IDENTIFICATION AND DISCRIMINATION OF CLOSELY SPACED SYNTHETIC VOWELS[1]

Peter Holtse

1. Introduction

It is commonly reported that human listeners will discriminate smaller differences in quality between vowels than between consonants. Further it is found that whereas there is a strong tendency for consonants to be perceived in a categorical way, i.e. discrimination is poor between qualities which cannot be identified absolutely, the tendency among vowels is towards continuous perception similar to the perception normally found with non-speech sounds. It has, however, been objected (Fischer-Jørgensen (1970-71)) that the relatively high discrimination between vowels might be caused by a difference in the auditory distances between vowel and consonant stimuli in the experiments. Thus it has been assumed by Stevens, Liberman, Studdert-Kennedy and Öhman (1969), and by a number of other authors who have used the same stimulus material, that if thirteen stimuli are placed along a continuum containing three phoneme categories the auditory distances between the stimuli must be approximately equal irrespective of the nature of the continuum. But it seems likely that the auditory distances between different vowel phonemes are actually somewhat larger than the distances between consonants since greater allowance must be made for social and personal variations in the pronunciation of vowels.

---

1) I am grateful to Eli Fischer-Jørgensen for much valuable criticism and many helpful suggestions during the writing of this article.

The aim of the present experiment was therefore to repeat the earlier identification/discrimination experiments with a set of closely spaced vowel stimuli in order to judge the possible influence from interstimulus distances. A second aim was an attempt to eliminate the effect from unevenly spaced stimuli which had been noticed in some earlier experiments (Hutters and Holtse (1972)).

2. The stimulus material

2.1. Formant frequencies

As material for the experiment was chosen a series of front unrounded vowels ranging in quality from [i] to [a]. The formant frequencies of the stimuli were chosen so that a line drawn through the stimuli in the F1/F2-diagram would pass through the areas normally taken up by the Danish long vowels /i:/, /e:/, /ɛ:/ and /a:/, (see Frøkjær-Jensen (1967) and Fischer-Jørgensen (1972)). In practice this was done by calculating the best fitting curves which would describe F2 and F3 of all the four vowels as polynomial functions of F1.

The stimuli were placed with equal logarithmic steps along F1. And the corresponding F2 and F3 frequencies were determined by the calculated functions. The formant frequencies of the stimuli are listed in table 1. And in fig. 1a F1, F2, and F3 of the stimuli are compared with mean formant frequencies from the data given by Fischer-Jørgensen (1972). In the calculation of these mean values those persons were excluded whose mean F2 of /i:/ was lower than the same person's F2 of /e:/. This particular formant pattern is characteristic

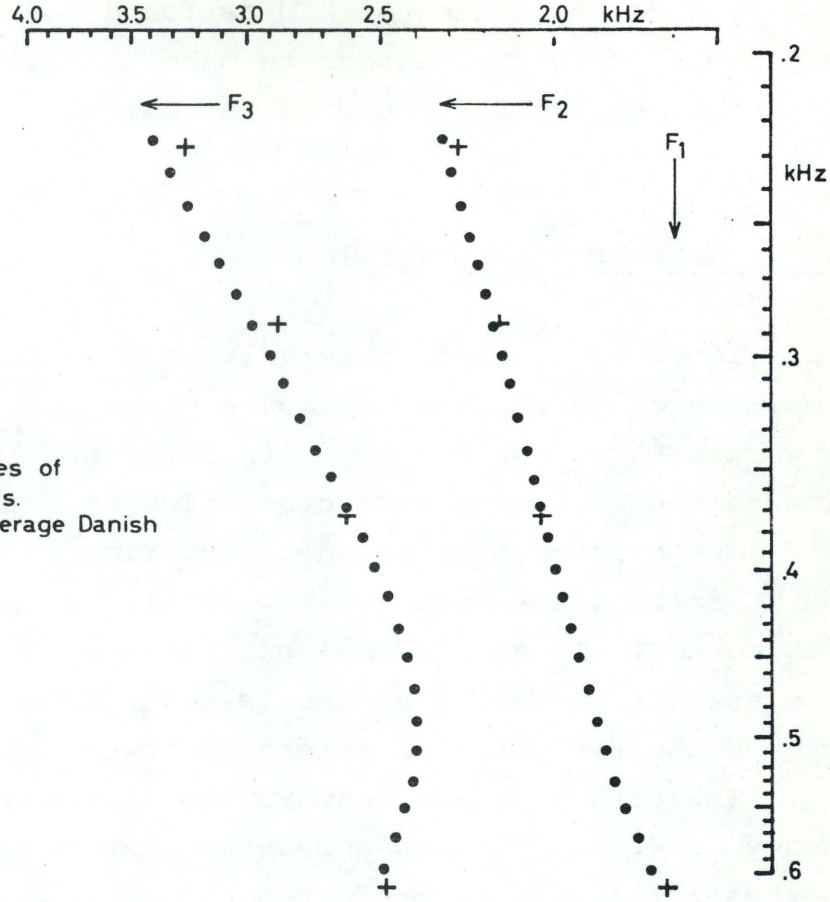


Figure 1a

Formant frequencies of 25 synthetic vowels. Crosses indicate average Danish vowel positions.

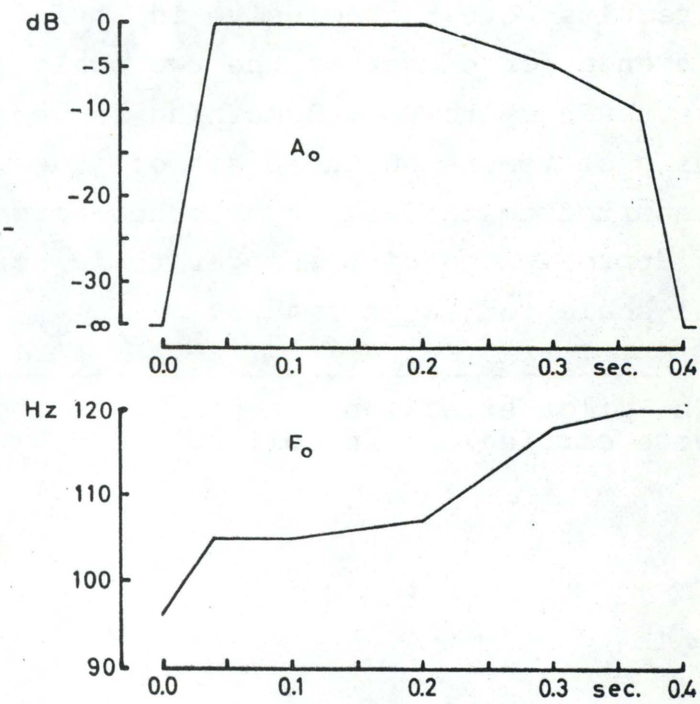


Figure 1b

Variations in total amplitude (A_o) and fundamental frequency (F_o)

of very close varieties of [i] as found in Swedish and Danish, and it was excluded in order to make the stimulus material applicable to a wider range of languages.

