



## THE IMPORTANCE OF SPECTRAL QUALITY OF VOWELS FOR THE INTELLIGIBILITY OF SENTENCES

Gitta P.M. Laan, Dick R. van Bergem & Florian J. Koopmans-van Beinum

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

### ABSTRACT

The intelligibility of Dutch sentences in which specific vowels were systematically replaced by the Dutch vowel /œ/ was tested. The prosodic features of the vowels were copied to their substitute by using the TD-PSOLA technique.

Plausibility of the sentences appeared to be the predominant factor in this experiment. Word frequency had only a small effect. For function words the number of correct responses was higher than for content words. A word stress effect could not be demonstrated.

### 1. INTRODUCTION

The spectral quality of vowels in a natural speech utterance is influenced by a lot of factors such as coarticulation, reduction, and prosody. The reduction of vowels, for instance, can result in a schwa-like sound. However, all this variation does not seem to disturb the listener in a natural speech situation. The importance of a high spectral quality of vowels seems to decrease with the increase of contextual information: Whereas spectral quality is the main cue for identifying an isolated vowel, high spectral quality of vowels is often redundant in a complete speech utterance.

In the present investigation it was hypothesized that the importance of spectral quality of vowels for the intelligibility of sentences is dependent on the following factors:

- Plausibility.  
Meaningful sentences are more easily recognized and may therefore need a less careful pronunciation than incoherent, illogical or nonsense sentences (e.g. Foss & Ross, 1978).
- Word frequency.  
For a listener it is more difficult to recognize low frequency words than high frequency words (e.g. Morton, 1969).
- Word class.  
Content words in a sentence contain the information that is essential for comprehending the message. Function words on the other hand, mainly regulate the grammar in speech utterances and only have a small contribution to meaning (Bolinger, 1975). Therefore, spectral quality of vowels may be of less importance in function words than in content words.

### - Word stress.

Statistical studies show that the number of word candidates is smaller when the search through the (English) lexicon is based on stressed syllables than when it is based on unstressed syllables. In particular the variety of vowels is greater in stressed syllables (Altman & Carter, 1989). Consequently, the spectral quality of vowels may be of less importance for unstressed syllables than for stressed syllables.

In order to find out how the *lack* of spectral quality of vowels affects the intelligibility of sentences with respect to these factors, a listening experiment was carried out with sentences in which vowels in specific syllables were systematically replaced by the (Dutch) central vowel /œ/. The experiment is discussed in more detail in Laan (1991).

### 2. EXPERIMENTAL DESIGN

The vowels that had to be replaced by /œ/-sounds should preferably be clearly distinguishable from the consonantal context. Therefore words with the phonemes /j, w, r, l/ were not suitable, nor were words beginning or ending with a vowel, because we wanted to keep word boundaries intact. With these restrictions we selected function words from a list of Dutch function words given by Van Wijk & Kempen (1980), and content words from the Dutch lexical database CELEX (1985). High frequency content words were defined to have a frequency greater than 19 (in a corpus of 42 million words), low frequency content words were defined to have a frequency below 3. The selected lists of function words and content words were used to construct four types of sentences:

1. Plausible sentences with high frequency content words (PH).
2. Plausible sentences with low frequency content words (PL).
3. Implausible sentences with high frequency content words (IH).
4. Implausible sentences with low frequency content words (IL).

Each group consisted of 10 sentences. The implausible sentences were made after the example of the plausible sentences: New function words and content words were put in place of the function words and content words from the plausible sentences, in such a way as to make nonsense sentences (that is,

meaningless but more or less grammatical sentences). In this way we tried to obtain an equal number of syllables and words of the same word class in each group of sentences. The lengths of the sentences varied between 4 to 7 words (and 7 to 11 syllables), so that the stimulus material was not too monotonous. For the same reason each group of sentences contained two interrogative sentences.

Vowel replacements in the sentences were done in all syllables of function words (F), all unstressed syllables of content words (U), all stressed syllables of content words (S), or combinations of these three possibilities. Adding the unchanged version (C) of each sentence gave the following 8 test conditions:

- C: No vowels replaced
- F: All vowels replaced in function words
- U: All vowels replaced in unstressed syllables of content words
- S: All vowels replaced in stressed syllables of content words
- FU: All vowels replaced in function words and unstressed syllables of content words
- FS: All vowels replaced in function words and stressed syllables of content words
- US: All vowels replaced in unstressed and stressed syllables of content words
- FUS: All vowels replaced.

