THE IDENTIFICATION OF VOWEL STIMULI FROM MEN, WOMEN, AND CHILDREN

David J.M. Weenink

INTRODUCTION

In this paper we give some preliminary results of a series of listening experiments which have been conducted as part of a larger study on speaker normalization. Of all possible aspects of normalization we want to consider especially the perceptual aspects which concern speaker variation. In a series of eight listening experiments we have investigated how well listeners can recognize vowels from different speakers when these vowels were presented in a mixed and in a blocked condition. In the mixed condition the listeners, on each vowel, encounter a voice that is unfamiliar and unpredictable, while in the blocked condition the listener hears a series of vowels produced by the same speaker. In this latter condition there is ample opportunity to become familiar with the voice and the speaker is fully predictable from one vowel to the next. In order to gain better insight in this speaker variation, the vowel stimuli we used were manipulated in terms of duration, consonantal context, and fundamental frequency. Several authors have directed their attention to the speaker's context effect in the recognition of vowels, a.o. Strange et al. (1976), Verbrugge et al. (1976), Macchi (1980) and Assmann et al. (1982). The experimental scheme generally is such that vowel stimuli are presented in two conditions, mixed and blocked, to subjects who are asked to identify the vowel. Although there are great differences in the absolute error rates in these experiments, they all reach the same conclusion: uncertainty about a speaker as is the case in the mixed condition, leads to more confusion errors than when the speaker is 'known' as in the blocked condition. This effect is persistent even if the vowels are gated to a duration of 100 ms. The influence of consonantal context on the perception of vowels is still under debate. Some experimenters (Strange et al., 1976; Verbrugge et al., 1976; Rakerd et al., 1984) report significantly less identification errors made by listeners for vowels presented in CVC's than for vowels in isolation. According to them the consonantal context aids vowel identification. Other investigators like Macchi (1980) and Assmann et al. (1982) do not support this hypothesis. On the contrary, they do state that consonant coarticulation is not a necessary condition for accurate identification of naturally produced vowels. The consonantal advantage found by the former groups is not a genuine perceptual effect but a mere methodological artifact. Diehl et al. (1981), using speech synthesis, did not find superior performance of listeners on CVC stimuli either. At first sight one could think that because the formant trajectories of consonant- bounded vowels often fail to reach the frequencies characteristic of vowels produced in isolation (Lindblom, 1963; Stevens and House, 1963; Koopmans-van Beinum, 1980), consonant-bounded vowels would appear to be acoustically less distinctive than isolated vowels. The experiments of a.o. Macchi (1980) show that vowels in CVC's are recognized just as well as

vowels in isolation which means that dynamic spectral features must compensate for loss in static distinction. There is additional evidence that the human auditory system can predict spectral targets based on the transitional information (Furui, 1986). This means, however, that, when we gate short segments out of the cental parts of vowels produced in isolation and in /p-t/ context, there will be a difference in listeners' performance because the CVC segments taken from the CVC's are acoustically less distinctive: they lack the transitional information. As for our final point, the influence of fundamental frequency and timbre on the quality of vowels, we can say that vowel quality is largely independent of fundamental frequency because the spectral envelope is determined by the shape and length of the vocal tract rather than by the vocal cords. This envelope does not shift when a vowel is produced at a different fundamental frequency. Slawson (1968) studied what effect changing the fundamental frequency and/or the formants has upon vowel quality. He showed that the perceptual distance between two vowels, whose fundamental frequencies differed by an octave, could be minimized by raising the formants of the vowel with the highest pitch by approximately 10 %. Because of the fact that higher formants as compared to lower ones show smaller variations from vowel to vowel (e.g. Weenink, 1985, table I), Fujisaki and Kawashima (1968) tested whether a normalization process could be based on higher formant frequencies. They showed that neither fundamental frequency nor higher formants by themselves are sufficient for perceptual normalization but that both are necessary in any succesful normalization theory. Van Bergem (1986), in his study on vowels in /p-t/ context, reports that it is the combined effect of (acoustical context,) pitch and timbre that is important in the normalization process, in such a way that pitch and timbre determine speaker category (men, women and children); after this precategorization the reference set of each category can be used for further classification. The global design for all 8 experiments we will describe was the same: the recordings of the speech material and its further processing were done once and served as a basis for all experiments. The preparation procedure of the stimulus tapes, the listening conditions and the subjects were the same; only the stimuli differed for each experiment. In the following paragraphs we will give a description of these parts.

SPEECH MATERIAL

Recordings were made from 10 male, 10 female and 10 children's voices. All were native speakers of Dutch and they were carefully selected on their ability to speak the standard Dutch language without dialect influences. The recordings were made in an anechoic room with a Sennheiser MD421N microphone and a Revox A77 taperecorder. The recordings consisted of series of sentences "V van pVt" (V from pVt), where V is one of the twelve Dutch vowels /u,y,a,ɑ,ɛ,æ,e,i,I,o,ɔ,ø/. These sentences were read from paper with normal intonation, each sentence was repeated at least twice. During the recordings of the children's voices there was always a person familiar to the child present in the anechoic room for reassurence. Sentences were repeated until they were correctly spoken but in general the children made few mistakes and hardly any repetitions were necessary.