2.1.1. The problem of scale

Equal logarithmic intervals were chosen rather than e.g. a mel-scale because the log-scale gives convenient figures to work with in conjunction with a reasonably close approximation to the frequency characteristics of the ear.

The mel-scale is based on experiments with simple sounds and it still seems doubtful how far it is applicable to complex sounds. In an informal experiment,[1] a group of listeners were asked to adjust a synthetic vowel to a quality halfway between the two pairs of reference vowels [œ] - [ɔ] and [ɑ] - [a]. The reference and test vowels differed only in the frequency of F2, and it was hoped that the experiment would show whether the listeners preferred a logarithmic or a mel-scale. The results were inconclusive in so far as they revealed no preference for either of the two scales but showed rather a vague tendency towards something between the two. The patterning of vowels on the basis of judgements of similarity, as reported tentatively by Fischer-Jørgensen (1970-71), are best approximated with a logarithmic scale rather than with a mel-scale for the formants.

---

1) The pilot experiments reported in sections 2.1.1. and 2.1.2. were carried out in collaboration with Birgit Hutters.

TABLE 1

Control parameters of 25 synthetic vowels

(a) Fixed parameters:

Bandwidths: B1 = 50 Hz

B2 = 65 Hz

B3 = 100 Hz

B4 = 120 Hz

F5 = 4400 Hz

L5 = -25 dB

(b) Variable parameters:

Vowel No.	F1	F2 (L2)	F3 (L3)	F4 (L4)
1	226	2326 -26.2	3391 -19.2	3800 -20.2
2	235	2302 -25.7	3320 -20.3	3800 -21.6
3	245	2279 -25.0	3249 -21.7	3800 -22.6
4	255	2255 -24.2	3178 -21.7	3800 -23.6
5	266	2231 -23.4	3108 -22.1	3800 -24.3
6	277	2207 -22.6	3039 -22.3	3800 -25.1
7	288	2183 -21.8	2971 -22.4	3800 -25.7
8	300	2159 -20.9	2905 -22.4	3800 -26.2
9	313	2135 -20.1	2841 -22.2	3800 -26.7

TABLE 1

(continued)

Vowel No.	F1	F2 (L2)	F3 (L3)	F4 (L4)
10	326	2112 -19.2	2779 -21.9	3800 -27.1
11	339	2088 -18.3	2720 -21.6	3800 -27.4
12	353	2066 -17.4	2665 -21.1	3800 -27.7
13	368	2043 -16.5	2613 -20.6	3800 -27.9
14	383	2021 -15.7	2566 -20.1	3800 -28.1
15	399	2000 -14.8	2524 -19.5	3800 -28.1
16	416	1979 -14.0	2487 -18.9	3800 -28.1
17	433	1958 -13.2	2457 -18.4	3800 -28.0
18	451	1937 -12.7	2433 -18.0	3800 -28.0
19	470	1917 -12.0	2416 -17.6	3800 -27.7
20	489	1896 -11.5	2407 -17.2	3800 -27.4
21	509	1874 -11.2	2405 -17.1	3800 -27.0
22	530	1850 -10.9	2411 -17.1	3800 -26.6
23	552	1825 -10.7	2426 -17.1	3800 -26.1
24	575	1796 -10.5	2447 -17.3	3800 -25.6
25	599	1763 -10.2	2476 -17.9	3800 -24.6

2.1.2. Inter stimulus differences

As a pilot experiment the stimuli of Stevens, Liberman, Studdert-Kennedy and Öhman (1969) had been tried in a discrimination test with Danish listeners. The test was a 4IAX test where the listeners are asked which of two pairs, AA or AB, are different. According to Pisoni (1971) this type of test yields a higher proportion of correct discriminations than the traditional ABX test. This was confirmed in our experiment where the discrimination score was almost 100 pct correct.

The results of this pilot experiment clearly indicated that if the auditory distance between the individual vowel stimuli were to be of the same order of magnitude as the distances between the consonant stimuli of other experiments, the physical differences between the vowels must be very small. In the end the frequency difference between F1 of the individual stimuli was set approximately equal to 4 pct, which turned out to be rather close to a just noticeable difference. This small distance ensured that even comparatively efficient test forms could be employed. It was found that twenty-five vowels with a 4 pct difference in F1 would cover the range from /i:/ to /a:/. This number is quite high but necessary if the whole range is to be covered in one test.

2.2. Synthesis

The speech synthesizer of the institute is of the parallel type in which not only the formant frequencies but also the formant levels must be specified separately. In order to make sure partly that the synthetic vowels were close approximations to natural vowels and partly that there was a continuous transition from one stimulus to the next, the levels of the formants relative to the first formant were precalculated after the formulae of Fant (1956). The results of these calculations are listed in table 1.

The source spectrum of the parallel synthesizer is essentially flat in order to allow for independent variation of the individual formant levels. The glottal spectrum of a normal voice was therefore simulated by making appropriate adjustments in the precalculated formant levels. The simulated glottal source had a fall of 14 dB per octave rather than the usual 12 dB per octave, since this appears to be a better approximation to the natural voice. In the frequency region below F1 of any given vowel the glottal spectrum was shaped by a suitably chosen subformant (which is of course no true subformant).

The synthetic vowels were made with a total duration of 400 msec and had a rising pitch as shown in fig. 1b.

The changing fundamental frequency in conjunction with the relatively narrow formant band widths caused some trouble. Thus in some cases the movements of the harmonics through the formants gave the impression of a slight diphthongization which could possibly make these vowels more easily distinguishable than others. This difficulty was taken care of by synthesizing five series of 25 vowels with slightly different F_o-contours. For the production of the test tapes was chosen that series of 25 vowels which on the whole was the most uniform. The chosen series was, however, not perfect in all respects. Thus a slight irregularity in the middle of stimulus no. 4 was most annoying. More will be said about this later.

3. Procedure

3.1. The identification test

The 25 test vowels were arranged in two different quasi random orders. Care was taken to avoid having two vowels of fairly similar quality follow immediately after each other. In order to further diminish any possible context effect a flute signal was played immediately before the presentation of a new vowel stimulus.

Every stimulus was played twice with a pause of about one second between followed by a pause of about four seconds. No numbers or other identification marks were spoken on the test tape, but the flute signal between every fifth and sixth vowel stimulus was of a different character. Before the test tape proper were put five dummy stimuli which were not counted in the scoring. The tape contained a total of 55 stimulus "units" and lasted twelve minutes.

The copying from master tape to test tape was done on semiprofessional Revox tape recorders. The test was presented to the listeners from the same recorders over earphones (AKG, type K58).

The listeners were four trained phoneticians (one male), two first year phonetics students (female), and one (male) technician. They were all speakers of Standard Danish with no known hearing defects. The subjects were asked to identify the vowels with the Danish long vowel phonemes /i:/, /e:/, /ɛ:/, /aː/ as they appear in the words "mile, mele, mæle, male" ['mi:lə, 'me:lə,'mɛ:lə 'mæ:lə]. The subjects noted their identifications on specially prepared answer sheets. Every subject listened ten times to the test, thus giving a total of twenty identifications per stimulus for every person.

