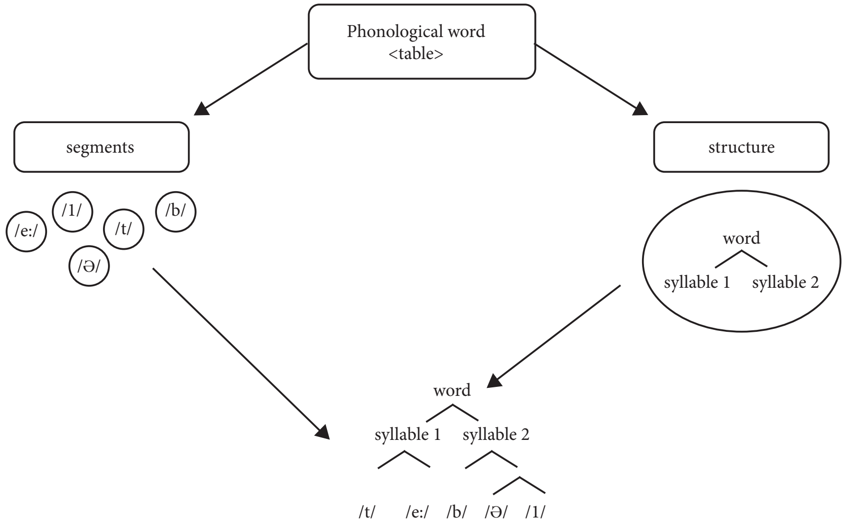


Figure 1.2 | Process of phonological encoding based on Hartsuiker et al. (2005; 6)

Spelling out the segmental and structural properties of the word form finally assembles and language-specific syllabification rules complete the process of phonological encoding. For example, in Dutch, the voiced /d/ in the word form of ‘hond’ (dog) changes in a voiceless /t/, due to the *final devoicing rule* in Dutch, that is, syllables cannot end in a voiced obstruent (/b/, /v/, /z/ or /d/), resulting in the pronunciation [hƆnt] (Bastiaanse, 2010).

Syllabification is a late process of phonological encoding because it depends on the word’s phonological environment. For example, the syllabification of the word form ‘demand’ is *de-mand*. However, the syllabification of the morphologically related word ‘demanding’, for instance, is different: *de-man-ding*, where the syllable *-ding* straddles the two morphemes *demand* and *-ing*. Levelt et al. (1999) assume that in the syllabification process, morphemes and phonemes become available simultaneously. The metrical template (i.e., the rhythmic pattern) may stay as it is, or may be modified in the context (see 1.2 for an elaborate description of this process). Syllabification follows universal as well as language specific rules (Levelt et al., 1999).

1.2.2 | Syllabary

Most languages have several thousands of syllables and a relatively small proportion is necessary to generate the majority of a language's lexical repertoire (Cholin, Levelt, & Schiller, 2006). For example, in Dutch, approximately 500 syllables, which is 5% of the entire syllable inventory, are used in everyday communication (Schiller, Meyer, Baayen, & Levelt, 1996). Since these syllables are used over and over again during a lifetime, it has been suggested that high-frequency syllables are stored in a repository, the so-called 'syllabary' (Levelt et al., 1999). Levelt and Wheeldon (1994) included the syllabary in their dual-route model to explain the process of syllable retrieval. Figure 1.3 represents this model.

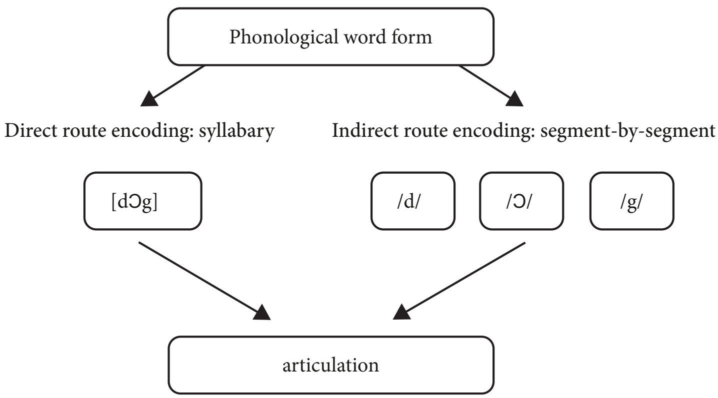


Figure 1.3 | The dual-route model of Levelt and Wheeldon (1994)

