ACOUSTIC VOWEL REDUCTION AS A FUNCTION OF SENTENCE ACCENT, WORD STRESS, AND WORD CLASS

Dick R. van Bergem

Institute of Phonetic Sciences, University of Amsterdam
Herengracht 338, 1016 CG Amsterdam, The Netherlands

Abstract. The effect of sentence accent, word stress, and word class (function words versus content words) on the acoustic properties of 9 Dutch vowels in fluent speech was investigated. A list of sentences was read aloud by 15 male speakers. Each sentence contained one syllable of interest. This could be a monosyllabic function word, an unstressed syllable of a content word, or a stressed syllable of a content word. The same syllable occurred in all three conditions. Sentence accent was manipulated with questions that preceded the sentences. A total number of 3465 vowels were segmented from the syllables and analysed. It was found that all three factors mentioned above had a significant effect both on the steady-state formant frequencies (F_1 and F_2) and on the duration of the vowels. Word stress and word class had a stronger effect on the vowels than sentence accent. A listening experiment showed the perceptual significance of the acoustic measurements. It appeared that spectral vowel reduction could be better interpreted as the result of an increased contextual assimilation than as the tendency to centralize. We also studied changes in the dynamics of the formant tracks due to the experimental conditions. It was found that formant tracks of reduced vowels became flatter, which supports the view of an increased contextual assimilation. Three simple models of vowel reduction are discussed.

Zusammenfassung. Die folgende Studie behandelt den Effekt von Satzakzent, Wortakzent und Wortklasse (Funktionswörter gegenüber Inhaltswörter) auf die akustischen Eigenschaften von 9 holländischen Vokalen in flüssiger Sprache. Eine Liste von Sätzen wurde vorgelesen von 15 männlichen Sprechern. Jeder Satz enthielt nur eine zu untersuchende Silbe. Das konnte ein monosilbisches Funktionswort sein, eine betonte Silbe in einem Inhaltswort, oder eine unbetonte Silbe in einem Inhaltswort. Der Satzakzent wurde manipuliert mit Fragen die den Sätzen vorhergingen. In total wurden 3465 Vokale segmentiert und analysiert. Die Ergebnisse zeigten, daß die oben genannten Faktoren alle einen deutlichen Effekt hatten auf die Dauer und auf die stabielen Formantfrequenzen (F_1 und F_2) der Vokale. Wortakzent und Wortklasse hatten einen stärkeren Effekt auf die spektrale Qualität der Vokale als der Satzakzent. In einem Hörexperiment wurde die perzeptive Bedeutung der akustischen Messungen aufgezeigt. Die spektrale Vokalreduktion konnte besser als das Resultat einer zugenommenen Assimilation erklärt werden als als eine Tendenz von Vokalen um zu zentralisieren. Wir untersuchten auch die Änderungen in der Dynamik der Formantspuren, verursacht durch die experimentellen Bedingungen. Die Ergebnisse zeigten, daß die Formantspuren von

reduzierten Vokalen flacher wurden, was die Idee einer zugenommenen Assimilation unterstützt. Drei einfache Vokalreduktion Modelle wirden diskutiert.

Résumé. Cette étude traite de l'effet de l'accent de phrase, de l'accent de mot et des classes de mots (mots fonction contre mots ...) sur les propriétés acoustiques de 9 voyelles hollandaises enoncées dans la parole. Une liste de phrases a été lue par 15 locuteurs masculins. Nous nous sommes seulement intéressés à une syllable dans chaque phrase. Ça pouvait être un mot fonction monosyllabique, une syllabe non-accentuée dans un mot ..., ou une syllabe accentuée dans un mot .... L'accent de phrase était manipulé grâce à des questions qui précédaient les phrases. Au total 3465 voyelles ont été segmentées et analysées. Les résultats ont montré que les facteurs linguistiques ont tous les trois eu un effet distinct sur la durée et les fréquences formantiques stables (F_1 et F_2) des voyelles. L'accent de mot et la classe de mot ont eu un effet plus fort que l'accent de phrase sur la qualité spectrale des voyelles. Une expérience d'écoute a démontré la signification perceptive des mesures acoustiques. Il nous est apparu que la réduction spectrale des voyelles se trouvait être mieux interprétée en tant que résultat d'une aasimilation augmentée que comme tendance des voyelles à centraliser. Nous nous sommes également penchés sur les changements dynamiques dans les traces formantiques causées par les conditions expérimentales. Il est apparu que les traces formantiques des voyelles réduites devenaient plus planes, ce qui soutient l'idée d'une assimilation augmentée. Trois modèles de réduction vocalique sont discutés.

Keywords. Acoustic vowel reduction, lexical vowel reduction, sentence accent, word stress, word class, centralization, increased contextual assimilation.

1. INTRODUCTION

A lot of information in written or spoken language is usually redundant for the proper understanding of the message. This becomes apparent if one thinks of, for instance, telegrams, the speech of young children, or the way you talk in a foreign country if you have little knowledge of the language. These examples show that the relevant information in written or spoken language can be compressed in only a few key words.
For an economic transfer of information in a natural speech situation a speaker will probably strive to pronounce only the important parts of an utterance clearly and relax his articulation in the less important parts, knowing that the listener is capable of filling in fuzzy or missing parts. Especially semantic and pragmatic knowledge sources can help the listener a great deal in restoring acoustically incomplete messages. But which parts of utterances are especially important for a proper transfer of the message and which parts are less important? The following three factors may play a role in this respect:

- Sentence accent.
  The most salient feature of sentence accent from a perceptual point of view is a sudden rise or fall in the pitch contour ('t Hart et al., 1990). Accentuation is used to draw the listener's attention to certain words 'in focus'. That is, a speaker will usually place an accent on a certain part of an utterance in order to highlight the information conveyed in it (Bolinger, 1972; 1985). This means that a speaker regards accented words as important parts of the message, so accented words may be more carefully pronounced than unaccented words.

- Word stress.
  In our view a word is perceived as a whole unit and not as for instance a string of phonemes which are recognized one by one and then merged into words. Following Klatt (1989), we believe that acoustic templates of words are matched with the acoustic input. The best match is established with a probabilistic strategy based on acoustic evidence as well as on other evidence, such as semantic and pragmatic sources of information. For a successful match it is probably not necessary that all parts of a word are pronounced equally well. Stressed syllables may serve as a kind of anchor points in this word recognition strategy (Van Bergem, 1990a). Hence stressed syllables may be more carefully pronounced than unstressed syllables.
  Another point of interest with respect to word stress is that several studies have shown an 'informational' advantage of stressed syllables over unstressed ones (e.g. Carter, 1987). This means that in most cases the number of word candidates is smaller when a search in the lexicon is based on stressed syllables of words than when it is based on unstressed syllables of words. Thus, the uniqueness of words in terms of their phonemic composition is mainly determined by their stressed syllables. A study by Altman and Carter (1989) has shown that this informational advantage of stressed syllables is caused by a greater variety of vowels used in these syllables. From an informational point of view, recognition of words could therefore be easier if a speaker would clearly pronounce the stressed syllables (and especially the vowels they contain).

- Word class (content words versus function words).

Content words are essential to comprehend the meaning of an utterance. Function words, on the other hand, mainly serve to specify the relations between content words and have only a small contribution to the meaning of an utterance (Bolinger, 1975). In addition, function words usually have a high frequency of occurrence. Many investigations have shown that words with a high frequency of occurrence are more easily recognized than words with a low frequency of occurrence (e.g. Morton, 1969; Forster, 1979). The high a priori probability of function words as it were lowers the 'recognition threshold'. This suggests that function words may be less carefully pronounced than content words.

The present experiment was restricted to the investigation of vowels, because they lend themselves well to spectral analysis in the sense that their spectral quality can be represented fairly well in a simple two-dimensional formant space (see e.g. Pols et al., 1973). Our main goal was to find out how the formant frequencies (F_1 and F_2) and the duration of vowels are influenced by sentence accent, word stress, and word class. Apart from this major question we were also interested in the relation between spectral vowel reduction and coarticulation, and in differences between speakers. Parts of the results from this experiment were reported earlier in Van Bergem (1990b) and Van Bergem (1991a).

Phoneticians often refer to vowel reduction as "a loss of vowel quality", whereas linguists usually describe vowel reduction as "the substitution of a full vowel with a schwa". Consider the following pairs of semantically related words in which the stressed syllable in the words on the left has been replaced by an unstressed syllable in the words on the right (word stress in small caps):

MAjor	maJORity	(1)
MAriner	maRIne	(2)
MInor	miNORity	(3)
CAPture	capTIvity	(4)

According to the English dictionary (Collins, 1989) the underlined vowels in the words on the right are replaced by a schwa /ə/ in word pairs (1) and (2), but remain unchanged in word pairs (3) and (4) with respect to their counterparts on the left. This means that a native speaker will aim at producing a *full* vowel in the words "minority" and "captivity". However, the absence of primary word stress in the first syllable of these words may to some extent cause a change in vowel quality. In general, not only stress, but also many other factors such as sentence accent, word class, or speech style may influence the vowel quality. Whenever a loss of spectral quality occurs in vowels that were *intended* to be full vowels, we will refer to this phenomenon as *acoustic vowel reduction*. In word pairs (1) and (2), on the other hand, the full vowel in words on the left is simply replaced by a schwa in words on the right. That is, according to the standards of the linguistic community of English speaking people, the words "majority" and "marine" are 'officially' pronounced with a schwa. We will refer to this phenomenon as *lexical vowel reduction*, for the schwa has become a characteristic part of the word.

