In: Studia gratulatoria dedicated to Hendrik Mol.
Amsterdam, Institute of Phonetic Sciences, 1979

SOME SPECTRAL AND PERCEPTUAL MEASUREMENTS ON DUTCH DIPHTHONGS

by Louis C.W. Pols

INTRODUCTION

The spectral information of isolated vowel sounds or vowel sounds in a fixed context can reasonably well be represented by single points in the formant space or any other dimensional spectral representation. The results of such an analysis for the 12 Dutch monophthongs /ɑ, a, ə, I, e, i, ɔ, o, u, oe, ø, y/ in a context of h(vowel)t have been published (Klein, Plomp and Pols, 1970; Pols, Tromp and Plomp, 1973; v. Nierop, Pols and Plomp, 1973). The words had been spoken once by 50 male and 25 female speakers. The first 100 msec of the most stationary part of the vowels were then gated out for a spectral analysis. The first three formant frequencies and formant levels were determined, as well as the one-third octave spectra of these vowel segments. The latter data were subjected to a data reduction procedure resulting in a two-dimensional principal-components spectral representation of the average spectral information in these vowel segments.

Both this information and the more traditional formant representation give us insight into the general spectral characteristics of Dutch vowels, as well as into the variation introduced by the different speakers. For more details we refer to the original publications.

However, such a representation tells us nothing about the

dynamic spectral behaviour of the vowel sounds. In a neutral or null context such as h(vowel)t this is probably not very important, but it certainly has to be taken into account when vowels are embedded in other consonant contexts in monosyllabic and multisyllabic words.

In this paper we do not intend to discuss such phenomena as coarticulation and vowel reduction. We should rather like to discuss another aspect of dynamic vowel behaviour: that being apparent in diphthongs.

Irrespective of whether the three Dutch diphthongs /ɑu,ɛi,ʌy/ have to be considered as vowels in themselves, combinations of two vowel nuclei, specific trajectories, or vowels with a peculiar dynamic character, the dynamic character itself is evident (Kaiser, 1948; Cohen, 1971). Up to recently it was not easy to measure this dynamic spectral behaviour and at best a few measurement points during the temporal course could roughly indicate the pattern. Such stylized patterns for the Dutch diphthongs in the formant space have been determined by Mol (1969). His results indicate that during the pronunciation of the diphthong the second-formant frequency F_2 only changes little, whereas F_1 changes in an avalanche-like way towards lower frequencies. Beginning and end are only roughly indicated and are said to change markedly for different speakers, and do not fit too well with the phonetic transcriptions of the diphthongs. Formant movement for American-English diphthongs /ɔI, aI, aʊ, eI, oʊ/ have been measured, for example by Holbrook and Fairbanks (1962). They found that the major movement of the diphthongs tends to occur during the last half of the utterance. For those diphthongs most of the variations is along the second formant, see also Lehiste and Peterson (1961). Also, on the basis of an intuitive articulatory description the diphthongs can be described as trajectories, or arrows in the vowel triangle, see for instance Jones (1972). For a phonological description, see Cohen et al. (1961).

Also, perceptual experiments can be done to study the dynamic behaviour of diphthongs. Van den Berg (1969) describes a method of playing back recorded diphthongs in reverse, in order to specify the various elements. A more sophisticated way to do that is by making use of a gating device, or segmentator, which allows detailed listening of any specific part in the utterance ('t Hart and Cohen, 1964; Gerber, 1974). Slis and v. Katwijk (1963) did experiments about perceptual tolerance regarding diphthongs synthesized as pairs of two-formant vowels. This results in idealized diphthong traces in the two-formant space. Similar perceptual experiments were done by Gay (1970) for American-English diphthongs. He tried to determine whether the phonetic identity of the target vowels, or the second formant transition course, served as the primary cue for identification. He concluded that the rate of change of the second formant for /əi, ai, au/ was more important than the onset and offset values. His own spectrographic measurements on naturally spoken diphthongs, however, show that the onset target formant position is also a fixed deature of the diphthong formant movement, together with the second-formant rate of change (Gay, 1968). 'Because of possible measurement error effects' the first formant rate of change was not measured.

In this paper we should like to extend somewhat the present knowledge of spectral diphthong movements in Dutch. We feel that our analysis procedure together with the subsequent data reduction and presentation allows a detailed analysis as a function of time.