3.2. The discrimination test

The testform chosen was the so-called AX-type where two vowels are presented to the listeners who are asked to judge whether the test items are the same or different. According to Pisoni (1971), who compares various test-procedures, listeners could be expected to respond to smaller stimulus differences with the AX-test than with any other procedures. This of course was what the experiment was intended to reveal. One major difficulty was, however, that with this procedure it is quite difficult to control what differences the listeners are detecting, i.e. any accidental physical peculiarities in the test tape will be given undue prominence. This did cause some problems as will become apparent later.

A test tape was prepared in which each vowel was paired (a) once with each of the vowels one and two steps removed and (b) twice with itself. The stimulus pairs were randomized according to the same principles as in the identification test. In the beginning of the tape five dummy stimulus pairs were recorded, giving a total of 105 pairs in the test. There was a pause of about one second between the two vowels in a pair, and the stimulus pairs were separated by flute signals in a manner similar to that employed in the identification test.

Six listeners (identical with those who took the identification test minus one (male) phonetician) wrote their answers on specially prepared answer sheets. The test was rather long (23 minutes) and the listeners were therefore allowed to pause at a certain point in the middle of the test. Every subject listened ten times to the test, thus giving a total of ten judgments on each one and each two step discrimination pair.

In their judgments of differences the listeners were asked to choose between four categories: (++) "certain they are different", (+) "don't know, believe they are different", (-) "don't know, believe they are the same", and (--) "certain they are the same". As it turned out the test was quite difficult which gave some bias towards the minus side, the range of answers was, however, sufficiently wide for all four points to be used in the calculations.

For each discrimination pair was calculated the area under the ROC-curve, P(A). In this case the ROC-curve is the function which relates the probability of a given answer if the vowels were different to the probability of the same answer if the vowels were identical.

This means that for any pair of vowels, AB, the score of correct discriminations is corrected for "false alarms", i.e. judgements "different", to the control pairs AA or BB. If the correction was not included in the scorings of the AX-test the best discrimination score would be found with the listener who simply answered "different" to all the vowel pairs. When the correction is included this strategy of the listener will yield a P(A) value of 0.5, exactly the same as one would get by answering at random. In order to obtain maximum P(A) = 1.0 the listener must not only discriminate correctly between all AB pairs but must also answer "same" to all AA or BB pairs. (See e.g. McNicol (1971) or Robinson and Watson (1972).)

In view of the very limited number of judgements on each pair the non-parametric measure P(A) was chosen rather than the more generally used d'.

4. Results

In fig. 2 are shown the individual results from the two listening tests for each of the seven listeners. In fig. 3 are shown the mean results for the seven listeners. Through the points of this identification curve have been drawn least squares approximations to the best fitting normal ogives. And in table 2 are listed the corresponding 50 pct cross-over points and the standard deviations of the estimated normal distributions.

TABLE 2

Average identification curves. The table shows 50 pct cross-over points (L) relative to stimulus numbers and standard deviation of the corresponding normal distribution (s) expressed in stimulus steps.

	L	s
/i/-/e/	6.85	1.33
/e/-/ε/	12.77	0.97
/ε/-/a/	18.96	0.44

4.1. The identification test

It is interesting to note that the dispersion of the identification scores is markedly smaller than found in earlier identification experiments, e.g. Stevens et al. (1969) or Pisoni (1971). This might of course be due to the choice of listeners, although there was no systematic difference

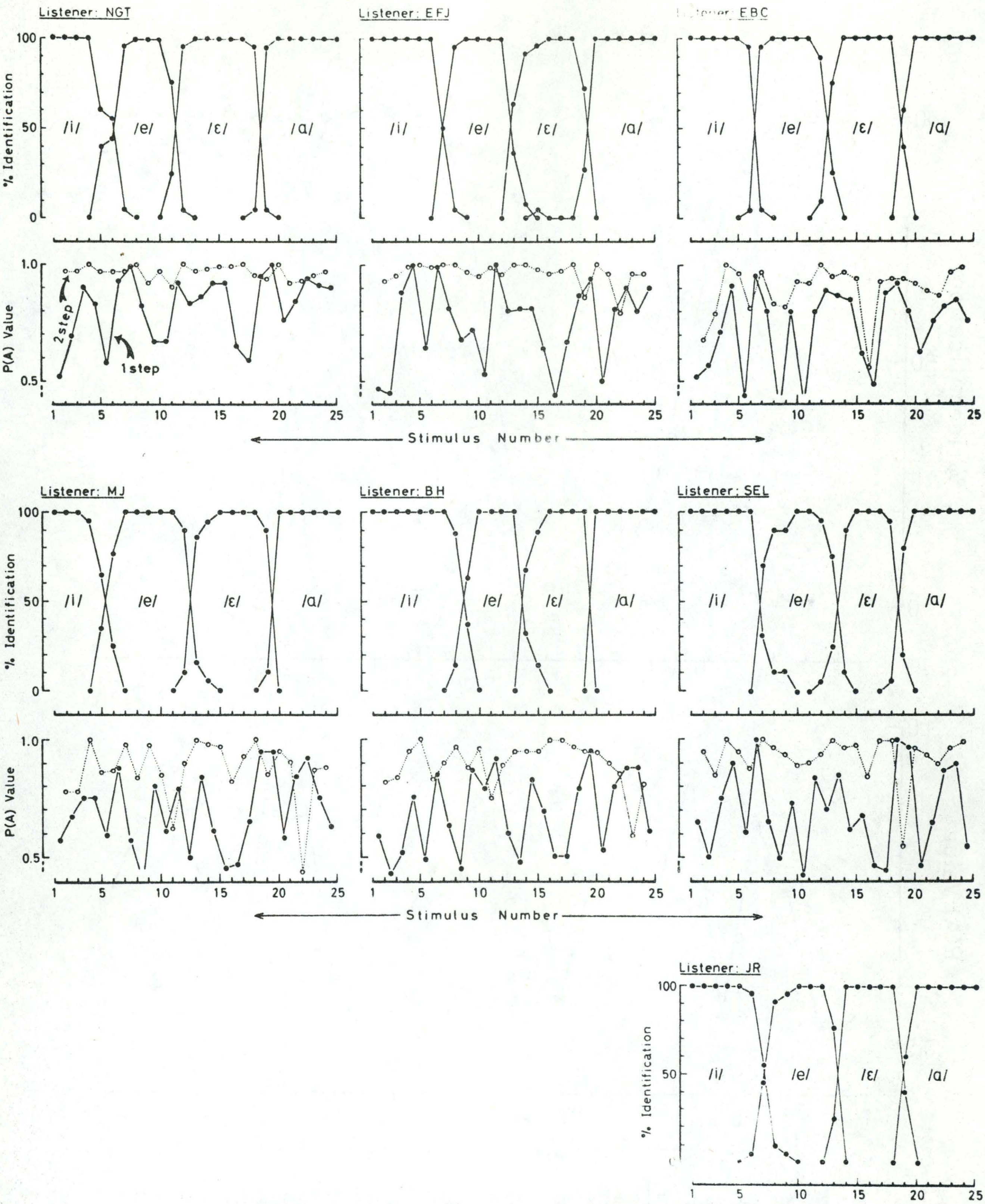


Fig. 2 Individual results from identification and discrimination tests for seven listeners. Listener JR did not take the discrimination test.

