Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 18 (1994), 95-110.

REFLECTIONS ON ASPECTS OF VOWEL REDUCTION*

Dick R. van Bergem

Abstract

In this paper several aspects of the phenomenon of vowel reduction are discussed. Two types of vowel reduction are distinguished: lexical reduction and acoustic reduction. The meaning of these terms is explained and subsequently a concise theory of vowel reduction is presented in which the schwa plays a key role. In the remainder of the paper vowel reduction is discussed in relation to mental representations, word stress, and the traditional distinction competence-performance. It is argued that acoustic and lexical vowel reduction are two intermediate stages in the process of the sound change 'full vowel -> schwa'. Finally, some suggestions are done for future research.

1. Introduction

Speech is probably the most important means of human communication. In our phonetic investigations we like to approach the phenomenon of speech from a functional point of view. That is, we believe that in general speakers strive to achieve optimal transfer of messages with a minimal amount of articulatory effort. Of course, there are bounds to the tendency for ease of articulation of a speaker. Recognition of the message will be harder, when speech is articulated more sloppily. On the other hand, listeners have several information sources at their disposal to help them in restoring acoustically fuzzy messages. Especially semantic and pragmatic knowledge sources may play an important role in this respect. A certain degree of sloppiness in the articulation can thus be tolerated. The phenomenon of vowel reduction, which is the subject of the present paper, can very well be interpreted as a tendency towards ease of articulation. In Van Bergem (1995a) two types of vowel reduction are distinguished: lexical reduction and acoustic reduction.

Lexical vowel reduction stems from linguistic research and is defined as the substitution of a full vowel with a schwa in specific words. An example of a Dutch word in which this phenomenon can occur is the word "baNAAN" (/ba:na:n/, banana). The unstressed vowel in the first syllable of the word "baNAAN" (word stress indicated with capitals) can be realized as a full vowel /a:/ (or its short countertpart /ɑ/), but also as a schwa /ə/. In the latter case the schwa has become a characteristic (generally accepted) part of the word and therefore we call this phenomenon *lexical vowel reduction*.

Suppose a speaker has the *intention* to produce a full vowel. This does not necessarily mean that a neatly articulated full vowel will be realized. There appears to be an

* This paper is an extended version of chapter 5 in Van Bergem (1995a).

enormous variability in 'vowel quality'. Some vowels are pronounced much more carefully than others. Speakers are usually not aware of these differences in pronunciation, and listeners normally do not notice the variation. However, if several different vowels are segmented from their natural context and presented to listeners, it appears that some specimens are identified much better than others (assuming that speakers intend to produce phonologically 'correct' vowels). Another way to demonstrate differences in vowel quality is through a spectral analysis.

This is often done by measuring the steady-state formant frequencies of vowels from normal speech utterances and comparing these with formant frequencies of vowels pronounced in isolation (which can be regarded as 'ideal' vowels). It appears that the formant frequencies of some vowels from normal speech are relatively close to their target position, whereas the formant frequencies of other vowels have shifted considerably away from their target position. These formant shifts can partly be ascribed to *coarticulation* effects: neighbouring sounds influence each other to a certain extent because of limitations of the articulators. However, the formant shifts are often much larger than would be expected on the basis of coarticulation effects alone. Such extra shifts emerge, for instance, when the vowel occurs in an unstressed syllable, or when a 'spontaneous' speaking style is used. In such cases we speak of *acoustic vowel reduction*.

There are numerous factors that are in one way or another related with the occurrence of acoustic or lexical vowel reduction. Among these factors are the following:

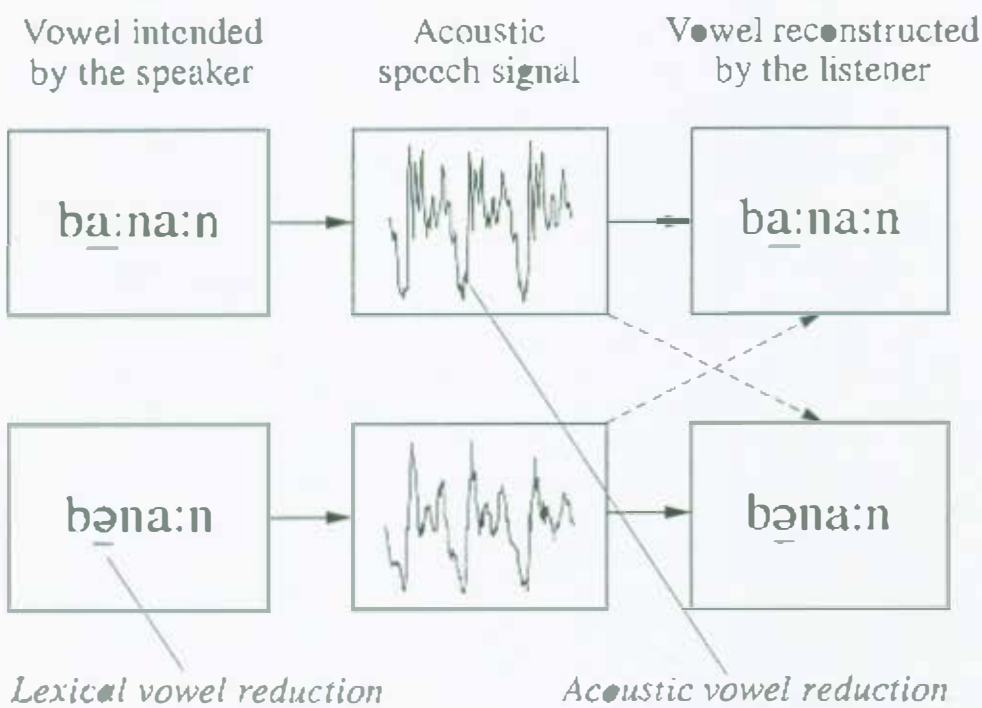

Figure 1. Two possible ways in which the word "baNAAN" (banana) can be encoded as a phoneme string in the speech signal by a speaker and subsequently decoded in the proper way by a listener. Acoustic vowel reduction can occur, if the speaker intends to produce a full vowel in the word "banaan" (upper part of the figure). Lexical vowel reduction occurs, if the speaker intends to produce a schwa in the word "banaan" (lower part of the figure). The dashed arrows indicate how the intended phoneme could be misidentified.

- word stress
- word class (content words - function words)
- frequency of occurrence of words
- sentence accent
- speaking style
- position of the vowel in the word
- vowel type
- phonemic context
- syllable type
- social background of the speaker
- geographical/dialectical background of the speaker

Notice that acoustic vowel reduction and lexical vowel reduction by definition cannot occur at the same time. If the speaker intends to produce a full vowel, acoustic vowel reduction may occur (when the ambition to actually reach the vowel target is small), but lexical vowel reduction does by definition *not* occur. If the speaker intends to produce a schwa, on the other hand, lexical vowel reduction occurs, but acoustic vowel reduction is impossible, for the schwa is the endpoint of acoustic reduction. The difference between acoustic and lexical vowel reduction is illustrated for the word "banaan" in Figure 1.