FURTHER PROCESSING OF THE SPEECH MATERIAL

The sentences on tape were digitized with a sample frequency of 10 kHz and 12 bits/sample. For each of the twelve vowels the best recording of each sentence from every speaker was stored on disk and was used for further processing. After selection and digitization our speech data base consisted of 360 sentences of the type "V van pVt" (30 speakers x 12 vowels) on disk. With the help of a speech editing program (Buiting, 1981), sentences were marked as the following figure shows.

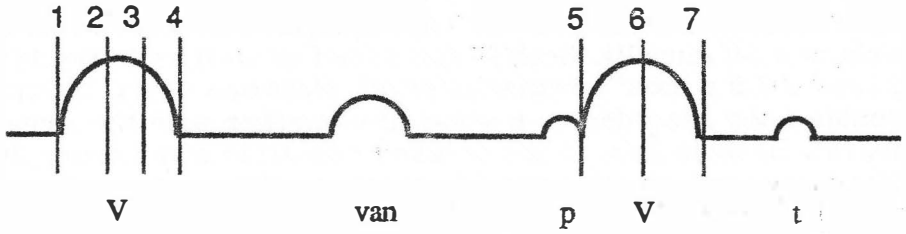

Fig.1. Position of marks in sentences 'V van pVt'.

Marks 1 and 4 bound the vowel produced in isolation, while marks 5 and 7 do so for the vowel produced in /p-t/ context. Mark 2, which is always at a stable part within the first 100 ms of the vowel, functions as a starting point for subsequent physical analysis, resynthesis and selection. Mark 6 has this function for the vowel in /p-t/ context and is placed approximately in the middle of this vowel, where the amplitude is most stable. The position of mark 3 is not of importance in this article. We can see from table I, where the mean lengths of intermark durations in ms are given, that the duration of the vowel in /p-t/ context is always smaller than the duration of the vowel produced in isolation. We performed a twelfth order linear prediction analysis and a software bandfilter analysis (Sekey and Hanson, 1984) on a 25.6 ms segment around mark 2 of all the vowels from our 30 speakers by means of a special computer program (Weenink, 1986). The results of this analysis were stored on disk to be used in subsequent listening experiments and further analyses. In these experiments we wanted to use all the vowels of a speaker twice in two listening conditions, mixed and blocked. Using all the speakers of our data base, the total amount of stimuli would have been 1440 (30 speakers x 12 vowels x 2 conditions x 2 repetitions), far too many for any practical listening experiment. We decided to select 5 male, 5 female and 5 children speakers out of the 30 speakers we had. This selection was made on the basis of the bandfilter analysis and the results of a pilot listening experiment with resynthesized vowels from the 30 speakers. From the categories man, woman and child we selected some 'extreme' and some 'mean' speakers and these 15 selected speakers were used in all the experiments we decribe in this article.

PREPARATION OF STIMULUS TAPES

For each of the eight experiments 5 audiotapes were prepared, each tape with a different random order of the stimuli. Each tape consisted of two

parts: the first part with the stimuli recorded in the mixed condition and the second part with the stimuli recorded in the blocked condition. The randomization procedure we used was as follows: in the mixed condition 360 stimuli (15 speakers x 12 vowels x 2 repetitions) were completely randomized under the constraint that maximally two adjacent stimuli came from the same speaker. The last 20 stimuli of this series of 360 were also put at the beginning of the tape and served as dummies to let the subjects get accustomed to this kind of stimuli. In the mixed condition we thus get a total of 380 stimuli. In the blocked condition the 15 speakers were randomized first, then for each speaker 24 stimuli (12 vowels x 2 repetitions) were randomized under the constraint that no two adjacent vowels were the same and the last six stimuli of this series were repeated at the beginning, summing to 30 stimuli for each speaker and 450 in this condition (15 speakers x 30 stimuli). Both in the mixed as well as in the blocked condition we used a 2.5 s inter stimulus interval. Between every 10 stimuli there was a double beep recorded as a separation marker with the same 2.5 s time interval. Besides this, in the blocked condition after every 30 stimuli a triple beep tone was recorded to separate different speakers.

LISTENING CONDITIONS

The identification tests were performed in a special acoustically isolated studio room at the Language Department (ITT) of the Faculty of Arts of the University of Amsterdam. In each session four subjects at a time could be handled, there were 5 sessions in every experiment. Test tapes were presented via a Revox A77 tape recorder, Sansui AU-22 amplifier, and a set of Sennheiser HD22 headphones at a comfortable listening level. Subjects were seated in front of a specially developed response unit which consisted of a monitor and a keyboard, and they responded by pushing a key on the keyboard (see fig. 2).
Twelve keys on the keyboard were marked with stickers, showing the orthographic symbols 'pVt', a thirteenth was labeled 'fout' (error). The remaining keys of the keyboard were covered with a special protection plate. The layout is shown in fig. 3.
Although we did not expect as much orthographic interference as in English, vowels which were expected to get confused orthographically like /y/ and /œ/ (pUUt and pUt) were placed as close as possible to each other, to attract special attention of the subject when responding. A subject's response was immediately displayed on his monitor, together with the response number as a confirmation. In case of a typing error or an incorrect response, subjects were able to correct their last given response by using the 'fout' button and then giving their intended response. This corrected response was displayed with the same response number as the previous one. The response units of the four subjects were connected to a central Apple IIe computer (Weenink, 1986). The responses of the four subjects were displayed on the Apple IIe's monitor together with the stimulus' number and type. In this way the experimenter had full control over the experiment and could intervene if necessary. He stopped the audiotape when a subject either forgot to respond or gave a double response to a stimulus, he then asked the subject who was in error to perform a certain action. The double beep between a series of ten successive stimuli served as a timer. Subjects made few mistakes, approximately once in every session the experimenter had to stop the tape to make a correction. Before a session started the subjects were instructed that the experiment was on vowels and that