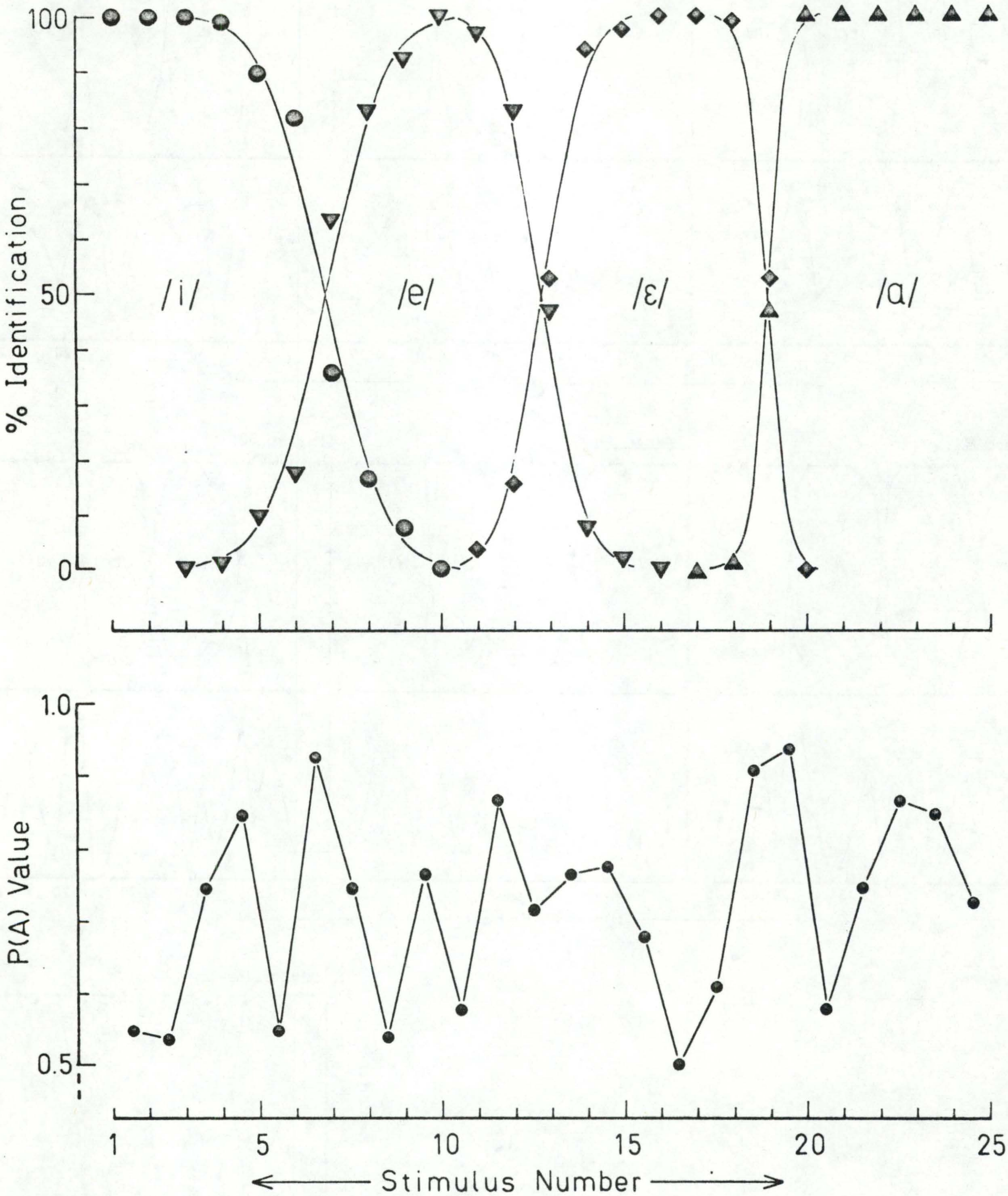


Fig. 3 Mean identification (top) and discrimination results (bottom) for seven listeners.

between the answers from the trained phoneticians and the other listeners. Another explanation could be that the stimuli of this experiment have sounded slightly more natural than the stimuli used with English speaking listeners. Thus in the experiments with English the difference in vowel length has generally been disregarded, although the difference in quality is always accompanied by a difference in duration. Therefore the short vowels /i/ and /ε/ may have been unnaturally long for their qualities and this may have disturbed the English listeners. In this respect Danish is an easier background to work with, since all the front vowels occur both long and short with the same qualities (Fischer-Jørgensen (1972)). The only exception is /a:/ which is a front vowel, while /a/ has a more retracted quality. The possible influence from this will be mentioned later.

In fig. 4 the 50 pct cross over points of the seven listeners are shown as lines in the F1/F2-F3 plot of the vowel stimuli. As in fig. 1 the crosses indicate average Danish vowel formant frequencies. Fig. 4 shows clearly that the listeners are in reasonably good agreement on the placing of their phoneme areas. Further it is evident that the phoneme areas of the identification experiment do not correspond 100 pct to the phoneme areas established on the basis of spectrographic measurements. Thus the border between identified /i:/ and /e:/ falls exactly in the middle of the measured /e:/-area. Similarly the /e:/-/ε:/ border coincides with the measured /ε:/-centre. Only the border between /ε:/ and /a:/ is situated in the expected area.

