

WHAT'S IN A SCHWA ?

Florien J. Koopmans-van Beinum

Abstract

Although the schwa sound in Dutch is by far the most frequently occurring vowel, it is up to now phonetically the most neglected vowel as well. We used an existing database of schwa sounds from focus words in spontaneous speech and in lexically the same text read aloud, by one male speaker to investigate durational and spectral characteristics of these schwa sounds and compared the results with data on schwa diphones used in Dutch text-to-speech synthesis. It turned out that, unlike what usually is thought, lexical schwa sounds in natural continuous speech are considerably shorter than the other short vowels, that no strong consonantal influence exists on schwa duration, that schwa sounds display a spectral spread larger than any other vowel and that surrounding consonants seem to play a role with respect to the midpoint formant distribution of the schwa within the whole vowel system. In no way the schwa can be considered as the 'bench mark' of a speaker's vowel system.

1. INTRODUCTION

Ten years ago, a Dutch journal for linguistics "Spektator", devoted a special issue to 'the schwa in the Dutch language'. In the editorial the guest editors (van Marle and Zonneveld, 1982) indicated that, concerning the schwa phenomenon, at least four main topics of research could be distinguished:

1. the schwa as optionally inserted vowel;
2. the schwa as reduction vowel;
3. the schwa as morphological phenomenon;
4. the word- or stem-final schwa.

Our own contribution to that special issue concerned topic 2, on vowel contrast reduction in Dutch (Koopmans-van Beinum, 1982), and was concluded then with the remark that the schwa, being the most 'worn out' Dutch vowel, could be considered phonetically as the 'bench mark' in the articulation of each speaker. When averaging over many vowel realizations in various speech styles, it is true that a clear centralization of all vowels within the vowel system is quite obvious. However, when carefully inspecting individual vowel realizations, the picture is much more complicated, and coarticulation effects turn out to play an important role (van Bergem, in press). In the present paper this phenomenon of acoustic-phonetic vowel contrast reduction, resulting in vowel realizations that more or less approach the schwa area, is left out of consideration now. Also topic 1 has been left out here, concerning the svarabhakti-vowel, in Dutch often inserted in final consonant clusters of which the first consonant is a liquid (e.g. 'melk' > /mɛlək/).

Whereas the schwa of topic 1 and 2 may be considered as produced optionally and the schwa of topic 4 is sometimes optional and sometimes obligatory, the schwa of topic 3 in Dutch, and in most other Germanic languages, is an obligatory phonemic schwa sound, also called a lexical schwa (Kager, 1989). It occurs only in unstressed position, for the greater part in Dutch suffixes with grammatical functions like constituting plural noun forms (e.g. 'boek'-'boeken', Eng. 'book'-'books'), diminutive forms (e.g. 'boek'-'boekje', 'book'-'booklet'), participle forms (e.g. 'loop'-'lopend'-'gelopen', 'walk'-'walking'-'walked'), and in articles (e.g. 'de', 'een', 'the', 'an'). Historically it might stem from either a svarabhakti-vowel, or an acoustic-phonetically reduced vowel, but in the present study we shall use the criterion that in the contemporary lexicon the specific vowel sound is indicated as an obligatory schwa (henceforth called 'lexical schwa').

In the present paper we will mainly concentrate on the acoustic properties of the phonemic or lexical schwa as mentioned above in the topics 3 and 4. We shall try to answer the question, whether there is any acoustic ground to consider this vowel indeed as the 'bench mark' in the articulation of each speaker. Or is it, as in English, to be seen as an 'indeterminate vowel' (Clark & Yallop, 1990), and what exactly does that mean? Henton (1990) mentions: "*The acoustic nature of /ə/ is known to be very variable depending on non-final or final position within a word; its articulation may range from half-close central to the most open central area.*" (p. 211). Does a schwa undergo any coarticulatory influence or is it a stable, fixed point in the vowel system? And is the schwa, as it occurs in read speech, comparable to the schwa in the same consonantal context in spontaneous speech, does it behave the same in both speech styles?

2. SPEECH MATERIAL

The schwa in Dutch is an important vowel, at least as far as its frequency of occurrence is concerned. Although this lexical schwa occurs only in unstressed position (cf. Fromkin et al., 1986), it takes about 30 % of all vowel phoneme realizations in Dutch (van den Broecke, 1988). Therefore it is quite surprising that little is known about the acoustic characteristics of this speech sound in normal continuous speech. That ignorance becomes the more obvious when we carefully listen to Dutch synthetic speech. At least in Dutch diphone speech synthesis, the schwa is unnaturally manifest. This unnaturalness is the more emphasized because of its high frequency of occurrence.

Since we have at our disposal the acoustic data of the Dutch diphones as used in the system "Spraakmaker" (van Leeuwen & te Lindert, 1991), spoken by a professional male speaker, and since furthermore we possess a database of spontaneous and read-aloud speech of the same speaker, it is challenging to use these data for further inspection with respect to the behaviour of the Dutch phoneme schwa.