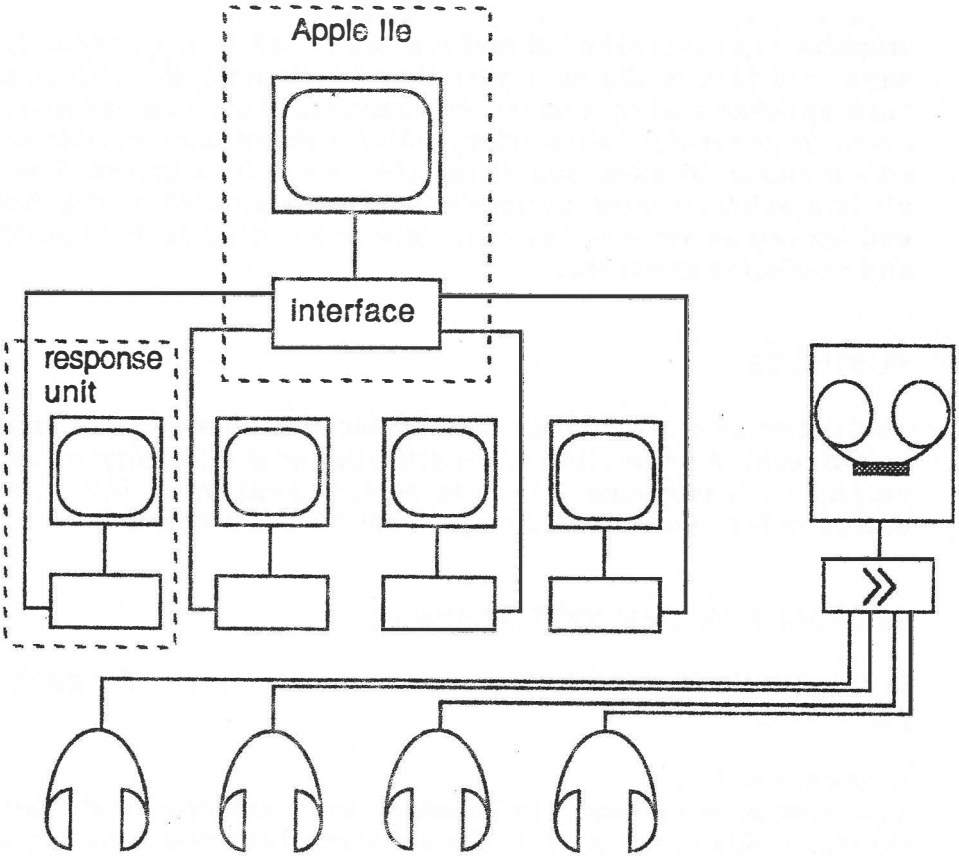

Fig.2. Listening configuration. Four response units are connected to a central unit (Apple IIe). A Revox A77 taperecorder and Sansui AU-22 amplifier provide the audio signals to the earphones.

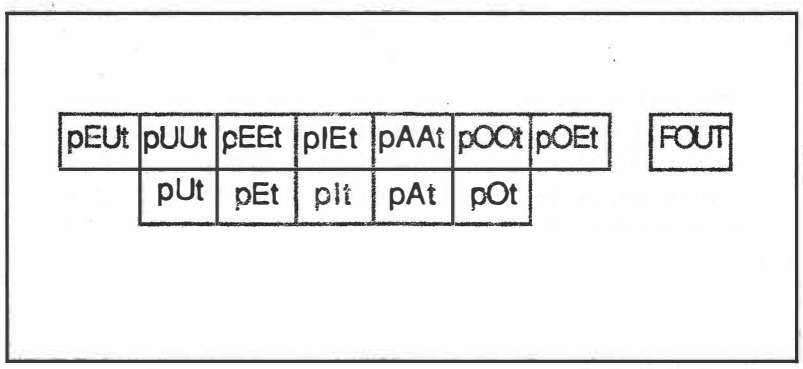

Fig.3. Layout of the keybord part of the response unit.

different vowels of different speakers were mixed in the first part of the experiment. After the 380 stimuli in the mixed condition had passed, the stimulus tape was stopped and a short break was granted. Then subjects were told that in the next part they would hear the stimuli blocked for each speaker and that every new speaker would be announced by a triple beep. In general a full session, which consisted of stimuli presented in mixed and in blocked condition, took about 50 minutes. The responses of all four subjects were gathered on the floppy disk of the Apple IIe computer and served as input for further data processing such as cumulative results and confusion matrices.

