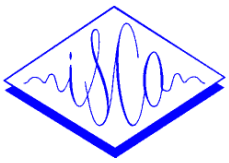
ISCA Archive
http://www.isca-speech.org/archive



3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

ON THE PERCEPTION OF ACOUSTIC AND LEXICAL VOWEL REDUCTION

Dick R. van Bergem

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

ABSTRACT

The present study was designed to investigate how well listeners are able to unambiguously categorize an unstressed vowel in a word as either a full vowel or a schwa. It was found that listeners disagree in many cases on the assignment of a vowel to either of these categories. This suggests that listeners cannot properly distinguish between acoustic reduction (the loss of spectral quality of a full vowel) and lexical reduction (the substitution of a full vowel with a schwa). Other points of interest in the present study were the frequency of occurrence of words and speech styles; both were found to have a considerable influence on the process of vowel reduction.

Keywords: *Vowel reduction, vowel categories, frequency of occurrence of words, speech styles.*

10.21437/Eurospeech.1993-164

1 INTRODUCTION

As a physical phenomenon speech is a *continuous* process, because it is produced by continuous movements of the articulators resulting in a continuously changing acoustic signal. In linguistic theories, on the other hand, speech is usually described in a *discrete* manner. It is assumed that a speaker has some kind of symbolic representation of a speech message in his mind, which he encodes in a series of articulatory movements that result in an acoustic speech signal. A listener receives this speech signal with his peripheral auditory system and subsequently has to decode it to recover the symbolic message that was intended by the speaker (see Figure 1).

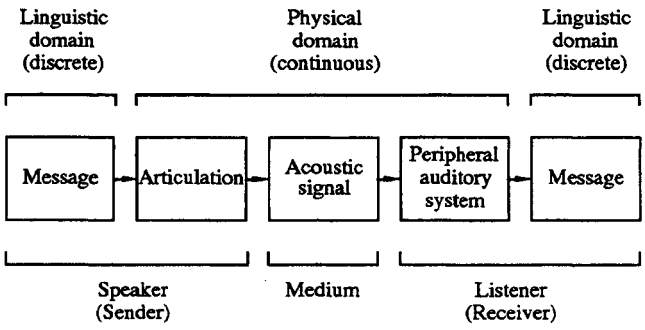


Figure 1. The speech communication process.

Whereas the physical speech signal can be studied in detail with many objective analysis techniques, the abstract representations of the speech message can actually only be guessed at.

The present study is concerned with vowel reduction. In phonology this phenomenon is described in a discrete manner: An unstressed vowel in a particular word can be either realized as a schwa or as a full vowel. Many experiments in the acoustic speech domain have shown, however, that vowel reduction can also be interpreted as a continuous process: The spectral quality of vowels can gradually decrease, see e.g. [1, 2, 3]. In order to distinguish between these two types of vowel reduction, we use the term "lexical vowel reduction" to refer to the abstract linguistic concept, and "acoustic vowel reduction" to refer to the physical phenomenon [3]. A speaker may have the *intention* to produce a schwa in a particular word that should be pronounced with a full vowel according to the standards of the linguistic community (lexical reduction). On the other hand, a speaker may also have the *intention* to produce a full vowel in a word, but the physical realization of this vowel can have a poor spectral quality due to a sloppy pronunciation (acoustic reduction). A listener only has the acoustic speech signal at his disposal to recover the vowel that was intended by the speaker. The present experiment was designed to find out if the acoustic signal contains enough cues for listeners to unambiguously categorize vowels as either a full vowel or a schwa. In addition, we wanted to investigate how the frequency of occurrence of words and speech styles influence the perception of vowel reduction.