Some researchers use the terms phonetic and phonological vowel reduction (e.g. Fourakis, 1991), which have the same meaning as our terms acoustic and lexical vowel reduction. The term phonological vowel reduction refers to the process described by

phonologists that causes unstressed vowels to be realized as schwas. This process applies to most unstressed vowels of (American) English, although there are exceptions (Ladefoged, 1982). However, as remarked by Nord (1986), for other languages such as Swedish there is no general phonological rule that changes a vowel phoneme into a schwa. This is also true for Dutch where the schwa only occasionally replaces a full vowel in unstressed syllables (It is, however, a very frequently occurring phoneme in Dutch affixes.). Therefore, we prefer to use the term lexical vowel reduction which indicates that a schwa replacement is specific to *particular words* rather than to an entire sound system. As regards the term phonetic vowel reduction, we think that it is rather vague and that the term acoustic vowel reduction more accurately describes the change in vowel quality.

Lexical vowel reduction is probably a systematic extension of acoustic vowel reduction. A vowel that is often affected by acoustic reduction may gradually be replaced with a schwa by the linguistic community and in the end this may lead to a permanent lexical substitution. During this process of language change two variants of the same word exist, the one with the original full vowel and the one with the schwa. Note that the lexical substitution of a full vowel with a schwa can only occur in languages that have accepted the schwa as a phoneme. The distinction between acoustic and lexical vowel reduction is more thoroughly discussed in Van Bergem (1991b).

The present study is exclusively concerned with acoustic vowel reduction. In all the test words that we used in the experiment that is going to be described, the vowel of interest could not be replaced by a schwa without the word sounding awkward and unnatural. This means that the speaker's intention would almost certainly be to produce a full vowel in the selected words and not a schwa. All reference that is made to vowel reduction in the following sections should therefore be read as *acoustic* vowel reduction.

The outline of this article is as follows: The experimental design and the measuring procedures are described in section 2 and 3, respectively. In section 4 the results are presented. In section 4.1 the effect of the experimental conditions on vowel quality is discussed. In section 4.2 we focus on the topics centralization and contextual assimilation. In section 4.3 differences between speakers are discussed. In section 4.4 the results of a listening experiment with the segmented vowels are given, and section 4.5 deals with some simple reduction models. A general discussion and the conclusions are given in section 5 and 6, respectively.

2. EXPERIMENTAL DESIGN

The main aim of this study was to investigate how the spectral characteristics and the duration of vowels are influenced by sentence accent, word stress, and word class. We wanted to do this in speech that was as natural as possible. In order to be able to create experimental conditions in which the effect of these three factors could be established, we opted for sentences containing existing words (no 'nonsense' words) that were read aloud by a number of speakers. Our test material consisted of CVC syllables that exist as:

- a monosyllabic function word
- an unstressed syllable in a content word
- a stressed syllable in a content word

It should be noted that all function words in this experiment are stressed, because monosyllabic words have by definition primary word stress. Although in the actual experiment only Dutch words were used, we will illustrate the experimental procedure with the English syllable "can" that exists as:

- a function word "can" (an auxiliary verb)
- an unstressed syllable in the content word "canteen"
- a stressed syllable in the content word "candy"

By using identical syllables in such triplets of words, coarticulation effects would be similar for the vowel of interest (the /æ/ in the English example) in all experimental conditions.
Each word was placed in a sentence that was read aloud by 15 male speakers. The sentences were constructed in such a way that each test syllable occurred in about the same position for all experimental conditions. All sentences were read twice, once with an accent on the test word and once with an accent elsewhere in the sentence. The place of the sentence accent was manipulated by means of questions, which had been recorded in advance. The speakers heard the questions through a loudspeaker and then answered by reading in an appropriate way the corresponding test sentences. Apart from the 6 test conditions created in this way (3 words with and without accent) a 7th condition was added in which the speakers uttered the test syllable in isolation. This condition was used to obtain a kind of 'ideal' pronunciation of the vowel in each test syllable. Test sentences for the (fictitious) English example could be as follows (For the sake of clarity word stress is printed in small caps, and sentence accent in bold in the examples below. This was not done in the reading list of the actual experiment.):

1. (What did you buy for your mother?)
   I bought **CAN**dy for my mother.
2. (For whom did you buy candy?)
   I bought CANdy for my **moth**er.
3. (Where do they sell beer?)
   In our can**TEEN** they sell beer.
4. (What do they sell in our canteen?)
   In our canTEEN they sell **beer**.
5. (What can your sister do for hours?)
   My sister CAN **talk** for hours.
6. (How long can your sister talk?)
   My sister CAN talk for **hours**.
7. CAN (spoken in isolation).

In the case of function words the sentence accent was never placed directly on the function word itself. In a normal speech situation accented function words are rather exceptional (Kruyt, 1985). In addition, it is very difficult to think of questions that could lead to an accentuation of function words. What question could for instance be posed that would lead in a natural way to the accented "can" in the sentence "My sister can talk for hours"? This problem could of course have been avoided by printing test words in bold

or in capitals in order to evoke an accentuation, but this would focus the speaker's attention too much on these words and would lead to a less natural pronunciation in our view. Instead of trying to make a condition with accented function words, we therefore attempted to create the same rhythmical pattern in the word sequence containing a function word as in content words with an unstressed test syllable. This was done by placing an accent on the word *following* (or sometimes preceding) the function word (Compare "can talk" and "canteen" in the examples given above.). In this way we could find out whether vowel quality in these cases is influenced by the rhythm in the utterance or that it is influenced by semantic aspects, as explained below.
The actual realization of the pitch movement causing the perception of accent (e.g. 't Hart et al., 1990), occurs on the syllable with primary word stress. Thus, for the word "canteen" the syllable "teen" gets a pitch accent, but nevertheless the entire word is highlighted. A pitch accent on the word "talk" of the sequence "can talk" gives rise to a similar rhythmical pattern as in the accented word "canteen", but only the word "talk" will be highlighted and *not* the function word "can". Does the similar rhythmical structure due to accentuation in "canteen" and "can talk" influence the quality of the vowel /æ/ in the same way, or does the different semantic structure cause differences in vowel quality?
In summary, apart from the control condition 'syllables spoken in isolation', the test conditions contain three different syllable types and presence or absence of sentence accent:

1.	stressed syllable in an accented content word	(S-A)
2.	stressed syllable in an unaccented content word	(S-NA)
3.	unstressed syllable in an accented content word	(NS-A)
4.	unstressed syllable in an unaccented content word	(NS-NA)
5.	function word followed (or preceded) by an accented content word	(F-A)
6.	function word followed (or preceded) by an unaccented content word	(F-NA)
7	syllable spoken in isolation	(ISO)

Appropriate test words were collected in the following way: First, all possible CVC syllables that exist as a function word in the Dutch language were selected. Although function words are said to form a closed set of words, there is no general agreement about which words belong to this set. In a strict sense, function words are all words that belong to the closed grammatical classes such as articles, prepositions, auxiliary verbs, etc. However, several of those words are rather formal or archaic and are hardly ever used. For practical goals some authors (e.g. Van Coile, 1987) therefore prefer to define function words on the basis of a high frequency of occurrence, regardless of the grammatical class they belong to. For the present experiment two inventories of function words were consulted. Van Wijk and Kempen (1980) list all the function words according to their strict definition. The inventory of Baart (1987) is smaller, but he incorporated a small number of high frequency verbs. We selected all the function words mentioned in these two inventories that had a CVC structure as well as a frequent usage. After this selection, the Dutch lexical database CELEX (1985) was searched for content words in which any of these CVC syllables occurs. For each CVC syllable two different types of content words had to be found, one with a stressed test syllable and another with an unstressed test syllable. A total of 33 appropriate triplets of words were gathered in this way. The content words always contained either two or three syllables. According to

metrical phonology, at least three degrees of stress can be assigned to the syllables in a word: primary stress, secondary stress, and no stress (Booij, 1981). The 'unstressed' syllables in the present experiment did not have primary stress and in almost all cases no secondary stress either.
For all these words test sentences and questions were constructed. The list of questions was read aloud by a male speaker who had no knowledge of the aims of the experiment. These prerecorded questions were subsequently presented to the 15 male speakers with intervals of 5 seconds, allowing them to respond by reading the test sentences. The order of the test sentences (and the corresponding preceding questions) was random within each test condition; the order of the test conditions was also random. The random ordering of test sentences and test conditions was different for each speaker. The speech was recorded on the audio channel of a Panasonic video recorder in an anechoic room. The total number of vowels to be analysed was 3465 (15 speakers × 7 test conditions × 33 test syllables). Not all 12 Dutch monophthongs occurred in the test material and some occurred more frequently than others (in agreement with their frequency of occurrence in Dutch). The vowels that were absent were /œ, y, ø:/ which are rather exceptional in Dutch words (together they constitute less than 5% of all vowel occurrences). A complete list of all triplets of words that were used in the experiment is given in Table 1.

TABLE 1 ABOUT HERE

3. MEASURING PROCEDURES

The test sentences were lowpass filtered at 4.5 kHz and digitized at a sample rate of 10 kHz with a 12-bit precision. The vowels that had to be analysed were segmented from the speech waveform together with a considerable part of the surrounding consonants. These segments were analysed with a 25-ms Hamming window that was shifted in steps of 2 ms. On each analysis frame a 10th-order LPC-analysis was performed and subsequently continuous formant tracks were obtained with a method due to Willems (1986) in which the Split-Levinson algorithm is used. All formant frequencies were transformed to a mel-scale which is more in line with the hearing process than a linear frequency scale. In this way formant frequencies were weighted in a proper way for the calculation of Euclidean distances in the F_1-F_2 plane. For the mel transformation we used the formula:

$$m = 2595 \log\left(1 + \frac{f}{700}\right)$$

where f is the formant frequency in Hz and m the mel-converted value (Makhoul and Cosell, 1976).
The onset and the offset of the vowels, which defined their duration, were determined by measuring the spectral variation over time. For this purpose we used cepstral coefficients which turned out to be more sensitive than formant frequencies. From the 10 LPC-coefficients 24 cepstral coefficients were calculated (Markel and Gray, 1976) and these 24 values were in turn transformed to 8 coefficients on a mel-like scale with a bilinear transformation (Oppenheim & Johnson, 1972). Such bilinear transformed cepstral

coefficients are successfully applied in automatic speech recognition systems (e.g. Lee, 1989). The spectral variation of the speech signal at the *k*-th frame was calculated over a $2L+1$ frame interval $[k-L, k+L]$. We chose $L = 5$, so that the frame interval was 22 ms. The spectral variation *V(k)* at the *k*-th frame was defined as the within standard deviation for the frame interval:

$$V(k) = \sqrt{\frac{1}{N}\sum_{i=k-5}^{k+5}\sum_{j=1}^{8}\left(c_{ij} - \bar{c}_j\right)^2}$$

in which *N* denotes the total number of cepstral coefficients in the frame interval $(8 \times 11 = 88)$, c_{ij} is the *j*-th mel-based cepstral coefficient of the *i*-th frame, and $\bar{c}_j$ is the mean value of the 11 *j*-th cepstral coefficients within the frame interval. The contour *V(k)* was subsequently smoothed by filtering it repeatedly (5 times) with a 3 point Hanning window. This procedure to measure spectral variation resembles the one described by Svendsen and Soong (1987).

