IDENTIFICATION AND DISCRIMINATION OF VOWEL DURATION

Niels Reinholt Petersen

Abstract: The identification and discrimination of vowel duration was investigated. The experimental results could not be unambiguously interpreted in favour of either categorial or continuous perception of the acoustic variable under study.

Furthermore, the subjects' response bias in the discrimination test was examined. It turned out that the number of "false alarms" (i.e. "different" responses to pairs of physically identical stimuli) varied in a systematic manner, being considerably higher near the phoneme boundary (as established by identification tests) than within the phoneme areas. It is attempted to explain the systematic variation of response bias in terms of Fujisaki and Kawashima's model of the decision process in discrimination tasks.

1. Introduction

It is commonly reported that consonants and vowels are perceived differently by human listeners. In a great number of experiments on the identification and discrimination of speech sounds it has been demonstrated that stop consonants in particular are perceived in a categorial manner, i.e. listeners can discriminate between different sounds only if they can identify them as belonging to different phonetic (or phonemic) categories. Vowels, on the other hand, seem to be perceived in a more continuous manner, similar to that of non-speech stimuli, i.e. listeners can discriminate also between sounds which are identified as belonging to the same category.

Whereas the categorial perception of consonants is fairly well established (see e.g. Abramson and Lisker 1968, Liberman, Harris, Hoffman, and Griffith 1957, Liberman, Harris, Kinney, and

Lane 1961, Liberman, Harris, Eimas, Lisker, and Bastian 1961, Pisoni 1971), the perception of vowels seems to be more susceptible to influence from experimental conditions. If vowels are presented in CVC syllables (Stevens 1968) or are followed by another vowel (Fujisaki and Kawashima 1968, Pisoni 1973) they tend to be perceived more categorially than isolated vowels. Similarly, short vowels seem to be less continuously perceived than long vowels (Pisoni 1971 and 1973, Fujisaki and Kawashima 1968).[1]

If the differences between the vowels to be compared are very small, this also seems to evoke a more categorial perception of the vowels (Holtse 1973).

With respect to the influence of test format Pisoni (1971 and 1973), who compared the commonly used ABX procedure to a four interval test of paired similarity (4IAX), found that the perception of vowels was more continuous with the former procedure than with the latter one, whereas the two procedures could not be shown to influence the perception of stop consonants differently.

The different modes of perception of vowels and consonants and the dependence of the perception of vowels on experimental conditions may be accounted for in terms of a model for the discrimination process developed by Fujisaki and Kawashima (here quoted from Pisoni 1971). According to their model the acoustic signal is converted to an auditory representation and a categorized representation in short term memory. When the listener is to decide whether the two stimuli in a pair are same or different he first compares the categorized representations. If they are different, i.e. belonging to different phonetic categories, he answers "different" right away. If they are same, i.e. belonging to the same category, he turns to the auditory representations and compares those. His answer will then depend on the extent to which the auditory representations are preserved in short term memory.

In the case of stop consonants it seems that their auditory representations are lost from short term memory as soon as the

---

1) However, from the discrimination curves given by these authors the tendency towards categorial perception does not appear to be very marked.

categorization has taken place, leaving only categorial information for the discrimination process. Auditory information on vowels, on the other hand, may or may not be available in the discrimination process, depending on experimental conditions: in some cases vowels seem to be perceived in a nearly continuous manner and in other cases in a more categorial manner.

The bulk of experiments on the perception of vowels has dealt with vowel quality, whereas the duration of vowels has been investigated to a very limited extent. Bastian and Abramson (1962) report briefly that Thai vowels of different duration presented to Thai and American listeners were perceived in a continuous fashion. The opposite is reported for Japanese by Fujisaki, Nakamura, and Imoto (1973), who found a clear discrimination peak at the boundary between short and long o.

The results of previous experiments carried out by the present author (Reinholt Petersen 1974) point in the same direction. In those experiments subjects adjusted the duration of the vowels i and ε in the synthetic words ['lV(:)sə] in accordance with the following instructions: a) adjust a "natural duration" of a phonemically short vowel, b) adjust a "natural duration" of a phonemically long vowel, c) reduce the duration of a long vowel until a phonemically short vowel is heard, and d) increase the duration of a short vowel until a phonemically long vowel is heard.

The adjustments made around the phoneme boundaries (i.e. instructions c) and d)) seemed to show a statistical variation which was smaller than the variation of the adjustments of "natural vowel duration". Although this finding may be taken to speak in favour of categorial perception of vowel duration, the evidence is, at best, indirect. It was considered desirable, therefore, to seek further information about the perception of vowel duration by repeating the earlier experiments in the field.

2. Stimuli and procedure

2.1 Stimuli

The vowels chosen were Danish /ɛ/ and /ɛ:/ (IPA [ɛ̞]). In order to avoid the stimuli to be perceived as non-speech sounds the vowels were embedded in the frame [ˈl_sə]. Since the differences between the formant frequencies of short and long ɛ are very small in normal spoken Danish (see e.g. Fischer-Jørgensen 1972), the Danish words /ˈlɛsə/ læsse 'to load' and /ˈlɛ:sə/ læse 'to read' could be generated by merely varying the duration of the vowel.

The words were synthesized on the parallel speech synthesizer of the Institute (see Rischel 1969 and Rischel and Lystlund 1972). The acoustic structure of the synthetic words were based on spectrograms of four Danish speakers. Information about the formant frequencies and levels of the vowel ɛ was kindly provided by Peter Holtse, and correspond to formant data which rendered 100% /ɛ/-identification in an identification test carried out by him (Holtse 1973). The fundamental frequency was rising throughout the stimulus word at a rate of 4 Hz per 100 ms, starting at 92 Hz at the beginning of the [l].

The duration of the vowel was defined as the distance between the programmed change of parameter values from [l] to [ɛ], i.e. at the start of the consonant-vowel transitions, and the programmed cessation of periodic energy in the vowel.

In the course of the experiments it turned out that the actual value of vowel duration may deviate from that specified by the function generator of the speech synthesizer, even if the latter performs correctly. This can be explained as an inherent source of error due to the fashion in which the voice source amplitude is controlled: the point in time at which the sound is programmed to start or a change in amplitude is to occur, i.e. the point at which the voice source amplitude gate is activated to produce an increasing amplitude, is independent of the repetition rate of the voice source pulses. Since the voice source amplitude gate is placed before the formant filters, the first pulse to produce an appreciable excitation of the formant filters or an amplitude change may occur with varying time lags relative to the programmed onset or change, i.e. signals differing up to one period in duration (e.g. 10 ms at F_o = 100 Hz) can be generated by exactly the same programming of the function generator. There is a similar possibility of error at the end of a synthesized vowel.