2 EXPERIMENTAL DESIGN

2.1 Speech material

Dutch words were selected in which an unstressed vowel is sometimes pronounced as a full vowel and sometimes as a schwa. Examples of such words are (word stress is indicated with capitals, the vowel of interest is underlined):

b<u>a</u>NAAN	(bananas)
mi<u>N</u>UUT	(minute)
choco<u>L</u>Ade	(chocolate)

Note that the stress pattern is different in the English translations of the Dutch words. According to the linguistic view, speakers have the intention to produce either a full

vowel or a schwa in these Dutch words. However, the acoustic realization of these vowels can show a considerable spread in spectral quality. Can listeners nevertheless unambiguously categorize the vowel as either a full vowel or a schwa, despite this variability in the acoustic signal?
Apart from this question, we were also interested in the relation between the frequency of occurrence of words and the perception of vowel reduction. Therefore, both words with a high frequency of occurrence and words with a low frequency of occurrence were included in the experimental design. The (high frequency) word examples given above were actually selected, because they have counterparts with a similar structure and a low frequency of occurrence, namely:

baNIER	(banner)
miNIEM	(insignificant)
akkoLAde	(brace)

In these words it is rather unlikely (although not impossible) that the underlined vowel is pronounced as a schwa. A complete list of the 20 word pairs that were used to test the role of frequency of occurrence of words on the perception of vowel reduction is given in Table 1. The vowel of interest has been underlined. Word frequencies in Table 1 were adopted from the Dutch database CELEX [4].

Table 1. Word pairs that were used in the present experiment to test the influence of the frequency of occurrence of words on the perception of vowel reduction. Word stress is indicated with capitals; the vowel of interest is underlined. From the similarly structured word pairs, Word 2 always has a lower frequency of occurrence.

Vowel	Word 1	freq	Word 2	freq
ɔ	microsCOOP	155	horosCOOP	119
o:	koloNEL	2787	koloNIST	287
o:	absoLUUT	3068	resoLUUT	339
o:	ecoNOmisch	5792	lakoNIEke	173
o:	chocoLAde	206	akkoLAde	17
a:	paTAT	82	paTENT	87
a:	baNAAN	260	baNIER	56
a:	acaDEmie	472	dekaDENtie	182
a:	etaLAge	489	fataLISme	53
a:	paraCHUTE	229	paraSIET	136
a:	vaKANtie	2319	vaKANte	37
ɛ	perSOON	8243	perVERS	187
ɛ	resPECT	1022	resPIJT	53
ɛ	terREIN	4259	terREUR	334
ɛ	anekDOte	269	anneXAtie	28
e:	meTAAL	833	meTHAAN	38
e:	deFENsie	445	deFLAtie	4
e:	redeNEren	655	reduCEren	466
i	miNUUT	7359	miNIEM	166
i	reliKWIE	87	peliKAAN	51

Another point of interest in the present study was the relation between speech styles and the perception of vowel reduction. Therefore, the selected words were pronounced in three different manners, namely "read from a word list" (W), "uttered as a word through the presentation of pictures" (P), and "read from a list of sentences" (S). The condition P was added to see how much the pronunciation of a word changes if it is not presented in its written form. Since only some of the words in Table 1 were suitable to be presented as pictures, some extra words were selected for condition P. The 20 words that were presented as pictures (6 words from Table 1 and 14 new words) are given in Table 2. These words also occurred in the conditions W and S, and could thus serve to test the influence of speech styles on the perception of vowel reduction.

Table 2. Words that were presented to speakers as pictures in the speech style condition P. Word stress is indicated with capitals; the vowel of interest is underlined.

Vowel	Word (Picture)	freq
u	KANgoeroe	42
ɔ	microsCOOP	155
ɔ	biosCOOP	701
o:	chocoLAde	206
o:	saxoFOON	21
o:	lokomoTIEF	302
o:	microFOON	419
a:	paTAT	82
a:	baNAAN	260
a:	paraCHUTE	229
a:	SInaasappel	352
a:	sigaRET	3135
a:	KAtapult	47
a:	CARnaval	133
a:	Ananas	97
ɪ	miLJOEN	2750
e:	meDAILLe	340
i	miNUUT	7359
i	biKIni	169
i	aspiRIne	127