2. A concise theory of vowel reduction

It is often assumed that the formant patterns of reduced vowels shift towards the centre of the vowel diagram. This central position, usually referred to as the schwa position, is associated with a neutral configuration of the vocal tract. The first two formant frequencies that arise from a vocal tract with a neutral form (the shape of a uniform tube) are approximately 500 Hz and 1500 Hz for a male speaker (Fant, 1960). The notion that the schwa is produced with a 'neutral' vocal tract (a straight lossless tube), that would require a minimal amount of articulatory effort, is clearly based on a *static* view on vowel production. Connected speech, however, requires a *dynamic* view on vowel production, because the influence of context on vowels can surely not be neglected. In Van Bergem (1994) it is shown that the F_2-tracks of schwas in various phonemic contexts move almost straight from the onset to the offset, suggesting an articulatory *path* that requires a minimal amount of effort. Put in a different way, this means that the schwa is *completely assimilated* with its phonemic context.

This view on the schwa forms the basis for our functional theory of vowel reduction. In order to enhance the economy of articulatory gestures in connected speech, speakers have a tendency to pronounce vowels in a 'schwa-like' manner. In terms of acoustic features this means that the formant frequencies of vowels shift to a position that the schwa would have *in an identical phonemic context*. That this position often does not coincide with the centre of the vowel triangle is demonstrated in the examples of Figures 2 and 3.

The vowels of interest in these figures are the /ɪ/ from the Dutch word "miLJOEN" (/miljun/, million) and the /ɔ/ from the Dutch word "bioscOOP" (/bijɔskoːp/, cinema). These words were uttered in three different speech conditions by 20 male speakers in an experiment described in Van Bergem (1995b). Formant frequencies were measured at the vowel centre. Because plots containing all 60 realizations (20 speakers × 3 conditions) of the vowels /ɪ/ and /ɔ/ were rather fuzzy, the data were split up in four groups of 15 vowel realizations on the basis of their sorted F_2-values (disregarding the three test conditions). The average formant frequencies of each group are plotted in Figures 2 and 3 (indicated with the numbers "1" to "4"). Also shown in the figures are

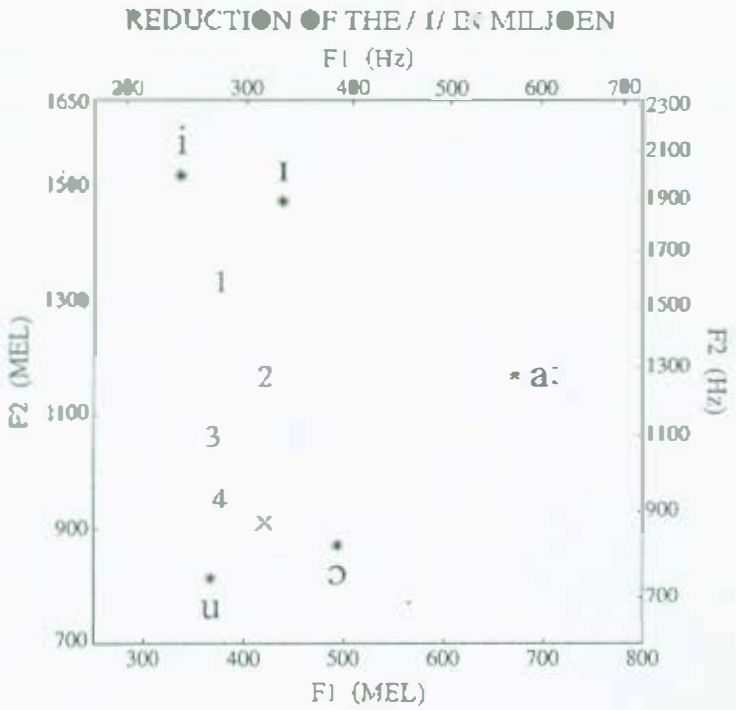

Figure 2. Four reduction stages (see text for more details) of the vowel /ɪ/ from the Dutch word "miljoen" (/miljun/, million), as spoken by 20 male speakers. Also shown are average formant frequencies of the vowels /i, ɪ, a:, ɔ, u/ spoken in the 'null' context /h-t/ by the same 20 speakers. The cross indicates the position of the schwa in the specific context /məljun/ according to a schwa model described in Van Bergem (1994).

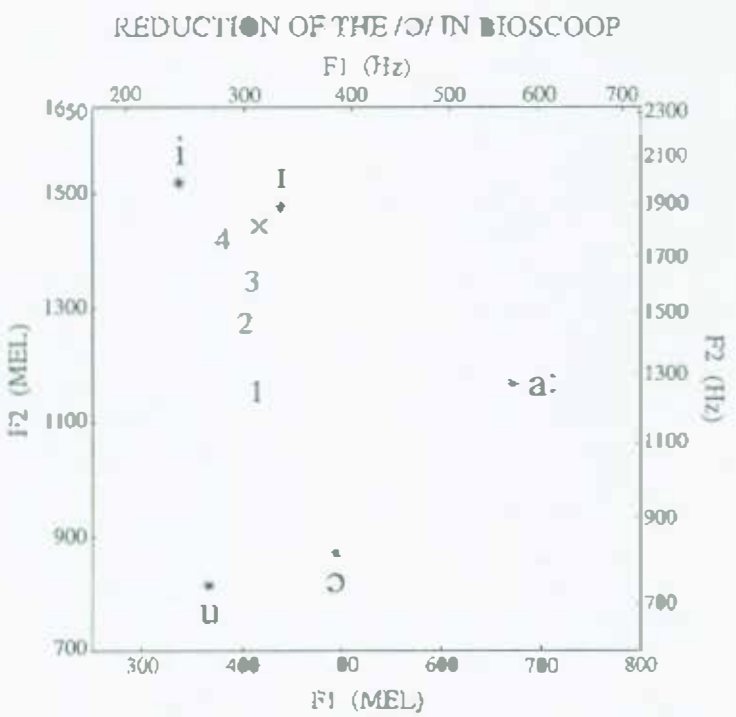

Figure 3. Four reduction stages (see text for more details) of the vowel /ɔ/ from the Dutch word "bioscoop" (/bijɔsko:p/, cinema), as spoken by 20 male speakers. Also shown are average formant frequencies of the vowels /i, ɪ, a:, ɔ, u/ spoken in the 'null' context /h-t/ by the same 20 speakers. The cross indicates the schwa position in the specific context /bijəsko:p/ according to a schwa model described in Van Bergem (1994).

average formant frequencies of the vowels /i, ɪ, a:, ɔ, u/ spoken in the often-used 'null' context /h-t/ by the same 20 speakers, which can serve as reference vowels.

In the fourth group (indicated with a "4" in Figures 2 and 3), where vowel reduction is strongest, the /ɪ/ from "miljoen" approaches an /ɔ/-like quality (or even an /u/-like quality), whereas the /ɔ/ from "bioscoop" approaches an /ɪ/-like quality. In Van Bergem (1994) a model of the schwa is described that can predict the spectral quality of schwas in various phonemic contexts. The formant frequencies of the reduced /ɪ/ from "miljoen" and the reduced /ɔ/ from "bioscoop" in the fourth group are close to the schwa positions that are predicted by this model (for the specific contexts /məljun/ and /bijəsko:p/). These positions are indicated with a cross in the figures.

