PHONETIC AND LINGUISTIC ASPECTS OF VOWEL REDUCTION

Dick R. van Bergem

1 INTRODUCTION

Suppose you are having dinner and you want the salt. There are various ways of asking your dish companions for the salt, but the shortest one is calling for "Salt!". Apparently, such a request, that will normally be expressed in several words, can also be said in one single word (if we disregard the aspect of decency). Other examples of condensed messages are telegrams, the speech of young children, or the way you talk in a foreign country when you have little knowledge of the language. These examples suggest that in a natural communicative situation the important information in spoken or written messages is conveyed in only a few key words. If the message is not condensed, it is often not necessary to receive it literally, as long as the key words are properly understood.

If we restrict our attention to spoken messages, we may ask ourselves how much of the acoustic information of an utterance is really needed by the listener to grasp its meaning and how much of it is redundant. In view of what has been mentioned above, the answer to this question must probably be that listeners need little acoustic information to understand a message. In a natural speech situation listeners can use knowledge sources at very different levels: phonetic, phonological, morphological, grammatical, semantic and pragmatic. The use of these knowledge sources is guided by a strategy of *prediction* and *elimination*.

A *prediction* of words is possible, because the listener knows what the conversation is about, so he can focus on a small set of words with a high probability of occurrence for the particular subject of conversation.

Elimination occurs in several stages going alternately 'bottom-up' and 'top-down'. That is, hypotheses of words based on acoustic information may be eliminated on the basis of constraints at a higher knowledge level which leads to the construction of alternative hypotheses. These are compared with the acoustic information which may lead in turn to new hypotheses and so on until the most likely sequence of words is selected.

The process of listening in my view is basically of a probabilistic nature: Prediction of words with a high probability and elimination of words with a low probability, given the acoustic information. Perhaps this explains the success of speech recognition with probabilistic models such as Markov chains (Van Alphen and Van Bergem, this volume), although the approach of these models may differ considerably from the human way of perception.

From the talker's point of view, the most economic way of communicating is to increase his articulatory effort just enough for words that contain important information, i.e. content words, and decrease his articulatory effort for function words. It has been shown (Zue, 1985; McAllister, 1989) that the amount of phonetic richness is much greater in stressed than in unstressed syllables. Therefore phonemes in unstressed syllables provide little lexical constraint compared to those in stressed syllables. This implies that a talker is allowed to restrict his articulatory effort in content words to the

stressed syllables. Apart from this communicative aspect of a talker's articulatory effort, the overall articulatory effort can be influenced by style of speech (formal - informal), speaking rate and social background. The phenomenon of vowel reduction, which is defined here as a reduction of vowel quality, is mostly determined by the talker's striving for an economic way of articulation. In chapter 2 and 3 two manifestations of the reduction phenomenon are discussed and chapter 4 gives some conclusions.

2 ARTICULATORY VOWEL REDUCTION

2.1 What is articulatory vowel reduction?

As we have seen a talker may reduce his articulatory effort under various circumstances. The resulting loss of vowel quality will depend on the way the articulators interact. In order to investigate this *articulatory reduction,* the effect of different consonant contexts on vowel quality was examined in CVC nonsense words (i.e. without linguistic message), spoken with high articulatory effort and with low articulatory effort. This experiment will be described in the following paragraphs.

2.2 Experimental design

All Dutch monophthongs /ɑ, ɔ, ɛ, I, œ, u, y, i, a, o, e, ɸ/ were spoken by one male talker (the author) in two conditions. In the first condition the vowels were pronounced in the 'symmetric' context CəCVCə, a construction that is also used to collect diphone sets. The following consonants were used: /p, t, k, f, s, χ, m, n, r, l, j, w/. In the second condition CVC words were embedded in the carrier sentence "Nu krijgt de CVC een beurt". The sentences were uttered in a monotonous way with a very weak accent on the CVC word. The nonsense words /ror/, /rer/ and /rɸr/ were omitted in both conditions because they are pronounced in the same way as /rɔr/, /rIr/ and /rœr/ (This is not so in meaningful words). The total number of vowels V was therefore 282 (12 vowels x 12 consonants x 2 conditions - 6). All recordings were made in an anechoic room with a Sennheiser MD421N microphone and a Revox A77 tape recorder. A similar experiment was performed one year ago (Van Bergem, 1988), but at that time we didn't find any differences between the two conditions. In the current experiment an attempt was made to pronounce the sentences with a total relaxation of the articulators as in 'normal' speech, whereas the isolated CəCVCə words were pronounced with a maximal articulatory effort.