The analysis procedure will shortly be described in the next section, together with the experimental results. For a more detailed description of the procedure we refer to Pols (in press). We will also present some identification results concerning the perceptual similarity between the different diphthongs.

SPECTRAL MEASUREMENTS ON DUTCH DIPHTHONGS

We have developed a spectral analysis system which allows real-time analysis of incoming speech with a parallel set of bandfilters. The outputs of these filters, after logarithmic amplification and envelope peak detection, are sampled once every 10 msec; the filter levels, together with overall level and the number of zero crossings, are stored in the computer and can be used for subsequent data processing. Also the fundamental frequency F_0 is measured by counting the number of zero crossings in the low-pass filtered signal of a throat microphone.

Subsequent 10-msec samples can be described as points in a 17-dimensional space, with coordinate values equal to the 17 filter levels, expressed relative to the overall level which is a simple level normalization. Because of interaction between sample points and between filters a considerable data reduction is possible.

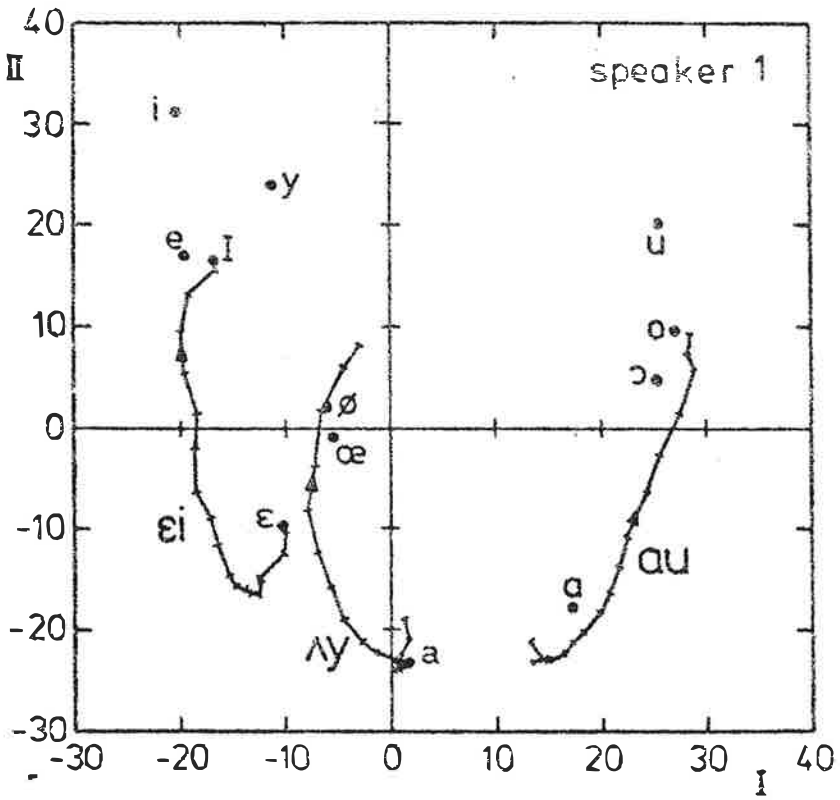

Figure 1.

Average traces for the diphthongs /ɑu/, /ʌy/ and /ɛi/ of speaker i in the two-dimensional principal-components of vowel subspace. All individual segments from the different words are linearly time normalized to 18 samples. For reference also the overall average vowel positions of this speaker are indicated.

The principal-components analysis (Harman, 1967) gives us a means of doing this. Eigenvalues and eigenvectors are extracted from a variance-covariance matrix based on all sample points. Successive eigenvectors describe the amount of variance explained by that new dimension, and the direction cosines of the corresponding eigenvectors define the new dimensions with respect to the original 17 dimensions. If only the first two new dimensions are used, the diphthongs can visually be represented as traces in that plane, similar to traces in the F_1-F_2 plane.

The axis orientations of the two-dimensional principal-components vowel space are highly correlated with the formant representation, which makes the interpretation more easily accessible to most readers. However, instead of single formant frequency values, these points represent different weightings of the complete bandfilter spectrum.