In the stimuli used in the tests the rise in amplitude from the [ɪ] to the vowel and the fall at the end of the vowel before [s] were generated partly by the voice source amplitude gate and partly by the formant amplitude gates. Since the latter are placed after the formant filters they may, to some extent, reduce the durational errors mentioned above.

In order to investigate the possible distortion of the identification and (particularly) discrimination results the durations of all stimuli were measured on mingographic tracings and compared with the discrimination results. The deviations from the programmed durations were generally very small, less than 2 ms, and they could not be shown to have influenced the shape of the discrimination functions.

2.2 Test procedure

2.2.1 The identification test

In the identification test the vowel duration was varied in 10 ms steps in the range from 100 to 200 ms. An informal pilot test taken by the same subjects who were to participate in the subsequent more formal tests showed that the phonemic boundary between /ε/ and /ε:/ in all cases was well within the limits of the range chosen.

The stimuli were arranged in ten different random orders on the test tape, and were presented one at a time with a response pause of 5 secs. between consecutive stimuli. The test was played back to the listeners via earphones (Sennheiser HD414) on semi-professional REVOX tape recorders. The listeners were asked to note on answer sheets whether they heard the word <u>læsse</u> (short vowel) or the word <u>læse</u> (long vowel). Each listener took the test twice, so that 20 answers were obtained for each stimulus value. There were six subjects, who were all phoneticians. One (NT) was a staff member of the Institute, the others (BH, EBC, MJ, PA, and JJ) were university students of phonetics.

The phoneme boundaries were computed for the subjects individually. The phoneme boundaries were defined as the arithmetic means of the cumulative distributions described by the identification functions (see Guilford 1954 p. 120ff). In contrast to the more commonly used linear interpolation between the data points adjacent to the point of 50 per cent identification this method takes advantage of all information available in the identification functions and in addition, it permits the standard deviation to be computed as a measure of the statistic reliability of the identification results.

2.2.2 The discrimination test

In a discrimination test like the present, where the discrimination is to be measured at various points along a physical continuum, the physical differences between stimuli to be compared should correspond to perceptually equal differences throughout that continuum. Furthermore, they should be of an order of magnitude approximately equal to the differential threshold for the variable in question. If the differences are too small or too large relative to the threshold they may fail to reveal possible discriminatory discontinuities along the continuum.

A survey of the literature on differential thresholds for sound duration did not contribute to the determination of differences between the stimuli in the present experiment, partly because most of the research in the field has been carried out on pure tones or isolated vowels, and partly because of a very great dispersion of the results.

Therefore, instead of attempting to conjecture some function describing the relation between physical and perceptual differences, it was decided to have physically equal steps between the stimuli throughout the 100-200 ms range, and then later to determine the step values in a pilot experiment. (As a matter of fact, this decision caused troubles later on, when the discrimination results were to be interpreted. However, an attempt to keep equal perceptual differences along the continuum might also have caused problems, problems which might not reveal themselves as such in the experimental results.)

In the pilot discrimination test only one reference duration, viz. 125 ms, was used. That duration was compared with test durations of 125 ms (control), 130 ms, 135 ms, 140 ms, and 145 ms.[1] The scores across the five listeners participating (i.e. those mentioned in section 2.2.1 above, minus JJ) appeared to be slightly above chance when the difference between the vowels was 5 ms, and almost 100 per cent correct when the difference was 20 ms. Intermediate score values were obtained for 10 ms and 15 ms discriminations, the scores of the latter being slightly higher than those of the former.

Judging from the results of the pilot experiment, an appropriate difference between stimuli seems to be 10 ms. Since it was considered undesirable to have only one value of interstimulus difference, and since, furthermore, differences of 20 ms might be too large (cp. above), it was decided to include 15 ms interstimulus differences in the discrimination test in addition to the 10 ms ones. Thus vowels of the durations 100, 110, ..., 190 ms were compared to vowels of the same durations (control condition) and to vowels being 10 ms and 15 ms longer.

---

1) The reference duration of 125 ms was chosen because this duration and the ones it was compared with were thought to fall in the vicinity of the phoneme boundary, i.e. in a duration range where discrimination may be expected to be most acute.

The stimuli were arranged for discrimination in a so-called AX test, where they are presented pairwise to the listeners, who are to judge whether the two stimuli in a pair are same or different. The reason for choosing this test format rather than the more frequent ABX format was that the latter procedure seems to make heavy demands on short term memory (see e.g. Harris 1952 and Pisoni 1971), so that what is actually measured is memory rather than discrimination.

On the test tape the 30 AX pairs, viz. 10 AA pairs, 10 AB_{10} pairs (i.e. 10 ms difference between A and B), and 10 AB_{15} pairs (i.e. 15 ms difference between A and B), were arranged in four different random orders. Each AX pair was presented three times in succession followed by a 5 secs. pause for response. Each listener took the test three times, thus giving 12 responses to each of the discrimination pairs.

The subjects, the same as those who took the identification test, were asked to rate the certainty of their responses by choosing between four response categories: (f+) 'I am certain they are different', (f) 'I am uncertain, but I believe they are different', (e) 'I am uncertain, but I believe they are same', and (e+) 'I am certain they are same'.[1] As the test was rather long (about 35 minutes) the listeners were instructed to take a break of at least five minutes in the middle. The initial five stimuli on the test tape were "dummies".

For each discrimination pair the area under the ROC-curve, P(A), was computed as a measure of the discriminability at that point in the continuum. The curve relates the probability of a given response (f+, f, e, e+) to AB pairs to the probability of that same response to the corresponding AA (control) pairs. Since correct discriminations of AB pairs and "false alarms" (i.e. the responses "different" to control pairs) alike are taken into account in the computation of the P(A) value the discrimination can be estimated independently of the listeners' response bias

---

1) The letter f stands for Danish forskellig 'different', and e stands for ens 'same'.

(see e.g. McNicol 1973 and Robinson and Wattson 1972).[1] Due to the limited number of responses the non-parametric measure P(A) was used rather than the parametric measure d' or Δ m.

3. Identification results

Identification functions for short and long ε are shown in fig. 1, and the phoneme boundaries and corresponding standard deviations are given in table 1. The listeners, except for NT, and perhaps JJ, seem to agree quite well on the position of the boundary. With respect to NT, the high value of her phoneme boundary is in agreement with her results in other experiments on the perception of vowel duration (see Reinholt Petersen 1974).