The dual-route model comprises two routes of encoding syllables: direct and indirect. The direct route is used for retrieving high-frequency syllables stored as pre-programmed units. The indirect route is used for encoding low-frequency syllables. These low-frequency syllables require online assembly each time they are used by a speaker. On the basis of reaction-time experiments, Levelt and Wheeldon (1994) suggest that direct-route encoding has shorter durations than the encoding via the indirect route. The duration of retrieving high-frequency syllables is shorter as a consequence of fewer computational steps. The retrieval of low-frequency syllables via the indirect route involves computing the phonetic representation of segment-to-segment assembly each time it is used.

While some researchers still refer to the syllabary, the evidence for its existence is limited. Brendel, Erb, Riecker, Grodd, Ackermann, & Ziegler (2011), for example, showed with their fMRI study, that there is no effect of syllable frequency on brain activity, but they emphasise the impact of syllable structure. Syllables with a complex onset (i.e., CCV) yield higher activation in motor execution areas than syllables with a simple onset (i.e., CV). Therefore, Brendel et al. (2011) suggest that syllables are not stored as equal holistic units in the syllabary as suggested by Levelt and Wheeldon (1994), but the process of encoding syllables depends on the complexity of syllable structure.

1.2.3 | Phonetic encoding

After the process of phonological encoding resulting in a phonological word, there is another process before articulation of the word: phonetic encoding. At this level of word production, articulatory gestures are assigned to the phonological word, specifying which patterns of articulatory movements are required. This still rather abstract representation needs to be modified at three different tiers (i.e., articulatory levels): (1) oral, (2) velar and (3) glottal. The oral tier contains lip and tongue (tip and body) structures. At this level, it may be necessary, for example, to close the lips for the production of the phoneme /b/ when starting to pronounce the word 'ball'. At the velar tier, there is activation for nasal consonants, for example, to lift the velum for the production of the phoneme /n/ in the word 'noise'. Finally, at the glottal tier, vocal tracts need activation for voiced consonants, for example, to produce the phoneme /d/ in the word 'dog'. Figure 1.4 illustrates the process of phonetic encoding of the word 'table'.

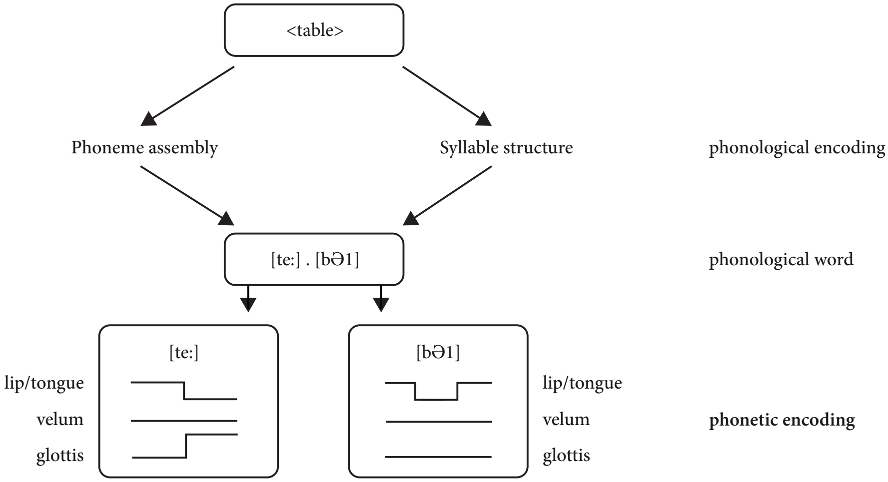


Figure 1.4 | Process of phonetic encoding adapted from Levelt et al. (1999)

The results of the process of phonetic encoding are phonetic gestural scores. There is either activation or no activation of the three different tiers for each phoneme in the syllable structure. However, relations between neighbouring phonemes are not specified in linear models. Other relevant components of phonetic encoding are underspecified in linear models as well. For example, supra-segmental aspects, such as prosody, are not accurately described in linear models. These aspects are better described in nonlinear models of speech motor control.

