PERCEPTUAL STUDIES OF DANISH STOP CONSONANTS

Eli Fischer-Jørgensen

The experiments upon which these studies are based were carried out in the years 1954-64, (mainly 1954-56), and most of the material was processed at that time. The publication was postponed because the studies were planned to form part of a monograph on Danish stop consonants including acoustic, physiological and perceptual descriptions, a plan which I later had to give up. A summary of the acoustic results was given in the paper "Acoustic Analysis of Stop Consonants" Miscellanea Phonetica II, 1954, p.42-59. The main part of the perceptual results are published on the following pages, and the results of the physiological measurements will be published at a later date.

I am grateful to the students who helped me with the processing of the data, particularly to Henny Pontoppidan Lauritzen, Brita Waage-Bech, Lise Preisz and Rune Jacobsen.

I also want to express my gratitude to Statens Videnskabsfond, whose support (and whose patience) has made this work possible.

PERCEPTION OF FOREIGN STOP CONSONANTS BY DANISH LISTENERS

Eli Fischer-Jørgensen

1. Introduction

Danish ptk and bdg are distinguished only syllable initially. Both sets are voiceless. ptk are articulated with somewhat less energy in the supraglottal cavities than bdg, but the main difference is one of aspiration. Earlier measurements (EFJ 1954) of the duration of the open interval (in the sense of the distance from the explosion to the beginning of the vowel, including affrication and aspiration) have given the result that the average in stressed position is 66 ms for p, 79 for t and 74 for k (with individual averages ranging from 53 for p to 98 for t) and 14, 17 and 23 ms for b, d, g respectively. The open interval of t is slightly longer than that of k in Danish because of the strong affrication of t. The order of the unaspirated stops (labial-dental-velar) is more normal. As in other languages, the open interval is normally somewhat longer before close vowels than before open vowels. It is rarely longer than 35 ms in bdg and shorter than 40 ms in ptk.

It might be expected that these durations would influence the perception of stops in foreign languages. This was tested by experiments carried out in 1958 and 1961. The results have not been published until now.

2. The material

The material used in the test consisted of 137 different CV syllables (where C is a stop consonant, and the vowel mostly i, a or u) cut out of words spoken by five Danish, one German, two British English, one French, one Dutch, one Chinese, and four speakers of Indian languages, viz. Urdu, Hindi, Marathi, and Malayalam. The recordings were made earlier for other purposes. The words were chosen so as to be representative of the languages (but the English speakers had relatively short aspirations) and, as far as the Danish stops and the Indian voiced aspirates are concerned, to cover a certain range of variation. The voiced aspirated stops from Indian languages represented various types. Most of them were totally voiced (though often with a very short cessation of vibrations at the explosion, a few had voiceless closure or voiceless aspiration. Some of the Danish examples were chosen to represent unusually short aspirations (35 and 40 ms for p).

The English, German and Danish words had been spoken in a context consisting of a preceding word ending in an unstressed vowel. The English bdg-sounds were voiced, the Danish and German ones voiceless. The different types of stops represented were: (a) voiced bdg 23 (English, French, Dutch, Indian), (b) voiceless bdg 29 (Danish, German), (c) unaspirated ptk 28 (French, Dutch, Indian, Chinese), (d) aspirated ptk 55 (Danish, English, German, Indian, Chinese), (e) voiced aspirated 18 (Indian), and, (f) aspirated affricates 2 (Chinese).

These syllables were combined into a test tape (each syllable was repeated three times). The signal to noise ratio was not quite good in the Danish, German and some Indian examples, but this does not seem to have influenced the results. The order was quasi-random. The French examples of ptk (nine in all) were repeated in different environments, once after a syllable with bdg, once after a syllable with aspirated ptk, in order to check the influence of the context. There was a slight influence in four of the nine examples, but only in one case was the difference significant.