4. Discrimination results

The discrimination functions, i.e. P(A) as a function of vowel duration, for the six listeners are shown in fig. 2. With all subjects there is a pronounced tendency for P(A) to decline as a function of vowel duration. This is, undoubtedly, caused by the physical equidistance between stimuli throughout the continuum. This equidistance, however, cannot fully explain the shapes of the curves. If it were so, monotonically declining functions should be expected, whereas the discrimination functions obtained here deviate from monotonicity by having discrimination peaks in the range 110 to 140 ms, and by showing tendencies to rising P(A) values at the longest vowel durations.

---

1) As a matter of fact, P(A) is not entirely independent of response bias. The P(A) turns out to be slightly lower if the subject is strongly biased towards either extreme of the rating scale than if he distributes his responses more evenly along the scale (see Robinson and Wattson 1972, p. 125).

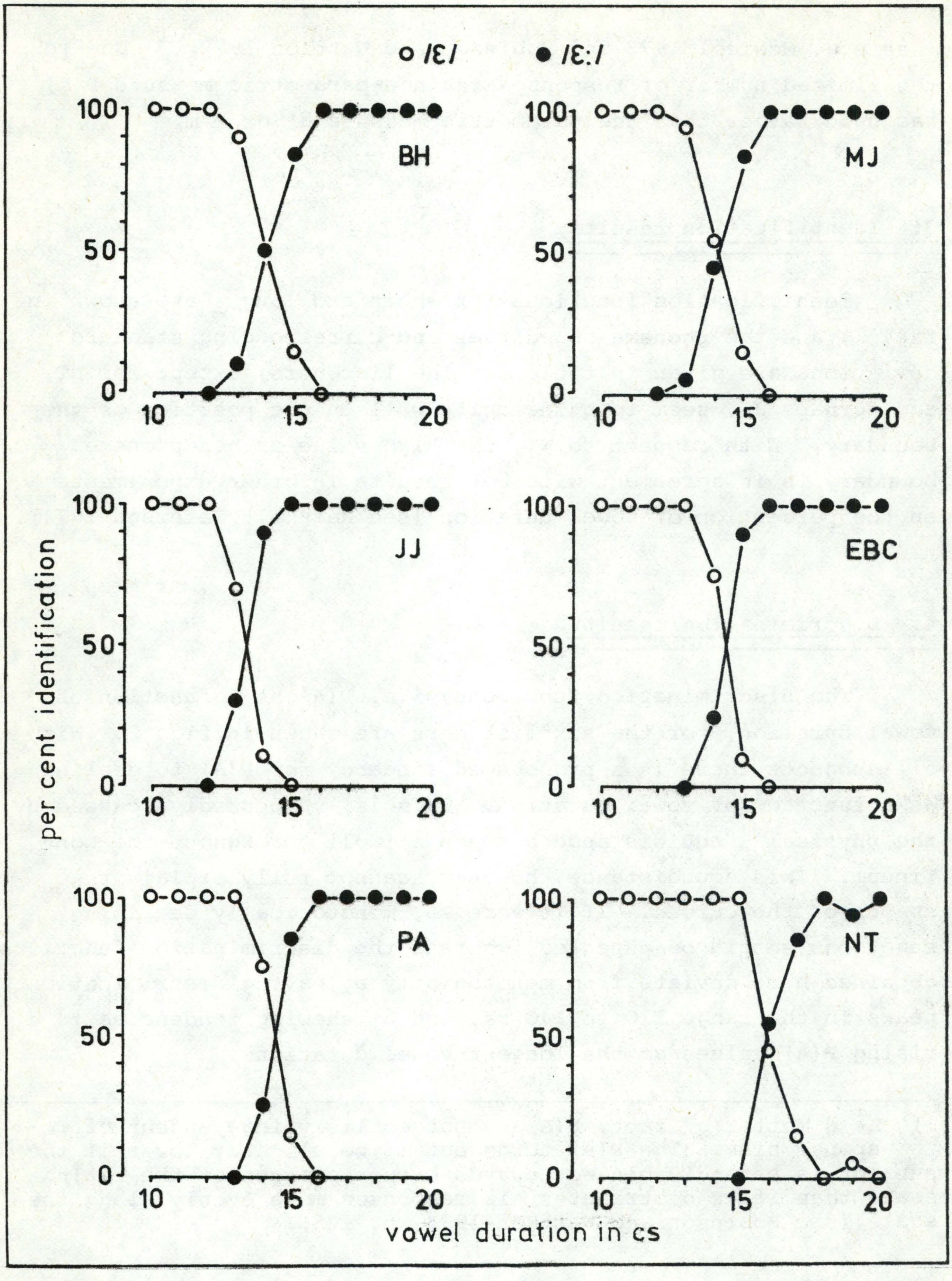

Figure 1

Identification functions for the six listeners.

Table 1

Boundary between /ɛ/ and /ɛ:/ and corresponding standard deviation for each of the six listeners.

subject	boundary	standard deviation
BH	140.5	8.65
JJ	133.0	6.00
PA	144.0	6.41
MJ	141.5	7.92
EBC	143.5	5.72
NT	161.5	10.70