1.3 | Nonlinear models of speech motor control

Linear models, such as Levelt et al.'s (1999) model, represent various levels of speech production. At the level of phonetic encoding, these models are restricted to local fragments of the speech stream. In such small fragments (the size of phonemes) there is little attention to *relations* between phonemes (Miller, 2000). Isolated movements are only employed in interaction with other movements. Lip movement, for example, in relation to velum and laryngeal movements determines whether the lip closure is perceived as /p/, /b/, or /m/ (Miller, 2000). Therefore, the question raised is whether the architecture of speech motor control can be considered as a linearly ordered string of units.

Ziegler, Thelen, Staiger, & Liepold (2008) examined speech errors from patients with Apraxia of Speech (AoS) to identify the primitives of linear representations. They studied the architecture of phonetic encoding modelled by a tree-like, nonlinear metrical structure (see Figure 1.5). Ziegler et al. (2008) described five different error-source models from where phoneme errors may arise. These models can be regarded as various levels in a metrical tree-structure (see Figure 1.5) of where phoneme errors may arise. The first model is a *phoneme* model (PHO), postulating that the source of phoneme errors is the segment: the phoneme. In this model, utterances are considered strings of phonemes. Effects of length can be predicted with this model: the more phonemes there are in a word, the more errors may occur. The second model is a *constituent* model (ONC), postulating that phonetic encoding operates on syllabic constituents involving onsets (O), nuclei (N) and codas (C). This model predicts that more errors occur in syllables in which constituent positions are filled (e.g., CVC) than in syllables with an empty onset (e.g., VC) and open syllables (e.g., CV) with an empty coda position. The third model is a *syllable* model (SYL) grounded on the assumption that phonetic encoding operates on units of syllable size. Syllables, in this approach, are considered as holistic units. This model relates to the proposed syllabary of Levelt and Wheeldon (1994), see 1.1.2 for discussion. According to this model, errors are sensitive to syllable frequency and accuracy is independent of syllable structure. The fourth model is a *metrical foot* model (MFT). In this model, stressed and unstressed syllables form a metrical foot that constitutes the core unit of phonetic encoding. Ziegler (2005) demonstrated the influence of metrical-foot structure in a word-repetition task with AoS patients, irrespective of the number of syllables or phonemes within a foot. The final model is a *word* model (WRD) with the phonological word as the core unit of phonetic encoding. Errors, in this model, are sensitive to the total number of words, irrespective of how many phonemes, syllables, or metrical feet these words contain. Figure 1.5 represents a template of the different models in a metrical tree.

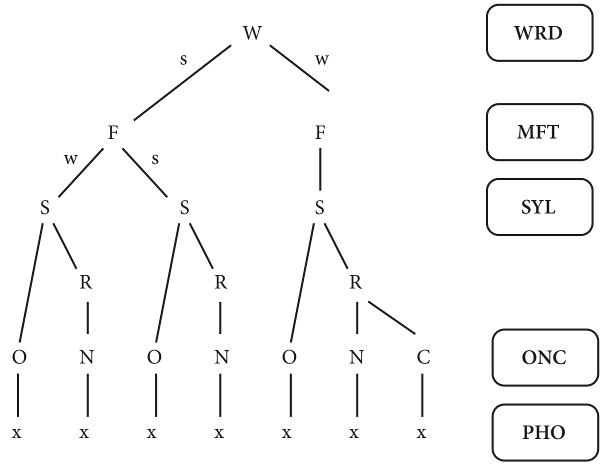


Figure 1.5 | Template of a metrical tree adapted from Ziegler, 2005 and Ziegler et al., 2008. W=word, s=strong, w=weak, F=foot, S=syllable, R=rhyme, O=onset, N=nucleus, C=coda, x=segment, WRD=word, MFT=metrical foot, SYL=syllable, ONC=onset, nucleus, coda; PHO=phoneme