Spectrograms and mingograms (including intensity and pitch curves) were taken of all examples. Moreover four phoneticians listened to the test and characterized the consonants in terms of voicing, aspiration and impression of fortis-lenis.

3. The listeners

The test tape was played over a loudspeaker in a sound-treated room to groups of listeners. The listeners, 82 in total, were Danish students of philology in the beginning of their first term. All of them were familiar with English, and most of them with German and French, but they had not yet learnt any phonetics. More naive subjects might have been preferable, but as it can be seen from the results, they were not much influenced by their knowledge of foreign languages. They were asked to identify the initial consonants of the syllables with one of the six Danish stops, and to write down the result on answer sheets.

4. The results

The main results (apart from the answers to voiced aspirates) can be seen in Fig. 1. The examples with a majority of bdg-answers are placed above the base line, the examples with a majority of ptk-answers below the line. The horizontal scale indicates the duration of the open interval and the vertical scale the number of examples. The various stop categories are distinguished on the graph.

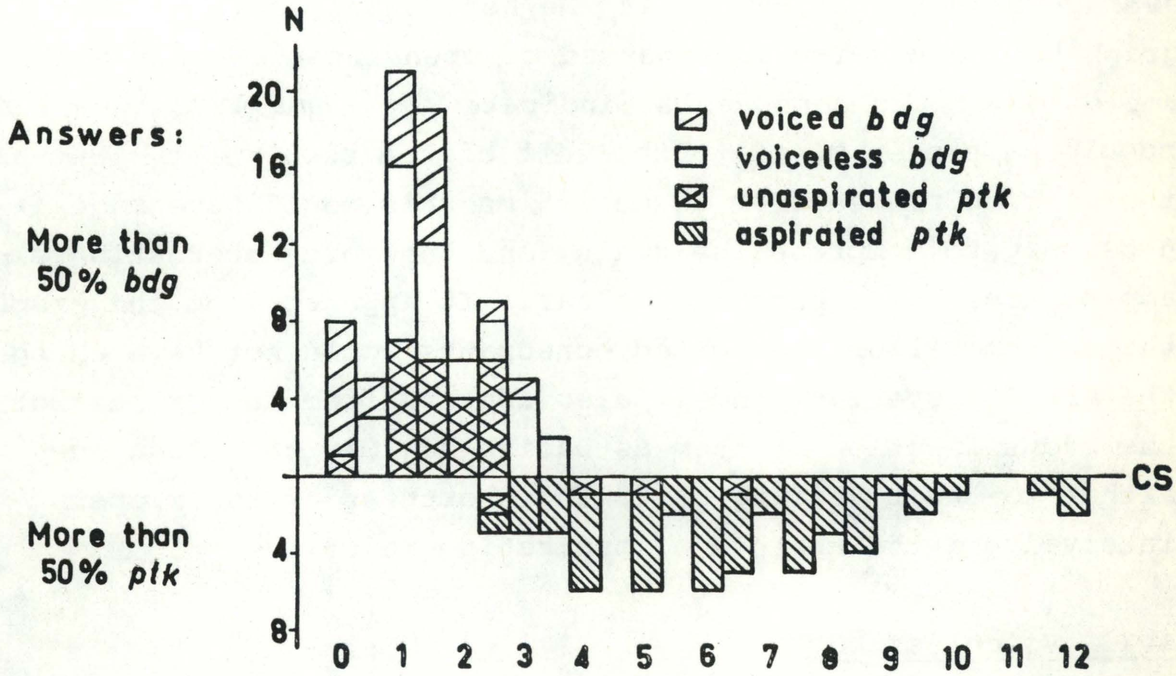


Fig. 1. Identification of foreign stop consonants by Danish listeners compared to the duration of the open interval.