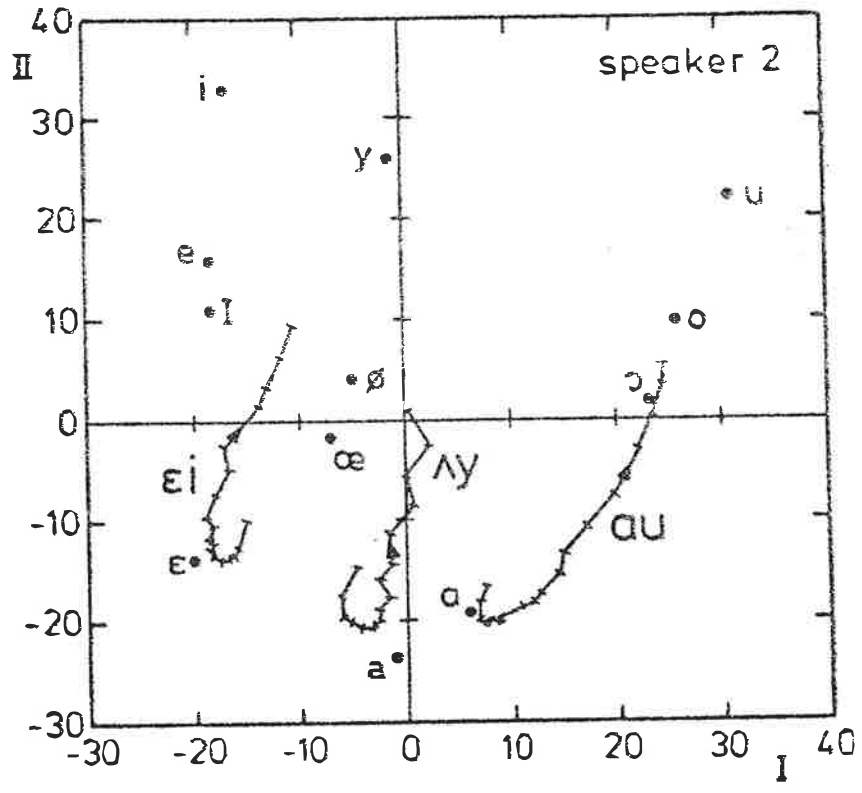

Figure 2.

Figure 2 similar to figure 1, but now for speaker 2.

The list of words used consisted of monosyllabic C_i. VC_f words with as initial consonants C_i all 18 possible Dutch consonants, including the condition without C_i. As vowels were used the three Dutch diphthongs /ɑu, ɛi, Ʌy/. Actually also the monophthongs were included, but for that part of the project see Pols (in press). Of all possible 14 final consonants C_f not many can actually be used in combination with the diphthongs. Table I gives the list of words in orthographic form. They were spoken by three different male speakers in a random order, as part of the larger word list with all vowels.

After the spectral analysis of these words, the diphthong segments had to be isolated from the words for a more detailed analysis. This segmentation was done by the experimenter on the basis of:

- a numerical display of all basic information per 10 msec;
- a graphical display of the word trace in a two-dimensional principal-components space;
- a synthesis of the whole word or any wanted part of it.

/ɑu/		/ʌy/		/ɛi/	
p au k		p ui		p ij p	
t ou w		th ui s	/tʌys/	t ij d	/tɛit/
k ou s		k ui p		k ij f	
b ou t		b ui t		b ij t	
d au w		d ui n		d ij k	
f au n		f ui k		f ij n	
s au l		s ui t		s ei n	
g ou d	/ɣɑut/	g ui t	/ɣʌyt/	g ei t	/ɣɛit/
v ou w		v ui l		v ij f	
z ou t		z ui l		z ei s	
h ou t		h ui d	/hʌyt/	h ij g	/hɛiɣ/
W ou t		w ui f		w ij n	
j ou w		j ui ch	/jʌyɣ/	j ij	
l au w		l ui t		l ij m	
r au w		r ui m		r ei n	
m ou t		m ui s		m ij n	
n au w		n ui s		n ij d	/nɛit/
ou d	/ɑut/	ui t		ij s	

Table I

List of meaningful $C_i VC_f$ words with the three diphthongs /ɑu, ʌy, ɛi/ as vowels, all 18 possible initial consonants, and some of the final consonants which are possible in combination with these diphthongs. Wherever the spelling may lead to confusion the phonetic transcription is also given.