SUBJECTS

The listeners were 10 male and 10 female, phonetically untrained, paid volunteers. Most of them were students at the Faculty of Arts of the University of Amsterdam. All were native speakers of Dutch, with no hearing deficiencies and ranging in age from 20 to 30 years.

STIMULI FOR THE EXPERIMENTS

In this section a description of the stimuli is given for each of the 8 experiments.

Experiment 1.
The vowels as produced in isolation were selected with their original length (in fig. 1 this is the part of the sentence between mark 1 and 4). This experiment is a replication of the experiments performed for English (American) vowels by a.o. Strange et al. (1967), Macchi (1980) and Assmann et al.(1982) and investigates how well natural, isolated, Dutch vowels are recognized when they are presented in mixed and in blocked condition to listeners. We expect listeners to make few mistakes, in accordance with the experiments of Assmann et al. and Macchi.

Experiment 2.
50 ms segments around mark 2 of the vowel produced in isolation were selected. The initial half of a cosine window was used to smooth the onset of the first 5 ms portion of the selected signal; this was followed by 40 ms at the original amplitude; the last 5 ms of the signal was smoothed by the second half of the cosine window. We choose this 50 ms length to have a duration which comes close to the duration of short vowels in conversational speech. A second reason was that we wanted to increase the number of confusions. We also wanted to avoid duration differences and dynamic features such as diphthongization. Because all segments are equalized in duration we introduce extra confusions between vowels where duration is the main cue for separating them, like between /ø/-/œ/, /o/-/ɔ/, /e/-/ɪ/ and /a/-/ɑ/ (Pols et al., 1973; Nierop et al., 1973; Nooteboom et al., 1980). We name confusions of this type 'long/short confusions' and we shall have to correct for them afterwards. Assmann et al. (1982) find a mixed/blocked effect in their experiment with vowel durations gated to 100 ms; we too expect this effect to happen despite our shorter duration of 50 ms because if speaker information is still present in the vowel, listeners can take advantage of this fact when the vowels are presented in a blocked condition.

Experiment 3.
50 ms segments around mark 6 of the vowel produced in /p-t/ context were selected and smoothed as described above. The importance of dynamic spectral information has been reported for vowel perception. In continuous speech, when vowels can be coarticulated with consonants the spectral pattern of the speech signal varies in such a way that the acoustic targets found in isolated vowels, may not be attained (Stevens and House, 1963; Koopmans- van Beinum, 1980). One refers to this phenomenon as target undershoot and it is determined by speaking rate, sentence and word stress, and individual style of speech (Lindblom, 1963). Because of this possible undershoot we expect our vowels taken from their /p-t/ context, to be acoustically less distinctive than their counterparts which were produced in isolation when both are gated to a fixed short duration and are presented in isolation.

Experiment 4.
Stimuli were 50 ms segments, resynthesized as a stationary signal from the linear prediction analysis of order 12 which was done on a 25.6 ms segment around mark 2 of the vowel produced in isolation. All pitch periods in this 50 ms resynthesized segment were the same, equal to the mean pitch of the corresponding analyzed segment. From pilot studies we got the impression that listeners made a precategorization of stimuli, mainly on the basis of pitch, into male-like, female-like and/or child-like. In order to manipulate with the fundamental frequency in a well defined way we had to use resynthesis. Because of the inherent smoothing performed by any analysis-resynthesis system we expect more confusion errors in this experiment than in the preceding ones. Although the spectral envelope of the resynthesized signal is smoothed we still expect enough speaker specific information to be present in this signal to be of help in the blocked condition, which means that their should be a difference in listeners performance in the mixed and blocked condition.

Experiment 5.
50 ms segments, resynthesized with a fundamental frequency of 135 Hz from linear prediction coefficients. Resynthesis was performed using the 12th order linear prediction coefficients from experiment 4. The chosen frequency is approximately the mean male fundamental frequency as was measured from the voices of our 10 male speakers. In resynthesizing all the analysed vowels of our 5 male, 5 female and 5 children speakers with the same fundamental frequency of 135 Hz ('male-like') we present to the listener partly conflicting vowel information: on the one hand a frequency envelope belonging to a certain speaker category and on the other hand a fundamental frequency which did not 'fit' (in this experiment this was the case for children and female voices). On the basis of investigations of - Fujisaki et al. (1968) and Wendahl (1959) we know that there is an interaction between fundamental frequency and spectral envelope. Therefore our expectation is that especially in the categories women and children the amount of confusions will rise.

Experiment 6.
50 ms segments, resynthesized with a fundamental frequency of 235 Hz from linear prediction coefficients. The prediction coefficients from experiment 4 were used. 235 Hz is approximately the mean female fundamental frequency as was measured from our 10 female voices. Again, like in experiment 5, there is conflicting information present in the resynthesized vowels,

but this time it should interact mainly with the vowels from the male and the children speakers.