In total, 20 male students were used as speakers. The order of words in condition W, and of test sentences in condition S, as well as the pictures in condition P was at random. For each speaker this random order was different. The order of the conditions W, S, and P was systematically varied for the 20 speakers. All recordings were made on the audio channel of a Panasonic video recorder in an anechoic room.
The speech recordings were low pass filtered at 4.5 kHz and digitized at a sample rate of 10 kHz with 12-bit precision. From this digitized speech entire words were segmented with the aid of a speech editing program, and each word was stored separately. Word boundaries were established by listening to selected parts of the speech signal and by visual inspection of the speech waveform.

2.2 Listening experiment

The segmented words were presented to 20 listeners in a blocked design. Most of the listeners were members of the staff and students of the Institute of Phonetic Sciences in Amsterdam; all listeners had at least a basic knowledge of linguistics and all were familiar with the schwa concept. In order to level out annoying loudness differences between the stimuli, they were scaled to a fixed maximum overall amplitude value. Artificial clicks were prevented by smoothing the word boundaries with half a Hanning window of 5 ms. All 54 test words (40 from Table 1 and 14 new ones from Table 2) occurred in the conditions W and S; in the condition P only 20 words occurred (all words from Table 2). In total, 2560 stimuli were presented to each of the 20

listeners (2 × 54 words × 20 speakers + 1 × 20 words × 20 speakers). Each listener attended 4 listening sessions (4 × 5 speakers); one listening session took about 45 minutes. The test words from all conditions were placed in a random order for each of the 20 speakers separately. The order of speakers was also randomized. For each subject in the listening experiment the random ordering of stimuli and speakers was different. In order to make the subjects familiar with the voice of each speaker, the 5 stimuli at the end of a speaker block were added at the beginning of the series.

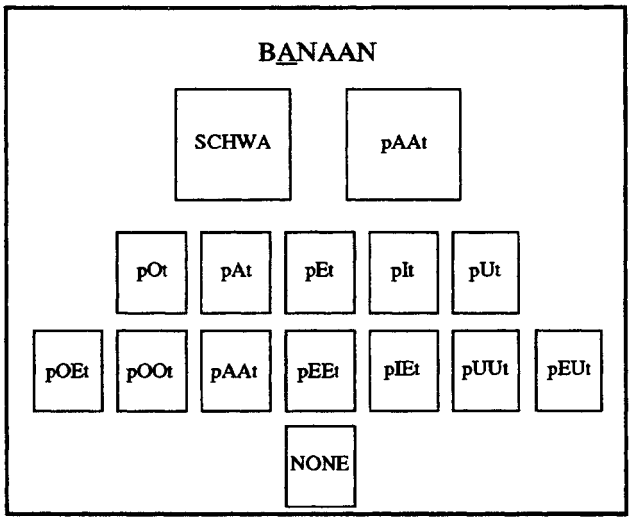


Figure 2. The lay-out of the terminal screen in the listening experiment

The listening test was done on line by each subject separately at the terminal of a VAX workstation. On the terminal screen a test word appeared in which the vowel of interest had been underlined. All response possibilities were also shown on the screen as exemplified in Figure 2. Response possibilities were all 12 Dutch monophthongs /ɔ, ɑ, ɛ, ɪ, œ, ʉ, o:, a:, e:, i, y, ø/, the schwa, and the category "vowel absent". The 12 monophthongs were shown in their orthographic form (which is unambiguous in Dutch), embedded in a p-t dummy context to clearly indicate which sound they represented. In order to speed up the listening test, the schwa and the phonologically correct vowel (i.e. the /a:/ in the example of Figure 2) as most probable response categories, were shown in somewhat larger blocks in the middle of the screen. It was explained to the subjects that this was merely done to speed up the test, and that they were free to choose any category they liked.
Subjects heard the stimulus word 0.25 sec. after its visual presentation. A second aural presentation of the stimulus word occurred 0.25 sec. after the first one. Subjects were instructed to carefully listen to the vowel of interest and to respond by clicking with the mouse on one of the response blocks. As soon as the subject had responded, the next word was presented to him. This design allowed listeners to establish their own pace and it prevented listeners from skipping responses.
In most related experiments on vowel recognition either synthetic vowels or natural vowels that were segmented from their context are presented to listeners. In the present experiment vowels were presented to listeners in their natural word context. This had the clear advantage that listeners could use all available information to properly identify the vowel, such as dynamic information about coarticulation, stress pattern, tempo, etc.