In Van Bergem (1995a) the schwa is described as *a vowel without target that is completely assimilated with its phonemic context*. In this view, acoustic vowel reduction should thus be interpreted as partial contextual assimilation. Centralization is not a phenomenon in its own right, but merely a by-product of contextual assimilation. That is, several forms of contextual assimilation lead to a more central position of the formant frequencies of vowels (particularly of F_2), but certainly not all. Figures 2 and 3 clearly show that centralization is out of the question in these examples.

Recently, Kondo (1994) suggested that the schwa is targetless in F_2, but targeted in F_1. We do not agree with this view. We think that a schwa is always produced with a minimal amount of opening similar to the amount of opening that is required for the high vowels /i, y, u/. The slightly lower F_1-values for these vowels (see Figures 2 and 3) are probably caused by the front tongue position in the case of the /i/ and the lip rounding in the case of the /y/ and the /u/ (Pickett, 1980). Even in the case of a minimal amount of opening for vowels (clearly *some* opening is required to produce vowel-like sounds), F_1-tracks are often parabolically curved, because the consonants surrounding the vowel are usually produced with narrow constrictions or complete closures in the vocal tract. The curving of F_1-tracks in the case of schwas thus does not necessarily point to the existence of an F_1-target as suggested by Kondo, but simply reflects the higher degree of opening for vowels (including the schwa) compared to consonants.

The vowels from the words "miljoen" and "bioscoop" were also presented to a group of 20 listeners in their original word context. Their task was to identify the vowels of interest. If listeners would have a *static* perception strategy, based on formant frequencies at the vowel nucleus, a number of /ɔ/-responses could be expected in the case of "miljoen" and a number of /ɪ/-responses in the case of "bioscoop". However, none of the 20 listeners ever responded with an /ɔ/ (or an /u/, or an /o:/) in the case of "miljoen" and none of the listeners ever responded with an /ɪ/ (or an /i/, or an /e:/) in the case of "bioscoop". This means that listeners more likely have a *dynamic* perception strategy; they take the influence of adjacent phonemes into account in their evaluation of vowel quality. Table 1 shows that the percentage of schwa responses given by the 20 listeners increases as the formant frequencies approach the schwa position for that specific context (see Figures 2 and 3), indicating

Table 1. Percentages of schwa responses for the vowels /ɪ/ and /ɔ/ in the 4 reduction stages that are shown in Figures 2 and 3

	% schwa responses			
	Group 1	Group 2	Group 3	Group 4
/ɪ/ from miljoen	5	36	60	69
/ɔ/ from bioscoop	15	43	69	74

that listeners are familiar with the /ɔ/-like quality of the schwa in the context /məljun/ and also with the /ɪ/-like quality of the schwa in the context /bijəsko:p/.

3. A closer look at lexical vowel reduction

3.1 Does lexical vowel reduction exist?

Are listeners able to unambiguously identify a vowel in a word as a full vowel or a schwa? This was one of the questions that was posed in a perception experiment described in Van Bergem (1995b). The answer to this question was: Sometimes they are, but more often they are not. Given the apparent ambiguity of vowels in terms of their classification as a full vowel or a schwa, one may wonder whether lexical vowel reduction exists at all. It could be that speakers in principle always aim to produce full vowels and that schwas emerge as the endpoint of acoustic reduction when the speaker's ambition to actually realize the full vowel is minimal.

Our belief in the existence of lexical vowel reduction stems from diachronic examples of schwa substitution as, for instance, in the French loan word "bəTON" (/bətɔn/, concrete). This word was originally pronounced with a full vowel /e:/, whereas the only 'correct' pronunciation of the first vowel of this word in modern Dutch is a schwa. It is very unlikely that Dutch speakers nowadays produce the word "beTON" with an 'underlying' full vowel /e:/. On the other hand, it is likely that at some transitional stage of the sound change /e:/ –> /ə/, two variants of the word "be'TON" existed: one with a full vowel and one with a schwa. It seems not too bold to suggest that such transitional stages of the sound change 'full vowel –> schwa' (i.e., lexical vowel reduction) also exist for several words at present. One of the clearest examples of a word with two variant forms in modern Dutch, supporting the theory of lexical vowel reduction, is in our opinion the word "bijvoorbeeld" (for example): /bɛivo:rbe:lt/ in full form, and /bəvo:rbe:lt/ in reduced form. In Van Bergem (1995b) it was shown that listeners agree very well in their judgement about which variant form of this word was used by speakers. Unfortunately, such good agreement among listeners was not achieved for many other words, indicating that in general a reliable determination of cases of lexical vowel reduction is hard to make.

3.2 Mental representations

How is lexical vowel reduction represented in a person's mind? In the traditional phonological view on vowel reduction the process is described as

full vowel –> schwa

under some specified conditions; it is assumed that words (or stems) are stored in the mental lexicon according to their most 'ideal' form. We think that this view is rather unlikely, because the phonological rule suggests that speakers 'deliberately' mutilate the ideal word form.

A second option would be that the mental lexicon contains words in a form that is most common, namely the one used in normal conversational speech. In this view lexically reduced vowels would be the norm and only some forms of 'hyperspeech' would lead to the transformation

schwa –> full vowel

The problem with this idea is that speakers cannot know which full vowel should be selected, if only the reduced word form is available.

A third option, that is most likely in our view, was proposed by Solé and Ohala (1991). They claimed that style-dependent variant forms are stored separately in the mental lexicon for words and word sequences that occur frequently. For the case of lexical vowel reduction this would mean that a frequently occurring word such as "mINUUT" (minute) would have (at least) two separate style-dependent templates in the mental lexicon, corresponding to the pronunciation variants /mɪ̄nyt/ and /mənyt/. This would mean that in normal conversational speech speakers 'automatically' select the (most common) reduced word form (/mənyt/). Only when they pay more attention to their speech, speakers tend to use the word form containing the full vowel (/mɪ̄nyt/).

3.3 Relation with the phoneme concept

In the description of lexical vowel reduction the phoneme concept plays an important role, because lexical vowel reduction is in fact the substitution of one vowel phoneme (a full vowel) with another (a schwa). Are phonemes psychologically real, or are they just invented by linguists as a means to conveniently describe language phenomena? There is certainly no unanimity among psycholinguists about the basic input unit of speech production and the basic units of representation in speech perception. (For an overview of the relevant literature, see Boucher, 1994.) Often phonemes and syllables are mentioned as units of representation in speech processing (Cutler, 1992).

Our view on this matter favours a combination of several units. We believe that *words* are the primary building blocks both in speech production and in speech perception, and that phonemes and syllables are secondary building blocks, especially used to model new unknown words. More specifically, we believe that speakers use complete words (or even word groups) retrieved from the mental lexicon as the input to the speech production mechanism and not, for instance, strings of phonemes that were derived from the lexical entries. Similarly, we believe that listeners match the acoustic flow of speech with entire words in their mental lexicon and they do not try to identify strings of phonemes in order to convert these into words. Normally, the relevant units in speech processing are thus words, although people are aware of the existence of syllables and phonemes. These smaller units might play an important role in the perception and production of new words.