FIGURE 1 ABOUT HERE

As an example Fig. 1 shows the spectral variation in the Dutch function word “kan” (the auxiliary verb “can”). In the lower part of Fig. 1 the oscillogram is shown, in the middle part the spectral variation contour *V(k)*, and in the upper part three formant tracks on a mel-scale. The onset and the offset of the vowels were found with a semi-automatic procedure that consisted of two steps:

1. The search for a starting point within the vowel part of the speech segment.
   The position in the speech segment with the highest instantaneous amplitude is proposed as a starting point (marker ‘M’ (maximum) in Fig. 1). In most of the cases this position is located somewhere within the vowel and is then accepted as a starting point. If not, the user must overrule the proposed starting point and indicate any position within the vowel part as a new starting point.

2. The search for the onset and the offset of the vowel.
   Boundaries between vowels and consonants are characterized by rapid spectral changes which result in peaks in the spectral variation contour *V(k)*. Spectral transitions within the vowel are more gradual and result in lower values in the spectral variation contour. Once a starting point within the vowel has been established, the procedure continues by tracing the spectral variation contour for peaks on either side of the starting point and defines these peak positions as the onset and the offset of the vowel (markers ‘O’ in Fig. 1). The peak value has to surpass a threshold value in order to be selected (We used a threshold value of 0.075, which was determined by trial and error). This threshold is indicated with a dashed line in Fig.1. If the threshold value is higher than all occurring peaks on one side of the starting point, it is lowered in small steps until a peak is found.

The proposed vowel boundaries were carefully checked by listening to selected parts of the speech signal and by visual inspection of the speech waveform, the contour *V(k)*, and

the formant tracks. In about 9% of all cases a proposed boundary was rejected. This occurred when the spectral variation within the vowel was large enough to produce a peak above the threshold or when the transition from consonant to vowel or from vowel to consonant was too gradual to produce a clear peak and a larger peak near the edges of the speech segment was chosen instead by the algorithm. In these cases the markers were placed by hand at a more appropriate local maximum in the spectral variation contour *V(k)*.
The steady-state vowel part was subsequently defined as that part of the vowel where a minimum value in the spectral variation contour occurred between the onset and the offset of the vowel (marker 'S' in Fig. 1). Only occasionally this position was hand-corrected (in about 6% of all cases). This occurred for vowels with a long duration that sometimes contained more than one steady-state part. In such cases the most appropriate steady-state part was determined by looking at global minima and maxima in the formant tracks.

4. RESULTS

4.1 Test conditions

4.1.1 Steady-state formant frequencies

In Fig. 2a the steady-state formant frequencies (F_1 and F_2) of the Dutch *short* vowels /u, ɔ, ɑ, ɛ, ɪ, i/ in each test condition have been plotted, averaged over all consonantal contexts and all speakers. In Fig. 2b the same has been done for the *long* vowels /o:, a:, e:/. The average positions of vowels from test syllables spoken in isolation (condition ISO), which we regard as 'ideal' vowels in those specific contexts, are indicated with an asterisk. It can be seen that in the other test conditions the mean formant frequencies of vowels shift away from the 'ideal' target position into the direction of the schwa. The position of the schwa was determined by averaging the formant frequencies of 300 schwas uttered by the 15 speakers. For this purpose the first 20 schwas that were encountered in the test sentences of each speaker were analysed. Since each speaker had a different random ordering of test sentences, the schwas were taken from different words for each speaker. The gathering of schwa items posed no problems, because the schwa is a frequently occurring phoneme in Dutch (especially in affixes).

FIGURE 2 ABOUT HERE

In order to determine the effect of the test conditions on the formant shifts, the mel-scaled Euclidean distances in an F_1-F_2 space were calculated between each vowel item and its 'ideal' counterpart. The mean distances for each test condition averaged over all test syllables and all speakers are shown in Fig. 3. This figure reveals that the distance to the 'ideal' vowel position increases in the order 'stressed syllable' - 'unstressed syllable' - 'function word'. It can also be seen that the distance to the 'ideal' vowel position is larger for unaccented vowels than for accented vowels.

FIGURE 3 ABOUT HERE

An analysis of variance with repeated measures was done on these distances with the factor 'speaker' as independent variable (15 levels) and two trials factors: 'accent' (2 levels) and 'syllable type' (3 levels). The factor 'speaker' which was significant ($F = 7.6,\ p < 0.001$) will be discussed in section 4.3. The factor 'accent' ($F = 127.0,\ p < 0.001$) and 'syllable type' ($F = 217.0,\ p < 0.001$) were also significant. In Fig. 3 it can be seen that the factor 'syllable type' shows larger shifts of formant frequencies than the factor 'accent'. This means that word stress and word class ('syllable type') have a greater effect on the spectral quality of vowels than sentence accent, especially if we consider that a 'spontaneous' accentuation of words by speakers, which is most common in conversational speech, is usually even weaker than an accentuation evoked by questions. The interaction between 'accent' and 'syllable type' was also significant ($F = 11.4,\ p < 0.001$). Tests for pairwise contrasts between conditions showed that the conditions NS-NA, F-A, and F-NA were not significantly different from each other ($p > 0.05$). From now on we will call the strongly reduced vowels from these three test conditions together the group RED.

As explained in section 2, a similar rhythmical pattern due to accentuation was created in the word sequence containing a function word ("can talk") as in content words with an unstressed test syllable ("canteen"). The significant difference between the conditions NS-A and NS-NA shows that the quality of the vowel in the unstressed test syllable of a content word (the /æ/ in "canteen") is influenced by accentuation of this word (i.e. presence or absence of an accent on the stressed syllable "teen"). The lack of a significant difference between the conditions F-A and F-NA, on the other hand, shows that presence or absence of an accent on the word following the function word ("talk") has *no* influence on the quality of the vowel in the function word (the /æ/ in "can"). This can be explained by the fact that the entire content word "canteen" as a semantic unit is highlighted by accentuation, whereas accentuation in the word sequence "can talk" only highlights the word "talk" and not the function word "can". So it appears that spectral vowel reduction is governed by semantics rather than by rhythm in this particular case.

4.1.2 Dynamics of the formant tracks

Apart from shifts in the steady-state formant frequencies, we also wanted to study the changes in the dynamics of the formant tracks (F_1 and F_2) as a result of the experimental conditions. In order to measure changes in the curvature, all formant tracks were modelled with a second-degree polynomial

$$F(t) = c_0 + c_1\, t + c_2\, t^2 \qquad (-1 \le t \le 1)$$

The coefficient c_2 is a measure of the amount of (parabolic) curvature of the tracks. For a proper comparison of the curvature in formant tracks from different vowels, each track was 'time normalized' by fitting it with the polynomial on the interval [–1,1].

FIGURE 4 ABOUT HERE

In Fig. 4a the coefficients c_2 for the formant tracks of the *short* vowels /u, ɔ, ɑ, ɛ, ɪ, i/ have been plotted, averaged over all consonantal contexts and all speakers. In Fig. 4b the same has been done for the *long* vowels /o:, a:, e:/. In order to get the usual configuration of vowels, we plotted $-c_2$ in the figures. For the sake of clarity we have only shown the coefficients for the condition ISO (indicated with asterisks) and the group RED (indicated with dots), because the complete plots with curvature coefficients from all test conditions were somewhat fuzzy. The appearance of the 'vowel triangle' in Figs. 4a and 4b indicates that the curvature of formant tracks is specific to each vowel. A similar observation was made by Van Son and Pols (1991). As might be expected, the curvature of F_1-tracks increases in more open vowels. The curvature of F_2-tracks is concave downward for front vowels and concave upward for back vowels. The figures show a reduction in the curvature of formant tracks for the group RED compared to the condition ISO (In the case that $c_2 = 0$ the track has no curvature at all.). This means that the formant tracks become flatter. For the schwa the curvature of F_1 and F_2 is almost zero, so the formant tracks of the schwa approach a straight line between the onset and the offset.

A comparison of Figs. 4a and 4b with Figs. 2a and 2b suggests that the shift of steady-state formant frequencies coincides with a reduced curvature of the formant tracks. Lindblom (1963) refers to this phenomenon as 'formant undershoot'. If the curvature is concave downward, we observe a decrease in steady-state formant frequencies of reduced vowels. If, on the other hand, the curvature is concave upward, we observe an increase in steady-state formant frequencies of reduced vowels. For F_1-tracks the curvature is always concave downward, which agrees with the fact that steady-state F_1-values of reduced vowels do not increase (see also Van Bergem, 1989; Van Son and Pols, 1990; Van Son and Pols, 1992). The reduction in formant track curvature for different vowels shown in Figs. 4a and 4b is thus clearly related to the shift of steady-state formant frequencies shown in Figs. 2a and 2b.