2.3 Measuring procedure

The isolated CəCVCə words and the sentences were lowpass filtered at 4.5 kHz and digitally stored in the computer with a sample frequency of 10 kHz. All vowels V and their immediate acoustic context were segmented manually using a speech editing program. These segments were analysed with a 25-ms Hamming window that was shifted in steps of 2 ms. Formants were extracted with a 10th-order LPC-analysis, more specifically the Split-Levinson algorithm (Willems, 1986). After formant extraction a computer program was run to display simultaneously the tracks of the first, second and third formant frequencies and the speech waveform. Markers defining the onset and the offset of the vowel were placed manually at positions that showed 'bends' in the

formant tracks and (energy) changes in the speech waveform. Although this was not always easy (especially for the consonant contexts /r, j, w/), all boundaries were chosen with the greatest possible consistency. Subsequently the position of the steady-state vowel part was determined with an automatic procedure. In this procedure a window is moved through three formant tracks (F1, F2, F3) simultaneously. The steady-state vowel part is defined as that vowel part where the pooled within-variance of the tracks is minimal. In order to give less weight to the variance in higher frequencies, a mel-scale is used. More details about this algorithm can be found in Van Bergem and Koopmans-van Beinum (1989).

2.4 Results

Before we started our data analysis all linear formant frequencies were transformed to a mel-scale using the formula

$$m = 2595 \cdot \log(1 + f/700)$$

where f is the formant frequency in Hz and m the mel-converted value (Makhoul and Cosell, 1976). This mel-scale is fairly close to the Bark-scale proposed by Sekey and Hanson (1984).
In a first test we wanted to find out whether the articulatory movements were less in the condition 'sentences' than in the condition 'words'. Therefore we measured the variance *within* the formant tracks of each vowel in both conditions, assuming that this variance would be proportional to the amount of articulatory movement. Table 1 gives the pooled within variance for the formant tracks of all vowels per condition.

Table 1. Pooled within variance in mel^2 for the formant tracks of all vowels per condition (W is the condition 'words' and S the condition 'sentences').

var F1 (mel^2)		var F2 (mel^2)		var F3 (mel^2)	
W	S	W	S	W	S
4023	1398	5067	3283	928	766

The large differences in variance between the condition 'words' and 'sentences' confirm that articulatory movements are considerably reduced in the condition 'sentences'.
In figure 1 the (steady-state) first and second formant frequencies of each vowel averaged over consonant contexts have been plotted. The open triangles indicate the positions of vowels from the condition 'words' and the dark squares those from the condition 'sentences'. The figure shows a considerable shift in formant frequency between vowels from the conditions 'words' and 'sentences'. The first formant frequency of all vowels decreases in the condition 'sentences' except for the high vowels /i, y, u/ where the F1-shift is only marginal. The second formant frequency of the front vowels /i, y, I, e, ɛ, ɸ/ decreases in the condition 'sentences', whereas the F2-shift is only marginal for the central and back vowels /œ, a, ɑ, o, ɔ, u/. The formant shifts in figure 1 are similar to the ones found in an earlier study in which the vowels /ɑ, ɔ, ɛ, I/ from free conversation were compared with those from read sentences (Van Bergem and Koopmans-van Beinum, 1989).