To clarify this view with a visual analogy, we think that the image of, for instance, a dog (a word-level unit) is perceived as a whole, although people are well aware of the different parts that a dog is composed of: a head, paws, a tail, etc. (syllable-level units), and at a smaller level ears, eyes, toes, etc. (phoneme-level units). Only when people see an unfamiliar object, they may try to grasp what it is by taking a closer look at its composing parts.

Although we regard words as primary building blocks in speech processing, we certainly do not exclude the existence of building blocks that exceed the word level. Idiomatic expressions or frequently occurring word sequences may be stored as single composite templates. This would explain why, for instance, assimilation processes across word boundaries (which arise as a natural consequence of articulatory constraints) are much stronger in frequently occurring word sequences than in word sequences that occur rarely (Solé and Ohala, 1991). Actually, the way in which the mental lexicon is organized may depend on the specific structure of a language, and it

may also depend on the (il)literacy of the language users (Aitchison, 1994). The state of affairs presented above may apply particularly to Western languages.

Evidence for the psychological reality of phonemes, whether as primary or as secondary building blocks in speech, is especially given by analyses of spontaneous errors in normal and aphasic speech (Fromkin, 1980). Vowels pronounced in isolation may be a reflection of the 'ideal' form in which these phonemes are mentally stored. Our interpretation of the schwa as a *targetless* vowel poses an interesting problem in this respect: How can a vowel without a target be mentally stored? People who claim to be able to produce a schwa in isolation are in our view actually producing the central vowel /œ/. (The same 'borrowing' of a central full vowel is done when people have to emphasize the Dutch articles "de" (/də/, the), and "een" (/ən/, a).) In Van Bergem (1995a) it was concluded that a schwa is produced (and identified) as a 'direct' articulatory path between two consonants. Since such a path is clearly dependent upon the specific nature of the surrounding consonants, the schwa cannot exist as a separate phoneme, but only as a phonemic element embedded in a syllabic structure.

4. Vowel reduction and word stress

4.1 Stress versus 'stressedness'

Stress is an *abstract* linguistic property of words. Each word has just one syllable with primary stress. In words containing three syllables or more, other syllables may have secondary stress. The remaining syllables are unstressed. If a word bears a sentence accent, the syllable with primary stress is realized with a pitch movement that causes the perception of accent ('t Hart et al., 1990). The stressed syllable in a word can thus fairly easily be detected if the word is pronounced in isolation, because in that case the one-word utterance gets a clear sentence accent.

Apart from studying word stress at an abstract linguistic level, one can also study the *realization* of stress in actual speech. In a study with words pronounced in isolation, Fry (1958) found that pitch movement was the most salient feature in the acoustic speech signal to represent word stress. In addition, the pitch cue turned out to be the most important perceptual correlate of stress. However, as mentioned above, one-word utterances automatically get a sentence accent, so it is not unlikely that the pitch cue in Fry's experiment was more related to accentuation than to word stress.

This assumption is confirmed by an investigation of Waibel (1988) who studied the relative impact of several acoustic cues on the automatic detection of word stress in the normal connected speech of 5 speakers (3 male and 2 female). It appeared that measures derived from pitch only marginally contributed to the detection of word stress; measures based on energy turned out to be most important. It also appeared that the acoustic realization of word stress could be better described on a gradual scale of 'stressedness' than by the dichotomy stress versus non-stress. Waibel also studied the perception of word stress. He found that the stress judgements made by linguistically untrained subjects varied widely, indicating that it is a rather subjective percept. Furthermore, the listeners tended to focus on sentence (pitch) accents rather than word stress. That pitch movements can be efficient cues for prominence is not surprising if one realizes that from a psycho-acoustic point of view only a few percent change in F_0 is sufficient to be supraliminal, whereas in actual speech F_0 changes are usually about ten times as large as these threshold values ('t Hart et al., 1990).

Following Lehiste and Peterson (1959), we believe that the realization of word stress is directly related to the physiological effort put into the production of a syllable. The amount of effort is reflected in four acoustic parameters: energy, pitch,

segmental duration, and phonetic quality (which includes the phenomenon of vowel reduction). The pitch is raised as a natural consequence of increased vocal effort (Picket, 1980). In fact, the practice of using pitch accents to mark words 'in focus' (and the use of pitch accents in tone languages as well) may very well originate from an 'exaggeration' of the pitch movements which naturally occur as a consequence of varying degrees of vocal effort in utterances. There are three factors that determine to a large extent the amount of effort put in the production of a syllable. These factors are the linguistic stress level, the speaking style, and the frequency of occurrence of the word in which the syllable occurs.

An important conclusion that comes forward from the discussion above is that a clear distinction should be made between the abstract *linguistic* notion of word stress and the actual *realization* of word stress in normal connected speech. The often-heard claim that function words are 'unstressed' is incorrect from a linguistic point of view. Just as any content word, function words have one syllable with primary stress, namely the one that is the potential bearer of a sentence accent. However, if function words are uttered in normal connected speech without a sentence accent, the degree of 'stressedness' in the realization of their stressed syllables is usually rather low (Waibel, 1988). Is this caused by their special grammatical status, or rather by their high frequency of occurrence?

4.2 The role of frequency of occurrence of words

Van Coile (1987) compared the performance of two possible predictors of vowel duration (which is one of the acoustic correlates of stress). These predictors were word class (function words versus content words) and frequency of occurrence of the word in which the vowel occurred (either above or below a specific optimal frequency value). It appeared that frequency of occurrence outperformed word class as a predictor of vowel duration.

Within the class of content words there is also a clear effect of the frequency of occurrence of words, both with respect to vowel duration and vowel quality (Van Bergem, 1995a). It is very likely that within the class of function words this frequency effect applies as well. Thus, function words which are rarely used will be pronounced with more articulatory effort than frequently occurring function words. So it seems that the degree of effort in the realization of words in normal speech is directly related to their frequency of occurrence and is not dependent on their word class.

If we focus on syllables, the observations made above can be summarized as follows: The degree of physiological effort ('stressedness') put in the production of a syllable is not dependent on the linguistic class it belongs to, but it is dependent on:

- its linguistic stress level (primary stress, secondary stress, or unstressed)
- the frequency of occurrence of the word in which the syllable occurs
- the speaking style in which the word is uttered.

The relation between 'stressedness' and the frequency of occurrence of words is schematically illustrated in Figure 4 for syllables with primary stress, secondary stress, and no stress; the exact shape of the curves is dependent on the speaking style. As an example, the figure indicates that the same degree of 'stressedness' applies to:

- syllables without stress occurring in words with frequency x_1
- syllables with secondary stress occurring in words with frequency x_2
- syllables with primary stress occurring in words with frequency x_3