In order to determine the effect of each test condition on the curvature, the Euclidean distances in a two-dimensional c_2-space were calculated between each vowel and its 'ideal' counterpart (in the same design as was used for the steady-state formant frequencies). The mean distances for each test condition averaged over all test syllables and all speakers are shown in Fig. 5. This figure shows the same trend in distances for the curvature of formant tracks as Fig. 3 did for steady-state formant frequencies. The distance to the 'ideal' vowel curvature increases in the order 'stressed syllable' - 'unstressed syllable' - 'function word'. It can also be seen that the distance to the 'ideal' vowel curvature is larger for unaccented vowels than for accented vowels. The factor 'syllable type' shows a greater flattening effect on the tracks than the factor 'accent'.

FIGURE 5 ABOUT HERE

An analysis of variance with repeated measures was done with the same design that was used in the former section. The factor 'speaker' which was significant ($F = 2.7$, $p < 0.001$) will be discussed in section 4.3. Both the factor 'accent' ($F = 23.9$, $p < 0.001$) and 'syllable type' ($F = 80.0$, $p < 0.001$) were significant. The interaction between 'accent' and 'syllable type' was not significant ($p > 0.05$). Tests for pairwise contrasts between conditions showed that the conditions NS-NA, F-A, and F-NA (the group RED) were again not significantly different from each other ($p > 0.05$).

4.1.3 Durations

The average durations of vowels in each test condition are shown in Fig. 6. This figure reveals that the durations of vowels decrease in the order 'stressed syllable' - 'unstressed syllable' - 'function word'. It can also be seen that the durations of unaccented vowels are smaller than the durations of accented vowels. The factor 'syllable type' has a larger influence on vowel durations than the factor 'accent'. Notice that Fig. 6 is the mirror image of Figs. 3 and 5.

FIGURE 6 ABOUT HERE

An analysis of variance with a similar design as in the former sections was done on vowel durations. The factor 'speaker' which was significant ($F = 3.9$, $p < 0.001$) will be discussed in section 4.3. Both 'accent' ($F = 60.3$, $p < 0.001$) and 'syllable type' ($F = 228.8$, $p < 0.001$) were significant factors. The interaction between 'accent' and 'syllable type' was also significant ($F = 6.3$, $p < 0.005$). Tests for pairwise contrasts between conditions showed that the conditions NS-NA, F-A, and F-NA (the group RED) were again not significantly different from each other ($p > 0.05$).
In Table 2 mean durations in all conditions are given for short vowels and long vowels separately. It appears that especially long vowels are affected by the test conditions. An analysis of variance on the durations of short vowels and long vowels separately revealed that for both vowel groups the factors 'accent' and 'syllable type' were significant ($p < 0.001$).

TABLE 2 ABOUT HERE

4.2 Centralization or contextual assimilation

Many acoustic studies on vowel reduction (e.g. Stålhammar et al., 1973; Koopmans-van Beinum, 1980; the present study) suggest that reduced vowels move towards a more neutral place in the vowel plane, usually referred to as the schwa position. Is this *centralization* effect a phenomenon in its own right, or is it the result of a *contextual assimilation* of the vowel with surrounding phonemes? Delattre (1969) suggests that all vowels have a tendency to become a schwa (i.e. to centralize), regardless of the effect of surrounding phonemes. Moreover, he claims that in some languages (for instance English) this tendency is stronger than in other languages (for instance Spanish). However, Van Bergem (1991b) showed that Delattre confused lexical vowel reduction with acoustic vowel reduction. In a study on vowel reduction in Swedish, Nord (1975) concluded that vowel reduction is caused by an increased contextual assimilation of the vowel with surrounding phonemes. In an extended version of his study this conclusion was further supported (Nord, 1986). The same view was taken by Lindblom (1963).

FIGURE 7 ABOUT HERE

In order to get a better view on the role of the surrounding consonants in the CVC syllables of the present study, the shift of steady-state formant frequencies was examined in detail for the vowel /ɛ/ in the six occurring consonantal frames. In Fig. 7 vowels from the condition ISO (indicated with asterisks) and the group RED (indicated with dots) are compared. Apparently, the shift of formant frequencies in each particular /ɛ/-item is dependent on the consonantal frame. Notice for instance that the F_2-value of the /ɛ/ from the syllable /wɛl/ shifts to a frequency of about 1100 Hz which gives an /ɔ/-like sound. Apparently, the formant pattern for this particular /ɛ/ is *not* centralizing. The observed effect is most likely caused by an increased assimilation of the vowel with the surrounding consonants /w/ and /l/. A similar strong assimilation effect in the syllable /wɛl/ was found by Pols (1977) and for English by Lindblom and Moon (1988). The F_2-value of the /ɛ/ from the syllable /hɛp/, on the other hand, shifts to a slightly higher F_2-value which makes it an /ɪ/-like sound. In general, both the amount and the direction of the shift of steady-state formant frequencies in reduced vowels seem to be specific to the consonantal frame in which they occur. The clear 'centralization' observed in Figs. 2a and 2b is the result of the *averaging* of different contextual influences on the vowels; the reduced formant frequencies of the vowels in one particular CVC syllable do not necessarily centralize.

Is centralizing the same as becoming a schwa? People who think it is, regard the schwa as a vowel that is produced with a completely neutral vocal tract (a uniform tube). For an average male such a vowel would have an F_1 and F_2 of 500 Hz and 1500 Hz, respectively (Fant, 1960). In this case the schwa position more or less coincides with the centre of the vowel triangle. However, recently we have measured the formant frequencies of schwas in systematically varying phonemic surroundings and it was found that the schwa can appear almost anywhere in the vowel plane, dependent on its context (Van Bergem, in preparation). In our view, the schwa is not a vowel with a target in the centre of the vowel triangle, but rather a vowel-like segment *without* target that is completely assimilated with its phonemic context. If reduced vowels are regarded as vowels that have been partially assimilated with their phonemic context, this means that the formant frequencies of reduced vowels are shifting to a position in the vowel plane that a schwa would have *in the same phonemic context*. As an illustration of this idea, we measured the F_1 and F_2 of the schwa in the word "gruwel" (/χrywəl/), where the schwa occurs in the same consonantal context as the /ɛ/ from the test syllable /wɛl/. For this purpose we asked 5 speakers who were subjects in the present experiment (and who were still available) to read a short sentence containing the word "gruwel" ("gruel" in English). The average F_1 for this schwa was 346 Hz and the average F_2 was 940 Hz. As discussed above, the reduced /ɛ/ from the test syllable /wɛl/ did not centralize, but shifted to an /ɔ/-like position in the vowel plane (more specifically, to an F_1 of 393 Hz and an F_2 of 1146 Hz). This shows that the /ɛ/ in /wɛl/ is reducing to the formant pattern of a schwa in the same consonantal context.

The covariation of shifts in steady-state formant frequencies and the flattening of the formant tracks supports the 'undershoot' concept suggested by Lindblom (1963). However, spectral undershoot can in principle be accounted for by three different models as shown in Fig. 8. The thick line in each model indicates the 'ideal' formant track of a vowel in a CVC syllable and the thin line the same formant track in its reduced form. For a proper understanding of the models, it should be noted that even in clearly pronounced CVC syllables (represented by the condition ISO in the present experiment) the vowel can

to some extent be assimilated with the consonants as well as the consonants with the vowel. That is, due to contextual influences the vowel and consonants may not reach their (context-free) canonical target positions (Daniloff and Hammarberg, 1973). This is the reason that the (quasi-) target formant frequencies of for instance the /ɛ/ in /wɛl/ are different from those in other syllables containing the /ɛ/ (see Fig. 7).

FIGURE 8 ABOUT HERE

In all three models shown in Fig. 8 the steady-state formant frequencies of the reduced vowels show 'undershoot' indicating an increased assimilation of the vowel with the surrounding consonants. However, the interpretation of the models is quite different with respect to the role that the surrounding consonants play in the reduction process. In model I the formant onset and offset of the reduced vowel shift *towards* the vowel 'target', which means that the surrounding consonants are assimilating more with the vowel. In model II the formant onset and offset of the reduced vowel stay put, indicating a stationary assimilation of the consonants with the vowel. In model III the formant onset and offset of the reduced vowel shift *away from* the vowel 'target', which means that the consonants are breaking away from the influence of the vowel and are probably moving to their own canonical target position. Although the effects on the formant onset and offset have been presented symmetrically in the models, they may of course in practice each be subject to a different model.
The plausibility of each model was tested with the present dataset. In order to get robust estimates of the onset and the offset of the formant tracks, the second degree polynomial approximations were used (see section 4.1.2). Since the polynomials were fitted on the interval [–1,1], the onset frequency could be determined by inserting $t = -1$ and the offset frequency by inserting $t = 1$ in the polynomial equation:

$$F_{onset} = c_0 - c_1 + c_2$$

$$F_{offset} = c_0 + c_1 + c_2$$

For all CVC syllables the onset and the offset of F_1-tracks and F_2-tracks were calculated in this way and subsequently they were averaged over subjects. The formant tracks of vowels from the condition ISO were defined as 'target' formant tracks and those of vowels from the group RED as 'reduced' formant tracks. For each CVC syllable the difference in formant frequency at the onset, steady-state part, and offset was calculated between target tracks and reduced tracks (both for F_1 and F_2). If the formant shift at a vowel boundary was less than 25 mel, it was assigned to model II. Otherwise, the formant shift at the steady-state part was compared with the formant shift at the vowel boundary. If these shifts had opposite directions model I was selected, if they had the same direction model III was selected.

TABLE 3 ABOUT HERE

In Table 3 the percentages of assignments to each model are given for F_1-tracks and F_2-tracks both at the onset and at the offset. It appears that model III occurs most frequently,

followed by model II. With respect to the onset frequency of F_2, Krull (1989) had similar results. The occurrence of each model may very well depend on the particular vowel and consonants under consideration, but the present dataset was too small to uncover systematic tendencies.