The database mentioned above includes 'spontaneous', i.e. free conversational speech, recorded in laboratory situation, and the same text material read out afterwards in an identical recording session, so that two parallel versions of the same lexical material were created. In this way the acoustic data for the material of the two speaking styles was completely comparable. This material has been used in a number of recent studies on focus words and the present paper is merely a result of the measurements done within that work (see Koopmans-van Beinum, 1992). From a passage of about 5 minutes of spontaneous speech and the concurrent read speech, all focus words had been selected, guaranteeing a pronunciation not too sloppy. Within these words all occurring vowels (stressed *and* unstressed vowels, *and* schwas) were segmented and measured for their duration and formant frequencies, resulting in data concerning 637 vowels including 179 lexical schwa vowels, occurring in both speech styles. In that

first stage the set of lexical schwa vowels had been left out of processing, but in the present paper it is exactly this set that we will discuss in the next pages.

Apart from our own acoustic data mentioned above, Drullman and Collier (1991) provided data about average formant frequencies of all vowels in the diphone sets ('full' and 'reduced'), and apart from that they made available the more detailed spectral and durational data of their investigation. In this way we could compare the schwa-data as occurring in that diphone set, with our own data from natural spontaneous and read speech, in order to understand the obvious unnaturalness of the lexical schwa vowels in our synthetic diphone speech.

3. ACOUSTIC MEASUREMENTS AND RESULTS

The speech material of both speech styles was lowpass filtered at 4.5 kHz and digitally stored in the computer with a 10 kHz sample frequency and a 12 bit precision. All vowels in the focus words of both speech styles were carefully segmented from their consonantal surroundings on the basis of the digitized speech waveform using visual and auditory feedback, in such a way that no consonantal transitions were included in the vowel part as far as possible. The segmented vowels were analysed with a 10-pole LPC analysis, and a 25 ms Hamming window shifted over the vowel segment in steps of 1 ms.

Since we want to compare the formant frequencies of the schwa vowels in our natural speech material with those of schwas in synthesized diphone speech and since these diphone values were given for about the midpoint of the vowel realizations, we will use here also the midpoint values of the vowel realizations in our further data processing. Therefore it might be assumed that at that point in the vowel segment consonantal effects would be minimum, and formant values might be as stable and homogeneous as possible. If desirable, dynamic analysis will be done at a later stage.

With respect to the results we shall first concentrate on schwa durations, pairwise comparing them in both speech styles. Next we shall discuss the distributions of the formant frequency values.

3.1. Schwa durations

In Table 1 a survey is given of schwa vowel durations in both speech styles. For reasons of comparison we included data concerning the other vowels in the focus words as well. A striking point is, that the mean durations of schwa in spontaneous as well as in read-aloud speech are considerably shorter than the mean durations of short vowels. However, in this Dutch diphone speech synthesis system schwa is considered to behave as the other short vowels with a corresponding duration. Evidently this will result in far too long vowel durations for schwa.

Table 1. Mean durations and standard deviations of schwa sounds, and of short and long vowels, in focus words from spontaneous speech and from the same text read aloud.

	spontaneous			read aloud		
vowels	schwa	short	long	schwa	short	long
mean duration	47	67	95	43	67	103
s.d. duration	17	22	31	12	18	34

In our previous study comparing focus words in spontaneous and read-aloud speech (Koopmans-van Beinum, 1992) we found, concerning the durations of the other vowels, a significant correlation ($r = .76$) between both speech styles, indicating a consistent consonantal coarticulation effect.

Fig. 1 displays the correlation between the durations of the 179 lexical schwas as they occur in the focus words in the two speech styles. It will be clear from this figure that the durational correlation of schwas in lexically identical word pairs in read and spontaneous speech is low ($r = .40$). This result concerning schwa durations agrees very well with data in van Son and Pols (1992), who compared vowel durations of the same speaker, when reading aloud at normal and at fast rate. Regarding schwa durations they did not find any correlation between their two speech conditions either.

Further inspection of our data in Fig. 1 reveals that those four schwas displaying very long durations in spontaneous speech, but normal durations in read-aloud speech, all occur in the final syllable of the last word within a tone unit. This points to the use of 'hesitation lengthening', characteristic of spontaneous speech. When leaving out these four data points from our processing, the durational correlation of schwas in the two speech styles increases only slightly ($r = .46$).

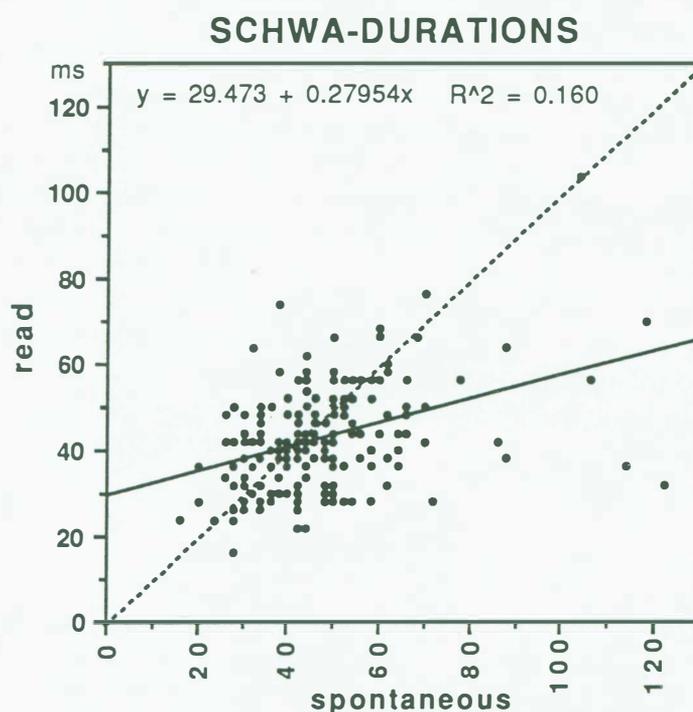


Fig. 1. Correlation between vowel duration of lexical schwas in identical consonantal context, from focus words in spontaneous and read speech, with regression line and diagonal.