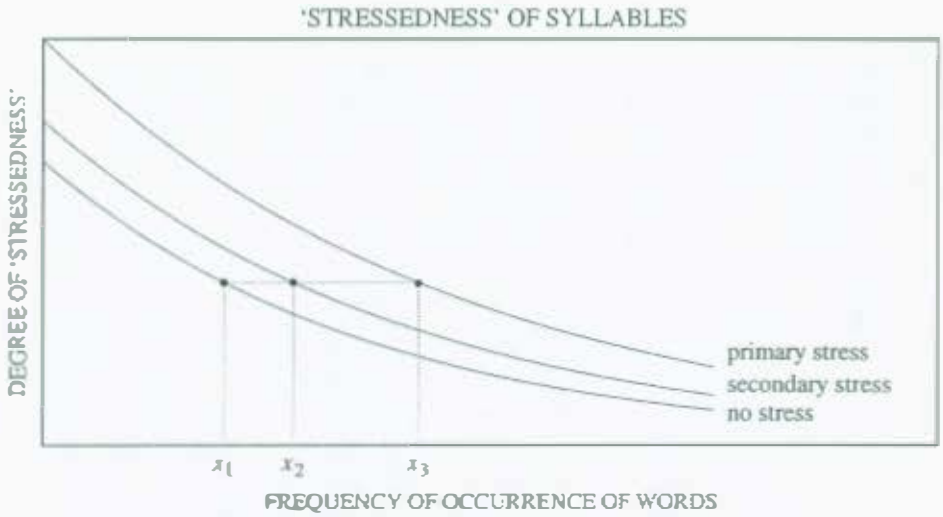

Figure 4. Schematic picture of the 'stressedness' of syllables as a function of their linguistic stress level (primary, secondary, unstressed) and the frequency of occurrence of the words in which they occur. The exact shape of the curves is dependent on the speaking style.

With respect to duration and spectral quality of vowels (both aspects of 'stressedness'), such an effect was demonstrated by Van Bergem (1993) for syllables with primary stress and syllables without stress. The syllables with primary stress were frequently used function words such as "kan" (/kɑn/, can), "moet" (/mut/, must), and "veel" (/ve:l/, much), whereas their unstressed counterparts occurred in less frequently used content words: "kantine" (/kɑntinə/, canteen), "weemoed" (/ʋe:mut/, melancholy), and "veelvuldig" (/ʋe:lʋœldəχ/, frequent). According to the Dutch lexical database CELEX (1990), the function words occur on the average about 500 times more frequently than the corresponding content words. All words were embedded in sentences and were read aloud by 15 speakers. It appeared that the duration and spectral quality of vowels in the syllables with primary stress (the monosyllabic function words) were comparable to the duration and spectral quality of vowels in the unstressed syllables from the content words.

5. Competence versus performance

5.1 Problems with competence models

Several decennia ago Noam Chomsky initiated a linguistic tradition that describes the structure of a language by means of formal recursive rules. The 'generative paradigm' that Chomsky introduced is still by far the most widely used basis of modern linguistic research, although several variants of Chomsky's theory have emerged in the course of time.

A central concept in the generative grammar is the distinction between the *competence* and the *performance* of a language user. Performance is what a speaker or hearer actually does, whereas competence is abstract knowledge of the language system, which can be interpreted as the *potential* performance of an idealized speaker or hearer (Chomsky and Halle, 1968). In other words, performance can be directly observed, whereas competence is 'hidden' in a person's mind. Chomsky's distinction 'competence-performance' corresponds closely with the distinction 'langue-parole'

that was made by De Saussure (1916). The aim of linguistic research in Chomsky's tradition is to discover the competence of language users by ignoring 'irrelevant detail' in their performance.

Scha (1990; 1992) mentions several problems that competence grammars run into. The most relevant ones for the present discussion are:

- *the problem of 'correctness'*
  Whereas the output of a formal grammar can be either 'correct' or 'incorrect', human language users do not have such a strict judgement of linguistic phenomena. For example, sentences can be judged as "almost grammatical", or "unusual", etc. Or, the substitution of a particular vowel in a word with a schwa can be judged as "unusual". In addition, different language users can have different judgements on linguistic phenomena.
- *the problem of language change*
  If a language is seen as a consistent and complete mathematical system at any point in time, it is hard to understand how such a system can be subject to change.
- *the problem of integrating competence and performance*
  A formal description of the competence of language users only partly accounts for actual language behaviour. Especially spoken language is more often than not characterized by imperfections and plain errors. The psycholinguistic processes, that should account for the gap between a competence grammar and real language use, usually do not get much attention in linguistic research.

Disregarding a performance theory poses methodological problems in our view. As mentioned before, the performance of language users can be directly observed, whereas the competence of language users has to be deducted from their performance by ignoring 'irrelevant detail'. However, clear criteria to define what 'irrelevant detail' is can only be established through an integrated study of competence and performance. Furthermore, a theory of competence without a complementary theory of performance is very hard to falsify, because all possible differences between the output of a competence model and actual language output can always be ascribed to the missing performance component.

5.2 Learning by example

Instead of focusing on a competence model with abstract linguistic rules, Scha (1990, 1992) focuses on concrete language data. According to Scha, all lexical elements, syntactic structures, and 'constructions' that a language user encountered *together with* their frequency of occurrence, can influence the processing of new input and output. By looking at the frequency of occurrence of language phenomena, the statistical properties of real language use are taken into account. Parsing of a new input sentence does not occur through application of grammatical rules, but through a matching process that tries to find an optimal analogy between the input sentence and a 'corpus' of already parsed sentences. Since each individual has a different inventory of language experiences, the performance grammar of each individual is different, which can account for differences in grammaticality judgements. A computer model of Scha's performance model using conventional rule-based parsing strategies showed encouraging results (Bod, 1992; 1993). In Scholtes and Bloembergen (1992) an implementation of Scha's model with neural networks is proposed.

Scha is one of the researchers who believes in the general principle of 'learning by example'. A relatively new approach to language (and cognitive processes in general) based on this principle is the connectionist theory (e.g., Kohonen, 1984; Rumelhart et

al., 1986; Elman, 1990). In the connectionist approach an attempt is made to simulate the structure of the human brain with mathematical models. The human brain contains billions of neurons and many more billions of synapses connecting them. Such 'neural networks' are simulated in the connectionist computer models. With these models an arbitrary pattern presented at the input nodes can be mapped to an arbitrary pattern at the output nodes by adjusting the 'weights' of the connections between the nodes. Positive weights indicate excitatory connections and negative weights indicate inhibitory connections. With the aid of simple algorithms the weights of connections can automatically be adjusted for optimal pattern recognition by presenting a large number of 'training' examples to the network.

Though up to now the connectionist models only have a weak resemblance to the complex chemical processes in the human brain, and though their size is still very limited (typically in the order of a few hundred nodes), their pattern recognition performance is remarkable. Neural networks can, for instance, learn to recognize the spectral pattern of the consonant and vowel part in the spoken syllables /bi, ba, bu, di, da, du, gi, ga, gu/ (Elman and Zipser, 1988); they can learn the past tenses of English verbs (Rumelhart and McClelland, 1986); or they can learn simple semantic relations between words (Ritter and Kohonen, 1989). In all these examples the networks do not need any explicit rule (whether phonetic, syntactic, or semantic) to perform their task. The relevant information is directly extracted from the training data and encoded in the weights of the connections between the network nodes. There are other (stochastic) models for language processing that can be trained with examples, such as Hidden Markov Models (e.g., Rabiner and Juang, 1986; Van Alphen and Van Bergem, 1989), but these models are purely mathematical and do not have any relation with the organization of the human brain.

5.3 A 'performance-based' approach to vowel reduction

Our view on the phenomenon of vowel reduction has much affinity with a 'performance-based' linguistic approach. As shown by Van Bergem (1995a), some very important factors related to the occurrence of vowel reduction are word stress, frequency of occurrence of words, and speaking style. Are these factors related to the competence or to the performance of language users?