The synthesis system is a newly developed vocoder-like device and is controlled by the analysis parameters stored in the computer. By specifying the sample numbers, any part of the utterance can be synthesized, if necessary repeatedly and/or stretched out. Together with the display of the word trace, these are very efficient tools for isolating the diph-

thong segments. It is difficult to make this an objective procedure, but all settings were independently controlled by a colleague. In the few cases when there was disagreement, a final decision was made by consultation.

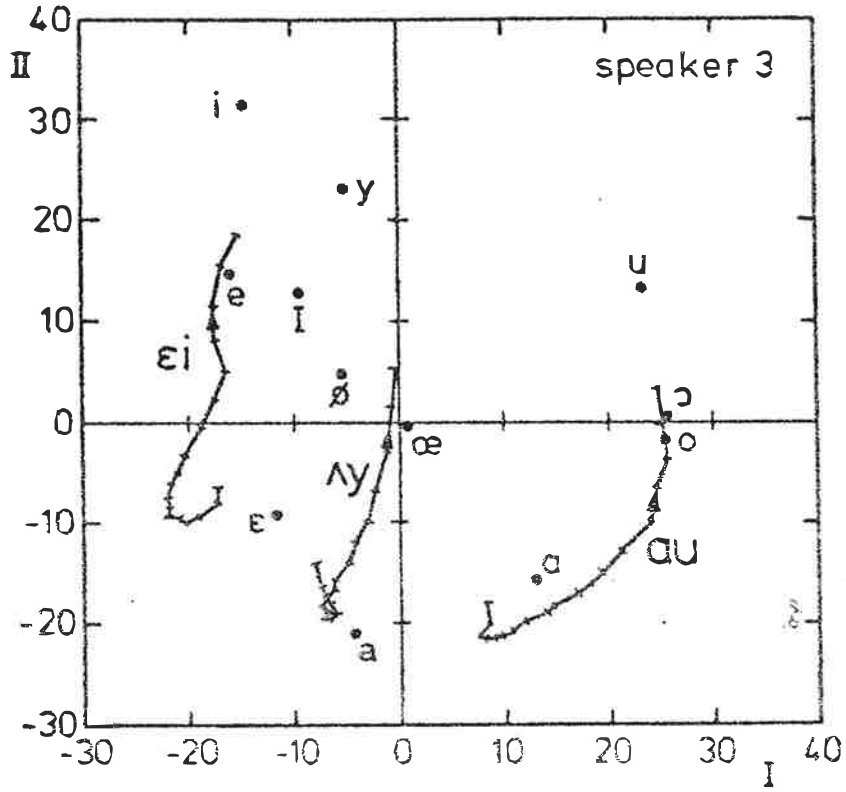

Figure 3.

Fig. 3 similar to figure 1, but now for speaker 3.

We have not tried with limited material to identify systematic consonant-specific effects in the diphthong traces, because the variation in the diphthong itself was already quite great. However, we did find consonant-specific effects for the monophthongs, especially for /e, o, ø/ in combination with a final /r/ (Pols, in press). Gay (1968) could not find any consistent initial or terminal consonant effect on target frequency values of diphthongs, but he did find a terminal consonant voicing effect on duration (longer if preceding a voiced consonant). A detailed study of durational properties of vowels in Dutch, including the diphthongs, is given by Nooteboom (1972).

In order to specify an average diphthong pattern we averaged the individual traces. The average duration of the diphthong segments was 180 msec, in other words 18 samples of 10 msec. The actual duration varied between 120 and 320 msec, but these extreme values were exceptional. Therefore, we accepted a linear time normalization to make the variable number of samples per segment equal to the average number of 18 samples. We then determined the average trace per diphthong per speaker in the two-dimensional principal-components vowel space. Figs. 1, 2, and 3 represent these traces for the three speakers respectively. A simple speaker normalization was applied by translating all sample points of one speaker in such a way that the origin represents the overall average vowel spectrum. The average vowel positions, also indicated in these three figures, are the average segment positions of the CVC words with monophthongs in the middle position.