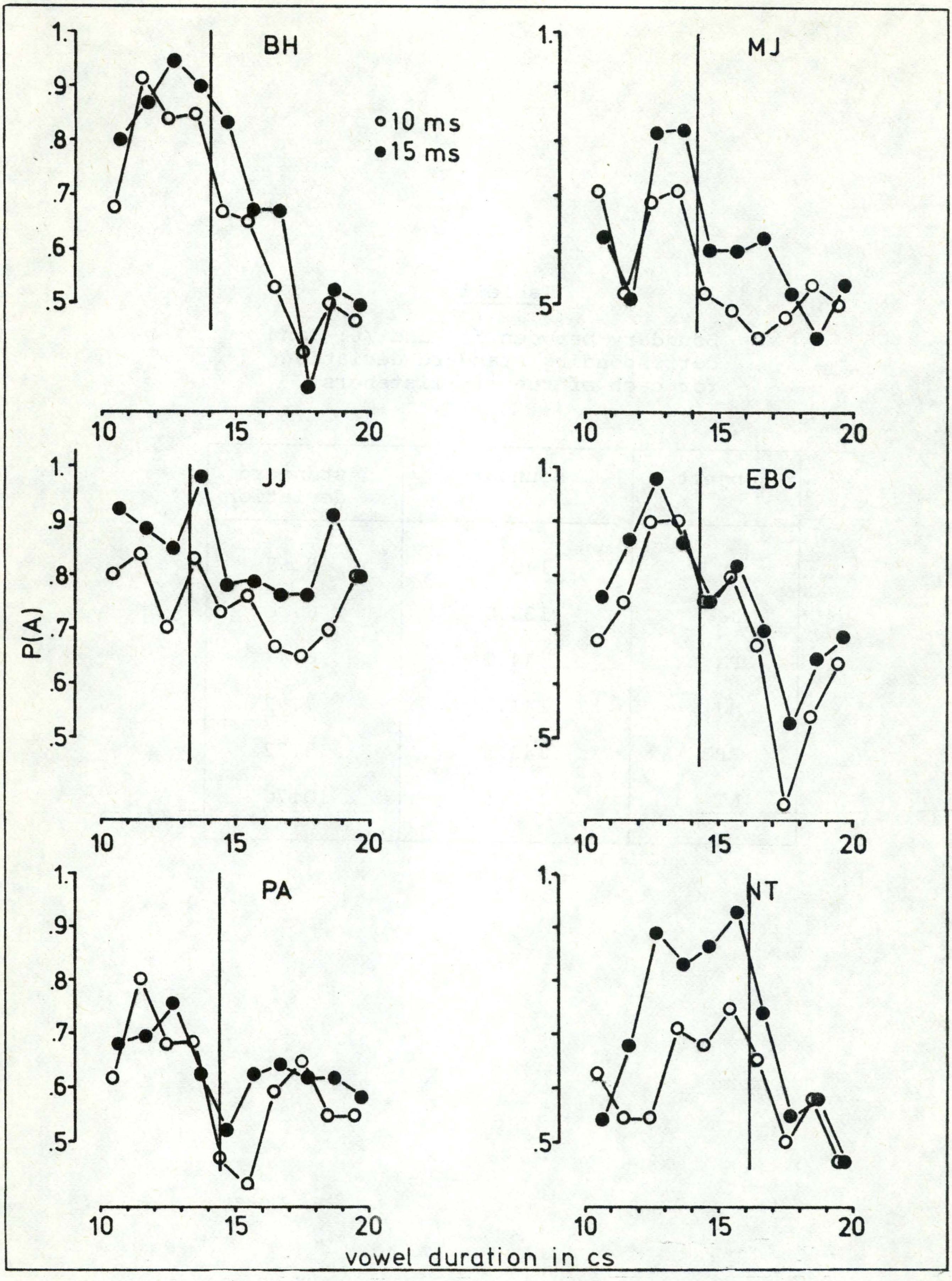

Figure 2
Discrimination curves (P(A) as a function of vowel duration) for the six listeners. The vertical lines in the graphs indicate the phoneme boundaries.

The peaks in the 110-140 ms range may be associated with the boundary between short and long vowel. (This matter will be treated more thoroughly in section 5 below.)

The upward slopes of the discrimination curves at the longest vowel durations may be explained by the possible existence of a third durational category, "unnaturally long". A similar phenomenon has been reported by Liberman, Harris, Eimas, Lisker, and Bastian (1961) for very long p-closures. In order to be tenable, this explanation presupposes, apart from vowel duration being perceived in a categorial fashion, that the range of durations from 150 to 170 ms corresponds to the range of what may be termed natural long vowels. However, in a previous experiment (Reinholt Petersen 1974) one of the tasks of the subjects (the same as those taking part in the present experiments) was to adjust the duration of the vowel /ɛ:/ in the word læse in accordance with a criterion of "naturalness". It appeared that the subjects preferred durations around and above 200 ms for a natural long vowel. On the other hand, the perception of vowel duration is presumably quite sensitive to external factors, such as experimental design, so the results referred to should not be considered to preclude with certainty the possibility of a third durational category.

5. Comparison between identification and discrimination

In the discrimination functions of fig. 2 the phoneme boundaries computed from the identification scores are indicated by vertical lines. In all cases (except for JJ's 15 ms discriminations) the discrimination peaks (defined as the highest P(A)) are found at lower stimulus values than the phoneme boundaries. It is not unusual, however, for discrimination peaks not to correspond exactly to the phoneme boundaries. But if the peaks are to represent a more acute discrimination across the boundary, a random distribution of peaks around the boundary should be expected. With the present results a one-tailed t-test ($H_0: \mu_{boundary} - \mu_{peak} \leq 0$) revealed the displacement of peaks towards lower stimulus values to be systematic ($p < .01$ for 10 ms discriminations and $p < .05$ for 15 ms discriminations).

Two explanations may be hypothesized to account for the displacement of the discrimination peaks relative to the phoneme boundaries:

1) The displacement is due to the use of physically equal differences between stimuli. The discrimination functions obtained may be the result of two factors, viz. an "equidistance effect" which causes the discrimination to decline as a function of stimulus duration, and a "phoneme boundary effect" which causes the discrimination to be sharpest at the phoneme boundary and to decline as a function of the distance to the boundary. Thus, in the range of durations beneath the phoneme boundary the two effects will counteract (the "equidistance effect" causing a falling and the "phoneme boundary effect" a rising discrimination as a function of vowel duration in that range) with the result that they may more or less neutralize each other, whereas at stimulus values above the boundary the two effects will act together to produce a steep decline in the discrimination functions.

If this is tenable in principle, the probability of the highest P(A) occurring at stimulus values below the phoneme boundary will exceed the probability of its occurrence at values above the boundary, i.e. at the steeply declining slope.

On the basis of the obtained discrimination curves it is, of course, not possible to evaluate the contributions of the two effects separately so as to obtain an estimate of the linguistically relevant information in the curves. As it seems pretty certain, however, that the "equidistance effect" will cause a monotonic decline of P(A) values, a normalization of the discrimination curves based on such a monotonic function might give a rough indication of the contribution from the "phoneme boundary effect", the greatest deviations from the function chosen assumedly reflecting the strongest phoneme boundary effect.

The function used for normalization was a straight line fitted to the data points of the discrimination functions by the least squares method. In fig. 3 the normalized curves are shown for 10 and 15 ms discriminations for the individual subjects. It must be emphasized that this normalization procedure is a very crude one, indeed, and theoretically unsatisfactory, too. But, supposedly, it gives at least a hint of the effect to be expected from more sophisticated procedures, preferably to be based on independent data on the influence from equal physical differences.

From fig. 3 it appears that the peaks are still located at stimulus values beneath the phoneme boundaries (indicated by vertical lines in the graphs). A one-tailed t-test ($H_0: \mu_{boundary} - \mu_{peak} \leq 0$) proved the mean difference between boundaries and peaks to be statistically significant for the 10 ms discriminations ($p < .05$) and non-significant for the 15 ms discriminations. Since in the non-normalized functions the differences were significant in both cases ($p < .01$ for 10 ms and $p < .05$ for the 15 ms discriminations) it can be tentatively concluded that a more sophisticated normalization procedure might further reduce or perhaps eliminate the displacement of the discrimination peaks relative to the phoneme boundaries.