In a competence model in Chomsky's tradition reduction phenomena are not incorporated in the lexicon; vowel reduction is obtained through a set of rules that can predict when the substitution 'full vowel –> schwa' occurs. Most competence models of vowel reduction take mainly the role of word stress into account; vowels in syllables without primary or secondary word stress are reduced to schwa (usually under some extra restrictions with respect to, for instance, vowel type or the position of the vowel in the word). It will be clear that such a competence model only accounts for lexical vowel reduction, because acoustic vowel reduction is a gradual process. Frequency of occurrence of words and speaking style influence lexical vowel reduction in a probabilistic manner: as the frequency of occurrence of words becomes higher or as the speaking style becomes more casual, lexical vowel reduction becomes more plausible. Because of their gradual nature these factors should probably somehow be incorporated in the performance component of the grammar. However, it is unclear how factors related to performance and competence should be integrated.

According to Chomsky and Halle (1968), the phonological component of a grammar consists of a sequence of ordered rules that convert initial classificatory (binary) representations into final phonetic (numerical) ones. In the case of vowel reduction the initial binary representation would be either full vowel or schwa. Adding phonetic

detail to full vowels seems possible (and desirable to account for their sometimes schwa-like character). On the other hand, it seems rather difficult to convert a schwa back to a (possibly schwa-like) full vowel by adding phonetic detail to it, which implies that the initial choice for a schwa at the competence level is irreversible. Hence, it is of crucial importance that the initial choice is correct, meaning that all possible influences on vowel reduction should be taken into account, including performance factors. Of course, this would make the distinction competence-performance superfluous. The same argument applies analogously to other phonological processes, such as assimilation, segment deletion, etc.

An alternative view, that we favour, would be that words are stored in the mental lexicon *including* phonetic detail. Such a view was advocated by, for instance, Vennemann (1972; 1974), Klatt (1979; 1989), and Van Bergem (1990a). We assume that there exists one type of word templates as a sort of 'acoustic echoes' for speech perception, and that there exists another type of word templates as a series of motor commands for speech production. In principle, the articulatory form of a word could be extracted from the auditory representation in the course of speech. In this case there would be no need for articulatory templates. However, if everyone would only pronounce words in the way they were perceived, it would be hard to understand how the sound structure of words could ever change. Therefore, we assume that apart from an auditory representation of a word, there also exists an articulatory word template that can be adjusted ('trained') in three different ways. In the first place the motor commands for an articulatory word template can 'mimic' the corresponding auditory word template; in the second place the motor commands for an articulatory word template can gradually become lax when a person uses the word very frequently; and in the third place the pronunciation of a word can more or less consciously be adapted by a group of persons to mark their social unity.

The second kind of adjustment can account for the sound change 'full vowel –> schwa'. That is, if a particular word is frequently used in a language community, people start pronouncing it with some more laxness. Subsequently, the corresponding auditory word template is adjusted, because newly perceived realizations of the word are characterized by a more lax pronunciation. The adjusted auditory word template will in turn reinforce a lax pronunciation of the word of interest through the 'mimicking' process. This spiralling could go on for some time and might eventually result in a sound change. In the case of the sound change 'full vowel –> schwa', the lax pronunciation would lead to a gradually increasing amount of acoustic vowel reduction, which could ultimately result in a smooth conversion into a lexical schwa.

As proposed by Solé and Ohala (1991), variant style-dependent forms of word templates may exist; perhaps auditory word templates also have variant forms for men, women, and children (Van Bergem et al., 1988). Since different individuals have different language experiences, their word templates also differ to some extent. Thus, bakers may, for instance, frequently use the word "yeast", and hence this word may be pronounced differently by bakers than by non-bakers. Different personal language experiences also explain different intuitions that people might have about the 'correctness' of word pronunciations.

The here presented 'performance-based' view preserves the strong relation between acoustic and lexical vowel reduction and also offers a plausible explanation for the sound change 'full vowel –> schwa'. We think that word stress is only implicitly present in the word templates, namely through more salient acoustic properties in auditory word templates and more definite motor commands in articulatory word templates. There is in principle no reason that vowel reduction should not occur in stressed syllables. That is, syllables are not marked as '+stress' to block the

occurrence of vowel reduction. Thus, in frequently used words vowel reduction can also occur in stressed syllables, as explained in section 4.2.

6. Vowel reduction as a sound change

What is the relation between acoustic vowel reduction and lexical vowel reduction? As already mentioned in section 5.3, we interpret acoustic and lexical vowel reduction as two intermediate stages in the process of the sound change 'full vowel –> schwa'. A vowel that is often subject to a strong acoustic reduction in a particular word may be confused with a schwa by listeners. In a next step, listeners may start pronouncing the word of interest with a schwa (lexical vowel reduction), which can be the initiation of a sound change. According to the linguist Caron (1972), the Dutch schwa was much less frequently used some centuries ago than nowadays. Perhaps the schwa did not even exist at all in the early days of the Dutch language. In modern Dutch almost one out of three vowels is a schwa.

Our phonetic investigations have concentrated on vowel reduction in Dutch. The results can probably to a large extent be generalized to most Western languages. However, it is not clear to what extent the results of our study apply to non-Western languages as, for instance, Mandarin Chinese (an 'isolating' tone-language), or Turkish (an 'agglutinative' language). Furthermore, the (il)literacy of language users may play a role in the occurrence of (lexical) vowel reduction. Our interpretation of vowel reduction as a natural articulatory tendency suggests that it could be a universal phenomenon. Of course, the sound change 'full vowel –> schwa' can only occur when the schwa is included in the phonological system of a language. If this is not the case (as, for instance, in Italian), the process does not come beyond the stage of acoustic vowel reduction.

7. Future research

Vowel reduction is in fact just one of many kinds of 'impoverishment' that can occur in the acoustic speech signal. In normal conversational speech it is, for instance, not unusual that vowels are completely absent, just as consonants, syllables, or even entire words. Nevertheless, this frequently occurring 'impoverishment' in the speech signal hardly seems to be a problem for listeners, probably because they have several 'higher' sources of information at their disposal to recover the speech message. Perhaps the 'impoverishment' of the speech signal is to some extent even *necessary* to get natural and intelligible speech. This might in part explain the inferiority of synthetic speech, which often sounds overarticulated.

In future research the influence of different degrees of vowel reduction on the naturalness and the intelligibility of speech could be investigated. With the term naturalness we refer to a broad range of human judgements about speech, such as its pleasantness, accessibility, appropriateness with respect to the speech style, etc. This could be done by studying the perceptual effect of systematically interchanging reduced and unreduced vowels in *natural* speech, for instance, by means of the TD-PSOLA-technique (Moulines and Charpentier, 1990; Laan et al., 1991). Alternatively, different degrees of vowel reduction could be introduced in *synthetic* speech by systematically adjusting the synthesis parameters for vowels, in order to investigate how this would influence the naturalness and the intelligibility.