Ziegler et al. (2008) tested the five models with data from AoS patients on a word-repetition test. They showed that the models were related to each other. There was not one type of error pattern in repeating words and the errors were influenced by various structural factors. For example, syllable-based errors were influenced by the word's foot-structure, and phoneme-based errors were modulated by metrical parameters. Ziegler et al. (2008), therefore, concluded that the domain of phonetic encoding incorporates several levels in a hierarchical architecture of word production, which asks for a nonlinear model.

Apart from the structural relationships between the levels in the metrical tree structure (Ziegler et al., 2008), there is additional evidence for nonlinear processing of speech motor control. Ziegler (2005) found that the workload of phonetic encoding strongly depends on structural properties of a word. Speech error data of a large sample (N=100 data sets) of patients with AoS showed that phonemes in a syllabic rhyme and trochaic foot contribute least to word-production errors. For example, more errors appear in the word 'straight' by adding the onset [str] to the rime [e:t] than in the word 'eight' with an empty onset. Likewise, more errors occur in the stressed syllable [tri] by adding an unstressed syllable [trəŋk] to form the trochee [tre:trəŋk] ('tree-trunk'), as compared with [tri], for example, in the word 'tree'.

According to Ziegler (2005), complexity of syllable retrieval depends on higher-order phonetic units. Rhymes and trochees are important motor programing units in phonetic encoding. With this finding, Ziegler (2005) emphasises the importance of rhythmicity in stress-timed languages, such as Dutch, German and English. Although word stress can vary in these languages, there is a preference for trochaic stress: first syllable strong, second weak (Domahs, Wiese, Bornkessel-Schlesewsky, & Schlesewsky, 2008; Brendel & Aichert, 2014). In contrast, in French, word stress is relatively fixed to the final syllable within multisyllabic words.

A paradox of Ziegler et al.'s (2008) model is the focus on isolated phonemes to explain mechanisms at the level of phonetic encoding. As stated before, relations between phonemes may be important in understanding the process of speech motor control (Miller, 2000). Therefore, Ziegler (2009) proposed to add gestural movements to the model describing transitions between neighbouring segments with sub-segmental units. He incorporated these gestures into the nonlinear metrical tree-model.

Gestures are directed to the articulatory levels as described in 1.1.3 (oral, velar and glottal tiers). Each single gesture is related to the rhythmically organised motor units in a word structure. Ziegler (2009) proposes a model that accounts for the probability of accurate word production in a series of steps. With this model, the likelihood of accurate word production can be calculated. Oral gestures form the basis of the model, including lip gestures, tongue-tip gestures and tongue-body gestures. These oral gestures can be combined with velar gestures (i.e., the transition between oral and nasal consonants) and glottal gestures (i.e., transitions between voiceless and voiced consonants). Subsequently, six coefficients have been added to the probalistic model resulting in a formula. These coefficients relate to transitions of gestures at various levels in the metrical template. With this formula, Ziegler (2009) examined whether the prediction of accurate word production in patients with AoS was better using a nonlinear model than using a linear model.

First, Ziegler (2009) predicted word accuracy according to a linear model of speech production. He used a list of words and non-words, repeated by patients with AoS, including one, two, three and four syllabic words resulting in a large set of data samples (N=120). The probability of a correct pronunciation for one-syllabic words was .51. In a linear model, syllables are stored as holistic units and, therefore, Ziegler (2009) could predict accurate word production of two, three and four syllabic words (see Figure 1.6). Then, he observed word accuracy in the same set of data (see Figure 1.6). Ziegler (2009) showed a clear discrepancy with higher repetition scores for the observed word accuracy. This is visualised in Figure 1.6.