2) Another reason for the displacement might be the difference in experimental design between the discrimination test and the identification test. In the discrimination test the stimuli in a pair were separated by .7 secs. of silence and the pause between repetitions of the pair was 1.7 secs., while in the identification test stimuli were presented one at a time, separated by the 5 secs. response pause. The relatively rapid repetition of stimuli in the discrimination test may have caused a shift of the perceptual phoneme boundaries towards lower stimulus values. In order to test this hypothesis an additional

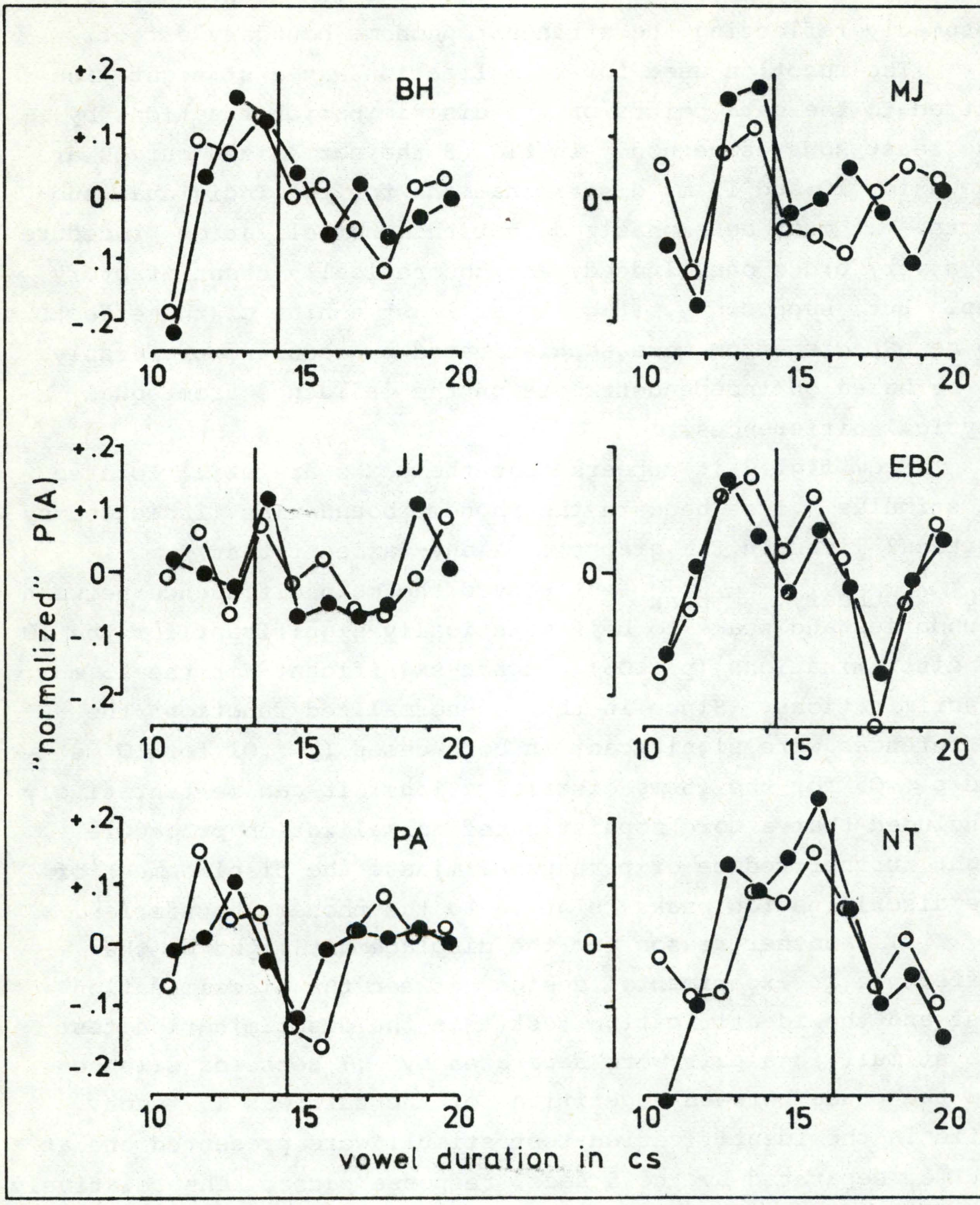

Figure 3

"Normalized" discrimination curves for the six listeners. The vertical lines indicate the phoneme boundaries. For further explanations, see the text.

listening test was arranged, in which three subjects (BH, JJ, and NT) were to identify the stimuli of the discrimination tape.

The results of the supplementary identification test did not, however, conform very well with predictions of the hypothesis, namely that the phoneme boundaries should be shifted to stimulus values lower than those of the original identification test. The listeners BH and JJ showed statistically significant shifts ($p < .05$) of 5 ms, but in opposite directions; for NT the phoneme boundary shift (1.5 ms in the direction contrary to the one predicted) could not be proved significant ($p > .05$).

If these results are valid for the other subjects as well the displacement of peaks in the discrimination test probably cannot be ascribed to the different manners of presentation in the discrimination test and the identification test.

6. Investigation of response bias

It was mentioned in section 2.2.2 above that the P(A) as a measure of discrimination is largely unaffected by the listener's response bias. However, information about response bias may contribute to the description of the listeners' perceptual behaviour. It was therefore considered desirable to collect such information, particularly with a view to revealing possible variations in response bias along the stimulus continuum.

The measure of response bias employed was the "false alarm" probability, i.e. the probability of obtaining an "f+" or "f" response to an AA stimulus pair (henceforth referred to as P("D"/s)).