So it can be concluded that the schwas in read speech turn out to be not systematically longer than in spontaneous speech, as little as this was the case in the other vowels. However, whereas the other vowels display a clear coarticulatory influence, for schwa this seems not to play a role, since we found only a slight correlation in spite of the identical consonantal contexts for each vowel in the two speech styles. In order to inspect this into more detail, we selected all schwas that were followed by an /r/, since it is generally known that this consonant causes the largest durational effect on preceding vowel sounds. However, mean durational values for this (-/ər/-) subset turned out to be 49 ms and 46 ms for spontaneous and read-aloud speech respectively.

These values do not deviate significantly from the overall mean durational values as given in Table 1. Although this is only a first step in the analysis of the durational distribution of schwas in natural continuous speech, our conclusion so far is that lexical schwas behave randomly with respect to their duration.

3.2. Schwa formant frequencies

Table 2 gives a survey of the values of first and second formant frequencies in both speech styles. For reasons of comparison we include here data concerning centroid values of the other vowels in the used focus words as well, calculated as the average of the midpoint values in Hz of all other vowels in the data set. Moreover, since it is our intention to understand why schwas in synthetic speech sound so unnatural and 'overarticulated', we also present here data concerning schwa sounds from the syllables that have been used to construct the schwa diphones (pronounced in the first syllable in isolated nonsense words of the type $C\text{ə}CVC\text{ə}$ with the accent on the second syllable, and C being one of the 19 phonotactically possible Dutch consonants), and data on the centroid of the diphone vowel system. From Table 2 it will be clear that in no case the centroid of the vowel system coincides with the location of the schwa within the vowel plane. Moreover the standard deviations of the schwa sounds in continuous speech are considerably larger than those of the diphone schwa vowels.

Table 2. Mean formant frequency values ($F1$ and $F2$) and standard deviations in Hz of schwa sounds, and of centroid values of the other vowels, in focus words from spontaneous speech and from the same text read aloud, together with values as used in Dutch diphone synthesis, based on speech material of the same speaker.

	spontaneous		read aloud		diphones	
	schwa ($n=179$)	centroid	schwa ($n=179$)	centroid	schwa ($n=19$)	centroid
$F1$	335	383	375	434	400	462
s.d. $F1$	60		61		28	
$F2$	1285	1375	1338	1439	1447	1456
s.d. $F2$	187		202		109	

In Fig. 2 and Fig. 3 formant frequencies ($F1$ and $F2$) are given of all lexical schwa sounds in focus words from spontaneous and from read speech respectively. For the sake of completeness we displayed also the mean formant frequencies of the other vowels in the focus words in the same speech style, all spoken by the same speaker. Surprising as that might be, mean vowel formant frequencies of reduced or unstressed diphones differ from the full diphone vowels only slightly (Drullman and Collier, 1991) and therefore are left out here.

In Fig. 4 we displayed the formant values of the schwa diphones, together with the formant values of the schwa sounds in the read aloud text. It will be clear from Fig. 2 to Fig. 4 that in natural continuous speech the spectral properties of schwa are very diffuse. Instead of being a 'bench mark' within the speaker's vowel system, in natural continuous speech this speech sound seems to be the most unstable vowel of all, occupying almost the whole vowel plane! The diphone schwas, however, seem to form a much more homogeneous group (apart from one token, see Fig. 4).