3 RESULTS

3.1 Agreement among listeners

Most of the listeners responded with either the schwa or the phonologically correct form of the vowel. In several words it was not clear whether the phonologically correct form would be a long vowel or a short vowel. In the word "banaan", for instance, some people think that the phonologically correct form of the underlined vowel is the long vowel /a:/, whereas others think it is the short vowel /ɑ/. In the listeners' responses both long vowels and short vowels occurred in these instances. If we disregard the long–short vowel 'confusions' for the time being, more than 96% of all vowels were identified as either "schwa" or "phonologically correct". The vowel was judged to be absent in less than 1% of all cases, and the vowel was judged to belong to one of the remaining response categories in about 3% of all cases.
The original responses were recoded into one of two categories: "schwa" or "full vowel". All responses other than "schwa" and "vowel absent" were recoded into the category "full vowel". The assignment of the "vowel absent" responses was less clear. They could have been regarded as 'missing values', but we found it more appropriate to assign the "vowel absent" responses to the "schwa" category, because the omission of a vowel can very well be interpreted as an extreme manifestation of vowel reduction.
As an indication of agreement among listeners, the intra-class correlation coefficient kappa [5, 6] was calculated for each speech style condition. The statistic kappa is a normalized measure of overall agreement between raters, corrected for the amount of agreement expected by chance alone. The calculated κ's and their standard deviations are given in Table 3. The value $\kappa/sd(\kappa)$ is approximately distributed as a standard normal variate. This value is given in the last column of Table 3.

Table 3. The values of the statistic κ for each speech style condition, indicating the amount of agreement among listeners on the assignment of vowels to the categories "full vowel" or "schwa".

Condition	κ	$sd(\kappa)$	$\kappa/sd(\kappa)$
W	0.52	0.0042	124.3
S	0.50	0.0017	298.4
P	0.49	0.0035	140.7

The values of κ for each speech style condition are comparable. Although all κ's are highly significant ($p < 0.001$), their value indicates only a very moderate amount of agreement among the listeners [7] in assigning a vowel to the category "schwa" or the category "full vowel".

3.2 Frequency of occurrence of words

In Table 4 the percentages of schwa responses for the word pairs from Table 1 are given for each of the speech style

conditions W and S (not all of these words occurred in the condition P) The differences in number of schwa responses between words with a high frequency of occurrence and words with a low frequency of occurrence are considerable.

Table 4. Percentages of schwa responses for the word pairs from Table 1 in the speech style conditions W and S.

Word 1	W	S	Word 2	W	S
microsCOOP	48	60	horosCOOP	39	59
koloNEL	13	27	koloNIST	5	25
absoLUUT	19	55	resoLUUT	4	33
ecoNOmisch	46	72	lakoNIEke	1	6
chocoLAde	40	63	akkoLAde	2	18
paTAT	24	75	paTENT	1	19
baNAAN	19	55	baNIER	1	1
acaDEmie	42	68	dekaDENtie	15	29
etaLAge	34	63	fataLISme	6	10
paraCHUTE	33	52	paraSIET	12	25
vaKANtie	32	80	vaKANte	10	24
perSOON	14	71	perVERS	4	10
resPECT	51	77	resPIJT	33	34
terREIN	39	64	terREUR	19	20
anekDOte	24	32	anneXAtie	18	37
meTAAL	0	6	meTHAAN	0	1
deFENsie	36	46	deFLAtie	0	4
redeNEren	3	11	reduCEren	0	0
miNUUT	2	70	miNIEM	1	10
reliKWIE	28	33	peliKAAN	0	9
Average	27	54	Average	9	19