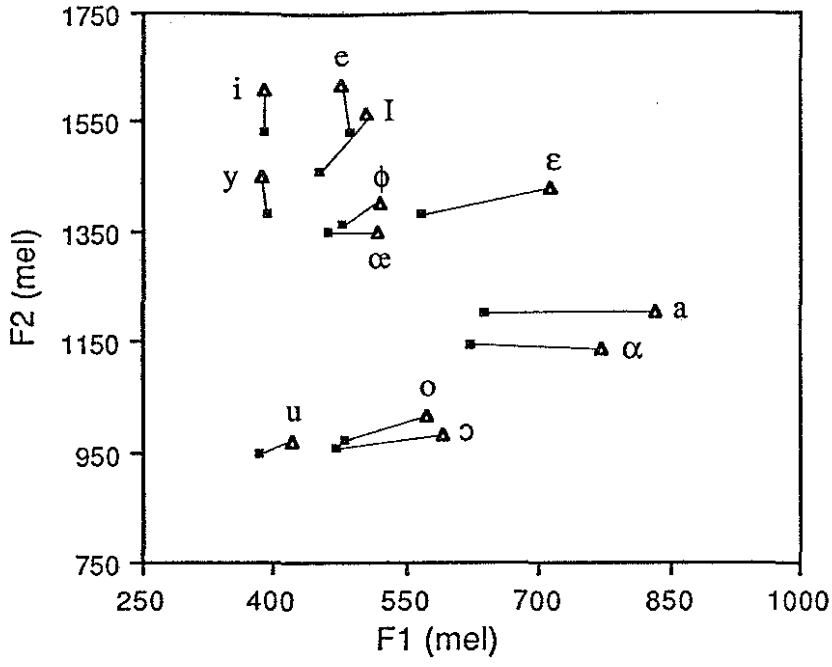

Figure 1. Plot of average (steady-state) formant positions in mel of vowels from the condition 'words' (open triangles) and the condition 'sentences' (dark squares).

2.5 Discussion of articulatory vowel reduction

The results of this experiment show that almost all F1-values decrease in the condition 'sentences' with regard to the condition 'words'. This effect is most apparent for the low vowels /a, ɑ, ɛ/ and only marginal for the high vowels /i, y, u/. This strongly suggests that a relaxation of the jaw movements tends towards a completely *unlowered* jaw. A talker does therefore *not* strive for a *'straight tube'* mouth position, as far as jaw movements are concerned, but rather to a closed mouth. This strategy can be most clearly observed in free conversation where the articulators are maximally relaxed. In Van Bergem and Koopmans-van Beinum (1989) it is shown that the first formant frequency of the vowel /ɑ/ from function words in free conversation can move past the 'neutral' jaw position and end up in the frequency region of 300-400 Hz.
For F2-values such a straightforward interpretation of reduction effects is as yet much more difficult, because the interactions of tongue- and lip-movements that are mainly responsible for F2-shifts are far more complicated. The results of this experiment show that especially F2-values of front vowels are shifted due to the reduction effect in the direction of a 'neutral' (schwa-like) F2-position. The F2-shifts of front vowels may point to a less extreme spreading of the lips (Pickett, 1980).
In Van Bergem (1988) a similar experiment was performed: CVC words were pronounced in isolation and in the carrier sentence "Nu krijgt de CVC een beurt". The sentence accent was placed on the word "beurt" in order to deemphasize the test word. The same vowels V and the same consonants C as in the present investigation were used. However, at that time we didn't find any reduction effects at all. This illustrates that it is far from being easy to study vowel reduction in strictly controlled experiments with nonsense words. The reduction effects found in the present study were obtained by 'imitating' a more spontaneous speech style. Although the present results are similar

to those found in real spontaneous speech (Van Bergem and Koopmans-van Beinum, 1989), this experimental method is not very practical. We therefore plan to continue our research with meaningful words in which reduction occurs in a natural way.