The characteristic diphthong behaviour is very well reflected in this straightforward spectral representation. We see that /ɛi/ starts close to /ɛ/ and terminates in the neighbourhood of /I/, that /ʌy/ follows a track from /a/, or perhaps /ʌ/, to a place somewhat beyond /œ/, and that /ɑu/ starts between /a/ and /ɑ/, or in the neighbourhood of /ɑ/ and goes to /o/. None of the three diphthongs reaches its final vowel region as represented in the phonetic transcription. This has also been found for the American English diphthongs, e.g. Gay (1968).

We also see in the figures that in their initial parts the diphthongs vary only slowly, but that indeed as Mol (1969) already mentioned, the changes go much faster towards the end. From the individual data this is even more evident than from the represented average data. Another measure for this variation is the spread σ, equal to the square root of the variance, of the 18 points of the average trace per diphthong. Fig. 4 gives this spread along the second dimension for the three diphthongs of speaker 1. The data for the other two

speakers are similar. The second dimension is represented because along this dimension the diphthongs have their greatest excursion. It is evident from the vowel representations in Figs. 1-3 that this second dimension is highly correlated with F_1. As can be seen in Fig. 4, the spread σ clearly increases towards the end of the diphthong, representing a greater variation in that part of the sound. This variation can be caused by different final diphthong positions, but it can also be a result of not reaching the final position.

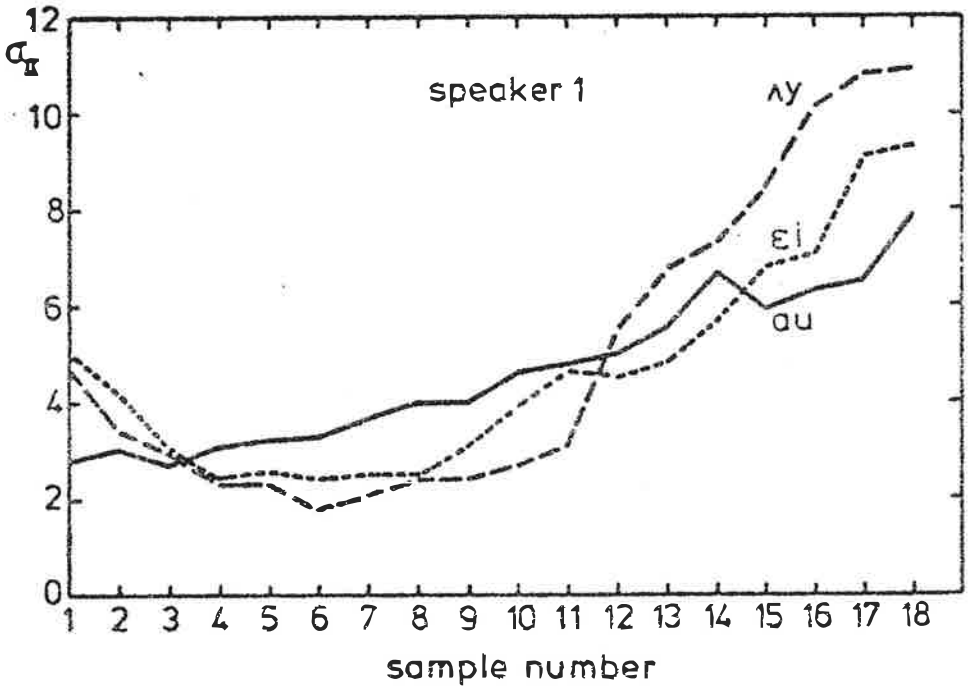

Fig. 4. The spread σ_{II} of the 18 points of the average trace per diphthong along the second dimension.

By looking at the individual data we get the impression that both effects play a part.

Our data confirm, also for Dutch, the finding of Gay (1968) for American English that onset vowel target position is a fixed feature of the spectral diphthong movement. Instead of

putting much emphasis on the rate of change for the dynamic part of the diphthong our data suggest that a diphthong can be described as quite a long and stable-state onset part followed by a fast specific transition into the direction of a larger offset area where no steady-state part is necessary. It is perhaps better to talk about the direction of change, rather than about the rate of change. Fig. 5 illustrates this with three examples of /ʌy/ traces varying in duration from 180 to 260 msec, but all three being almost unanimously identified as /ʌy/, as we will see in the next section.