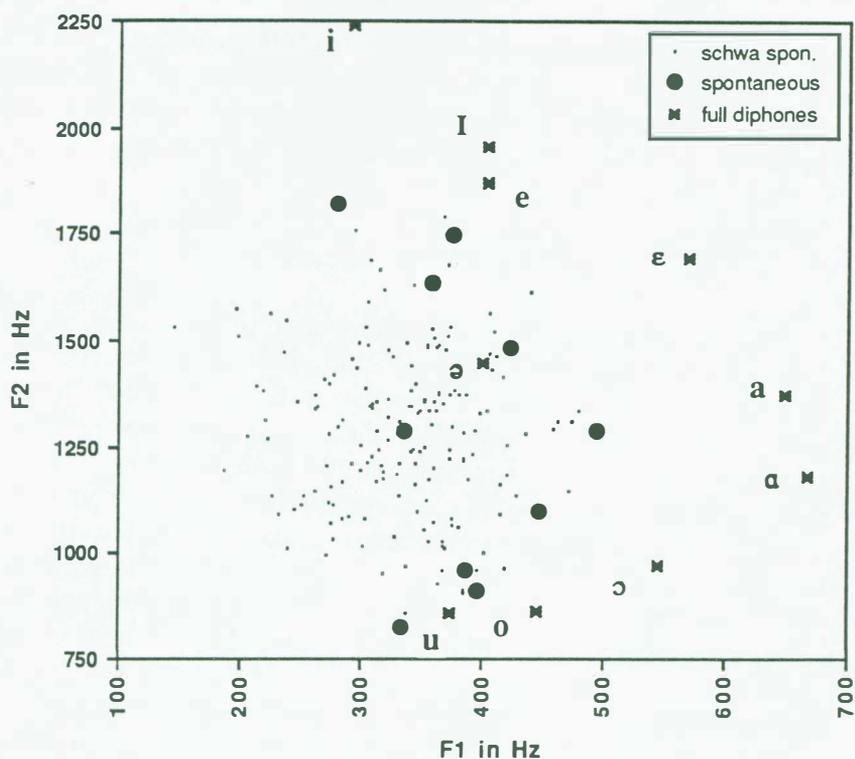


Fig. 2. Formant frequencies ($F1$ and $F2$) of all lexical schwas in focus words from **spontaneous** speech, together with mean values of all other vowels in the focus words in the same speech style, and of vowels in full diphones, all spoken by the same speaker.

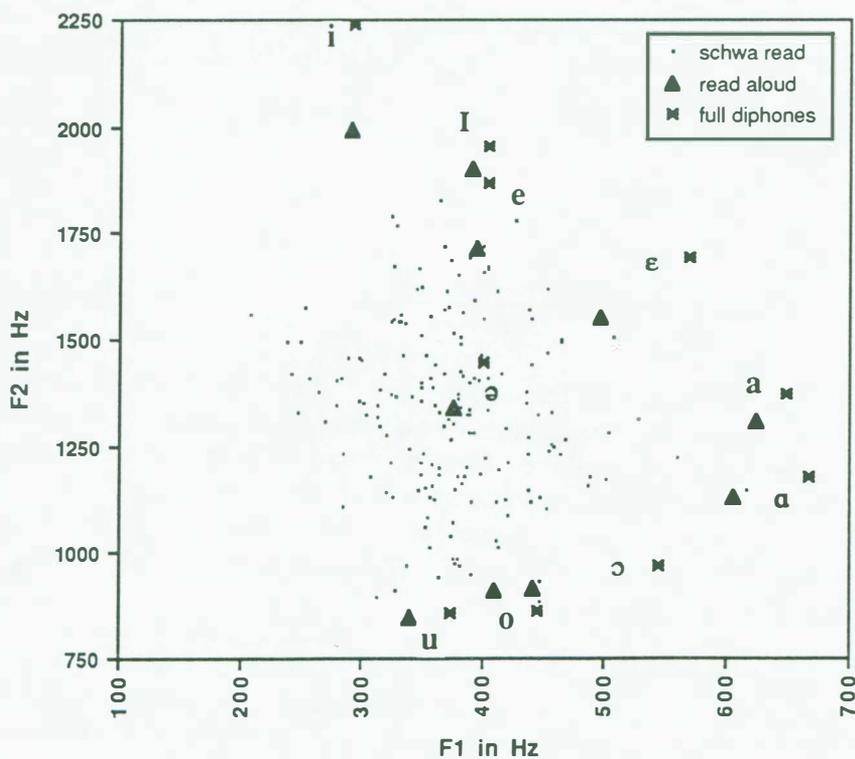


Fig. 3. Formant frequencies ($F1$ and $F2$) of all lexical schwas in focus words from **read-aloud** speech, together with mean values of all other vowels in the focus words in the same speech style, and of vowels in full diphones, all spoken by the same speaker.