Experiment 7.
50 ms segments, resynthesized with a fundamental frequency of 335 Hz from linear prediction coefficients. This frequency is approximately the mean children fundamental frequency as was measured from our children voices. The same prediction coefficients were used as in experiments 4, 5 and 6. This time we expect the male and female vowels to have the greatest interaction because their vowels are resynthesized with the greatest shift in fundamental frequency with respect to their 'normal' fundamental frequency.

Experiment 8.
50 ms segments, resynthesized with noise from the linear prediction coefficients.The same prediction coefficients were used as in all the above resynthesis experiments. Because of the fact that a very important indication of speaker category, the fundamental frequency, is absent we expect more confusion errors in this experiment than in experiment 4 where the vowels are resynthesized with their 'own' fundamental frequency. If, on the other hand, the information about speaker category is still present in the spectral envelope in another way, listeners performance should be comparable.

RESULTS AND DISCUSSION

In table II results of the 8 listening experiments are presented. This table contains the mean error percentages for each experiment, in the mixed and the blocked condition averaged over all subjects and vowels, both for all speakers as well as for the separate speaker categories men, women and children. Table III presents the data corrected for long/short confusions. This correction means that a short vowel response given to its long counterpart stimulus is considered to be a correct response. The reverse, a long vowel responded to its short counterpart stimulus, is considered as a false response. In figures 4, 5 and 6 the data from these tables are visualized in histograms.
From experiment 1 we can conclude that vowels produced in isolation and presented in a mixed condition, can be recognized very well by listeners, only 10.9 % errors. This result is significantly better in the blocked condition: only 4.4 % errors. These percentages are close to the percentages that Macchi (1980) and Assmann et al. (1982) report. See table IV for an overview. We want to emphasize that the differences in error percentages between the mixed and the blocked condition were statistically significant ($p<0.01$) in all 8 experiments. Reducing the length of the stimuli to 50 ms (experiment 2) has increased considerably the number of incorrect responses: 35.6 and 31.2 % respectively for mixed and blocked conditions. When we correct our data for long/short confusions, results become much better: 18.7 and 15.1 % confusion errors for mixed and blocked condition respectively. These error scores are somewhat higher than the percentages that Assmann et al. report for their experiment on gated vowels, but only relatively because the duration of our gated vowels is half the duration of theirs. The number of confusion errors in experiment 3 (segments from vowels produced in /p-t/ context) has increased as compared to experiment 2 (segments from vowels produced in isolation), see table III.

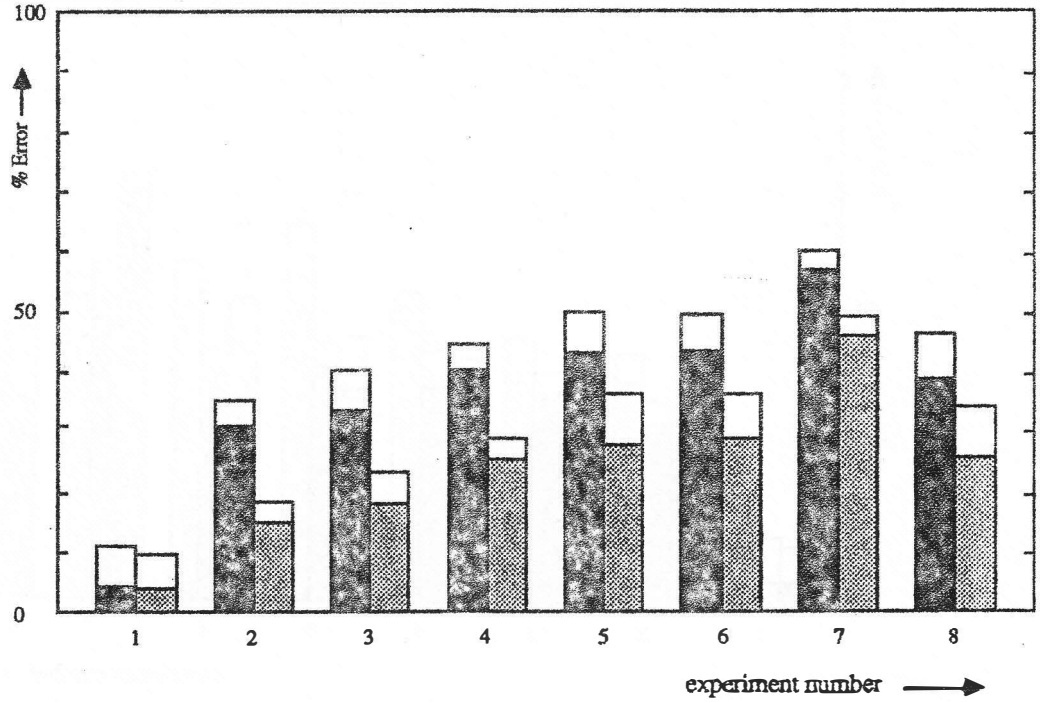

Fig.4. Error percentages averaged over subjects (20), vowels (12) and speakers (15) for experiments 1 to 8 (see text). Each column contains the error percentages in the mixed (open) and in the blocked condition (shaded). In each pair of the columns the left column contains the uncorrected data while in the right column these data are corrected for long/short confusions.