An analysis of variance with repeated measures was done on the proportion of schwa responses per word item with the trial factor "frequency" and the grouping factor "speech style". Since the data consisted of proportions which are not normally distributed, an inverse sine transformation was applied [8]. The factor "frequency" turned out to be significant ($F = 514.9, p < 0.001$), and the factor "speech style" as well ($F = 133.2, p < 0.001$).

3.3 Speech style

A comparison between the conditions W and S for the word pairs from Table 1 was already made in the former section. For a fair comparison of all three speech style conditions, the words from Table 2 were used, because only these words occurred in all conditions. The average percentages of schwa responses in the conditions W, P, and S were 33%, 39%, and 60%, respectively.

An analysis of variance with repeated measures was done on the proportions of schwa responses for each word item. On these proportions an inverse sine transformation was applied. The trial factor "speech style" turned out to be significant ($F = 130.5, p < 0.001$). Tests for pairwise contrasts revealed that all three conditions differed significantly from each other ($p < 0.005$). The condition P thus evokes a somewhat more 'spontaneous' pronunciation of words (with more vowel reduction) than the condition W. The largest amount of vowel reduction occurs when words are pronounced in fluent speech, i.e. in condition S.

4 DISCUSSION AND CONCLUSIONS

The statistic kappa indicated only a moderate agreement among listeners on the assignment of vowels to the categories "full vowel" or "schwa". We also looked at the distribution of the number of listeners who responded with a schwa. In the case of a perfect agreement on the category of a vowel, either *none* of the listeners should have responded with a schwa (perfect agreement for the "full vowel" category), or *all* listeners should have responded with a schwa (perfect agreement for the "schwa" category). However, there was a considerable area under the middle part of the distribution, indicating that the listeners often disagreed to some extent about the vowel category. We also looked at the total percentage of schwa responses for each listener separately. These percentages ranged from 18% to 49%, showing a large difference in the number of schwas that each subject heard (or thought he heard). Thus, it seems justified to conclude that in many cases listeners are unable to properly distinguish between acoustic vowel reduction and lexical vowel reduction.

Another conclusion that comes forward from the experimental results is that both the frequency of occurrence of words and the speech style play a very important role in the vowel reduction process. The number of schwa responses increased dramatically for the unstressed vowels in words with a high frequency of occurrence compared to similar words with a low frequency of occurrence. A fluent speech style (read sentences) evoked far more schwa responses than uttering words through the presentation of pictures or reading a word list.

In future research we want to analyse the results from the listening experiment in more detail, and we also plan to relate the perception data to acoustic measurements on the vowels.

ACKNOWLEDGMENTS

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.

REFERENCES

[1] Lindblom, B.E.F. (1963). 'Spectographic study of vowel reduction', J. Acoust. Soc. Am. 35, 1773-1781.

[2] Koopmans-van Beinum, F.J. (1980). Vowel contrast reduction: An acoustic and perceptual study of Dutch in various speech conditions, Doctoral Dissertation, University of Amsterdam.

[3] Bergem, D.R. van (1993). 'Acoustic vowel reduction as a function of sentence accent, word stress, and word class', Speech Communication 12, 1-23.

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

[5] Cohen, J. (1960). 'A coefficient of agreement for nominal scales', Educational and Psychological Measurement 20, 37-46.

[6] Fleiss, J.L. (1971). 'Measuring nominal scale agreement among many raters', Psychological Bulletin 76, 378-382.

[7] Landis, J.R. & Koch, G.G. (1977). 'The measurement of observer agreement for categorical data', Biometrics 33, 159-174.

[8] Kirk, R.E. (1981). Experimental Design: Procedures for the Behavioral Sciences (Wadsworth, Belmont, CA).