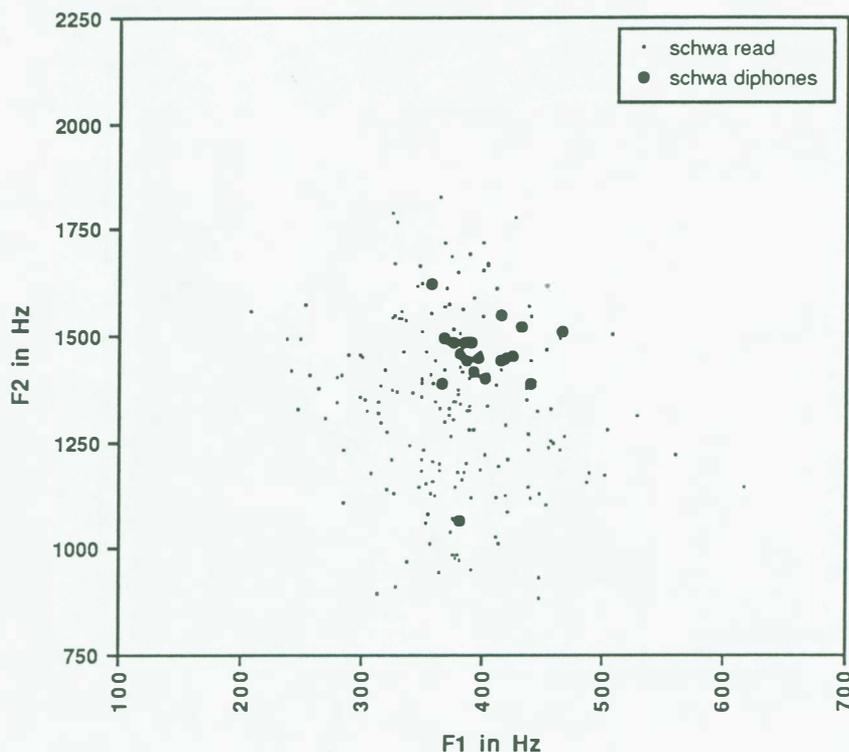


Fig. 4. Formant frequencies ($F1$ and $F2$) of all lexical schwas in focus words from read speech, together with formant values of schwa diphones, all spoken by the same speaker.

Our next questions to be answered concern the spread of the schwa formant values. Is this spread comparable to the spread of the other vowels? And is it completely at random or is it influenced by coarticulatory effects?

To answer the first question we calculated standard deviations of $F1$ (Fig. 5) and of $F2$ (Fig. 6) in both speech styles, after transforming the individual formant frequencies in Hz to a mel scale, using the formula

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

where f is the formant frequency in Hz and m the mel-transformed value (Makhoul & Cosell, 1976). In this way a more accurate comparison of the spread within the whole vowel system can be made. Vowels occurring less than $n=20$ within the data set, have been left out of the graph. The vowels are ordered according to ascending values for spontaneous speech. From Fig. 5 it will be clear that the standard deviations for $F1$ of the schwa occupy an average position in relation to the other vowels. On the whole the values increase according to the greater degree of openness of the mouth. There is no reason to expect that schwa will behave differently from the other vowels. Maximum adaptation will mean a medial position with respect to the openness of the mouth. As for $F2$, however, Fig. 6 shows that the standard deviations of all vowels are smaller than the standard deviation of schwa. This can be explained by the fact that a differentiation between front and back vowels is caused by movements of the tongue body to the front or to the back part of the vocal tract. Maximum adaptation to the other vowels will mean that for schwa the tongue body has to move permanently or has to be very inert and stable in a neutral position. Since in natural, continuous speech the latter condition is quite unlikely, only a very mobile behaviour with a maximum spread can be the result.

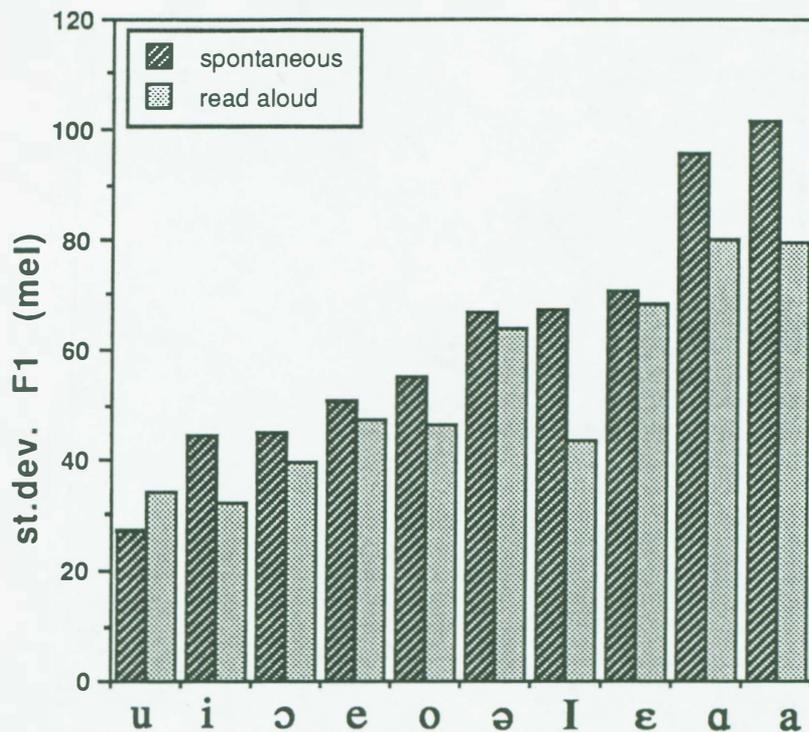


Fig. 5. Standard deviations of $F1$ (mel) for all vowels ($n \geq 20$) and for the lexical schwas in focus words from spontaneous and read-aloud speech, represented in ascending order.