Syllables with 'natural' schwa's were not replaced. An example of a plausible and corresponding implausible Dutch test sentence is (Syllable types (F, U, S) that could be replaced are indicated above the sentences):

F S S F U S  
P: Dat beest vangt soms een konijn.  
(That animal sometimes catches a rabbit.)

F S S F U S  
I: Zijn maand snapt vaak een minuut.  
(His month often understands a minute.)

The /œ/-sounds that replaced the original vowels in the words were taken from nonsense words in which they had consonantal contexts identical to those of the original vowels. Apart from the original vowels, their immediate consonantal contexts were also replaced in order to avoid coarticulatory cues that could lead to the identification of the original vowel.

The 40 sentences and the list of /œ/-words embedded in a carrier sentence were read aloud by a male speaker in an anechoic room and recorded with a Sennheiser MKH 105T microphone and the audio channel of a Panasonic NV-F70HQ video recorder. The test material was lowpass filtered at 4.5 kHz and digitized at a sample rate of 20 kHz with a 12 bit resolution.

In order to maintain a high speech quality the TD-PSOLA technique (Moulines & Charpentier, 1990) was used to replace the vowels by /œ/-sounds. Pitch periods were automatically marked (Van Bergem, 1990) both in the syllables in which vowels had to be replaced and in the corresponding /œ/-words. The

period markers in the original speech utterance served as synthesis parameters for the /œ/-sounds. The manipulated /œ/-sounds then took the place of the original vowels in the test sentences. The pitch contour, duration, and energy of the original vowel were simply copied to its substitute, so that prosodic features of the speech utterance were not affected.

A total of 8 listening tests were composed in such a way that each test contained a particular sentence in only one of the 8 conditions. In the listening experiment 32 subjects participated (2 males and 2 females for each test). The task of the subjects was to listen only once to a stimulus sentence and then to report what they thought was being said. They were instructed to use existing words in their responses (no phonetical transcriptions) and mark unrecognized words with an '-' only if they really could not think of anything to respond with. Directly after typing their response the subjects were asked to indicate on a five-point scale the 'ease of recognition' they had experienced when hearing the sentence. The subjects could use as much time as they needed to respond to each stimulus sentence. Each subject heard the test in a different random order.

The score on a sentence was determined by the proportion 'correctly' recognized words as well as the indicated ease of recognition. A word was considered to be correctly recognized when it literally matched the intended word and was placed at the right position in the sentence.

### 3. RESULTS

It will be clear that the percentages of correct words will decrease when the number of replaced vowels in the sentences is increasing. Tabel 1 shows the relation between the mean percentages of replaced vowels, the mean percentages of correct words, and the mean scores for ease of recognition in this experiment.

In figure 1a mean percentages of correct words are shown in each condition for plausible sentences and implausible sentences. Plausibility of the sentences kept the recognition results above 80 % for a long

Tabel 1. Percentages of replaced vowels, percentages of correct words, and scores for ease of recognition (1=difficult to recognize, 5=easy to recognize), in each condition averaged over the four groups of sentences together with their orders.

	Condition							
	C	U	F	S	FU	US	FS	FUS
% replacements	0	14.7	23.3	25.4	38	40.1	48.7	63.4
order	1	2	3	4	5	6	7	8
% correct words	94.3	86.3	76.6	80.3	64.5	68.4	59.1	51.2
order	1	2	4	3	6	5	7	8
ease of recogn.	4.1	3.3	2.9	3	2.6	2.5	2.2	2
order	1	2	4	3	5	6	7	8