4.3 Differences between speakers

In order to compare the vowel space of different speakers, we wanted to measure the discriminability of all their vowels in the 6 'natural' test conditions. For a clear separation of vowel clusters, (phonologically) identical vowels should be close together in a formant space and different vowel clusters should be far apart. We chose the following definition for vowel separation σ, based on Rabiner et al. (1983), which resembles the F-ratio in an analysis of variance:

$$\sigma = \frac{\sum_{i=1}^{M} N_i \; \mathrm{d}\left(\overline{F}_i, \overline{F}\right)}{\sum_{i=1}^{M} \sum_{j=1}^{N_i} \mathrm{d}\left(F_{ij}, \overline{F}_i\right)}$$

in which M denotes the number of vowel clusters (in our case 9), N_i is the number of vowels in cluster i (varying between 6 for /u/ and 42 for /ɑ/), $\mathrm{d}\left(\overline{F}_i, \overline{F}\right)$ is the Euclidean distance in mels (in two dimensions F_1 and F_2) between the mean formant vector of cluster i and the overall mean formant vector, and $\mathrm{d}\left(F_{ij}, \overline{F}_i\right)$ is the Euclidean distance between the j-th formant vector in cluster i and the mean formant vector of cluster i. The nominator of the formula denotes the (weighted) sum of between-cluster distances, and the denominator denotes the sum of within-cluster distances. The vowel separation for each speaker is shown in Fig. 9.

FIGURE 9 ABOUT HERE

FIGURE 10 ABOUT HERE

It can be seen that the σ-ratios of the speakers do not differ very much except for speaker S_1 who has an extremely good vowel separation and speaker S_{15} whose vowel separation is rather poor. It will not come as a surprise that speaker S_1 is a professional speaker (a newsreader) who is well known for his clear pronunciation. As an illustration of the differences in vowel space between speakers, the clusters of short vowels from speaker S_1 and speaker S_{15} are shown in Figs. 10a and 10b (ellipses of one standard deviation around mean values). The professional speaker S_1 has a very large vowel triangle and his compact vowel clusters are clearly separated. Speaker S_{15}, on the other hand, has a rather small vowel triangle and his different vowel clusters are large and overlapping.

FIGURE 11 ABOUT HERE

In Fig. 11 the average durations of all vowels from the 6 'natural' test conditions are shown for each speaker. It will be clear that there is hardly any relation between the vowel separation and the average vowel durations of speakers (the rank correlation is 0.23). The average vowel duration of, for instance, speakers S_2 and S_{15} is comparable, but their vowel separation is clearly different.
The analyses of variance discussed in section 4.1 showed that the factor 'speaker' was significant for spectral distances ($F = 7.6,\ p < 0.001$), for curvature distances ($F = 2.7,\ p < 0.001$), and for durations ($F = 3.9,\ p < 0.001$). In order to determine if for *all* speakers the vowel quality was affected by the experimental conditions, an analysis of variance on spectral distances, curvature distances, and durations was done for the vowels of each speaker separately. The significance level was set to a value of 0.003 $(= 0.05 / 15)$ because of the large number of simultaneous tests. At this level of significance the factor 'syllable type' was significant for all 15 speakers in the analysis of variance on spectral distances, for 8 speakers in the analysis of variance on curvature distances, and for 14 speakers in the analysis of variance on durations. The factor 'accent', on the other hand, was significant for only 6 speakers in the analysis of variance on spectral distances, for only 1 speaker in the analysis of variance on curvature distances, and for only 4 speakers in the analysis of variance on durations. This supports the view that sentence accent has a minor influence on vowel quality compared to word class and word stress.

4.4 Listening experiment

In order to test the perceptual significance of the acoustic measurements, we performed a listening experiment with part of the data set. All segmented vowels from three speakers were used for this experiment. The speakers were selected on the basis of their vowel separation as shown in Fig. 9. Speaker S_1 and speaker S_{15} were chosen, because they represent extremes in terms of their vowel separation and speaker S_8 was added to represent the group of speakers with a mediocre vowel separation.
The segmented vowels were presented to 24 listeners (12 male and 12 female) in a blocked design. They had their original duration as established with the semi-automatic segmentation procedure described in section 2. The listeners were for the greater part students from the University of Amsterdam. In order to level out annoying loudness differences between the stimuli, they were scaled to a fixed maximum amplitude value. Artificial clicks were prevented by shifting the vowel boundaries to the sample value closest to zero in an interval of 2 ms around the boundaries. All vowels from the 7 different test conditions were placed in a random order for each of the three speakers separately. The order of speakers was also randomized. The random ordering of stimuli and speakers was different for each subject in the listening experiment. In order to make the subjects familiar with the voice of each speaker, the 10 stimuli at the end of a speaker block were added at the beginning of the series as 'trial' items (without feedback). In this way a total number of 723 stimuli were presented to the 24 listeners (3 speakers × 7 conditions × 33 test syllables + 30 trial items).
The listening test was done on line by each subject separately at the terminal of a μVAX computer. One listening session took about 45 minutes. Subjects heard a stimulus and had to respond by pressing one of 12 buttons on the keyboard of the terminal,

representing the 12 Dutch monophthongs. These were indicated in their orthographic form (unique for Dutch) on stickers that were placed on the buttons in the following format:

	O		A	E	I	U
OE	OO	AA	EE	IE	UU	EU

In this format vowels that might be confused because of similar orthographic symbols, or similar sounds, have been placed close to each other. On the top row the vowels /ɔ, ɑ, ɛ, ɪ, œ/ are shown in their orthographic form, and on the bottom row the vowels /u, o:, a:, e:, i, y, ø:/. Although only 9 monophthongs were actually present in the test material, all 12 Dutch monophthongs were offered as response possibilities, because the vowels that were intended by the speakers could actually have been pronounced as any Dutch vowel. As soon as subjects had responded, the next stimulus was presented to them. This design allowed listeners to establish their own pace and it prevented listeners from skipping responses.

FIGURE 12 ABOUT HERE

In Fig. 12 the mean error percentages of the 24 subjects for the 6 'natural' test conditions are shown, averaged over test syllables and speakers. A response was defined as incorrect if it did not agree with the vowel that was (or should have been) intended by the speaker. Notice the similarity of Fig. 12 with Figs. 3, 5, and 6. The error percentages increase in the order 'stressed syllable'-'unstressed syllable'-'function word'. Furthermore, the error percentages are higher for unaccented vowels than for accented vowels. The factor 'syllable type' causes larger differences in error percentages than the factor 'accent'.

An analysis of variance with repeated measures was done on the error percentages with the same design as used before. Since the data consisted of proportions which are not normally distributed, an inverse sine transformation was applied (Kirk, 1982). These transformed scores are more suitable for an analysis of variance than the original scores. Both the factor 'accent' ($F = 34.1$, $p < 0.001$) and 'syllable type' ($F = 58.0$, $p < 0.001$) were significant. Tests for pairwise contrasts between conditions showed that the conditions NS-NA, F-A, and F-NA (the group RED) were not significantly different from each other ($p > 0.05$). In addition, the condition NS-A was not significantly different from the condition F-A ($p > 0.05$) for this limited data set (only 3 speakers).

The factor 'speaker' was also significant ($F = 9.5$, $p < 0.001$). The error percentages in the test conditions for each speaker separately are shown in Table 4. As might be expected, the vowels from speaker S_1 are more often correctly identified than those from speaker S_8 and speaker S_{15}, for whom the error percentages are highest. The rather high error percentages for the condition ISO are mainly caused by confusions of long vowels with their short counterparts (and vice versa).

TABLE 4 ABOUT HERE

TABLE 5 ABOUT HERE

Despite the fact that the acoustic reduction can be considerable, especially in the vowels from the group RED, listeners are yet often capable to correctly identify them. Table 5 shows a confusion matrix in percentages for the vowels in the test syllables from the group RED pooled over the three speakers. Most of the errors that are made in the vowel identifications are confusions with neighbouring vowels or confusions of long vowels with their short counterpart (and vice versa). The vowel /ɛ/ from the syllable /wɛl/ was frequently confused with the vowels /ɔ/ and /ɑ/, and the vowel /ɛ/ from the syllable /hɛp/ was frequently confused with the vowel /ɪ/. This is in agreement with the observed shifts of formant frequencies discussed in section 4.2. The 'centralization' effect is evidenced by the confusions with the central vowels /œ/ and /ø:/, which were not present in the actual experiment, but only existed as response possibility. About 15% of all vowels from the group RED were confused with either the /œ/, or the /ø:/.

4.5 Models of spectral vowel reduction

The present dataset, which is based on fairly natural speech, can be used to find models that predict the amount of spectral reduction in vowels (i.e. the shift of steady-state formant frequencies). In order to eliminate noise caused by physiological differences between speakers and idiosyncratic peculiarities, the average formant values (and curvature values) of the 15 speakers will be used. In this way we can model the vowels of an 'average male speaker'. The shifts of steady-state formant frequencies with respect to those of 'ideal' vowels (condition ISO) will be considered in the following four conditions which turned out to be significantly different in section 4.1.1:

1. Stressed syllables in accented content words (S-A)
2. Stressed syllables in unaccented content words (S-NA)
3. Unstressed syllables in accented content words (NS-A)
4. Unstressed syllables in unaccented content words, and all function words (RED)

The first model that could be used to predict spectral vowel reduction is described with the following formula:

$$\Delta F = a + b\, F_{target}$$

The input to this model is the steady-state target formant frequency F_{target} (either F_1 or F_2) of a vowel in a specific consonantal context. Formant frequencies of vowels from the condition ISO are considered to be target formant frequencies. The output of the model is the amount of formant shift for this vowel (in the same consonantal context) due to one of the four experimental conditions mentioned above:

$$\Delta F = F_{target} - F_{reduced}$$

where F_{target} denotes the steady-state target formant frequency and $F_{reduced}$ the steady-state formant frequency of the reduced vowel in a particular test condition. The constants

a and *b* of this 'static' model were obtained from the present data set through a least squares fit. They are given in Table 6 for F_1 or F_2 in each condition. The fit of the various versions of the model indicated with the correlation coefficient turns out to be fairly good.