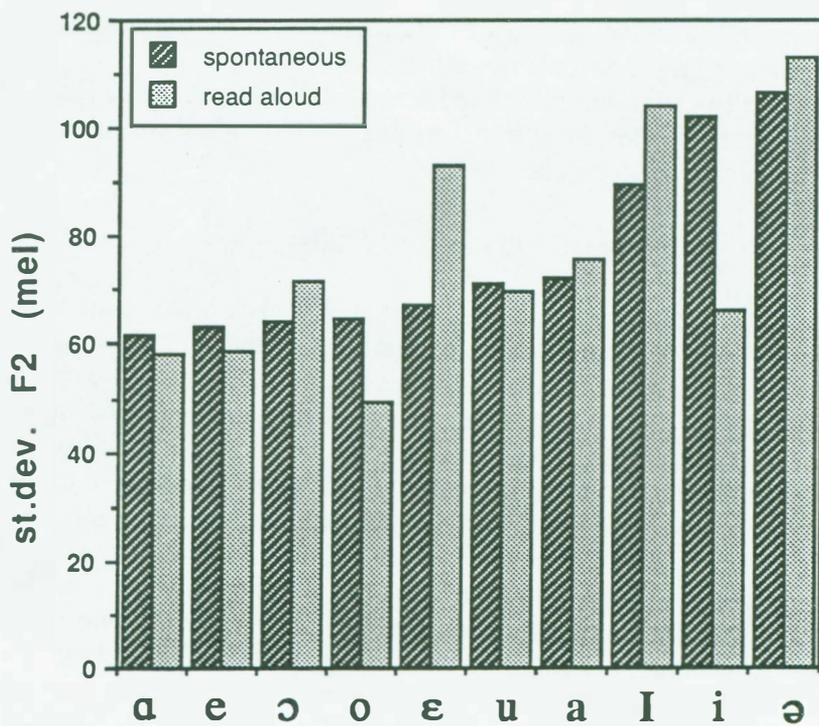


Fig. 6. Standard deviations of $F2$ (mel) for all vowels ($n \geq 20$) and for the lexical schwas in focus words from spontaneous and read-aloud speech, represented in ascending order.

To investigate our second question with respect to the spread of the schwa (at random or influenced by consonantal context), we correlated formant values of $F1$ as well as of $F2$ (both in Hz) of all schwa sounds in spontaneous speech with the values of the schwa sounds in the concurrent consonantal context of the same text read aloud (see Fig. 7 and Fig. 8).

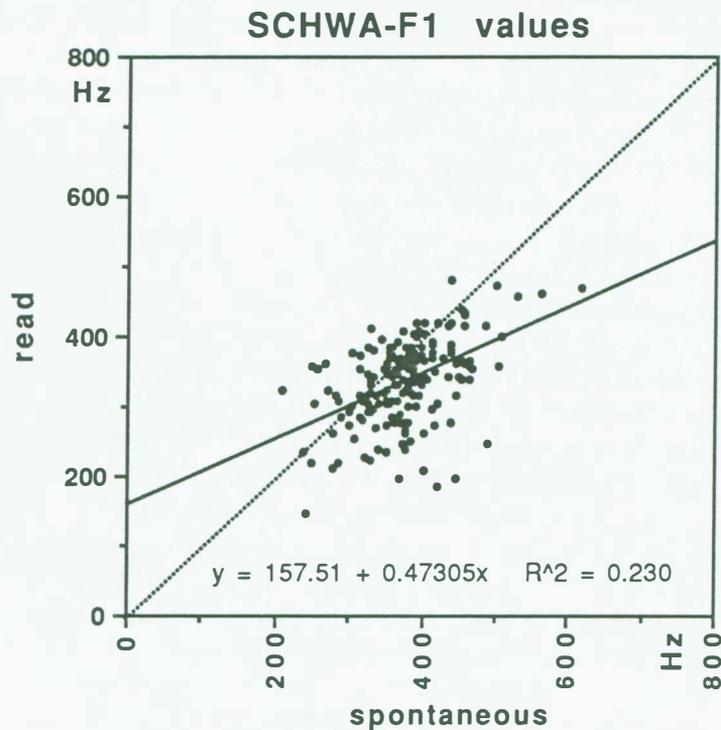


Fig. 7. Correlation between $F1$ (HZ) of lexical schwas in identical consonantal context, from focus words in spontaneous and read speech, with regression line and diagonal.

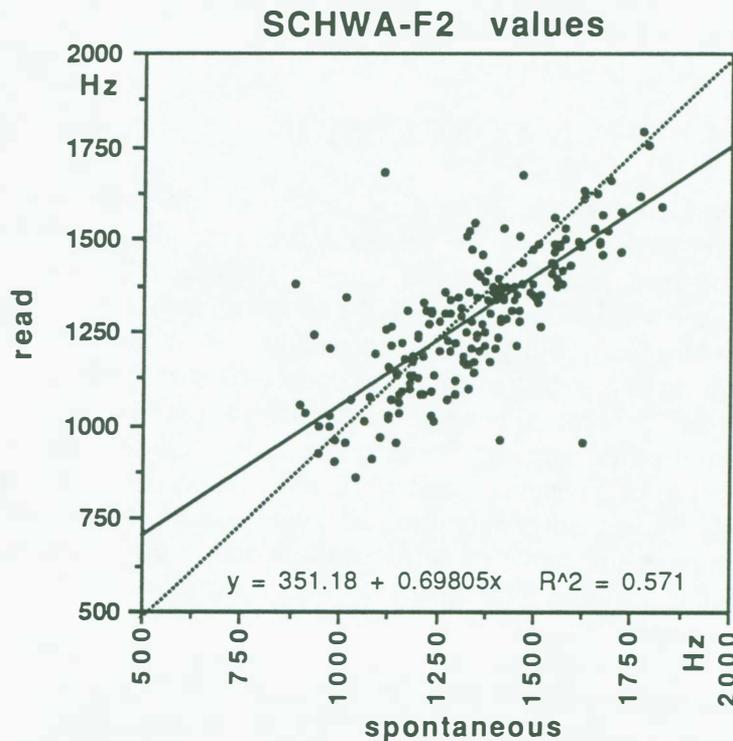


Fig. 8. Correlation between $F2$ (Hz) of lexical schwas in identical consonantal context, from focus words in spontaneous and read speech, with regression line and diagonal.