This difference in percentage confusion errors proved to be statistically significant, which confirms the hypothesis that the center part of vowels in /p-t/ context is acoustically less clearly defined than the center part of vowels produced in isolation.
The only difference between the stimuli of experiments 4 to 8 is the fundamental frequency of the source used for the resynthesis. Because of the fact that the percentages error in fig 4 are not the same for all these experiments, we can conclude that indeed their is an interaction between source and spectral envelope. This effect is strongest in experiment 7 where we resynthesized with a fundamental frequency of 335 Hz. This impression of the interaction becomes even stronger if we look at figures 5 and 6 where the speakers were split up in the separate categories men, women and children. We see that the error percentages in these experiments differ considerably for these categories. In general one could say that the error percentages are lowest when a category is resynthesized with its 'proper' fundamental frequency (in experiment 5: men; in experiment 6: women; in experiment 7: children).
We further note that the children's stimuli, according to the performances of the listeners, are not as well defined as the stimuli of the men and women. This is already clear in experiment 2 where we see that the 50 ms male and female vowel stimuli are much better recognized than the children's stimuli. Because we use the analysis of the vowel segments for further processing, this effect proceeds in the resynthesis experiments.

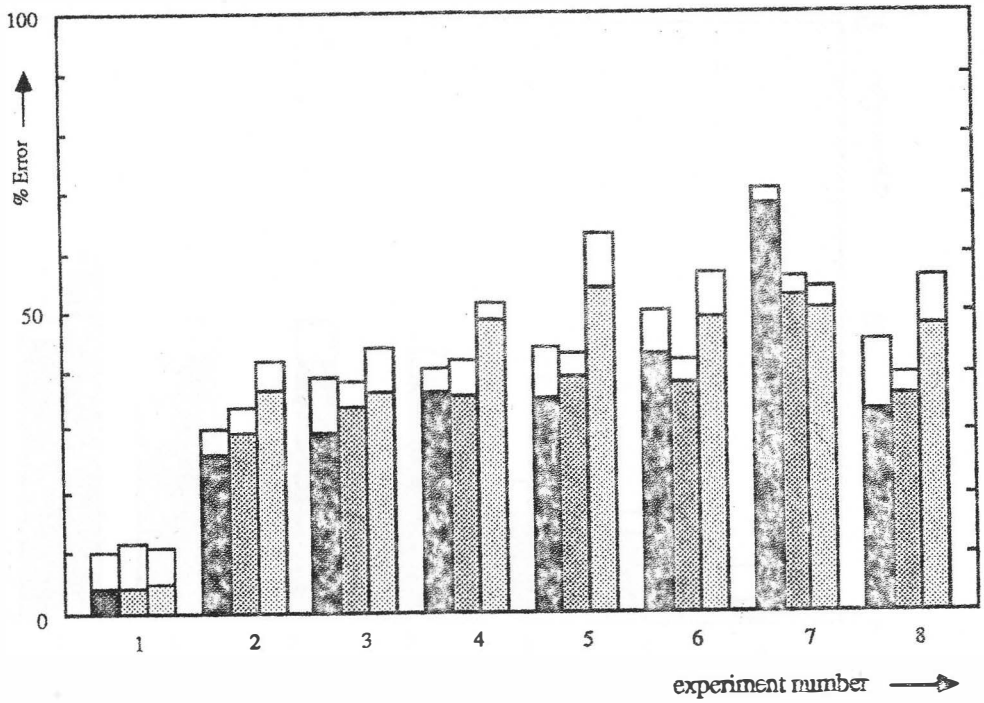

Fig.5. Error percentages averaged over subjects (20) and vowels (12) but split up into speaker categories men (left column), women (centre column) and children (right column) for experiments 1 to 8 (see text). Each column contains the percentage error in the mixed (open) and in the blocked condition (shaded).

There are several explanations why the children's segments are not as clearly defined:

- the limitation of the bandwith to 5000 Hz can have a greater degrading effect on the children's vowels. The high frequency components of the children's voices seem to be stronger than the corresponding components of the female and male voices.
- in the children's vowels there are more amplitude variations than in the vowels of the men and women, probably because children have less control over their voices. These amplitude variations can, in the subsequent linear prediction analysis, be the cause of some more spectral smoothing.
- the high fundamental frequency of the children's vowels makes their spectral envelope less clearly defined. This also has a degrading effect on the linear prediction analysis because the 'effective' time interval for the analysis becomes shorter.
- maybe the listener is in need of more dynamic spectral variation to compensate for the loss in static resolution in the children's vowels.