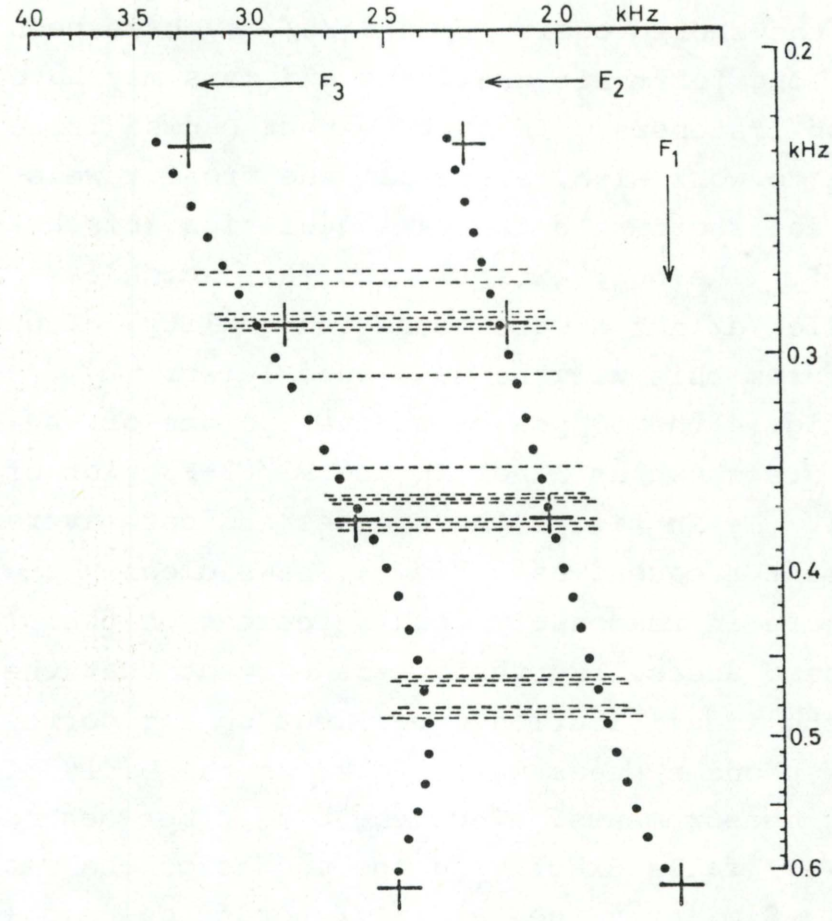


Fig. 4 Comparison between mean formant frequencies of natural vowels and identification of synthetic vowels. The broken lines in the F1/F2-F3 plot indicate 50 pct cross-over points of seven listeners. The crosses indicate average formant frequencies of Danish /i:/, /e:/, /ε:/, and /a:/.

4.1.1. The deviation of identified areas from measured areas

There are several possible explanations to this discrepancy between auditory and acoustic results. One would be that the listeners (of which the majority were women) are influenced by their own formant frequencies in their interpretations of the acoustic signal. This does not sound particularly likely since in natural speech people will interpret correctly almost any sort of voice no matter what their own voice is like. However, in order to investigate this point spectrograms were made of the four vowels in question as spoken by five of the listeners. The results of the measurements are shown in fig. 5. together with the 50 pct cross over points of the same persons.

For two persons, NGT and EFJ, the first formants of the identified and measured areas are in good agreement, for the remaining three persons the discrepancies are at least as great as between identified areas and average formant frequencies. Besides, the comparison of formant frequencies is, strictly speaking, only meaningful with the two male listeners SEL and JR, since the series of synthetic vowels do not pass through the phoneme areas of the female listeners. The fact that both male listeners have identified a series of synthetic vowels as /e:/ which in their own pronunciations would have been divided between /e:/ and /ɛ:/ and on the whole show the greatest discrepancies between identified and measured areas clearly refutes the hypothesis that identification should be affected by the listener's own formant positions.

Another possible explanation was of course a systematic bias in the formant frequency measurements. This seemed a likely hypothesis since the discrepancy between measured and identified vowels is greatest among the close vowels which are

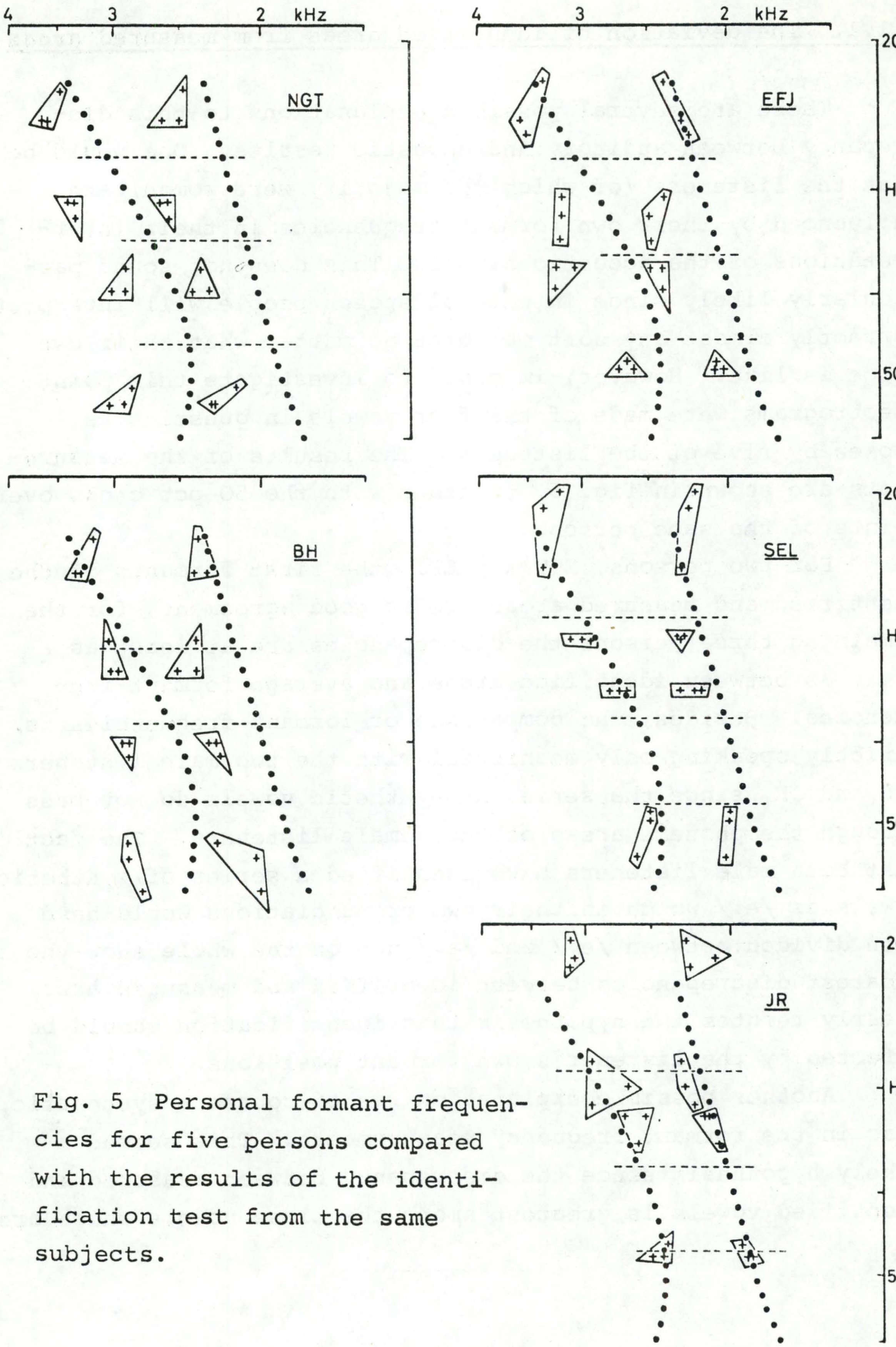


Fig. 5 Personal formant frequencies for five persons compared with the results of the identification test from the same subjects.

known to be difficult to measure accurately. In order to test this hypothesis Professor Eli Fischer-Jørgensen, who had measured the spectrograms on which the synthetic vowels were based, kindly consented to measure the formant frequencies of the stimulus vowels. The deviations of these measurements from the original formant frequencies were very small indeed and showed no systematic variations which could have caused the discrepancy between acoustic and auditory data.