In fig. 4 the computed P("D"/s) values are shown as a function of vowel duration. It is evident that the response bias varies along the continuum, the variation being more pronounced with the subjects JJ, PA, MJ, and EBC, and less pronounced with BH and NT. Moreover, the variation seems to be systematic in

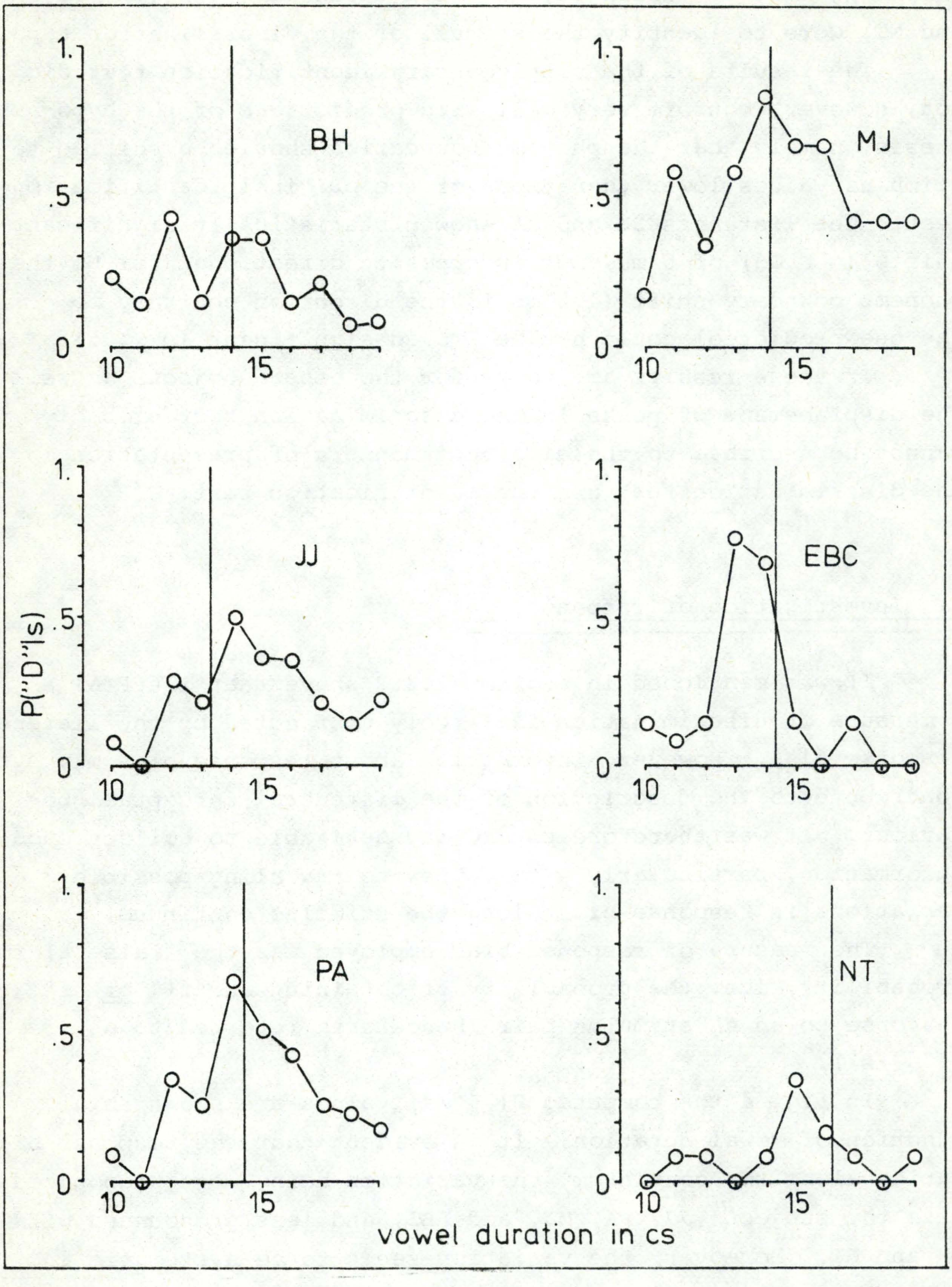

Figure 4

P("D"/s) as a function of vowel duration. The vertical lines indicate phoneme boundaries.

the sense that P("D"/s) is generally low at the marginal stimulus values and high at stimulus values in the vicinity of the phoneme boundary (indicated by vertical lines in the graphs). Only one subject, BH, does not conform to this pattern.

The question is whether the systematic variation of response bias is part of the subjects' perceptual behaviour or whether it is simply an artifact introduced by the stimulus inaccuracies described in section 2.1 above: Measurements revealed that the vowels in the AA pairs were not exactly equally long in all cases. However, the differences between stimuli in the AA pairs were rather small (between 1.0 and 1.5 ms at all stimulus values except the 190-190 ms pair, where the difference was 3 ms), and in order to explain the P("D"/s) variation the differences between stimuli in AA pairs should have been large in the range 130 to 150 ms and small at the margins of the continuum. As this was not the case the systematic variation in response bias can hardly be an experimental artifact due to stimulus inaccuracies.

If the variation is part of the subjects' perceptual behaviour a similar systematic variation may be expected to occur in the results of other identification/discrimination experiments. The only data available to me in a form permitting P("D"/s) to be computed were the raw data from a series of experiments on identification and discrimination of vowel quality in Danish carried out by Peter Holtse (Holtse 1973). The format of Holtse's experiments was essentially identical to the design of the present tests, and four of his subjects (BH, MJ, EBC, and NT) also served as listeners in the present experiment. Thus, Holtse's material seems very well suited for a comparison.

The P("D"/s) values computed from Holtse's data are shown in figs. 5 and 6 as a function of stimulus number. In this case P("D"/s) indicates the probability of obtaining a "++" or "+" response (corresponding to "f+" and "f", respectively) to an AA stimulus pair. It appears that variations in response bias along