Not only formant frequencies and vowel durations could be varied as synthesis parameters, but also formant *bandwidths*. Studies on formant bandwidths are very

detail to full vowels seems possible (and desirable to account for their sometimes schwa-like character). On the other hand, it seems rather difficult to convert a schwa back to a (possibly schwa-like) full vowel by adding phonetic detail to it, which implies that the initial choice for a schwa at the competence level is irreversible. Hence, it is of crucial importance that the initial choice is correct, meaning that all possible influences on vowel reduction should be taken into account, including performance factors. Of course, this would make the distinction competence-performance superfluous. The same argument applies analogously to other phonological processes, such as assimilation, segment deletion, etc.

An alternative view, that we favour, would be that words are stored in the mental lexicon *including* phonetic detail. Such a view was advocated by, for instance, Vennemann (1972; 1974), Klatt (1979; 1989), and Van Bergem (1990a). We assume that there exists one type of word templates as a sort of 'acoustic echoes' for speech perception, and that there exists another type of word templates as a series of motor commands for speech production. In principle, the articulatory form of a word could be extracted from the auditory representation in the course of speech. In this case there would be no need for articulatory templates. However, if everyone would only pronounce words in the way they were perceived, it would be hard to understand how the sound structure of words could ever change. Therefore, we assume that apart from an auditory representation of a word, there also exists an articulatory word template that can be adjusted ('trained') in three different ways. In the first place the motor commands for an articulatory word template can 'mimic' the corresponding auditory word template; in the second place the motor commands for an articulatory word template can gradually become lax when a person uses the word very frequently; and in the third place the pronunciation of a word can more or less consciously be adapted by a group of persons to mark their social unity.

The second kind of adjustment can account for the sound change 'full vowel –> schwa'. That is, if a particular word is frequently used in a language community, people start pronouncing it with some more laxness. Subsequently, the corresponding auditory word template is adjusted, because newly perceived realizations of the word are characterized by a more lax pronunciation. The adjusted auditory word template will in turn reinforce a lax pronunciation of the word of interest through the 'mimicking' process. This spiralling could go on for some time and might eventually result in a sound change. In the case of the sound change 'full vowel –> schwa', the lax pronunciation would lead to a gradually increasing amount of acoustic vowel reduction, which could ultimately result in a smooth conversion into a lexical schwa.

As proposed by Solé and Ohala (1991), variant style-dependent forms of word templates may exist; perhaps auditory word templates also have variant forms for men, women, and children (Van Bergem et al., 1988). Since different individuals have different language experiences, their word templates also differ to some extent. Thus, bakers may, for instance, frequently use the word "yeast", and hence this word may be pronounced differently by bakers than by non-bakers. Different personal language experiences also explain different intuitions that people might have about the 'correctness' of word pronunciations.

The here presented 'performance-based' view preserves the strong relation between acoustic and lexical vowel reduction and also offers a plausible explanation for the sound change 'full vowel –> schwa'. We think that word stress is only implicitly present in the word templates, namely through more salient acoustic properties in auditory word templates and more definite motor commands in articulatory word templates. There is in principle no reason that vowel reduction should not occur in stressed syllables. That is, syllables are not marked as '+stress' to block the

occurrence of vowel reduction. Thus, in frequently used words vowel reduction can also occur in stressed syllables, as explained in section 4.2.

6. Vowel reduction as a sound change

What is the relation between acoustic vowel reduction and lexical vowel reduction? As already mentioned in section 5.3, we interpret acoustic and lexical vowel reduction as two intermediate stages in the process of the sound change 'full vowel –> schwa'. A vowel that is often subject to a strong acoustic reduction in a particular word may be confused with a schwa by listeners. In a next step, listeners may start pronouncing the word of interest with a schwa (lexical vowel reduction), which can be the initiation of a sound change. According to the linguist Caron (1972), the Dutch schwa was much less frequently used some centuries ago than nowadays. Perhaps the schwa did not even exist at all in the early days of the Dutch language. In modern Dutch almost one out of three vowels is a schwa.

Our phonetic investigations have concentrated on vowel reduction in Dutch. The results can probably to a large extent be generalized to most Western languages. However, it is not clear to what extent the results of our study apply to non-Western languages as, for instance, Mandarin Chinese (an 'isolating' tone-language), or Turkish (an 'agglutinative' language). Furthermore, the (il)literacy of language users may play a role in the occurrence of (lexical) vowel reduction. Our interpretation of vowel reduction as a natural articulatory tendency suggests that it could be a universal phenomenon. Of course, the sound change 'full vowel –> schwa' can only occur when the schwa is included in the phonological system of a language. If this is not the case (as, for instance, in Italian), the process does not come beyond the stage of acoustic vowel reduction.

7. Future research

Vowel reduction is in fact just one of many kinds of 'impoverishment' that can occur in the acoustic speech signal. In normal conversational speech it is, for instance, not unusual that vowels are completely absent, just as consonants, syllables, or even entire words. Nevertheless, this frequently occurring 'impoverishment' in the speech signal hardly seems to be a problem for listeners, probably because they have several 'higher' sources of information at their disposal to recover the speech message. Perhaps the 'impoverishment' of the speech signal is to some extent even *necessary* to get natural and intelligible speech. This might in part explain the inferiority of synthetic speech, which often sounds overarticulated.

In future research the influence of different degrees of vowel reduction on the naturalness and the intelligibility of speech could be investigated. With the term naturalness we refer to a broad range of human judgements about speech, such as its pleasantness, accessibility, appropriateness with respect to the speech style, etc. This could be done by studying the perceptual effect of systematically interchanging reduced and unreduced vowels in *natural* speech, for instance, by means of the TD-PSOLA-technique (Moulines and Charpentier, 1990; Laan et al., 1991). Alternatively, different degrees of vowel reduction could be introduced in *synthetic* speech by systematically adjusting the synthesis parameters for vowels, in order to investigate how this would influence the naturalness and the intelligibility.

Not only formant frequencies and vowel durations could be varied as synthesis parameters, but also formant *bandwidths*. Studies on formant bandwidths are very

scarce, probably because it is difficult to obtain good estimates of them (Klatt, 1980). It is generally assumed that formant bandwidths are relatively unimportant in the perception of vowel qualities, which is probably the reason that many synthesis systems use a fixed value for formant bandwidths (or a fixed percentage of the formant frequency). Van Bergem (in preparation), however, found indications that the formant bandwidths of reduced vowels are much larger than those of unreduced vowels, especially regarding F_1. In a study of the naturalness and intelligibility of synthetic speech it thus seems worthwhile to also investigate the role of formant bandwidths.

Although superficially it seems plausible that the intelligibility of speech is optimal if only 'ideal' vowels occur, this is not generally true. Listeners often *expect* reduced vowels in specific words, and in these cases vowel reduction can be a cue for the proper identification of these words. This becomes most apparent in minimal word pairs that only differ in their stress pattern, such as the English words "reBEL" and "REbel". Apart from cues provided by context and prosody, vowel reduction can be an extra cue to disambiguate these words. The intelligibility of speech is thus probably only optimal, if all vowels have a 'proper degree of reduction'.

References

Aitchison, J. (1994). *Words in the mind. An introduction to the mental lexicon*, Blackwell, Oxford.

Bod, R. (1992). "A computational model of language performance: Data oriented parsing", *Proceedings of the International Conference on Computational Linguistics*, Nantes: 855-859.