4.1. Voiced stops

All voiced stops were identified as Danish bdg with an average majority of 99 %. If the horizontal scale of the graph had been VOT-value instead of open interval, the examples of voiced consonants (indicated by a small oblique line) should have been moved to the left of the zero point. But as the voicing ranges from 30 to 195 ms this would have made it necessary to compress the scale, and the more interesting examples would have been less clear. It appears from the graph that a removal of the voiced consonants would not have changed the area of overlap, and it also appears from the graph that the identification of them as Danish bdg may have been due either to their voicing or to the shortness of their open interval and the absence of aspiration noise.

4.2. Voiceless bdg

All voiceless bdg-stops were identified as Danish bdg. The average for all examples was 93 % bdg (with the exception ga left out it was 95 %). This means that voicing is not a necessary cue; the small difference in open interval is sufficient for explaining the slight difference in the majority (99 and 95 %) for the two categories. If the voiceless bdg-sounds are divided into two groups (5-20 ms and 25-35 ms) the averages are 99 and 87 %, respectively.

The only exception to the bdg-answers is a Danish example of ga (subject OT) with an open interval of 25 ms. We shall return to this problem below.

4.3. Unaspirated ptk

The unaspirated ptk-sounds were heard by the majority as Danish bdg in 24 out of 28 examples. It appears from the graph that three of the four exceptions can be explained by a much longer open interval. These were Dutch ki (65 ms, 99 % k), French ku (50 ms, 98 % k) and Indian ka (40 ms, 59 % k). Only one example (French pu 25 ms, 77 % p) has overlapping with the examples identified as bdg. The average majority for those identified as bdg is 90 %. The average bdg-majority for all unaspirated stops is 80 %.

4.4. Aspirated ptk

All instances of aspirated ptk were identified with Danish ptk, and as it can be seen in Fig. 1, their open interval in 48 out of 55 examples was longer than that of the other stops. 7 examples overlap with the stops heard as bdg. The two consonants with 12 ms open interval were Chinese affricated aspirates.

4.5. Overlapping cases

The great majority of answers can be explained by the length of the open interval. But there is a number of overlapping cases between 25 and 35 ms. The individual examples are given together with the answers in tables I and II.

Table I

Overlapping of Labials

answers	25 ms	30 ms	35 ms
b		(Da.) by 98 % (G.) bu 65 %	
p	(Fr.) pu 77 % (E.) pa 70 %	(E) pi 100% (Da.) pa 100% (Da.) pa 100%	(Da.) pi 82 % (Da.) pi 90 %

Table II

Overlapping of velars

answers	25 ms	30 ms	35 ms
g	(E.) gi 94 %	(E.) gu 82 %	(G.) gu 95 %
	(Da.) gy 76 %	(Da.) ga 90 %	(Da.) gu 88 %
	(G.) ga 71 %	(G.) gi 96 %	
	(I.) ka 100%		
	(Ch.) ka 98 %		
	(Du.) ku 99 %		
	(Fr.) ka 81 %		
	(Fr.) ke 52 %		
k	(Da.) ga 74 %		(E.) ka 93 %

The 15 consonants heard as <u>b</u> or <u>g</u> in this area are all sounds generally characterized as <u>bdg</u> or unaspirated <u>ptk</u>. The 9 consonants heard as <u>p</u> and <u>k</u> are, with two exceptions, sounds normally characterized as aspirated <u>ptk</u>. This seems to indicate that a supplementary criterium besides length of open interval must be aspiration noise. It cannot be tenseness in the sense of fortis articulation, since Danish <u>ptk</u> are not fortis in this sence, whereas French <u>ptk</u> are, and since French <u>ptk</u> are normally heard as <u>bdg</u>, whereas Danish <u>ptk</u> are heard as <u>ptk</u>.

As for the labials the Danish examples were chosen as extreme values both of <u>b</u> and <u>p</u>, and I wanted to see whether they were identified correctly. As a matter of fact they were. A check on the stimuli by inspection of curves and by listening reveals that the p-explosions are followed by a certain aspiration, whereas the Danish <u>b</u>-sounds are not (moreover, <u>by</u> had some voicing at the beginning of the closure).