It can be concluded that with respect to $F1$ values the correlation of $r = .48$ is only moderate, explaining 23% of the variance, but that for $F2$ values the correlation of $r = .76$ is highly significant, explaining 57% of the variance. This is reasonably well in agreement with our above mentioned explanation concerning the standard deviations. (Transformation to a mel scale reduces the correlations slightly to .47 and .75). With respect to $F1$ and $F2$ of all other vowels together, actually representing the spread of the total vowel system, the correlation coefficients between formant values of spontaneous and read-aloud speech were .74 and .96 for $F1$ and $F2$ respectively. For each vowel individually the correlation coefficients are comparable to those of schwa, i.e. moderate for $F1$ and high for $F2$. These results are in agreement with those of van Son & Pols (1992), although their fast rate condition caused a shift towards a more open pronunciation for this specific speaker, the same one in both studies.

Our next question to be answered is, whether there is any systematicity within the spread of the formant frequencies of these schwa sounds. In other words, will it be possible to indicate any consonantal coarticulatory tendencies within the distribution as represented in Fig. 2 (spontaneous) and Fig. 3 (read aloud)?

From the database of all schwa realizations we therefore selected some subgroups based on identical preceding or following consonant. Fig. 9 displays a number of plots of these subgroups (all consisting of more than 20 schwa realizations), together with the whole schwa distribution in that specific speech style. Although the consonantal effects are not quite striking, there are some tendencies that, at least partly, coarticulatory influences are cause of the great variation in schwa realizations. E.g. initial velar consonants / χ / and / γ / have a lowering effect both on $F1$ and $F2$, easily to be explained by the closure in the back part of the vocal tract. Final / r / causes a lowering of $F2$, possibly to be explained by the fact that our speaker often uses a velar / r /, so also in the back part of the vocal tract. Final / n / causes the lowest $F1$ values, but does not display any systematicity apart from that. An interesting subgroup is the group with final schwa. It is obvious that $F2$ values show a clear increasing tendency. This results in positions approximately coinciding with those of the diphone schwa sounds. The cause of this deviant behaviour might be a lexical or morphological one, since these items can be ranged under topic 4, representing the schwa in word- or stem-final position. Clearly more detailed investigations are needed here.

4. CONCLUSION

This study has to be considered as a first step to obtain more insight into the behaviour of a speaker with respect to the realizations of the most frequently occurring vocalic speech sound in natural, continuous-speech situations. To start with, as for formant frequencies we confined ourselves to the more or less stable midpoints of the schwa. Even there a large variability in formant positions was found, for $F2$ partly to be explained by fronting and backing influences caused by surrounding consonants. A systematic acoustic and perceptual analysis of the schwa, mainly with respect to the influences of surrounding consonants on its realization, including dynamic investigations as well, seems to be a necessary and worthwhile enterprise. These aspects have been studied recently by van Bergem (submitted), who indeed found clear systematic effects. Investigating the influence of formant track shape on the identification of synthetic schwa sounds, as done for the other vowels by van Son & Pols (submitted), will be worthwhile as well.

So far we do not know yet whether the variable character of the schwa, if applied in synthetic speech, will influence intelligibility. In the meantime it has become a little bit more clear why schwa vowels in Dutch synthetic (diphone) speech sound so unnatural and overarticulated: in no way justice has been done to its shortlived, volatile character.