Similar differences between formant measurements and identification tests can be seen in the results of Fry, Abramson, Eimas and Liberman (1962) where the border between English /I/ and /e/ in the identification test is found at about 500 Hz (Holtse (1972)). J. Rischel (unpublished) has found an identification border between Danish /i/ and /e/ at 270 Hz and between /e/ and /ε/ at 375 Hz. These findings are in good agreement with the results of the present experiment and equally hard to explain.

The apparent shift in the identification may have been caused by the balance of the frequency spectrum of the synthetic vowels. Thus too much energy in the frequency region below F1 would have the effect of shifting the perceived centre downwards, especially with low first formants. Another possibility, suggested to me by E. Fischer-Jørgensen, is that the listeners simply divide the whole series of stimuli into four parts. These problems will be looked into some time.

4.2. The discrimination test

The individual P(A) values for the six listeners are found in fig. 2. And the average values are shown in fig. 3. Generally speaking the two-step discrimination is good with a P(A) very close to 1.0 in most cases. There are no marked peaks to be seen, but some listeners have a few unmotivated valleys. However, these valleys are found in different places with different listeners. And except for one case (EBC) they do not correspond to valleys in the one-step discrimination. It seems likely therefore that these valleys are artifacts of the test. Because of the very high discrimination score of the two-step test this part of the results has been left out in the following discussion.

The one-step discriminations are more varied but also rather difficult to interpret. The curves show both peaks and valleys but one very disturbing observation is that the agreement seems to be better between the discrimination functions of the six listeners taken together than between the discrimination and identification functions of each individual listener. This may be due either to unnoticed inaccuracies in the test tape (close examination of the tape has failed to reveal any such errors) or to some inherent universal constraint as suggested by Stevens et al. (1969). The material of the present experiment is, however, limited and further experiments appear to be needed before this question can be answered, although the former solution seems the more probable. Especially since the peaks and valleys are not found in the same places in the two experiments. One possible source of error would be some sort of context effect in the test tape. Obviously the experiment should be repeated with several different orders of test items.

4.2.1. Learning effect

Two listeners were examined for learning effects in the discrimination test. This was done by calculating the average P(A) value of all the stimulus pairs from each listening session. The results of these calculations are listed in table 3.

TABLE 3

Average P(A) values from individual listening sessions for two listeners.

Listening session	Listener EFJ	NGT
1	0.71	0.81
2	0.80	
3	0.71	0.75
4	0.68	
5	0.72	0.86
6	0.72	
7	0.72	0.81
8	0.79	
9	0.80	0.83
10	0.82	

No systematic trend is revealed from table 3 but the variation in the average P(A) values is an indication of the maximal amount of reliability to be expected from the individual discriminations.

5. Discussion

5.1. The connection between discrimination and identification

In the average discrimination curves of fig. 3 there is a marked peak about stimuli 18-19. This peak corresponds well with the border between identified /ɛ/ and identified /a/. The peak is quite steep corresponding to the good agreement in the identification.

Between /i:/ and /e:/ there is another peak in the discrimination function: stimuli 6-7-8. But the picture is complicated by an extra peak between stimuli 3-4-5. However, as previously mentioned there was a small irregularity in stimulus No. 4. This irregularity has made No. 4 so easily recognizable that it would produce a peak in the discrimination.

There is no evident peak in the discrimination between /e:/ and /ɛ:/ but the discrimination is generally good from stimulus 9 to 15. Similarly the discrimination is rather poor in the middle of the phoneme areas in several cases, approaching P(A) = 0.5. However, we are left with a few unexplained peaks, e.g. between stimuli 9 and 10, and between 22 and 25. There appears to be no explanation for the peak between 9 and 10 which again may have been caused by external disturbances. The peak between 22 and 25 is situated in the area where the [a]-quality changes from that associated with long /a:/ to the quality of short /a/. Therefore this particular peak may well have been caused by influence from the identification system. Because of the questions asked in the identification test this peculiarity does not show in the identification.

5.1.1. The number of identifiable stimuli

In this connection it should be noted that the theory of categorical perception predicts that it is impossible to discriminate between qualities which cannot be identified in isolation. It has, however, always been tacitly assumed that only phonemes could be identified absolutely. This has never been proved. On the contrary it is a common observation that, at least for vowels, people will identify several different variants of the same phoneme, especially if these variants serve some social function.

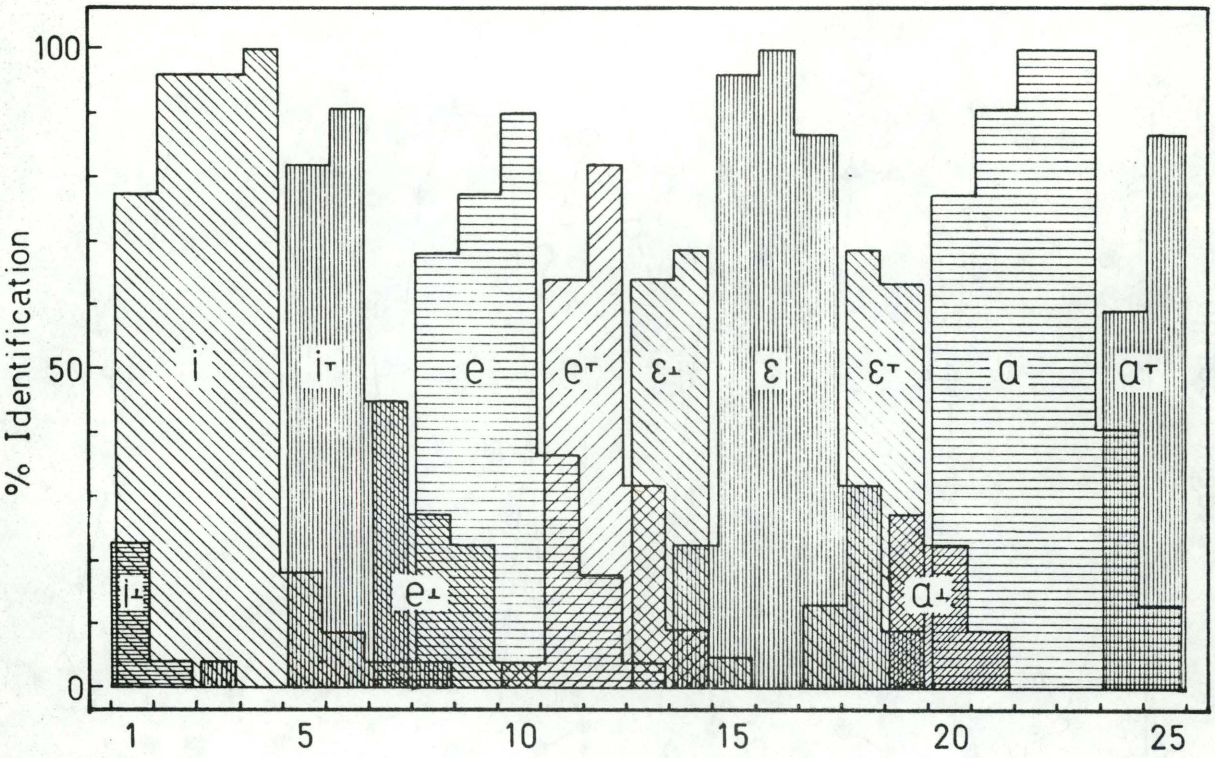
As a very preliminary investigation of this problem one listener (a trained phonetician) identified both a raised and a lowered quality of each of the four vowel phonemes. The results of this identification experiment are shown in fig. 6a.