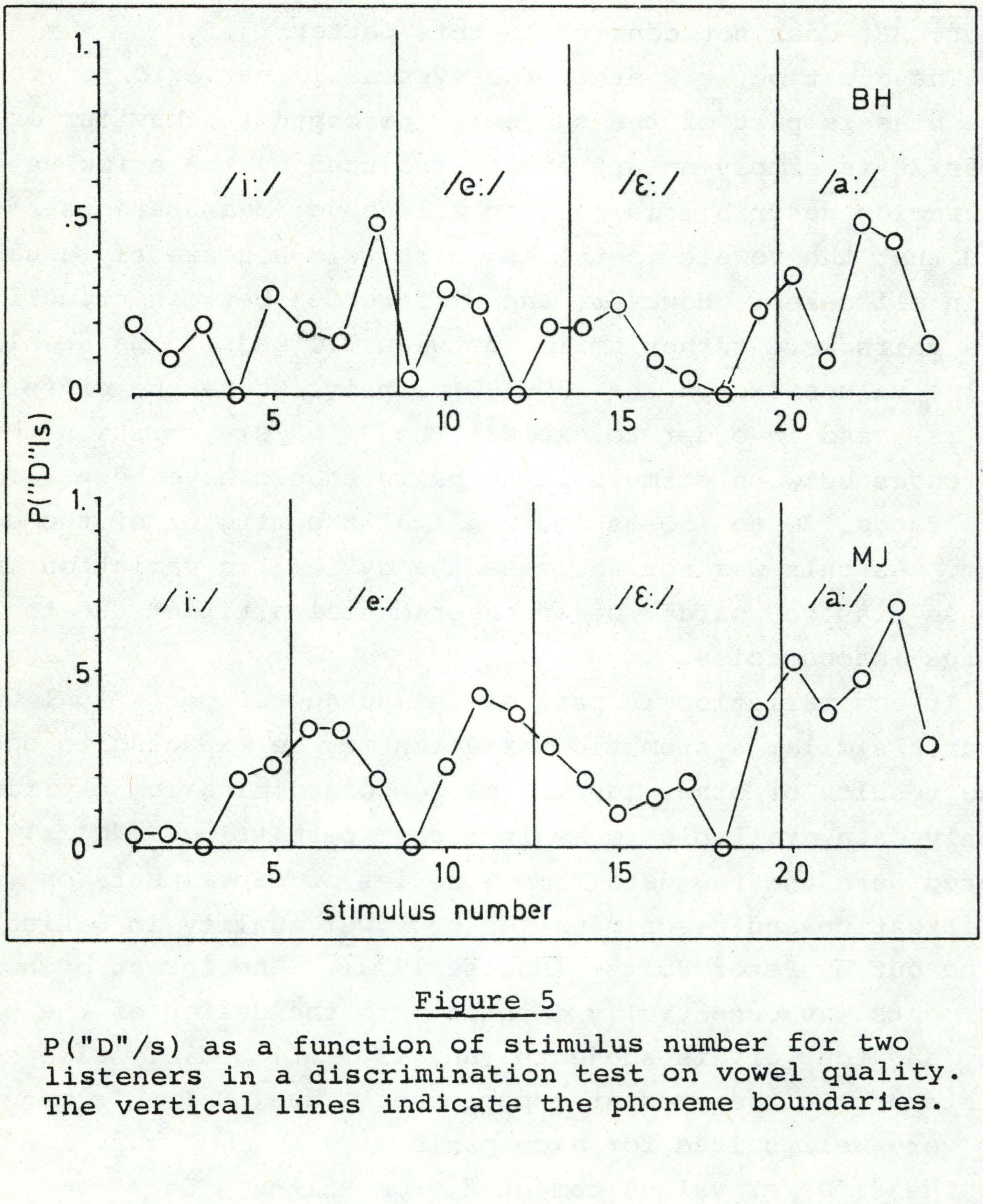

Figure 5

P("D"/s) as a function of stimulus number for two listeners in a discrimination test on vowel quality. The vertical lines indicate the phoneme boundaries.

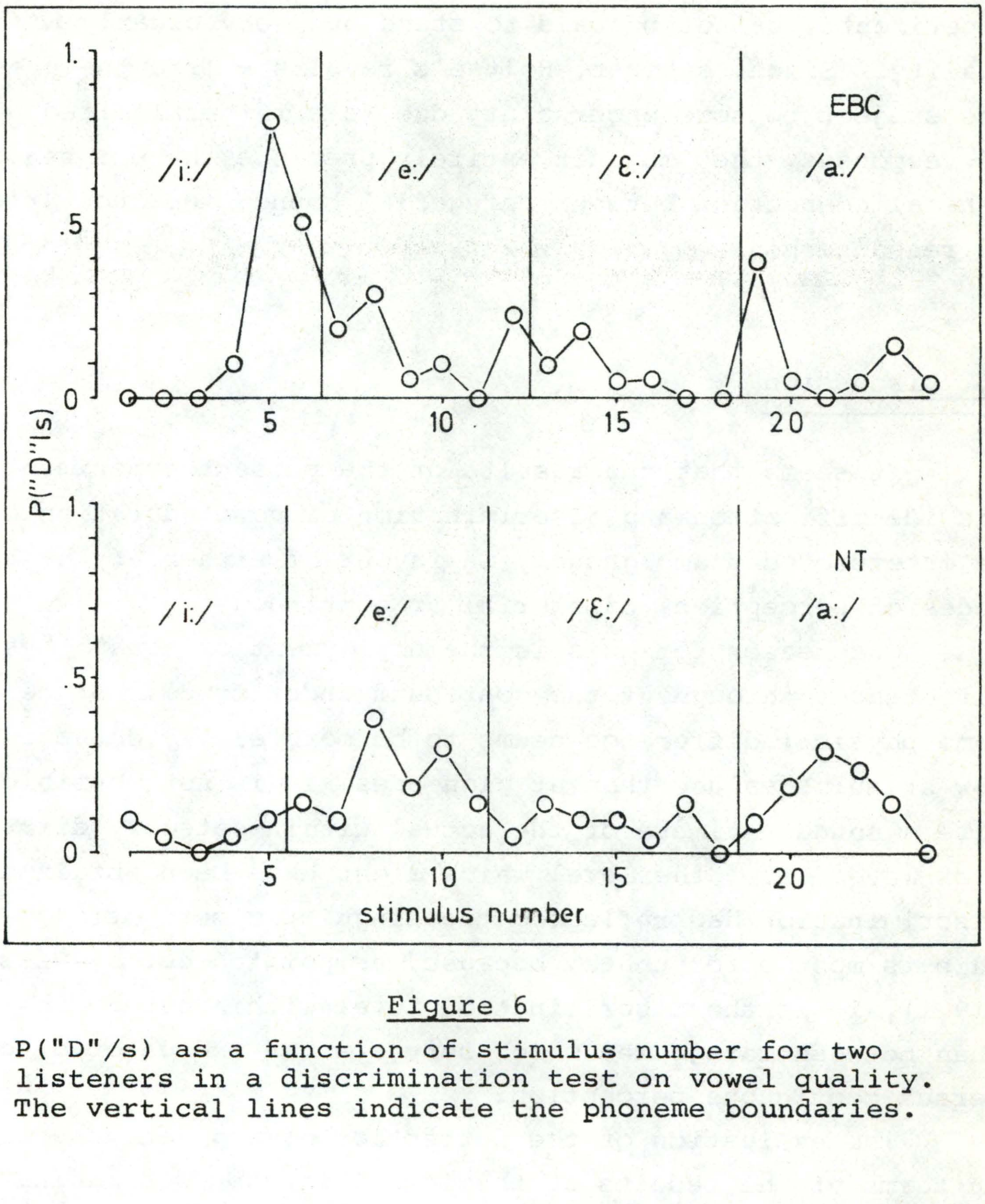

Figure 6

P("D"/s) as a function of stimulus number for two listeners in a discrimination test on vowel quality. The vertical lines indicate the phoneme boundaries.

the continuum are present also when the variable discriminated is vowel quality. But the strong tendency for the response bias to vary systematically, which was found in the vowel duration experiments, cannot be said to stand out very clearly with vowel quality. Since, however, Holtse's results - like the present - are subject to some uncertainty due to a rather limited number of responses, they may not entirely preclude the existence of a general connection between categorial boundaries and variations of response bias.