Bod, R. (1993). "Using an annotated corpus as a stochastic grammar", *Proceedings of the European Chapter of the Association for Computational Linguistics*, Utrecht: 37-44.

Boucher, V.J. (1994). "Alphabet-related biases in psycholinguistic enquiries: Considerations for direct theories of speech production and perception", *Journal of Phonetics* 22: 1-18.

Caron, W.J.H. (1972). "De reductievocaal in het verleden", In: *Klank en teken: verzamelde taalkundige studies*, Groningen: 131-146.

CELEX (1990). *A program for retrieval of lexical information (for Dutch, English, German)*. Centre for lexical information, University of Nijmegen.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English*, Harper & Row, New York.

Cutler, A. (1992). "Psychology and the segment", In: Kingston, J. & Beckman, M.E. (Eds.), *Papers in Laboratory Phonology II*, Cambridge University Press: 290-295.

De Saussure, F. (1916). *Cours de linguistique generale*, Lausanne.

Elman, J.L. & Zipser, D. (1988). "Learning the hidden structure of speech", *Journal of the Acoustical Society of America* 83: 1615-1626.

Elman, J.L. (1990). "Representation and structure in connectionist models", In: Altman, G.T.M. (Ed.), *Cognitive models of speech processing. Psycholinguistic and computational perspectives*, MIT Press, Cambridge, Massachusetts: 345-382.

Fant, G. (1960). *Acoustic theory of speech production*, Mouton & Co., The Hague.

Fromkin, V.A. (Ed.) (1980). *Errors in linguistic performance: Slips of the tongue, ear, pen and hand*, Academic Press, New York.

Fry, D.B. (1958). "Experiments in the perception of stress", *Language and Speech* 1: 126-152.

Harmegnies, B. & Poch-Olivé, D. (1992). "A study of style-induced vowel variability: Laboratory versus spontaneous speech in Spanish", *Speech Communication* 11: 429-437.

Hart, J. 't, Collier, R. & Cohen, A. (1990). *A perceptual study of intonation. An experimental-phonetic approach to speech melody*, Cambridge University Press.

Klatt, D.H. (1979). "Speech perception: a model of acoustic-phonetic analysis and lexical access", *Journal of Phonetics* 7: 279-312.

Klatt, D.H. (1980). "Software for a cascade/parallel formant synthesizer", *Journal of the Acoustical Society of America* 67: 971-995.

Klatt, D.H. (1989). "Review of selected models of speech perception", In: Marslen-Wilson, W.D. (Ed.), *Lexical representation and process*, MIT Press, Cambridge, Massachusetts: 169-226.

Kohonen, T. (1984). *Self-organization and associative memory*, Springer-Verlag, Berlin.

Kondo, Y. (1994). "Phonetic underspecification in schwa", *Proceedings of the International Conference on Spoken Language Processing*, Vol. 1: 311-314.

Laan, G.P.M., Van Bergem, D.R. & Koopmans-van Beinum, F.J. (1991). "The importance of spectral quality of vowels for the intelligibility of sentences", *Proceedings of Eurospeech '91*, Genua: 1129-1132.

Lehiste, I. & Peterson, G.E. (1959). "Vowel amplitude and phonemic stress in American English", *Journal of the Acoustical Society of America* 31: 428-435.

Moulines, E. & Charpentier, F. (1990). "Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones", *Speech Communication* 9: 453-467.

Pickett, J.M. (1980). *The sounds of speech communication: A primer of acoustic phonetics and speech perception*, University Park Press, Baltimore.

Rabiner, L.R. & Juang, B.H. (1986). "An introduction to Hidden Markov Models", *IEEE Transactions on Acoustics, Speech, and Signal Processing* 3: 4-16.

Ritter, H. & Kohonen, T. (1989). "Self-organizing semantic maps", *Biological Cybernetics* 61: 241-254.

Rumelhart, D.E. & McClelland, J.L. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, MIT Press, Cambridge, Massachusetts, Vol. 1 and 2.

Rumelhart, D.E. & McClelland, J.L. (1986). "On learning the past tenses of English verbs", In: Rumelhart, D.E. & McClelland, J.L. (Eds.). *Parallel distributed processing: Explorations in the microstructure of cognition*, MIT Press, Cambridge, Massachusetts, Vol. 2: 216-271.

Scha, R. (1990). "Taaltheorie en taaltechnologie: competence en performance", In: De Kort, R. & Leerdam, G.L.J. (Eds.), *Computertoepassingen in de Neerlandistiek*, Almere: LVVN: 7-22.

Scha, R. (1992). "Virtuele grammatica's en creatieve algoritmen", *Gramma/ TTT*, tijdschrift voor taalkunde 1: 57-77.

Scholtes, J.C. & Bloembergen, S. (1992). "The design of a neural data-oriented parsing (DOP) system", *Proceedings of the International Joint Conference on Neural Networks*, Baltimore.

Solé, M.J. & Ohala, J.J. (1991). "The phonological representation of reduced forms", Proceedings of the ESCA Workshop *"Phonetics and phonology of speaking styles: Reduction and elaboration in speech communication"*, Barcelona: 49:1-49:5.

Stålhammar, U., Karlsson, I. & Fant, G. (1973). "Contextual effects on vowel nuclei", *KTH Speech Transmission Laboratory Quarterly Progress and Status Report* 4, Stockholm: 1-18.

Van Alphen, P. & Van Bergem, D.R. (1989). "Markov models and their application in speech recognition", *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 13: 1-26.

Van Bergem, D.R., Pols, L.C.W. & Koopmans-van Beinum, F.J. (1988). "Perceptual normalization of the vowels of a man and a child in various contexts", *Speech Communication* 7: 1-20.

Van Bergem, D.R. (1990). "In defense of a probabilistic view on human word recognition", *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 14: 53-66.

Van Bergem, D.R. (1993). "Acoustic vowel reduction as a function of sentence accent, word stress, and word class", *Speech Communication* 12: 1-23.

Van Bergem, D.R. (1994). "A model of coarticulatory effects on the schwa", *Speech Communication* 14: 143-162.

Van Bergem, D.R. (1995a). *Acoustic and lexical vowel reduction*, Doctoral dissertation, University of Amsterdam, Studies in Language and Language Use, Vol. 16.

Van Bergem, D.R. (in preparation). "Perceptual consequences of adjusting formant bandwidths of synthesized 'reduced' vowels".

Van Bergem, D.R. (1995b). "Perceptual and acoustic aspects of lexical vowel reduction, a sound change in progress", to appear in *Speech Communication* 16.

Van Coile, B.M. (1987). "A model of phoneme durations based on the analysis of a read Dutch text", *Proceedings of the European Conference on Speech Technology, Edinburgh*, Vol. 2: 233-236.

Vennemann, T. (1972). "Phonetic detail in assimilation: Problems in Germanic phonology", *Language* 48: 863-892.

Vennemann, T. (1974). "Words and syllables in natural generative grammar", In: Bruch, A., Fox, R.A. & La Galy, M.W. (Eds.), *Papers from the Parasession on Natural Phonology*, Chicago Linguistic Society, Chicago: 346-374.

Waibel, A. (1988). *Prosody and speech recognition*, Pitman, London.