3 LINGUISTIC VOWEL REDUCTION

3.1 What is linguistic vowel reduction?

Compare the following Dutch word pairs

minuut'	mineur'
banaan'	banier'
regentes'	regisseur'

The members of each word pair have the same number of syllables and the same stress pattern. Furthermore, the underlined vowels on both sides occur in the same consonantal context. Therefore these vowel pairs are subject to the same articulatory reduction rules. However, the underlined vowels from words on the left side can be replaced by a schwa (/ə/) without any loss of naturalness, whereas the underlined vowels from words on the right side cannot. Clearly there are linguistic rules specifying this substitution of a 'full' vowel by a schwa and therefore we will call the loss of vowel quality in this way *linguistic vowel reduction.* The crucial difference between this kind of reduction and articulatory reduction is the *intention* of the talker. In the word "minuut" the talker can try to pronounce an /i/ which may then still suffer from a certain loss of quality due to articulatory reduction. On the other hand, the talker can also choose to substitute the /i/ by a schwa.
Whereas all vowels are more or less subject to articulatory reduction, there are only a limited number of words (usually Roman loanwords) in which one or more vowels can be replaced by a schwa. The vowels in for instance the function words "voor", "en", "dat" can definitely not be replaced by a schwa. If these function words are spoken in isolation, the vowels will probably be close to their 'target' position, although in running speech their quality is often strongly affected by reduced articulatory movements. On the other hand, the underlined vowels in the words "academie", "regie" and "terrein" spoken in isolation will probably be pronounced as a schwa by many talkers. This means that the pronunciation of the mentioned function words is stored 'correctly' in a talker's lexicon, whereas the pronunciation of the mentioned content words may be stored in the lexicon in a reduced form (or in several forms that are used in different speech styles).

3.2 The occurrence of linguistic vowel reduction

Under what circumstances does linguistic vowel reduction occur? Before we answer this question, we want to emphasize the diachronic character of linguistic vowel reduction. Each Dutchman pronounces the vowels of the suffixes "-lijk" and "-ig" in for instance the words "vrolijk" and "aardig" as a schwa. However, the linguist Caron argues on the basis of ancient phonetic writings and ancient poetry that these suffixes may have been pronounced with 'full' vowels some centuries ago (Caron, 1972). In fact, he believes that the 'neutral' vowel (/ə/) didn't occur at all in former days. The

French loanword "beton" is probably pronounced as /bətɔn/ by all Dutchmen. However, once it must have been pronounced as /betɔn/, which is still the way Frenchmen pronounce it.
Most Dutch people are probably not aware of the fact that a lot of words they pronounce with a schwa were originally pronounced as 'full' vowels. There is no reason to believe that this language change has stopped. Especially French loanwords are subject to the (linguistic) reduction phenomenon. According to Martin (1968) the vowels /ɛ/ and /e/ in more than 800 French loanwords (he examined a total of about 2800 loanwords) are occasionally or often substituted by a schwa in Dutch. However, in many cases the original unreduced versions of these French loanwords are still known (and used) by a lot of people at the same time. After a transitional stage in which both the unreduced and reduced forms of these words are used, it may happen that only the reduced form will remain like for the above mentioned example "beton". Note that Frenchmen are less inclined to substitute the vowels in the same words by a schwa which clearly indicates the linguistic (language specific) nature of this reduction process.
Are there any circumstances that will cause an increase (or decrease) of the probability that linguistic reduction occurs? The linguists Martin (1968) and Booij (1981, 1982) mention several of these circumstances. The most important ones are:

1. Only vowels in syllables that have no primary or secundary word stress can reduce to a schwa.

2. Vowels at word initial position do not reduce to a schwa.

3. The high vowels /i, y, u/ and the rounded vowels /ɔ, o, u, y/ are less frequently reduced to a schwa than other vowels.