TABLE 6 ABOUT HERE

The second model that could be used to predict spectral vowel reduction is described with the formula:

$$\Delta F = a + b\, cur_{target}$$

The input to this 'dynamic' model is the parabolic target curvature cur_{target} for the formant track (either F_1 or F_2) of a vowel in a specific consonantal context, calculated as described in section 4.1.2. Curvature values of vowels from the condition ISO are considered to be target curvature values. The output is again the amount of formant shift for this vowel (in the same consonantal context) due to one of the four experimental conditions. Table 7 gives the model parameters for F_1 or F_2 in each condition. Since formant curvature and formant frequency shifts are relative measures without reference to absolute formant values, this model can also be fitted to F_1 and F_2 simultaneously. The model parameters for this combined model are given in the last column of Table 7. The dynamic approach of this model may be more realistic than the static approach of the former one, but the performance of both models is comparable.

TABLE 7 ABOUT HERE

The third model we want to discuss is Lindblom's (1963) model, which is described by the following formula:

$$\Delta F = a \left(F_{target} - F_{onset} \right) \mathrm{e}^{-b\, dur}$$

This model has two different inputs. The first input is the 'locus-target' distance $F_{target} - F_{onset}$ which is the difference between the target formant frequency and the formant frequency at the initial vowel boundary. This is a measure for the coarticulatory effect between the vowel and the preceding consonant and resembles the curvature measure of the 'dynamic' model mentioned above. The second input is the total duration *dur* of the vowel segment. According to Lindblom, these two measures determine the amount of formant shift independent of linguistic conditions like, for instance, stress or sentence accent.

The model parameters *a* and *b* were estimated by Lindblom for three different consonantal contexts /b-b/, /d-d/, and /g-g/ with the data of one male speaker. These data consisted of 24 CVC nonsense syllables (8 short vowels × 3 consonantal contexts) uttered either at the beginning or at the end of a carrier sentence and either with or without a 'sentence accent'. The experimental conditions generated vowel durations in the range of approximately 100-250 msec. Consequently, Lindblom's models are based on this range of durations. In more natural speech, however, the durations of vowels are usually

shorter. In the present experiment, for instance, the durations of the short vowels in the 6 'natural' conditions varied in the range 50-95 msec (one standard deviation around the mean). It is interesting to see that the F_2-shifts predicted by Lindblom's equations can become rather extreme as the vowel durations get shorter. The predicted F_2-shifts for the /ʊ/ and the /a/ in the syllables /dʊd/ and /gag/, for instance, are larger than 1000 Hz for vowel durations of 30 msec. Such durations were not unusual for the short vowels in the present experiment, but the formant shifts we observed were never larger than about 500 Hz. Krull (1989) who investigated vowel reduction in spontaneous speech also reports F_2-shifts that do not become larger than about 500 Hz. So it seems unlikely that Lindblom's models, based on vowels with a rather long duration in nonsense syllables, can simply be extrapolated to vowels with shorter durations that occur in more natural speech.

TABLE 8 ABOUT HERE

In the present experiment we did not systematically vary the vowels in one particular consonantal context as Lindblom did, so it was not possible to reliably determine the parameters of his model. Nevertheless, we could check if there exists a monotonic relation between vowel duration and spectral reduction for each of the 33 CVC syllables separately. For this purpose we calculated the rank correlation between the duration of a vowel and its steady-state formant frequency (F_1 or F_2) for each CVC syllable. Each rank correlation was based on average measurements from all test conditions ($n = 7$). In Table 8 the rank correlations per vowel (averaged over consonantal contexts) are given for both F_1 and F_2. In most cases there seems to be a fair relation between vowel duration and formant undershoot. However, we do not agree with Lindblom's view that spectral reduction is *caused* by a decreasing vowel duration. We believe that both duration and spectral quality usually decrease at the same time as the pronunciation gets sloppier.

5. DISCUSSION

In the present experiment it was found that the spectral quality of vowels can be influenced by a variety of factors. An important one is *stress*. In polysyllabic words the vowels from stressed syllables are generally more clearly pronounced and hence closer to their target form than the vowels from unstressed syllables. Stressing and destressing are properties that can only be assigned to existing words; it is difficult to simulate the natural destressing of syllables in nonsense words. A demonstration of this point is the attempt to gather 'reduced' diphones for a Dutch speech synthesis system through the simulation of natural stress patterns in nonsense words (Drullman and Collier, 1991). It appeared that the spectral quality of the vowels from 'reduced' and 'unreduced' diphones hardly differed. In addition, Koopmans-van Beinum (1992) showed, that the vowels from these 'reduced' diphones formed a *larger* vowel triangle than the vowels from stressed syllables of 'focus words' in a text that was read aloud by the same speaker. Apparently, the lack of a clear pronunciation of unstressed syllables can only naturally occur in existing words that were uttered thousands of times by a speaker.

Apart from stress, it was found that spectral reduction is influenced by *surrounding phonemes* (see Fig. 7). In general, vowels that are surrounded by 'compatible' phonemes will retain a rather high quality under various different linguistic and extralinguistic circumstances. Vowels that are surrounded by 'incompatible' phonemes, on the other hand, will show effects of contextual assimilation that can rapidly increase when the speech gets sloppier. This view on the relation between 'compatibility' and the amount of spectral reduction is supported by the fairly good fit of the 'dynamic' model discussed in section 4.5 (if we regard the amount of curvature of formant tracks in clearly pronounced vowels as a measure of 'compatibility' between the vowel and the surrounding consonants).

The factor *word class* can also play an important role. In the present experiment it was found that the vowels from monosyllabic function words with primary stress had the same poor quality as unstressed vowels from content words. This shows that the effect of stress on vowel reduction can be overruled by the factor word class. It is not clear whether the poor articulation of the vowels in function words is mainly related to their small 'semantic load', or to their high frequency of occurrence. Both factors may play a role, but we suspect that *frequency of occurrence* is of greater importance, because the frequency effect is also clearly present across word classes. Van Coile (1987) found, for instance, that vowel durations in read aloud speech could be better predicted by a word dichotomy based on frequency of occurrence than by a word dichotomy based on word class (in its strict sense).

Another factor that can influence the amount of spectral vowel reduction is *sentence accent*, which is used to highlight parts of utterances. However, the results of the present experiment suggest that this effect is of minor importance compared to word stress and word class. This may be explained by the fact that sentence accent is only assigned temporarily to words in specific situations, whereas word stress, word class, and frequency of occurrence are fixed properties of words.

A final factor of importance with respect to vowel reduction, that was not explicitly investigated in the present study, is *speech style* which covers many related phenomena. Each person speaks in his own way due to idiosyncratic features, and to his geographic and sociolinguistic background. All speakers from the present experiment spoke 'standard Dutch'. However, some had a rather acting style of reading, whereas others read in a more neutral style. Different speech styles can also be used by one and the same speaker in different situations; a conversation with friends requires a different speech style than, for instance, an application for a job, or a lecture. A less formal speech style will usually lead to a sloppier pronunciation and more vowel reduction (Koopmans-van Beinum, 1980; Van Bergem and Koopmans-van Beinum, 1989; Krull, 1989; Duez, 1991). We regard differences in tempo also as differences in speech style. Tempo is often influenced by the emotional state of the speaker. Someone who is, for instance, excited about something may speak fast and rather sloppily, resulting in strongly reduced vowels. However, someone who is, for instance, angry may speak very fast and yet have a clear articulation. Thus, a higher speaking rate can cause vowel reduction, but only if it evokes a sloppy pronunciation.

According to Lindblom (1963) the degree of spectral reduction of a vowel is, apart from the influence of surrounding phonemes, determined by the duration of the vowel segment. His hypothesis was that the articulators are unable to reach an 'ideal' vowel target position if there is not enough time. Although superficially attractive, the clear

relation between vowel duration and spectral quality has been contradicted by several other investigations (e.g. Gay, 1978; Nord, 1986; Engstrand, 1988; Van Son and Pols, 1990). In the present study it was found that there is no clear relation between the vowel separation of a speaker and the duration of his vowels (section 4.3). Some speakers produce vowels with relatively long durations and have a poor vowel separation, whereas others produce vowels with relatively short durations and have a clear vowel separation. This is most likely dependent on the particular speech style a certain speaker uses and not on physiological constraints of his articulatory organs. In our view spectral reduction is not caused by a decreasing vowel duration, but both duration and spectral quality usually decrease at the same time if the pronunciation gets sloppier. This covariation may occur frequently, but not necessarily. The sloppy pronunciation can in turn be caused by a great variety of factors such as stress, word class, emotion, etc.

The amount of vowel reduction is dependent on the specific combination of factors mentioned above. However, these factors are only related to vowel reduction in a statistical sense. No combination of factors is able to block the occurrence of a clear vowel articulation; each speaker has the possibility to pronounce any speech sound as clearly as he likes, regardless of reduction factors. For all the combinations of reduction factors there only exists a certain *probability* that sloppy speech will occur and that the vowels will be reduced to some extent. In general, this will be the case if the particular speech part is of minor importance for a proper understanding of the message by the listener.

The present study was exclusively concerned with acoustic vowel reduction. In future research we are planning to investigate the relation between acoustic and lexical vowel reduction, especially with regard to word stress and frequency of occurrence of words.

6. CONCLUSIONS

In the present experiment it was found that:

1. Sentence accent, word stress, and word class have a significant effect on the steady-state formant frequencies and on the duration of vowels. The effect of sentence accent is of minor importance compared to the effect of word stress and word class.

2. Spectral vowel reduction is not a tendency of vowels to centralize, but rather the result of an increased contextual assimilation. The observed centralization of formant patterns is a frequently occurring and probably natural consequence of this assimilation process. In reduced vowels formant tracks become flatter.

3. The perceptual significance of the acoustic measurements was confirmed in a listening experiment. There is a clear influence of the factors sentence accent, word stress, and word class on the identification of vowels.