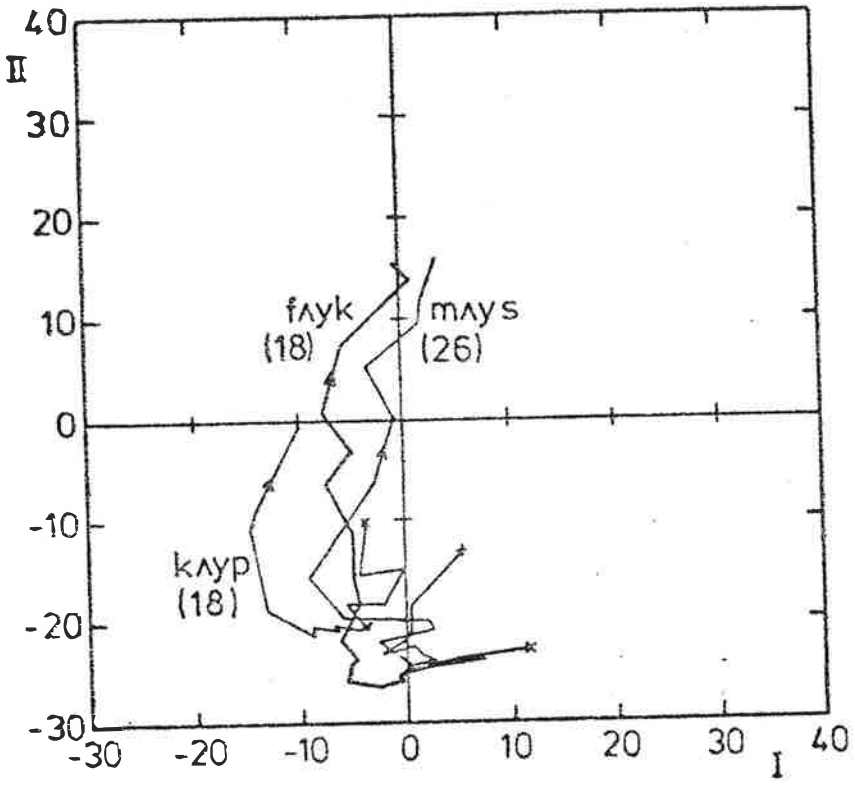

Figure 5

Three examples of /ʌy/-traces, varying in duration from 180 to 260 msec, but all three being unanimously identified as /ʌy/.

The segments from the words /fʌyk/ and /mʌys/ have a similar tail but a different number of samples in the onset area, whereas the tail of /kʌyp/ is shorter but apparently distinctive enough in its direction to also cause a clear /ʌy/ impression.

IDENTIFICATION MEASUREMENTS ON RESYNTHESIZED DIPHTHONGS

In order to verify the relevance of the spectral differences found, we also presented all isolated vowel segments of speaker 1, including the diphthongs, to a total of 27 listeners for identification. In order to have better control over the stimulus parameters, and in order to ease stimulus preparation we opted for resynthesizing the original diphthong segments. While interpreting the identification results one has to be aware of this point. The identification results for the three diphthongs, cumulated over the listeners, are given in Table II. Response categories were the 12 Dutch vowels plus the three diphthongs presented in orthographic form on top of the response buttons and grouped in alphabetic order. The average correct score was 74.4%, with a standard deviation of 7.3% over the 27 subjects.

	au	ɛi	ʌy	ɑ	a	ɛ	ɔ	o	ø	y	no	TOTAL	%CORR.
au	472	3	5		1		1	3			1	486	97.1
ɛi	1	313	111			7	2		24		1	459	68.2
ʌy	22	8	419	1	9			1	24	2		486	86.2
TOTAL	459	324	535	1	10	7	3	4	48	2	2	1431	84.1

Table II

Results in terms of a confusion matrix of the identification of the 18 different segments of three diphthongs by 27 listeners. The data are cumulated over the segments and the listeners. By accident one segment of the /ɛi/ stimuli was missing which makes the total for this diphthong 459 instead of 486.