4. Vowels in words that have a high frequency of occurrence have a higher probability to reduce to a schwa.

A remarkable fact is that linguists always restrict their attention to content words in their analysis of vowel reduction. Some highly frequent Dutch function words have two forms of pronunciation, a 'full' one and a reduced one. An example is the possessive pronoun "mijn" (mine) that is usually pronounced as /mən/, unless it has sentence accent. Clearly such a word is an exception on rule 1 mentioned above, because monosyllabic words have by definition primary word stress.
The reason for the linguistic reduction in the words "minuut" and "banaan", mentioned at the beginning of paragraph 3.1, is probably their high frequency of occurrence; the words "mineur" and "banier" are rather uncommon and therefore they are not reduced.
For the third word pair "regentes"-"regisseur", mentioned at the beginning of paragraph 3.1, the diachronic character of linguistic reduction becomes apparent. According to Martin (1968) the linguistic reduction in the word "regentes", a title given to the Dutch queen Emma about a century ago, occurred by analogy with the reduction in the word "regent" (/rəχɛnt/), the title of a Dutch governor in former centuries. Nowadays the words "regent" and "regentes" are hardly used anymore. The word "regisseur", on the other hand, has become a very common word, especially with regard to films, radio and television. However, it is a relatively new word and hasn't suffered from linguistic reduction so far, but this may very well happen in the future.

3.3 Other forms of linguistic vowel reduction

Apart from a substitution of 'full' vowels by a schwa, there are two other forms of linguistic reduction. One is the substitution of the long vowels /a, o, e/ by their short counterparts /a, ɔ, I/, respectively. Here are some examples of Dutch words in which the underlined long vowel is often replaced by its short counterpart:

- papier, machine, categorie
- tolerant, komedie, kolossaal
- secretaris, generaal, telefoon

The second one is the substitution of diphthongs by monophthongs. This kind of reduction is mainly found in dialects, for instance the substitution of /hʌys/ (house) by /hys/ and the substitution of /dɛik/ (dyke) by /dik/.
All these types of reduction can be explained as a striving for an economic way of articulation. Although this is an important phenomenon, it should be noted that there are other tendencies in a natural language that may counteract the reduction process. Every natural language is influenced by social, economic and geographic factors (Dittmar, 1978). This means that certain sociolects or dialects may arise in which for instance a number of short vowels are replaced by long ones or a number of monophthongs by diphthongs in order to establish a group identity. In the dialect of Groningen the word "goed" (/χut/) is for instance pronounced as /χɑut/.

3.4 The relation between linguistic and articulatory vowel reduction

It will be clear that articulatory and linguistic reduction are closely related. Both kinds of reduction occur in vowels that contain relatively little linguistic information. As we have seen in chapter 1 this is especially valid for vowels in unstressed syllables. In addition, words with a high frequency of occurrence can be easier predicted by listeners (see chapter 1), so that a talker is allowed to reduce one or more vowels in these words (in an articulatory way or a linguistic way).
What is the relation between the two kinds of reduction? In my view linguistic reduction is a consequence of articulatory reduction. If 'full' vowels are strongly affected by articulatory reduction, their quality may be degraded to a schwa-like vowel. If the words to which these vowels belong occur frequently, people may eventually decide to abandon their striving for a 'full' vowel and simply substitute it with a schwa. Figure 1 shows that this is less likely to happen to the high vowels /i, y, u/, because their reduced F1-value does not shift towards a 'neutral' (schwa-like) F1-position. It is also less likely to happen to the rounded vowels /ɔ, o, u, y/, because their reduced F2-value does not shift towards a 'neutral' (schwa-like) F2-position. This is in agreement with the observations made by linguists (see paragraph 3.2). The substitution of long vowels by short ones can be explained in a similar way. The duration of long vowels in running speech tends to become smaller in unstressed syllables. This will result in a quality shift towards their short counterparts, which may eventually lead to a lexical substitution.
It is interesting to note that languages such as Japanese and Italian do not have the schwa (/ə/) in their vowel system. This implies that substitution of 'full' vowels with a schwa, i.e. linguistic vowel reduction, cannot exist in such languages. However, the communicative rules specifying the amount of articulatory effort (see chapter 1) are of course valid for *all* languages. Koopmans-van Beinum (1983) and Den Os (1988) have shown that articulatory reduction is also very common in Japanese and Italian.