Further we note the especially good identification of the stimuli resynthesized with noise: the error percentages in experiment 8, where the vowels were resynthesized with noise, are approximately the same as the percentages in experiment 4, where the stimuli were resynthesized with their original fundamental frequency. Because of the fact that no direct fundamental frequency information is present in the stimuli from experiment 8, a major

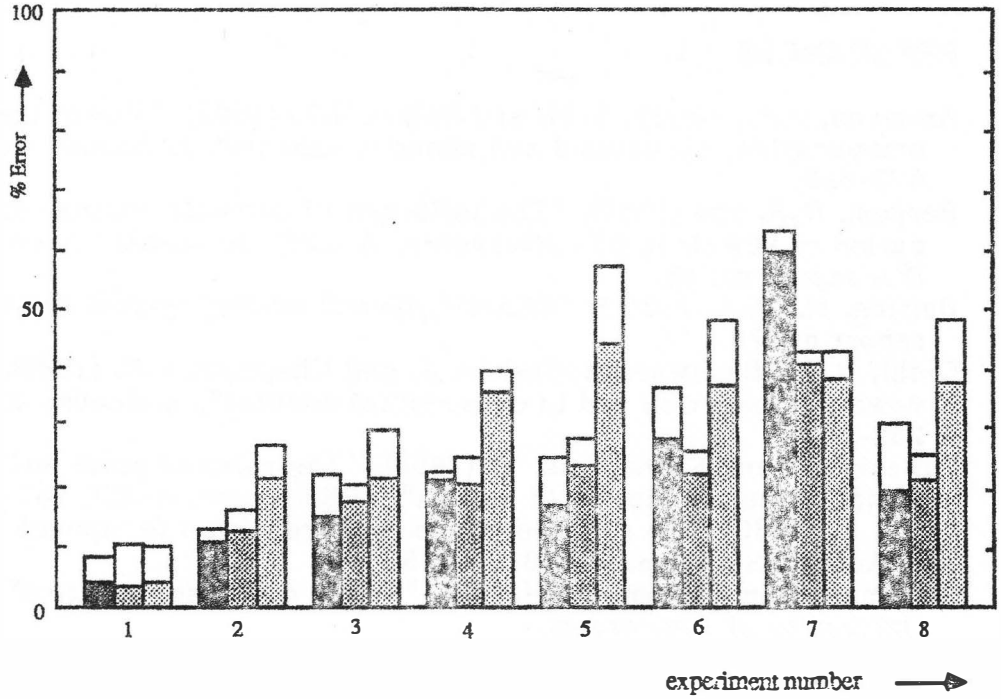

Fig.6. Same as fig.5, all data have been corrected for long/short confusions.

cue for speaker precategorization is not present. This means that besides pitch there must be spectral cues in the signal from which the listener can nevertheless extract relevant normalization information. Detailed physical analyses have been performed on the spectra to gather data for this information. The physical analysis of all the stimuli used in our listening experiments will hopefully shed some light on which spectral cues the listener might use for his normalization. With this information we will be able to predict listener's behaviour in our listening experiments and we will have gained a better insight in the perceptual and physical proces of normalization. The data on the physical analysis will be presented in a following paper.

ACKNOWLEDGEMENTS

This investigation was supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.), project nr. 300-161-030. I thank Louis Pols en Florien Koopmans- van Beinum for critically reviewing this paper and for stimulating discussions.

REFERENCES

Assmann, P.F., Neary, T.M. and Hogan, J.T. (1982), "Vowel identification: orthographic, perceptual and acoustic aspects", J. Acoust. Soc. Am. 71, 975-989.

Bergem, D.R. van (1986), "The influence of acoustic context on the identification of vowels in pVt utterances. A study on speaker normalization", IFA report nr. 88.

Buiting, H.J.A.G. (1981), "SESAM, speech editing system Amsterdam", IFA report nr. 70.

Diehl, R.R., Buchwald McCusker, S. and Chapman, L.S. (1981), "Perceiving vowels in isolation and in consonantal context", J. Acoust. Soc. Am. 69, 239-248.

Fujisaki, H. and Kawashima, T. (1968), "The roles of pitch and higher formants in the perception of vowels", IEEE Trans. ASSP, vol AU-16, 73-77.

Furui, S. (1986), "On the role of spectral transition for speech perception", J. Acoust. Soc. Am. 80, 1016-1025.

Koopmans-van Beinum, F.J. (1980), "Vowel contrast reduction", Ph. D. Thesis, University of Amsterdam.

Lindblom, B.E.F. (1963), "Spectrographic study of vowel reduction", J. Acoust. Soc. Am. 35, 1773-1781.

Macchi, M.J. (1980), "Identification of vowels spoken in isolation versus vowels spoken in consonantal context", J. Acoust. Soc. Am. 68, 1636-1642.

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

Nooteboom, S.G. and Doodeman, G.J.N. (1980), "Production and perception of vowel length in spoken sentences", J. Acoust. Soc. Am. 67, 276-287.

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

Rakerd, B., Verbrugge, R.R. and Shankweiler, D.P. (1984), "Monitoring for vowels in isolation and in consonantal context", J. Acoust. Soc. Am. 76, 27-31.

Sekey, A. and Hanson, B.A. (1984), "Improved 1-Bark bandwidth auditory filter", J. Acoust. Soc. Am. 75, 1903-1904.

Slawson, A.W. (1968), "Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency", J. Acoust. Soc. Am. 43, 87-101.

Stevens, K.N. and House, A.S. (1963), "Perturbation of vowel articulations by consonantal context: An acoustical study", J. Speech Hear. Res. 6, 111-128.