7. Discussion

It seems that the results of the present experiments on the identification and discrimination of vowel duration cannot be interpreted unambiguously in favour of either of the two modes of perception, categorial or continuous.

One reason for this is the employment of equal physical differences throughout the continuum under study. Since the same physical difference seems to be more easily detected at low stimulus values than at high ones, it is not possible to give a sound estimate of the actual within-category discrimination level, i.e. the level which might have been obtained if the discrimination had reflected the linguistic perception alone. This is most unfortunate, because, as pointed out by Pisoni (1971), it is the discrimination level within categories rather than between categories which serves as an indicator of categorial versus continuous perception.

The evaluation of the perception of a given acoustic variable in terms of the results of the identification/discrimination test paradigm rests on the assumption that peaks in the discrimination functions do reflect the between-category discrimination level. In the present experiments it was found that the peaks did not coincide with the phoneme boundaries established in the identification test, but were displaced towards lower stimulus values.

The displacement could not be attributed to stimulus inaccuracies or to the different manners of stimulus presentation in the two tests. However, an attempt to compensate for the influence of equal physical differences between stimuli seemed to reduce the degree of peak displacement relative to the phoneme boundaries.

If, on this basis, the discrimination peaks can be assumed to reflect the between-category discrimination level, it may be concluded that vowel duration is not perceived in a continuous manner. On the other hand, the lack of a reliable measure of the actual within-category discrimination level prevents an estimation of the degree to which the perception of vowel duration approaches the categorial perception. It may be said with some certainty, however, that vowel duration is not perceived in the same nearly-categorial manner as stop consonants. Were that the case one might have expected a less pronounced influence from physical equidistance between stimuli, since the perception of stop consonants apparently is highly resistant to the influence of experimental conditions.

Apart from the analysis of vowel discrimination as such, the subjects' response bias was examined. It turned out that the listeners had, in a number of cases, responded "different" to AA stimulus pairs. That is, of course, not very remarkable. A number of false alarms should always be expected. What is more interesting was the tendency for the false alarm probability, P("D"/s), to be considerably higher near the phoneme boundaries than within the phoneme areas. This is surprising; one might expect essentially the same rate of false alarms at all stimulus values.

The systematic variation of P("D"/s) may perhaps be explained in terms of Fujisaki and Kawashima's model of the decision process in discrimination tasks (see section 1 above). According to their model the acoustic signal is translated into an auditory representation and a categorized representation in short term memory.

Now, if we expect P("D"/s) to be invariant along the continuum the underlying assumption is that the two stimuli in AA pairs are always given identical categorized representations in short term memory. This assumption, however, is doubtful: Errors of categorization (i.e. the two AA stimuli being categorized as different) may occur, and in such cases the P("D"/s) will increase. This can be illustrated by an example:

Let us assume first that the AA stimuli are never categorized as different. Under these circumstances comparison of categorized representations will never provoke "different" responses, and the decision will be based entirely on the auditory representations. Here some "different" responses are likely to occur in, say, 25 per cent of the trials. The percentage of false alarms will be 25 per cent in all, and because of the non-categorial nature of the auditory representations that percentage will be invariant along the continuum.

If, on the other hand, the two stimuli in an AA pair are categorized as different in, say, 20 per cent of the trials, comparisons of categorized representations will contribute that percentage to the total false alarm rate. In the remaining 80 per cent of the trials the false alarm rate will depend on the auditory representations. If, as above, a quarter of the remaining 80 per cent of the trials, i.e. 20 per cent of the total number of trials, give rise to "different" responses, the total percentage of false alarms amounts to 40 per cent (20 + 20).

If categorizations of AA stimuli as different occur at all they assumedly occur more frequently near the phoneme boundary than well within the phoneme areas.

In an informal identification experiment on vowel duration in which subjects were to label both stimuli in AA and AB pairs it sometimes happened that the two stimuli in AA pairs were given different labels. And this different labelling was found only near the phoneme boundary, never within the phoneme areas. However, the number of AA pairs in the test was very small and it was not possible to test statistically the reliability of the evidence.

The explanation outlined here for the systematic variation of response bias along the duration continuum is weakened by the fact that a similar systematic variation could not be unambiguously established in the case of another acoustic variable, namely vowel quality. Thus, it will require further investigations to decide whether the systematic variation of response bias reported here can be said more generally to be part of the perceptual behaviour of listeners in the discrimination of linguistic stimuli.

References

Abramson, A.S. and L. Lisker 1968: "Voice timing: Cross-language experiments in identification and discrimination", Haskins SR 13/14, p. 49-63

Bastian, J. and A.S. Abramson 1962: "Identification and discrimination of phonemic vowel duration", JASA 34/5 (abstract), p. 743-44

Fujisaki, H. and T. Kawashima 1968: "The influence of various factors on the identification and discrimination of synthetic speech sounds", Reports of the 6th international congress on acoustics

Fujisaki, H., K. Nakamura and T. Imoto 1973: "Auditory perception of duration of speech and non-speech stimuli", Paper presented at the Symposium on Auditory Analysis and Perception of Speech, Leningrad 1973, p. 1-27

Guilford, J.P. 1954: Psychometric Methods (New York)

Harris, J.D. 1952: "Remarks on the determination of a differential threshold by the so-called ABX technique", JASA 24, p. 417

Holtse, P. 1973: "Identification and discrimination of closely spaced synthetic vowels", ARIPUC 7, p. 235-64

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. 54,5, p. 358-68

Liberman, A.M., K.S. Harris, J.A. Kinney and H. Lane 1961: "The discrimination of relative onset-time of the components of speech and nonspeech patterns", J. Exp. Psychol. 61,5, p. 379-88

Liberman, A.M., K.S. Harris, P. Eimas, L. Lisker and J. Bastian 1961: "An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance", LS 4,4, p. 175-95

McNicol, D. 1972: A Primer of Signal Detection Theory (Sydney)

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

Pisoni, D.B. 1973: "The role of auditory short term memory in vowel perception", Haskins SR 34, p. 89-118

Reinholt Petersen, N. 1974: "The influence of tongue height on the perception of vowel duration in Danish", ARIPUC 8, p. 1-10

Rischel, J. and S.-E. Lystlund 1972: "A formant coded speech synthesizer", ARIPUC 6, p. IX-XXX

Robinson, D.E. and C.S. Wattson 1972: "Psychophysical methods in modern psychoacoustics", Foundations of Modern Auditory Theory, vol. II (J.V. Tobias, ed.), p. 99-132

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