ACKNOWLEDGEMENTS

I would like to thank Louis Pols, Florien Koopmans-van Beinum, and Gitta Laan for their critical discussions about the experimental design of this investigation and for their useful comments on earlier versions of this article. I would also like to thank Björn Lindblom for his very thorough review and many valuable suggestions.

REFERENCES

Altman, G. & Carter, D. M. (1989), 'Lexical stress and lexical discriminability: Stressed syllables are more informative, but why?', Computer Speech and Language 3, 265-275.

Baart, J.L.G. (1987), Focus, Syntax, and Accent Placement, Doct. diss., University of Leiden.

Bergem, D. R. van (1989), 'Phonetic and linguistic aspects of vowel reduction', Proc. of the Institute of Phonetic Sciences Amsterdam 13, 97-105.

Bergem, D. R. van & Koopmans-van Beinum, F. J. (1989), 'Vowel reduction in natural speech', Proc. Eurospeech '89, Paris, Vol. 2, 285-288.

Bergem, D. R. van (1990a), 'In defense of a probabilistic view on human word recognition', Proc. of the Institute of Phonetic Sciences Amsterdam 14, 53-66.

Bergem, D. R. van (1990b), 'The influence of linguistic factors on vowel reduction', Proc. of Linguistics and Phonetics '90, Charles University, Prague, 427-436.

Bergem, D. R. van (1991a), 'The influence of sentence accent, word stress, and word class on the quality of vowels', Proc. Eurospeech '91, Genova, Vol. 3, 1455-1458.

Bergem, D. R. van (1991b), 'Acoustic and lexical vowel reduction', Proc. of the ESCA Workshop 'Phonetics and phonology of speaking styles: Reduction and elaboration in speech communication', Barcelona, 10:1-10:5.

Bolinger, D. (1972), 'Accent is predictable (If you're a mind-reader)', Language 48, 633-644.

Bolinger, D. (1975), Aspects of language, Harcourt Brace Jovanovich, New York.

Bolinger, D. (1985), 'Two views of accent', J. Linguistics 21, 79-123.

Booij, G.E. (1981), Generatieve fonologie van het Nederlands, Het Spectrum, Utrecht.

Carter, D. M. (1987), 'An information-theoretic analysis of phonetic dictionary access', Computer Speech and Language 2, 1-11.

CELEX-report (1985), Proposal for creating a national, multilingual, lexical database, University of Nijmegen.

Coile, B.M. van (1987), 'A model of phoneme durations based on the analysis of a read Dutch text', Proc. European Conference on Speech Technology, Edinburgh, Vol. 2, 233-236.

Collins Dictionary of the English Language (1989), Second Edition, Collins, London.

Daniloff, R.G. & Hammarberg, R.E. (1973), 'On defining coarticulation', J. Phon. 1, 239-248.

Delattre, P. (1969), 'An acoustic and articulatory study of vowel reduction in four languages', IRAL 7, 295-325.

Drullman, R. & Collier, R. (1991), 'On the combined use of accented and unaccented diphones in speech synthesis', J. Acoust. Soc. Am. 90, 1766-1775.

Duez, D. (1991), 'Some factors affecting F_2-patterns in spontaneous speech in French', Proc. of the ESCA Workshop 'Phonetics and phonology of speaking styles: Reduction and elaboration in speech communication', Barcelona, 21:1-21:5.

Engstrand, O. (1988), 'Articulatory correlates of stress and speaking rate in Swedish CVC utterances', J. Acoust. Soc. Am. 83, 1863-1875.

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

Forster, K.I. (1979), 'Levels of processing and the structure of the language processor', In: Cooper, W.E. & Walker, E.C.T. (Eds.), Psycholinguistic studies presented to Merrill Garrett, 27-85.
Fourakis, M. (1991), 'Tempo, stress, and vowel reduction in American English', J. Acoust. Soc. Am. 90, 1816-1827.
Gay, T. (1978), 'Effect of speaking rate on vowel formant movements', J. Acoust. Soc. Am. 63, 223-230.
Hart, J. 't, Collier, R. & Cohen, A. (1990), A perceptual study of intonation. An experimental-phonetic approach to speech melody, Cambridge University Press, Cambridge.
Kirk, R.E. (1982), Experimental design: Procedures for the behavioral sciences, Wadsworth, Inc., Belmont.
Klatt, D.H. (1989), 'Review of selected models of speech perception', In: Marslen-Wilson, W.D. (Ed.), Lexical representation and process, Cambridge, Mass.: MIT Press, 169-226.
Koopmans-van Beinum, F.J. (1980), Vowel contrast reduction: An acoustic and perceptual study of Dutch in various speech conditions, Doct. diss., University of Amsterdam.
Koopmans-van Beinum, F.J. (1992), 'The role of focus words in natural and in synthetic continuous speech: Acoustic aspects', Speech Communication 11, 439-452.
Krull, D. (1989), 'Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech', PERILUS X, 87-108.
Kruyt, J.G. (1985), Accents from speakers to listeners, Doct. diss., University of Leiden.
Ladefoged, P. (1982), A course in phonetics, Harcourt, Brace, Jovanovitz, New York.
Lee, K.F. (1989), Automatic speech recognition: The development of the SPHINX system, Kluwer Academic Publishers, Boston.
Lindblom, B.E.F. (1963), 'Spectographic study of vowel reduction', J. Acoust. Soc. Am. 35, 1773-1781.
Lindblom, B. & Moon, S.J. (1988), 'Formant undershoot in clear and citation-form speech', PERILUS VIII, 21-33.
Makhoul, J. & Cosell, L. (1976), 'LPCW: An LPC vocoder with linear predictive spectral warping', Proc. ICASSP 1976, 466-469.
Markel, J.D. & Gray, A.H. (1976), Linear prediction of speech, Springer-Verlag, New York.
Morton, J. (1969), 'Interaction of information in word recognition', Psychological Review 2, 165-178.
Nord, L. (1975), 'Vowel reduction - centralization or contextual assimilation?', In: Fant, G. (Ed.), Speech Communication, Vol. 2, Almqvist & Wiksell Int., Stockholm, 149-154.
Nord, L. (1986), 'Acoustic studies of vowel reduction in Swedish', KTH-QPSR Stockholm 4, 19-36.
Oppenheim, A.V. & Johnson, D.H. (1972), 'Discrete representation of signals', Proc. IEEE 60, 681-691.
Pols, L.C.W. (1977), Spectral analysis and identification of Dutch vowels in monosyllabic words, Doct. diss., Free University of Amsterdam.

Pols, L.C.W., Tromp, H.R.C. & Plomp, R. (1973), 'Frequency analysis of Dutch vowels from 50 male speakers', J. Acoust. Soc. Am. 53, 1093-1101.

Rabiner, L.R., Levinson, S.E. & Sondhi, M.M. (1983), 'On the application of vector quantization and Hidden Markov Models to speaker-independent, isolated word recognition', The Bell System Technical Journal 62, 1075-1105.

Son, R.J.J.H. van & Pols, L.C.W. (1990), 'Formant frequencies of Dutch vowels in a text, read at normal and fast rate', J. Acoust. Soc. Am. 88, 1683-1693.

Son, R.J.J.H. van & Pols, L.C.W. (1991), 'The influence of speaking rate on vowel formant track shape as modelled by Legendre polynomials', Proc. of the Institute of Phonetic Sciences Amsterdam, 15, 43-61.

Son, R.J.J.H. van & Pols, L.C.W. (1992), 'Formant movements of Dutch vowels in a text, read at normal and fast rate', J. Acoust. Soc. Am. 92, 121-127.

Stålhammar, U., Karlsson, I. & Fant, G. (1973), 'Contextual effects on vowel nuclei', KTH-QPSR Stockholm 4, 1-18.

Svendsen,T. & Soong, F.K. (1987), 'On the automatic segmentation of speech signals', ICASSP87, 77-80.

Wijk, C. van & Kempen, G. (1980), 'Funktiewoorden - Een inventarisatie voor het Nederlands', ITL, Review of Applied Linguistics 47, 53-68.

Willems, L.F. (1986), 'Robust Formant Analysis', IPO-APR 21, 34-40.

Figure 1 Example of a vowel segmentation: the /ɑ/ is segmented from the function word /kɑn/. The bottom figure shows the oscillogram, the middle figure the spectral variation contour *V(k)* (the segmentation threshold at a value of 0.075 is indicated with a dashed line), and the top figure three formant tracks on a mel scale. For an explanation of the segmentation procedure, see text.

Figure 2a Formant plot (F_1 and F_2) in mel and in Hz of the *short* vowels for 7 test conditions (see text) averaged over 15 speakers and all consonantal contexts. Formant frequencies were measured at the steady-state vowel part.

Figure 2b Formant plot (F_1 and F_2) in mel and in Hz of the *long* vowels for 7 test conditions (see text) averaged over 15 speakers and all consonantal contexts. Formant frequencies were measured at the steady-state vowel part.

Figure 3 Average Euclidean distances (F_1 and F_2) in mel between vowels from each test condition and their 'ideal' counterpart (vowels from syllables spoken in isolation). Syllable types are: stressed syllables (S), unstressed syllables (NS), and function words (F). These can occur in accented words (A), or in unaccented words (NA).

Figure 4a Curvature of formant tracks (F_1 and F_2) of the *short* vowels in the conditions ISO (asterisks) and RED (dots) averaged over 15 speakers and all consonantal contexts. The curvature is indicated with the coefficient $-c_2$ from the polynomial fit of the formant tracks: $F(t) = c_0 + c_1\, t + c_2\, t^2$.

Figure 4b Curvature of formant tracks (F_1 and F_2) of the *long* vowels in the conditions ISO (asterisks) and RED (dots) averaged over 15 speakers and all consonantal contexts. The curvature is indicated with the coefficient $-c_2$ from the polynomial fit of the formant tracks: $F(t) = c_0 + c_1\, t + c_2\, t^2$.