4 CONCLUSIONS

In this article two forms of vowel reduction were presented: Articulatory vowel reduction and linguistic vowel reduction. The loss of vowel quality due to a low articulatory effort and dependent on the way the articulators interact is called articulatory reduction. The substitution of a 'full' vowel by a schwa, a long vowel by a short one, or a diphthong by a monophthong is called linguistic reduction. It was argued that linguistic reduction is a consequence of articulatory reduction.
In order to increase the naturalness of (Dutch) *speech synthesis* both types of reduction should be carefully examined. Words in which linguistic vowel reduction is common should be stored in the lexicon in their reduced form. Articulatory reduction should be studied in a detailed way, preferably in speech that is as natural as possible. The quality of for instance the vowel /i/ in the word "mimiek" (embedded in a sentence) can be compared pairwise with the quality of the vowel /i/ in the nonsense word /məmimə/, which is one of the building blocks of the Dutch diphone system. Such comparisons should lead to a set of (articulatory) reduction rules for vowel durations and vowel spectra that could contribute to the naturalness of synthesized speech. This line of research is the aim of my future work.
For (Dutch) *speech recognition* it seems preferable to focus on *word* recognition. A vowel that has been (linguistically) reduced to a schwa can only be recognized as a schwa. Vowels of which the quality has been strongly affected by articulatory reduction will almost certainly not be properly recognized. However, this need not be any problem, if recognition at for instance the phoneme level is followed by recognition at the word level. Such an approach is much more robust with regard to (phoneme) distortions. If the word recognition could be supported by higher level knowledge sources (especially semantic and pragmatic sources), computers might eventually be able to reach the same recognition performance as human listeners, despite reduction phenomena.

ACKNOWLEDGEMENTS

I am grateful to Louis Pols, Florien Koopmans-van Beinum and Gitta Laan for their comments on earlier versions of this article.

REFERENCES

Alphen, P. van & Bergem, D. R. van (1989). "Markov models and their application in speech recognition", Proc. of the Institute of Phonetic Sciences Amsterdam 13 (this volume).
Bergem, D. R. van (1988). "The first step to a better understanding of vowel reduction", Proc. of the Institute of Phonetic Sciences Amsterdam 12, 61-75.
Bergem, D. R. van & Koopmans-van Beinum, F. J. (1989). "Vowel reduction in natural speech", Proc. Eurospeech '89, Paris, Vol. 2, 285-288.
Booij, G. E. (1981). Generatieve fonologie van het Nederlands, Utrecht: Het Spectrum, 232 pages.
Booij, G. E. (1982). "Fonologische en fonetische aspecten van klinkerreductie", Spektator 11, 295-301.
Caron, W. J. H. (1972). "De reductievocaal in het verleden", In: Klank en teken; verzamelde taalkundige studies, Groningen, 131-146.

Dittmar, N. (1978). Handboek van de sociolinguïstiek, Utrecht: Het Spectrum, 468 pages.
Koopmans-van Beinum, F. J. (1983). "Systematics in vowel systems", In: M. P. R. van den Broecke, V. J. van Heuven and W. Zonneveld (Eds.), Sound structures: Studies for Antonie Cohen, Foris Publications, Dordrecht, 159-171.
Makhoul, J. & Cosell, L. (1976). "LPCW: An LPC vocoder with linear predictive spectral warping", Proc. ICASSP 1976, 466-469.
Martin, W. (1968). "De verdoffing van gedekte en ongedekte e in niet-hoofdtonige positie bij Romaanse leenwoorden in het Nederlands", Nieuwe Taalgids 61, 162-181.
McAllister, J.M. (1989). "The processing of stressed syllables in connected speech", Proc. Eurospeech '89, Paris, Vol. 2, 686-689.
Os, E. A. den (1988). Rhythm and tempo of Dutch and Italian; a contrastive study, Doct. diss., University of Utrecht, 117 pages.
Pickett, J. M. (1980). "The sounds of speech comunication. A primer of acoustic phonetics and speech perception, University Park Press, Baltimore, 249 pages.
Sekey, A. & Hanson, B. A. (1984). "Improved 1-Bark bandwidth auditory filter", J. Acoust. Soc. Am. 75, 1902-1904.
Willems, L. F. (1986). "Robust formant analysis", IPO annual progress report 21, 34-40.
Zue, V. W. (1985). "The use of speech knowledge in automatic speech recognition", Proc. of the IEEE, Vol. 73, 1602-1615.