The German bu has a very slight aspiration noise and has only 65 % majority. There is also aspiration noise after the two English p-examples, and French pu. Labials outside the area of overlap have normally more than 90 % majority in the answers. There is however an example of Dutch pu (15 ms) with only 57 % majority for b. Two Danish examples of pa (35 and 25 ms) were left out because they were so weak that they were mostly heard as f.

There is no overlapping of t and d, which may partly be due to the fact that there are no examples of 30 and 35 ms. All answers to dental stops below and above these values show more than 90 % agreement on d and t respectively, except for French tu (25 ms, 52 % d), and French te (20 ms,52 % d). There is one other example of unaspirated t at 20 ms (Chinese ta heard as d with 91 % majority);it has a weaker explosion.

As for the velars, the boundary seems to be slightly higher than for the labials (about 35 for velars and 25-30 for labials), which is in agreement with the difference observed in the actual length of the open interval. English ka (35 ms) has some aspiration. The Danish example ga (25 ms, 74 % k) is astonishing. It is not aspirated, but it has a strong explosion and is characterized by the phoneticians as fortis in contradistinction to ga (30 ms), spoken by a different Danish speaker and heard as ga. It has a pitch fall in the start, but so has Dutch and Chinese ka, which are heard as ga. Similarly, the Danish ga of 30 ms, heard as ga, has a slight fall in pitch at the start. On the whole, it does not seem possible to explain the overlapping cases on the basis of pitch differences.

Only the French speaker has a clear difference. Nor can they be explained by differences in formant transitions. In the exceptional Danish ga-example the intensity of the explosion may play a role. (But the explosion of German ga (25 ms) heard as ga seems to be just as intense).

4.6. The answers of the phoneticians

Four phoneticians listened to the test tape and characterized all examples as ptkbdg and moreover as aspirated or unaspirated, voiced or unvoiced and fortis or lenis. In view of the small number of listeners it has no sense to give a detailed account of the answers.

The most interesting question is whether the phoneticians were able to distinguish between voiceless bdg and unaspirated ptk. It turned out that they made a distinction, but it coincided only partly with the expected distinction between German and Danish bdg on one hand and French, Dutch, Indian, and Chinese ptk on the other (the Chinese sounds are, by the way, often described as voiceless bdg). Of the 112 answers to Danish and German bdg there were 23 % ptk-answers and 77 % bdg-answers, and of the 116 answers to unaspirated ptk there were 53 % ptk-answers and 47 % bdg-answers (for the French stops 69 % ptk-answers). The answers to fortis-lenis were not quite the same, since particularly one of the listeners often combined bdg-answers with fortis-answers. There were 37 % fortis- and 62 % lenis-answers to German and Danish bdg, and 60 % fortis- and 40 % lenis-answers to unaspirated ptk (for the French stops there were 80 % fortis-answers). This means that there was a difference in the answers to the two groups, both in the designations as bdg or ptk and in the reaction to fortis-lenis, and this difference is statistically significant, but there is great overlapping.

As regards the designations as bdg or ptk the length of the open interval has some influence. There are hardly any ptk-answers to sounds with 5-10 ms open interval, and no bdg-answers to the three sounds with 40, 50 and 65 ms open interval, but from 15 to 35 ms there is complete overlapping for the unaspirated ptk, and from 25 to 35 for voiceless bdg. For the values 15 and 20 ms there are very few ptk-answers to voiceless bdg, but many to unaspirated ptk. There must therefore be some further criteria. On the whole, k-answers are more frequent than p- and t-answers. This means that the longer open interval of velars influences the judgement. (One might have expected a compensation in the perception because the longer duration of the open interval in velars is a universal, mechanical phenomenon).

The answers to the fortis-lenis question are much less dependent on VOT-values for the consonants in question.[1] There are evidently other cues. The number of examples and listeners, and the irregularities in a material consisting of natural speech do not permit any clear decision concerning these cues. There do not seem to be differences in pitch within the vowel or in formant transitions, corresponding to the answers, but there seems to be a certain correlation with the intensity of the explosion. This needs further investigation.