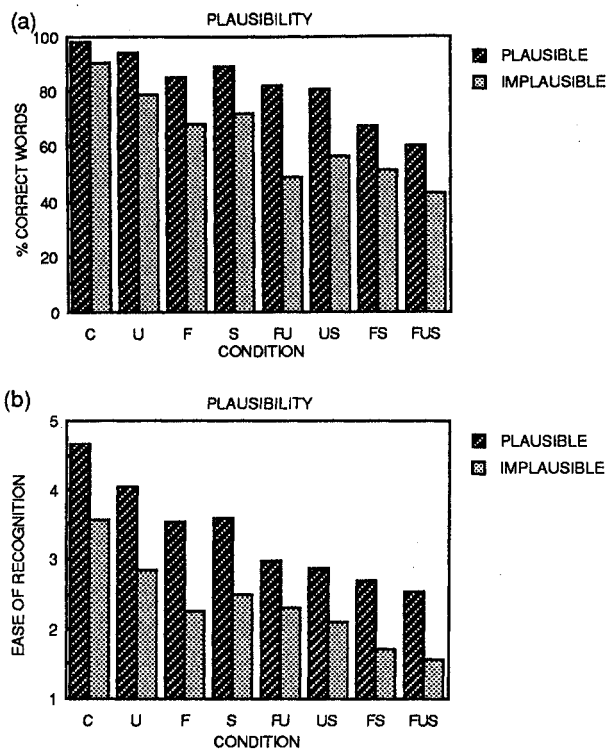


Figure 1. Percentages correct words (a) and scores for ease of recognition (b, 1=difficult to recognize, 5=easy to recognize) in each condition for plausible sentences (PH and PL) and implausible sentences (IH and IL) separately.

time. When almost all vowels were replaced (in the conditions FS and FUS) plausibility could not help the recognition process any longer. The percentages correct words collapsed to the mean level for implausible sentences (66%). This interaction of plausibility of the sentence and condition was not found in the scores for ease of recognition on plausible sentences and implausible sentences, shown in figure 1b.

Figure 2a shows the percentages of correct words in each condition for the sentences with high frequency content words (H) and low frequency content words (L). It will be clear that the frequency of the content words used in the sentences did not have any effect on the results of recognition. However, the scores for ease of recognition on H-sentences and L-sentences, which are shown in figure 2b, revealed a small but consistent word frequency effect.

An analysis of variance was done on the percentages of correct words with the factors 'plausibility', 'frequency', and 'condition'. Both the factors 'plausibility' ( $F=163.9, p<0.001$ ) and 'condition' ( $F=50.3, p<0.001$ ) were significant, but the factor 'frequency' was not significant. The interaction between 'plausibility' and 'condition' was also significant ( $F3.0, p<0.01$ ). A pairwise test for contrasts on the 8 conditions revealed that, apart from the pairs

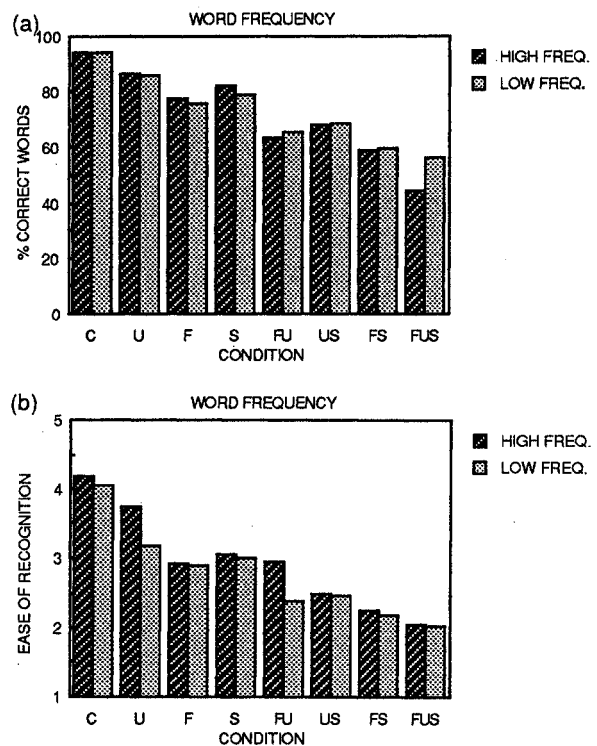


Figure 2. Percentages correct words (a) and scores for ease of recognition (b, 1=difficult to recognize, 5=easy to recognize) in each condition for sentences with high frequency content words (PH and IH) and sentences with low frequency content words (PL and IL) separately.

F-S, FU-FS, and FU-US, all conditions differed significantly from each other.

A similar analysis of variance was done on the scores for ease of recognition. From this analysis it appeared that, besides the significant factors 'plausibility' ( $F=197.4, p<0.001$ ) and 'condition' ( $F=49.4, p<0.001$ ) the factor 'frequency' was also significant ( $F=8.1, p<0.01$ ). A pairwise test for contrasts showed that, apart from the pairs F-S, F-FU, FU-US, and FS-FUS, all pairs of conditions differed significantly. Note that the differences in the orders in table 1 can be neglected because the pairs of switched conditions did not significantly differ (see the tests for contrasts above). With regard to a word class effect, the higher percentage correct words in condition F (function words) with respect to condition US (content words) might merely be the result of the fact that fewer vowels were replaced in condition F (see table 1). Instead of looking at the overall word scores on complete sentences, we looked at the responses on function words and content words separately. Based on table 2 a test for the difference of two proportions revealed that function words were significantly better recognized than content words ( $Z=23.2, p<0.001$ ). With regard to word stress the higher percentage correct words in condition U than in condition S might again merely be the result of the fact that fewer vowels

Table 2. The number of correctly and incorrectly recognized function words and content words in condition F and US respectively are shown.