Strange, W., Verbrugge, R.R., Shankweiler, D.P. and Edman, T.R. (1976), "Consonant environment specifies vowel identity", J. Acoust. Soc. Am. 60, 213-224.

Verbrugge, R.R., Strange, W., Shankweiler, D.P. and Edman, T.R. (1976), "What information enables a listener to map a talker's vowel space ?", J. Acoust. Soc. Am. 60, 198-212.

Weenink, D.J.M. (1986), "QQ, een programma voor analyse, resynthese en herkenning van klinker segmenten", IFA report nr. 82.

Weenink, D.J.M. and Wempe, A.G. (1986), "Communicatie tussen een Apple IIe en vier Commodore Vic 20's", IFA report nr. 83.

Wendahl, R.W. (1959), "Fundamental frequency and absolute vowel identification", J. Acoust. Soc. Am. 31, 109-110 (A).

Table I. Intermark durations in ms. See figure 1 for position of marks. Means and standard deviations are given in subsequent entries.

	4-1		2-1		7-5		6-5	
vowel	mean	std.	mean	std.	mean	std.	mean	std.
u	174	48	47	14	133	32	54	34
y	186	49	60	16	134	31	57	26
a	225	55	58	21	218	33	71	28
ɑ	162	45	44	18	131	23	50	19
ɛ	167	34	50	17	130	23	53	18
œ	166	48	53	17	130	29	53	21
e	234	66	53	14	199	43	60	23
i	175	46	48	12	126	32	48	27
I	164	46	48	16	125	33	47	21
o	238	73	52	16	218	37	75	31
ɔ	165	45	52	17	131	36	55	31
ø	245	70	58	14	215	43	68	31

Table II. Error percentages over subjects (20), vowels (12) and speakers (15). The speakers are also split up into categories men, women and children for both mixed as well as blocked condition. See text for a further specification of the experiments.

exp. nr.	short characterization of the experiment	averaged/total		men		women		children	
		mixed	blocked	mixed	blocked	mixed	blocked	mixed	blocked
1	vowel V	10.9	4.4	9.9	4.2	11.5	4.0	11.4	5.0
2	50 ms from V	35.6	31.2	30.7	26.4	34.0	30.0	42.0	36.8
3	50 ms from pVt	40.6	33.6	39.2	30.1	38.5	34.0	44.0	36.6
4	50 ms, mean F0	44.6	40.3	40.3	36.3	41.8	36.2	51.7	48.5
5	50 ms, F0=135	49.9	42.8	44.0	35.3	42.9	38.9	62.8	54.0
6	50 ms, F0=235	49.5	43.3	50.1	43.0	41.8	38.1	56.6	48.9
7	50 ms, F0=335	59.9	57.0	70.3	68.0	55.4	52.7	53.9	50.5
8	50 ms, noise	46.7	39.2	45.1	33.7	39.5	36.1	55.5	47.7

Table III. Same as table II. All data have been corrected for long/short confusions.

exp. nr.	short characterization of the experiment	men		women		children		averaged/total	
		mixed	blocked	mixed	blocked	mixed	blocked	mixed	blocked
1	vowel V	8.5	3.8	10.6	3.6	9.8	4.0	9.6	3.8
2	50 ms from V	12.9	11.1	16.1	12.7	27.1	21.6	18.7	15.1
3	50 ms from pVt	22.0	15.1	21.0	17.7	29.6	21.5	24.2	18.1
4	50 ms, mean F0	22.6	20.8	24.9	20.7	39.3	35.9	29.0	25.8
5	50 ms, F0=135	25.0	17.3	28.0	23.3	57.0	44.0	36.7	28.2
6	50 ms, F0=235	36.3	28.0	26.1	22.0	47.9	37.0	36.7	29.0
7	50 ms, F0=335	62.8	59.3	42.5	40.4	42.4	37.8	49.2	45.8
8	50 ms, noise	30.5	19.6	25.3	21.2	47.8	37.4	34.5	26.0

Table IV. Comparison of percentage error in the mixed and the blocked condition for different experiments. From left to right the subsequent columns (indicated within brackets) are respectively: the experimenters (1), the year of the publication (2), the number of speakers in the categories men, women and children (3), the number of vowels used (4), the duration of the stimuli (in ms), 'full' means that no segmentation has taken place (5), stimulus type (6), percentage error in the mixed (7) and in the blocked condition (8).

Experiment	year	speakers	#V	length	type	mixed (%)	blocked (%)
Verbrugge et al.	76	5,5,5	9	full	pVp	17.0	9.5
Strange et al.	76	5,5,5	9	full	pVp	17.0	9.0
	76	5,5,5	9	full	V	42.6	31.2
	76	4,4,4	9	full	CVC	23.0	22.0
Macchi	80	5,5,5	11	full	V	7.8	1.5
	80	5,5,5	11	full	tVt	8.6	2.0
Assmann et al.	82	5,5,0	10	full	V	5.4	4.1
	82	5,5,0	10	100	V	13.8	9.5
Weenink	86	5,5,5	12	full	V	9.6	3.8
	86	5,5,5	12	50	V	18.7	15.1
	86	5,5,5	12	50	pVt	24.2	18.1