The three diphthongs by themselves were 84.1% correct. /ɑu/

has a very high correct score, /Ʌy/ a somewhat lower score, confusions being mainly with /œu/ and /ø/. The /Ʌy-ø/ confusions can perhaps be attributed to the diphthongal nature of /ø/, which could be written as /œʏ/, coming close to /Ʌy/. Many /ɛi/ confusions are with /Ʌy/. Although the two average traces are relatively close, this amount of confusion is somewhat unexpected. The /ɛi/ segment out of /tɛil/ was only correctly identified 4 out of 27 times, which is by far the lowest score for all diphthong segments, and is caused by the tail of the transition part which first goes into the correct direction but then turns away under influence of the following /l/. The same effect can be seen in the /Ʌy/ segments out of /vƱyl/ and /zƱyl/, again causing many identification errors.

In general it can be said that strong deviations from the average diphthong pattern cause many identification errors.

CONCLUSION

In this paper we have tried to give some more detailed spectral information about the dynamic course of the Dutch diphthongs. The findings in general confirm the description given by Mol (1969). Together with the results from the identification experiment, we come to the conclusion that the spectral information in diphthongs can be described in such a way that the sound starts with a long and stable speaker-specific, steady-state onset part followed by a relatively fast specific transition into the direction of a larger offset area where no steady-state part is necessary.

Soesterberg, november 1976.

SUMMARY

Up to now the dynamic spectral behaviour of the Dutch diphthongs /ɑu, ʌy, ɛi/ was only known in terms of stylized patterns or arrows in the formant plane. In the present study we have analyzed with a bandfilter analysis system the actual spectral variation in 10-msec steps in diphthong segments in a number of words spoken by three different speakers This information could be visualized as traces in a two-dimensional principal-components vowel space. The same isolated segments were also presented to a group of listeners for identification. Strong deviations from the average diphthong pattern caused many identification errors.

Title received 1976

REFERENCES

Berg, B. van den (1969). Foniek van het Nederlands, Van Goor Zonen, Den Haag.

Cohen, A. (1971). "Diphthongs, mainly Dutch", in Form and Substance, L.L. Hammerick, R. Jacobson, and E. Zwirner, (eds.), Odense, 277-289.

Cohen, A., Ebeling, C.L., Fokkema, K., and Hoek, A.G.F. v. (1961). Fonologie van het Nederlands en het Fries, M. Nijhoff, 's-Gravenhage.

Gay, T. (1968). "Effect of speaking rate on diphthong formant movements". J. Acoust. Soc. Amer. 44, 1570-1573.

Gay, T. (1970). "A perceptual study of American English diphthongs", Language and Speech 13, 65-88.

Gerber, S.E. (1974). "Categorical perception of segmented diphthongs", Proc. Speech Comm. Seminar, Stockholm, Vol. 3, pag. 131-136.

Harman, H.H. (1967). Modern factor analysis, The University of Chicago Press, Chicago.

Holbrook, A., and Fairbanks, G. (1962). "Diphthong formants and their movements", J. Speech Hear. Res. 5, 33-58.

Jones, D. (1972). An Outline of English Phonetics, ninth ed., W. Heffer & Sons Ltd., Cambridge.

Kaiser, L. (1948). "Diphthongs in Dutch". Lingua 1, 303-305.

Klein, W., Plomp, R., and Pols, L.C.W. (1970). "Vowel spectra, vowel spaces, and vowel identification", J. Acoust. Soc. Amer. 48, 999-1009.

Koopmans- v. Beinum, F.J. (1969). "Nog meer fonetische zekerheden", Nieuwe Taalgids 62, 245-250.

Lehiste, I., and Peterson, G.E. (1961). "Transitions, glides, and diphthongs", J. Acoust. Soc. Amer. 33, 268-277.

Mol, H. (1969). "Fonetische zekerheden", Nieuwe Taalgids 62, 161-167.

Nierop, D.J.P.J. v., Pols, L.C.W., and Plomp, R. (1973). "Frequency analysis of Dutch vowels from 25 female speakers", Acustica 29, 110-118.

Nooteboom, S.G. (1972). "Production and Perception of Vowel Duration. A Study of Durational Properties of Vowels in Dutch". Doctoral thesis, University Utrecht.

Pols, L.C.W. (in press). "Spectral and perceptual differences between Dutch vowels in monosyllabic words".

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

Slis, I.H., and Katwijk, A.F.V. v. (1963). "Onderzoek naar Nederlandse tweeklanken", IPO report nr. 31.