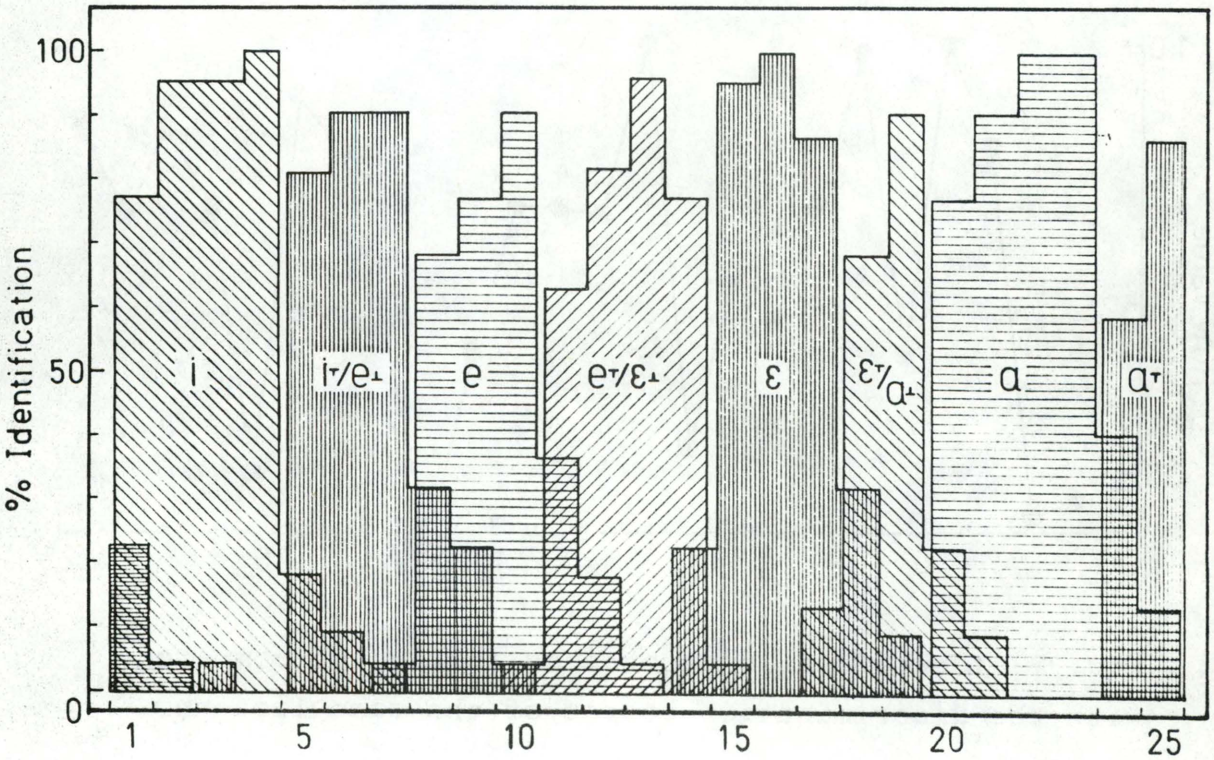
On the whole the identification of twelve different categories appears to have been quite difficult, and some categories have not been much used. Thus category [e˔] has a maximal identification score of 45.0 pct (stimulus 7), while [a˔] only reaches 27.2 pct (stimulus 19). The other between-phoneme categories all exceed 50 pct identification in their maxima but [ɛ˕] and [ɛ˔] do not reach 75 pct. (Category [i˔] is not included since the low score obtained was caused by inconsistent use of this category. The other extreme quality [a˕] was more frequently used, suggesting that, for this listener at least, stimuli 24 and 25 could well have been dispensed with in the original identification test.)

There appears to be a general tendency to use the raised category less often than the lowered category of the neighbouring phoneme, cf. e.g. [i˕] and [e˔]. In fig. 6b the neighbouring raised and lowered categories are combined. This results in a series of homogeneous identification categories all of which have maxima of at least 90 pct identification. This suggests that the listener was unable to identify consistently more than one category between phonemes. This observation may have implications for fine phonetic transcription but the problem is very much in need of further investigation. The possible effect of extending or reducing the range of stimuli should for instance be examined.

A comparison between the maximal identification function and the discrimination function (fig. 7) reveals little connection between these two functions. This is another point in need of investigation.

Fig. 6 a (top): Identification scores for twelve vowel categories, i.e. raised ($\perp$) and lowered ($\top$) and normal variants of Danish /iː/, /eː/, /ɛː/, and /aː/.
b (bottom): Combination of neighbouring raised and lowered categories of fig. 6a.