Figure 5 Average Euclidean curvature distances (F_1 and F_2) between vowels from each test condition and their 'ideal' counterpart (vowels from syllables spoken in isolation). Syllable types are: stressed syllables (S), unstressed syllables (NS), and function words (F). These can occur in accented words (A), or in unaccented words (NA).

Figure 6 Average durations in ms for each of the 6 'natural' test conditions. Syllable types are: stressed syllables (S), unstressed syllables (NS), and function words (F). These can occur in accented words (A), or in unaccented words (NA).

Figure 7 Shift of formant frequencies (F_1 and F_2) for the vowel /ɛ/ in 6 different consonantal frames. Vowels from the condition ISO (asterisks) and the group RED (dots) are compared. The vowels /i, u, a:/, adopted from Figs. 2a and 2b, have been added for reference only.

Figure 8 Schematic picture of three 'undershoot' models (see text). The thick line in each model indicates the 'ideal' formant track of a vowel in a CVC syllable and the thin line the same formant track in its reduced form.

Figure 9 Sigma ratio for each speaker, indicating the amount of vowel separation (see text).

Figure 10a Clusters of short vowels for speaker S_1 who had the best vowel separation according to Fig. 9.

Figure 10b Clusters of short vowels for speaker S_{15} who had the worst vowel separation according to Fig. 9.

Figure 11 Average duration in ms of the vowels from each speaker in the 6 'natural' test conditions.

Figure 12 Mean error percentages for the 6 'natural' test conditions from a listening experiment with 24 subjects in which the vowels from 3 speakers had to be identified.

Table 1 Complete list of all 33 triplets of words that were used in the present experiment.

Table 2 Mean durations in ms of short vowels and long vowels separately in the 6 'natural' test conditions averaged over 15 speakers.

Table 3 Percentages assignments to each of the three models shown in Fig. 8 for the vowels in all test syllables at the onset and the offset of F_1 and F_2 (see text).

Table 4 Mean error percentages for all test conditions from a listening experiment with 24 subjects in which the vowels from 3 speakers had to be identified. The error percentages are given for each speaker separately.

Table 5 Confusion matrix in percentages for all vowels in the test syllables from the group RED pooled over three speakers. Correct scores are indicated in grey.

Table 6 Parameters for the 'static' reduction model in 4 different conditions (see text). The intercept of the model is *a*, and the slope is *b*. The fit of each model is indicated with the correlation coefficient *r*.

Table 7 Parameters for the 'dynamic' reduction model in 4 different conditions (see text). The intercept of the model is *a*, and the slope is *b*. The fit of each model is indicated with the correlation coefficient *r*.

Table 8 Rank correlations per vowel between steady-state formant frequencies (F_1 and F_2) in mel and vowel durations averaged over test syllables.

TABLE 1

Vowel	IPA	Function word	Unstressed syllable	Stressed syllable
u	mut	moet	weemoed	vermoedt
ɔ	kɔm	kom	kompas	komma
ɔ	kɔn	kon	consult	consul
ɔ	tɔχ	toch	hertog	tochtig
ɑ	dɑn	dan	danseres	pedante
ɑ	kɑn	kan	kantine	kandelaar
ɑ	lɑχ	lag	glimlach	gelach
ɑ	mɑχ	mag	magneet	machtig
ɑ	pɑs	pas	pastoor	passer
ɑ	vɑn	van	vandalen	havanna
ɑ	wɑs	was	afwas	gewas
ɛ	bɛn	ben	benzine	bendelid
ɛ	hɛp	heb	hebzuchtig	hebzuchten
ɛ	mɛn	men	mentaal	mentor
ɛ	pɛr	per	perceel	perzik
ɛ	wɛl	wel	weldadig	weldaden
ɛ	zɛt	zet	inzet	bezet
ɪ	lɪχ	lig	voorlichting	verlichting
ɪ	wɪl	wil	knotwilgen	verwilderd
ɪ	zɪχ	zich	inzichten	voorzichtig
ɪ	zɪt	zit	bijzit	bezit
i	lit	liet	getralied	satelliet
i	nit	niet	deugniet	graniet
o:	do:r	door	doortastend	doorsnede
o:	vo:r	voor	voordelig	voordelen
a:	da:r	daar	standaarden	bedaarde
a:	la:t	laat	uitlaat	verlaat
a:	na:r	naar	evenaar (noun)	evenaar (verb)
a:	wa:r	waar	waardering	waardeloos
e:	he:l	heel	heelkundig	heelkunde
e:	me:r	meer	zeemeermin	kalmeerde
e:	ve:l	veel	veelzijdig	veelvouden
e:	we:r	weer	onweer	geweer

TABLE 2

Syllable type	Short vowels		Long vowels	
	+ accent	– accent	+ accent	– accent
+ stress	82	82	150	131
– stress	76	71	104	93
func	70	67	91	88

TABLE 3

Model	F_1		F_2	
	onset	offset	onset	offset
I	0	3	15	24
II	39	45	21	12
III	61	52	64	64

TABLE 4

Condition	Speaker			
	S_1	S_8	S_{15}	Average
ISO	13	18	23	18
S-A	21	22	28	24
S-NA	21	26	41	30
NS-A	29	41	57	42
NS-NA	35	54	71	53
F-A	31	50	66	49
F-NA	40	51	70	54

TABLE 5

Intended syllable	Vowel responses											
	u	ɔ	ɑ	ɛ	ɪ	i	y	œ	o:	a:	e:	ø:
mut	77	2	1	0	0	2	8	7	0	0	0	1
kɔm	20	64	2	0	2	6	2	2	1	0	0	0
kɔn	20	73	1	0	1	1	0	0	3	0	0	0
tɔχ	1	62	13	1	0	0	0	12	9	0	0	2
dɑn	1	8	68	0	1	0	0	8	3	7	0	3
kɑn	2	19	63	2	0	0	0	9	0	2	0	1
lɑχ	0	21	74	0	0	0	0	3	0	2	0	0
mɑχ	2	10	84	0	0	0	0	1	0	4	0	0
pɑs	2	48	23	0	1	0	0	20	0	1	0	3
vɑn	8	8	62	1	0	2	5	13	0	0	0	0
wɑs	6	51	24	1	0	0	0	10	5	2	0	2
bɛn	0	0	1	49	8	6	6	22	0	3	1	4
hɛp	0	0	0	42	49	0	3	4	0	0	2	0
mɛn	3	0	2	62	2	2	15	11	0	2	0	0
pɛr	0	1	1	13	3	0	0	75	0	0	0	7
wɛl	0	48	34	7	1	0	0	6	2	0	0	0
zɛt	0	0	1	47	18	0	0	27	0	0	1	6
lɪχ	0	1	0	6	49	1	6	29	0	0	3	5
wɪl	2	6	0	0	19	1	9	53	1	1	2	6
zɪχ	0	0	0	6	70	0	4	13	0	0	6	0
zɪt	0	0	0	0	86	7	1	2	0	0	3	0
lit	0	0	0	2	25	56	6	5	0	0	2	2
nit	2	0	0	0	6	85	5	1	0	0	0	0
do:r	23	14	0	0	0	0	2	19	38	0	0	3
vo:r	38	21	0	0	1	0	5	17	15	0	0	1
da:r	0	0	5	3	0	0	1	7	0	70	0	13
la:t	0	4	46	1	0	0	0	7	0	41	0	0
na:r	2	0	19	2	1	0	1	7	0	63	0	4
wa:r	0	7	41	2	0	0	0	19	1	18	0	10
he:l	2	0	13	33	25	0	1	12	0	4	10	0
me:r	0	0	0	7	29	20	8	4	0	0	31	0
ve:l	0	5	6	7	19	0	1	37	2	0	12	10
we:r	0	0	0	9	34	2	16	3	0	0	28	7

TABLE 6

Condition	F_1			F_2		
	a	*b*	*r*	*a*	*b*	*r*
S-A	– 62	0.13	0.61	– 138	0.13	0.77
S-NA	– 161	0.38	0.88	– 241	0.22	0.87
NS-A	– 158	0.39	0.79	– 388	0.33	0.85
RED	– 202	0.50	0.87	– 483	0.41	0.87

TABLE 7

Condition	F_1			F_2			F_1 and F_2		
	a	*b*	*r*	*a*	*b*	*r*	*a*	*b*	*r*
S-A	– 15	0.18	0.56	19	– 0.24	0.83	4	– 0.14	0.56
S-NA	– 23	0.52	0.77	28	– 0.40	0.92	14	– 0.33	0.80
NS-A	– 24	0.61	0.79	24	– 0.63	0.92	6	– 0.52	0.85
RED	– 28	0.77	0.86	30	– 0.78	0.95	7	– 0.65	0.89

TABLE 8

	u	ɔ	ɑ	ɛ	ɪ	i	o:	a:	e:
F_1	0.56	0.80	0.81	0.80	0.50	0.57	0.16	0.82	0.50
F_2	0.71	0.80	0.36	0.84	0.82	0.80	0.92	0.59	0.85

(20)

ɛ ɛ ɛ ɛ ɛ ɛ ɛ ɛ ɛ ɛ

(16)

bɛn	mɛn	pɛr	hɛp	zɛt	wɛl
bɛn	mɛn	pɛr	hɛp	zɛt	wɛl
bɛn	mɛn	pɛr	hɛp	zɛt	wɛl
bɛn	mɛn	pɛr	hɛp	zɛt	wɛl

(28)

u	ɔ	ɑ	ɛ	ɪ	i	o:	a:	e:	ə
u	ɔ	ɑ	ɛ	ɪ	i	o:	a:	e:	ə
u	ɔ	ɑ	ɛ	ɪ	i	o:	a:	e:	ə
u	ɔ	ɑ	ɛ	ɪ	i	o:	a:	e:	ə
u	ɔ	ɑ	ɛ	ɪ	i	o:	a:	e:	ə
u	ɔ	ɑ	ɛ	ɪ	i	o:	a:	e:	ə