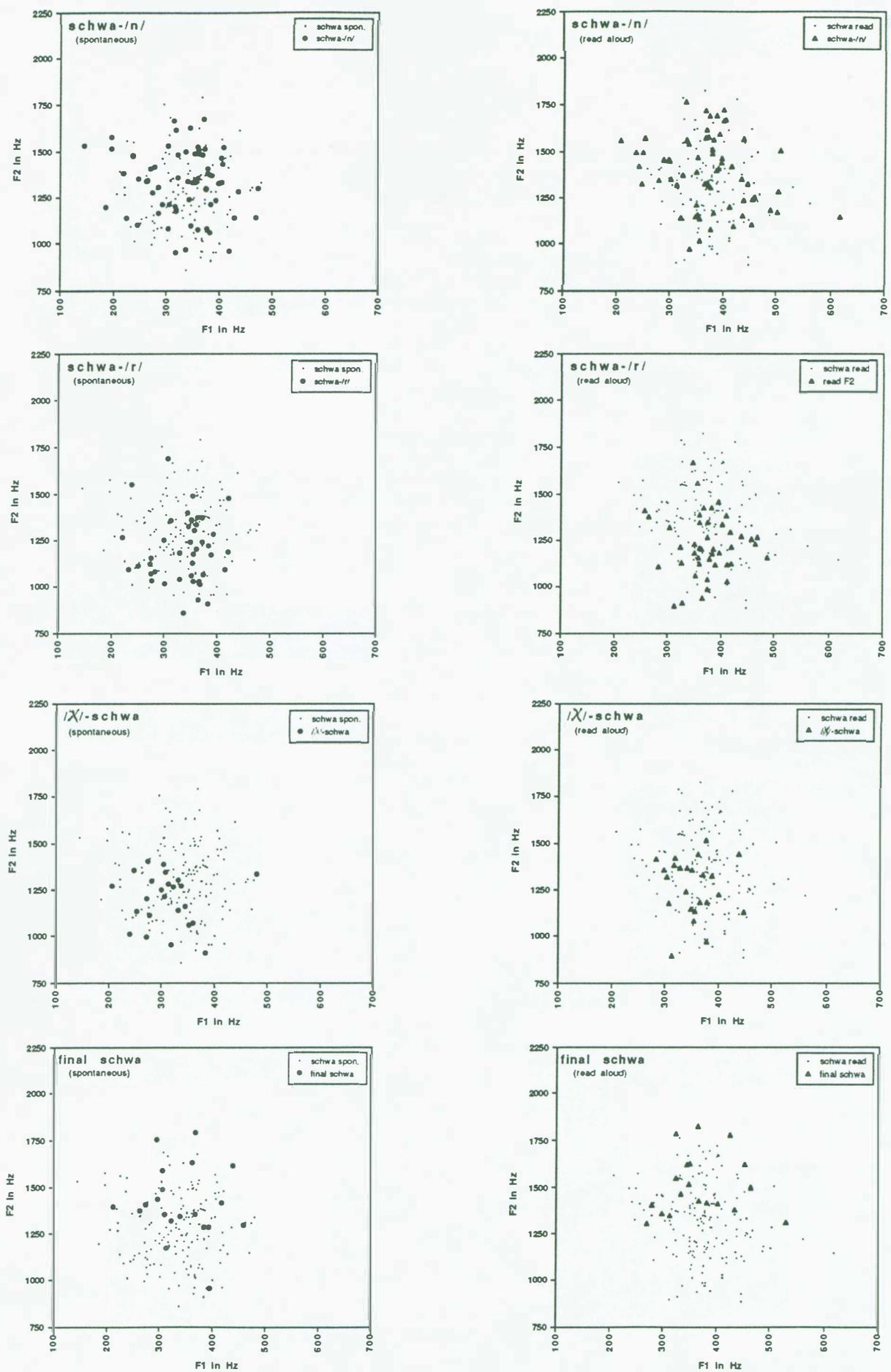


Fig. 9. Formant frequencies ($F1$ and $F2$) for all lexical schwas in focus words from spontaneous and read speech, together with formant values of schwa in subgroups determined by preceding or following consonants.

REFERENCES

- Bergem, D.R. van (in press): "Acoustic vowel reduction as a function of sentence accent, word stress, and word class". To appear in *Speech Communication*.
- Bergem, D.R. van (submitted): "A model of coarticulatory effects on the schwa". Submitted for publication to *Speech Communication*.
- Broecke, M.P.R. van den (1988): "Frequenties van letters, lettergrepen, woorden en fonemen in het Nederlands". In: Broecke, M.P.R. van den (Ed.), *Ter Sprake. Spaak als betekenisvol geluid in 36 thematische hoofdstukken*, Foris, Dordrecht: 400-407.
- Clark, J. & Yallop, C. (1990): *An introduction to Phonetics and Phonology*, Basil Blackwell Ltd, Oxford.
- Drullman, R. and Collier, R. (1991): "On the combined use of accented and unaccented diphones in speech synthesis". *J. Acoust. Soc. Am.* **90**: 1766-1775.
- Henton, C. (1990): "One vowel's life (and death) across languages: the moribundity and prestige of /ʌ/". *Journal of Phonetics* **18**: 203-227.
- Kager, R.W.J. (1989): *A metrical theory of stress and destressing in English and Dutch*; Ph.D. thesis, University of Utrecht.
- Koopmans-van Beinum, F.J. (1980): *Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions*, Ph. D. thesis, University of Amsterdam.
- Koopmans-van Beinum, F.J. (1982): "Akoestische en perceptieve aspecten van klinkercontrastreductie en de rol van de fonologie". *Spektator, Tijdschrift voor Neerlandistiek* **11**: 284-294.
- 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.
- Leeuwen, H.C. van & Lindert, E. te (1991): "Speech Maker: text-to speech synthesis based on a multi-level, synchronized data structure". *Proc. ICASSP 91 Toronto* **2**: 781-784.
- Makhoul, J. & Cosell, L. (1976): "LPCW: An LPC vocoder with linear predictive spectral warping". *Proc. ICASSP 1976*: 466-469.
- Marle, J. van & Zonneveld, W. (1982): "Bij een themanummer over de schwa". *Spektator, Tijdschrift voor Neerlandistiek* **11**: 279-283.
- 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.
- Son, R.J.J.H. van & Pols, L.C.W. (submitted): "The influence of formant track shape on the identification of synthetic vowels". Submitted for publication to *J. Acoust. Soc. Am.*