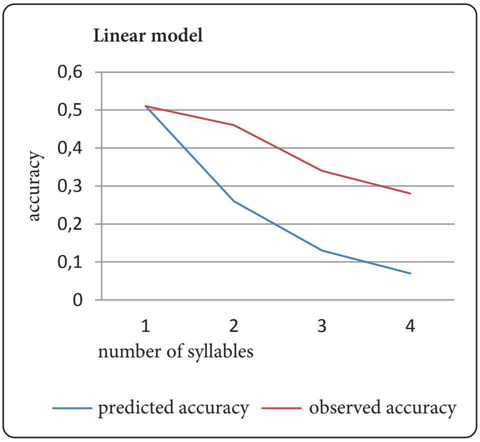


Figure 1.6 | Linear model of word accuracy in relation to the number of syllables adapted from Ziegler (2009)

Next, Ziegler (2009) predicted word accuracy according to a nonlinear model. He predicted word accuracy with his above-mentioned formula, considering the gestural movements at the various levels in the metrical template. In the nonlinear model, the predicted and the observed accuracy were almost identical. This is visualised in Figure 1.7. This study was replicated with Dutch data (Van den Eynde, Temmink, Kooi, Willemsen, Timmerman, Jonkers, & Feiken, 2010). Comparable findings were revealed for the German and Dutch data suggesting that this nonlinear approach is specifically suitable for one language.

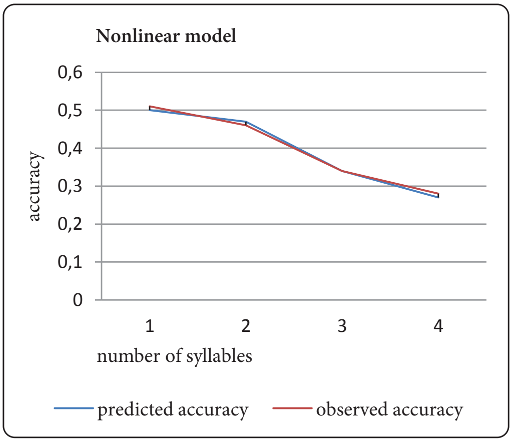


Figure 1.7 | Nonlinear model adapted from Ziegler's (2009)

It can be concluded that, from a psycholinguistic perspective, the process of speech motor control is a complex, nonlinear, hierarchical organisation of motor units extending from the level of articulatory gestures to the level of metrical feet.

In addition to the nonlinear model of speech production, processes of speech motor programming and planning are needed to complete the description of the process of speech motor control. These processes are described in speech motor theories and will be discussed in the following paragraph.

1.4 | Speech motor theories: Speech motor programming and planning

The Schema Theory of Schmidt (1975; 2003) is a theory on motor control and learning. It assumes that production of rapid complex movements involves the determination of related groups of motor actions (i.e., motor programs) rather than a series of individual movements. These motor actions are retrieved from memory and then adapted to the actual situation. Motor programs, according to this theory, are never produced exactly the same. Therefore, Schema Theory presumes that the programs are generalised. A Generalised Motor Program (GMP) captures the timing and force of the movement and specifies relative timing and force of muscle contraction, which is an abstract movement pattern. The GMP defines the shape of the movement. For individual movement patterns, *parameters* are activated for the absolute timing and force. With parameters it is possible to determine the movement pattern in context. Therefore, values are assigned to the parameters. The result is a variation in duration and amplitude of timing and force. For example, serving a ball in tennis involves a basic pattern of a backswing and a forward swing motion; the GMP governs these movements (Maas, Robin, Austermann Hula, Freedman, Wulf, Ballard, & Schmidt, 2008). However, the speed and the amplitude of the movement are varied by assigning different values to the parameters of the GMP for each specific action (Clark & Robin, 1998).