100
50
0
% Identification
i
iᵀ
e
eᵀ
ε⊥
ε
εᵀ
a
aᵀ
i⊥
e⊥
a⊥
1
5
10
15
20
25


100
50
0
% Identification
i
iᵀ/e⊥
e
eᵀ/ε⊥
ε
εᵀ/a⊥
a
aᵀ
1
5
10
15
20
25

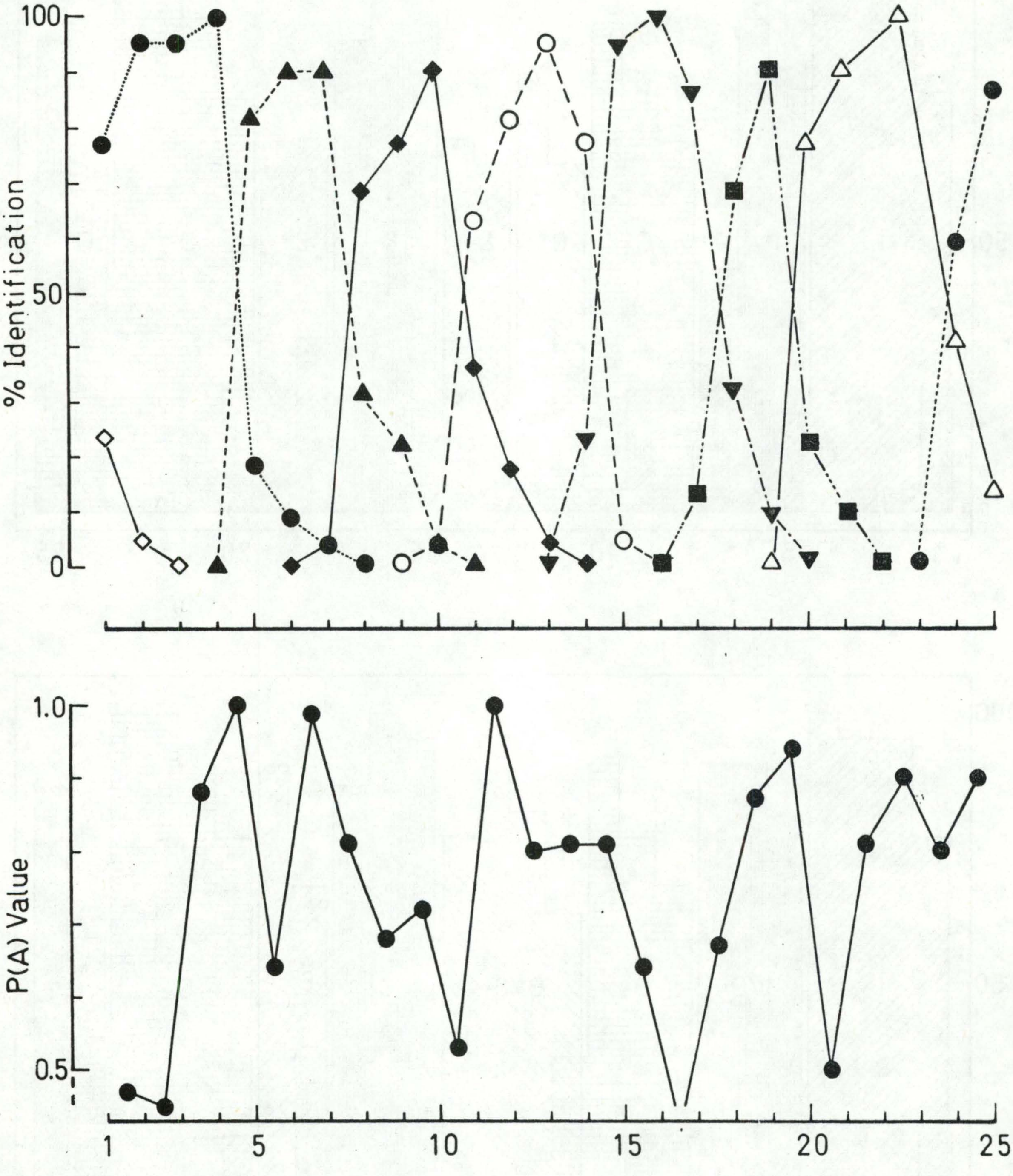


Fig. 7 Identification scores as in fig. 6b compared with the discrimination scores of the same listener.

5.2. Categorical perception

One interesting observation is that the valleys of the discrimination curve of fig. 4 are quite deep. In this respect the results of the present experiment differ from earlier discrimination experiments. Thus in the cases where the ABX-test had been used the percentage "correct discrimination" was usually about 70 for the valleys of vowels and considerably lower for the valleys of the consonants. The theory was therefore advanced that only consonants are perceived categorically. Pisoni (1971) quotes a theory by Sawashima and Fujusaki which suggests that the question whether a given sound continuum shows signs of categorical or continuous perception is determined by the way the incoming information is stored and processed by the brain. Thus most of the acoustic signal is kept for some time in a sort of temporary analogue storage. Some of the information, more particularly the information concerning consonants, is immediately converted to categorical information and only stored in this form. Now, in a discrimination task the listener will, unless he is forced to do otherwise, compare the stored signals which are closest to the original acoustic signal. In the case of consonants the closest approximation appears to be the categorized signal, which of course means that the results of a discrimination test will show signs of categorical perception. In the case of vowels the closest approximation is apparently the signal kept in temporary analogue storage. And the discrimination function will only be determined by the general difference limen for quality differences (if any such thing exists).

In order to make the discrimination of vowels show signs of categorical perception it would be necessary to block admission to the analogue storage. This seems to be what happens when vowels are put in context, as shown by Stevens

(1968). Another way would be to make the difference to be examined so small that even the signal stored in analogue fails to reveal this difference. In this case it will only be possible to detect a difference if the two vowels to be compared have been given different representations in the categorized storage.

The latter procedure is the one adopted for the present experiment. The procedure was successful in so far as there were tops and valleys in the discrimination function. But the expected picture of categorical perception was seriously contaminated by minute physical irregularities in the test tape. In the suggested model any such irregularity will of course be available in the analogue signal and thus produce a high discrimination without passing through the categorized storage. It appears therefore that for the procedure of the present experiment to be successful a very high precision is required in the synthesized material. At present it is difficult to distinguish categorization from minor accidents in the synthesis process.

References

Fant, G. 1956: "On the predictability of formant levels and spectrum envelopes from formant frequencies" FRJ p. 109-120

Fischer-Jørgensen, E. 1970-71: Resume af forelæsninger over psykoakustik og auditiv fonetik, (Abstract of lectures on psychoacoustics and auditory phonetics) Institute of Phonetics, Copenhagen (mimeographed)

Fischer-Jørgensen, E. 1972: "Formant frequencies of long and short Danish vowels" Studies for Einar Haugen, The Hague, p. 189-213

Fry, D.B., A.S. Abramson, P.D. Eimas, and A.M. Liberman 1962: "The identification and discrimination of synthetic vowels" LS 5, p. 171-189

Frøkjær-Jensen, B. 1967: "Statistic calculations of formant data" ARIPUC 2, p. 158-169

Fujisaki, H. 1971: "A model for the mechanisms for identification and discrimination of speech sounds" Proc. Acoust. 9, p. 56-59

Guilford, J.P. 1954: Psychometric Methods, London

Holtse, P. 1972: "Spectrographic analysis of English vowels" ARIPUC 6, p. 1-48

Hutters, B. and P. Holtse 1972: "On universals in vowel perception" ARIPUC 6, p. 177-184

Lane, H.L. 1965: "The motor theory of speech perception: A critical review" Psychological Review 72, p. 275-309

Liberman, A.M., K.S. Harris, H.S. Hoffman, and B.C. Griffith 1957: "The discrimination of speech sounds within and across phoneme boundaries" J. Exp. Psychol. 54, p. 358-368

Mattingly, I.G., A.M. Liberman, A.K. Syrdal, and T. Halwes 1970: "Discrimination in speech and non-speech modes", Haskins SR 21/22, p. 99-131

McNicol, D. 1970: A primer of signal detection theory London

Pisoni, D.B. 1971: "On the nature of categorical perception of speech sounds", Supplement to Haskins SR

Robinson, D.E. and C.S. Watson 1972: "Psychophysical methods in modern psychoacoustics", Foundations of modern auditory theory vol. II (J.V. Tobial ed.), p. 99-131

Stevens, K.N. 1968: "On the relations between speech movement and speech perception", Zs.f.Ph. 21, p. 102-106

Stevens, K.N. A.M. Liberman, M. Studdert-Kennedy, and S.E.G. Öhman 1969: "Cross-language study of vowel perception", LS 12, p. 1-23

Studdert-Kennedy, M., A.M. Liberman, K.S. Harris, and F.S. Cooper 1970: "Motor theory of speech perception: A reply to Lanes's critical review", Psychol. Review 77, p. 324-249