	Correct	False	Total
Function words (F)	178	122	300
Content words (US)	124	156	280

were replaced in condition U (see table 1). Therefore we restricted the analysis of the responses in condition U and S to only the 29 two-syllabic content words in the test sentences, in which either the unstressed or the stressed syllable was replaced. The two-syllabic content words had a somewhat better recognition score in condition U (71.6 %) than in condition S (66.4 %), but this difference tested with a t-test was not significant.

#### 4. DISCUSSION

The results of this experiment clearly showed that even without a proper identity of vowels a speech utterance can be recognized. Especially the plausibility of a sentence was of great influence on the results. The recognition percentages of plausible sentences were on the average about 19 % higher than the percentages of the implausible sentences. Even when all vowels were replaced (FUS), the percentages of correct words in the plausible sentences was still 60 %. The subjects also indicated to have more difficulty in recognizing the implausible sentences than the plausible sentences. A natural speech situation provides a lot of extralinguistic information which may even further reduce the need for a high spectral quality of vowels.

The word frequency effect which is often reported in the literature occurred as a small effect in the scores for ease of recognition only. The ease of recognition in this experiment can be compared with often used reaction times: Low frequency words are properly recognized by the subjects, but more time (effort) is needed than for high frequency words.

Function words were more often correctly recognized than content words. This may have been caused by the fact that function words form a small set of words and therefore are rather predictable. The responses of the subjects suggested that function words are processed differently than content words. Whereas content words are retained literally, function words merely leave a global impression that is sufficient to comprehend the message. The function words "the", "not", "in", and "her" for instance were sometimes substituted by "a", "nothing", "within", and "his", respectively. We considered all these often good alternatives as incorrect responses. The problem is in fact that true understanding of a sentence, the *comprehension* of the message, should actually be determined in order to investigate a word class effect. A perceptual effect caused by vowel replacements in

either the stressed syllable or the unstressed syllable(s) of content words could not be demonstrated. The existence of such an effect is suggested by e.g. Van Bergem (1991) who found more vowel reduction in the production of unstressed syllables than in stressed syllables. A more sensitive experimental design is needed to establish a (probably small) perceptual effect of word stress.

#### 5. CONCLUSIONS

A listener can still fairly well perceive a complete speech utterance if the spectral quality of a lot of vowels is highly reduced. The following conclusions can be drawn in particular:

- Especially the plausibility of a sentence was an important factor in this experiment, and may perhaps be of even greater importance in a natural speech situation.
- Word frequency has only a relative small effect in speech perception. Subjects did experience word frequency, but recognition results were not influenced.
- Function words with replaced vowels were more often recognized than content words with replaced vowels. The importance of spectral quality of vowels in function words and content words might be better investigated by measuring in some way the comprehension of a speech utterance.
- The present experiment was not sensitive enough to demonstrate a (probably small) perceptual effect caused by vowel replacements in either stressed or unstressed syllables of content words.

#### 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.
- Bergem, D.R. van (1991). 'The influence of sentence accent, word stress, and word class on the quality of vowels', This volume.
- Bergem, D.R. van (1990). 'Pitch period estimation by filtering the fundamental frequency out of the speech waveform', *Proceedings of the Institute of Phonetic Sciences Amsterdam* 14, 17-26.
- Bolinger, D. (1975). *Aspects of language*, Hartcourt Brace Jonavich, New York.
- CELEX-report (1985). 'Proposal for creating a national, multilingual, lexical database', University of Nijmegen.
- Foss, D.J. & Ross, J.R. (1978). 'Great expectation: Context effects during sentence processing', Paper presented at the 12th International Congress of Psychology.
- Laan, G.P.M. (1991). 'The importance of spectral quality of vowels in speech perception', IFA-report 116, 34 pp. (In Dutch).
- Morton, J. (1969). 'Interaction of information in word recognition', *Psychological Review* 2, 165-178.
- Moulines, E. & Charpentier, F. (1990). 'Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones', *Speech Communication* 9, 453-467.
- Wijk, C. van & Kempen, G. (1980). 'Functiewoorden - Een inventarisatie voor het Nederlands', *ITL, Review of Applied Linguistics* 47, 53-68.