Within an articulatory motor program, a GMP corresponds to the motor commands associated with a phoneme, syllable, word or even a frequently produced phrase (Varley, Whiteside, Windsor, & Fisher, 2006). Schema Theory presumes that a series of GMPs that occur in a serial order, such as for the production of phonemes, become integrated into a single GMP (Schmidt & Lee, 2005). Two phonemes in isolation have a different GMP than those two phonemes in a cluster or word. Further, speakers can realise phonemes in many ways. The bilabial plosive

/b/, for instance, can be produced by either moving both lips or only one lip, with or without jaw movement. Moreover, it remains to be specified which aspects of speech movements are to be considered as GMPs (Ballard, Granier, & Robin, 2000). Ballard, Maas and Robin (2007) suggest that features of articulation are the distinctive aspect. Syllables with the same *place* of articulation (such as 'my' and 'pie'; labial) are governed by the same GMP. In contrast, syllables with a different *manner* of articulation (such as 'my 'and 'sigh'; full closure versus narrow constriction) are, according to Ballard et al. (2007), governed by different GMP's.

Within the Schema Theory, parameters correspond to motor planning. Planning speech production is a speaker's constant task during articulation; it functions as a control system. Therefore, motor planning interacts with motor programming. Motor planning can adjust speech production if necessary. Speech rate, for example, can be reduced to enhance speech accuracy and fluency by producing a long and infrequent word with recurring phonemes or phonemes with analogous features of articulation. For instance, in uncommon names, such as 'Eyjafjallajokull', which is a volcano in the Republic of Iceland, Dutch speakers will reduce speech rate for a fluent articulation. However, there are more parameters to adapt the process of speech production by speech planning, such as the prosodic features pitch, loudness, duration and intensity. Motor planning also controls various emotional prosodic features, such as in anger, happiness and sadness.

Figure 1.8 represents a model in which the various processes of speech motor control are incorporated as discussed in the above described psycholinguistic models and speech motor theories. This model will be used in the subsequent chapters on AoS and the treatment of AoS. In the final chapter of this thesis, the model will be used to relate various musical elements to the process of speech motor control, in order to explain the underlying mechanisms of SMTA leading to an understanding of which therapeutic musical elements contribute to its success.

Visualising the nonlinear process of speech motor control in a two-dimensional way is hardly possible. The two-way directions of the arrows between the level of phonetic encoding and the processes of motor programming and motor planning, modulated by the metrical tree and the parameters, is an attempt to represent the assumed continuous interaction between these abstract processes and, therewith, the dynamic nature of speech motor control.

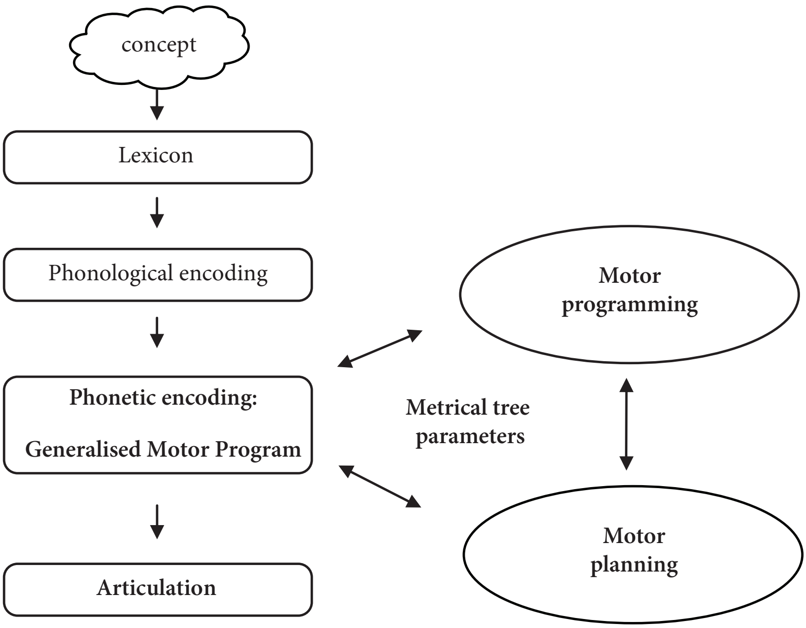


Figure 1.8 | Model of speech production (based on Levelt et al. 1999) with an extension of the level of phonetic encoding including the process of speech motor control