4.7. Aspirated b d g

The answers to aspirated Indian bdg have not been included in Fig. 1 because they would spoil the neat distribution.

There are 18 examples. Five of these were heard as bdg, but not with a very great majority (from 51 to 79 % with an average of 62 %). The remaining 13 were identified as Danish ptk, with a majority varying from 56 to 100 and an average of 83 %).

The duration of the open interval (i.e. the aspiration) is not decisive here. The five examples with a bdg-majority have

---

1) The aspirated stops are characterized as fortes. This may depend on the VOT-value, but it is more probable that it is due to the intensity of the aspiration noise.

an aspiration ranging from 40 to 100 ms with an average of 77 ms, and the thirteen examples heard as ptk have an aspiration ranging from 30 to 110 with an average of 64 ms. The two examples with a clear bdg-majority have 100 and 40 ms aspiration, the four with unanimous ptk-identification have 45, 60, 70 and 80 ms.

Voicing may be of some importance. The five examples with a bdg-majority are fully voiced (except for an interval of 10 ms at the explosion in one of the examples).

As for the 13 examples which gave ptk-answers, three are fully voiced, six have a short voiceless pause at the explosion of about 10-20 ms, three have voiceless closure, and one has voiceless aspiration. But there is no evident correlation between the degree of voicing and the percentage of bdg-answers. There are seven examples with more than 88 per cent ptk-answers; two of these are fully voiced.

The intensity of the aspiration noise seems more important. Of the seven examples with the highest number of ptk-answers (above 88 %) six are characterized by strong noise, of the other cases only one has strong noise. (The intensity has not been measured but only judged from spectrograms and intensity curves.)

4.8. Conclusion

When Danish listeners are asked to identify foreign stops as Danish ptk or bdg, their main criterion is aspiration, which plays a role both in terms of the duration of the open interval and in terms of the degree of noisiness. Voicing plays a very subordinate role, if any.

Apart from the Indian voiced stops, for which the aspiration noise seems to be the most important cue, the duration of the open interval shows a rather clear boundary, which is located somewhere in the range 25-35 ms according to the place of articulation. This is in agreement with measurements of natural

Danish consonants. For voiceless sounds this measure is the same as VOT-value, and the Danish values are in good agreement with the VOT-values found by Lisker and Abramson for American English in experiments with synthetic speech (see e.g. Abramson and Lisker 1965 and Lisker-Abramson 1970). The phoneticians were, partly, able to distinguish voiceless bdg and unaspirated ptk.

References

Abramson, A.S. and Lisker, Leigh 1965: "Voice Onset Time in Stop Consonants: Acoustic Analysis and Synthesis", SR. Haskins 1, 1-7.

Lisker, Leigh and Abramson, A.S. 1970: "The Voicing Dimension. Some Experiments in Comparative Phonetics", Proceedings 6, Int. Congr. Phon. Sc. Prague 1967, p. 563-67.

Fischer-Jørgensen, Eli 1954: "Acoustic Analysis of Stop Consonants", Miscellanea Phonetica II, p. 43-59.

IDENTIFICATION OF UNRELEASED DANISH STOP CONSONANTS

Eli Fischer-Jørgensen

1. Experiments 1 A and B (1954-56)

1.1. Material and listeners

A series of Danish nonsense syllables of the structure VC and CVC, where V was short i, ε, a, u, and C was b, d or g, were spoken by me and recorded on tape, in 1954. In experiment 1A there was one example of each of the four vowels before each of the three consonants, thus 12 VC syllables in all. In experiment 1B the same four vowels were found after and before the same three consonants, which gives 36 different CVC syllables. Initial bdg were voiceless, final bdg almost voiceless, with a few periods of weak voicing in the start. The vowel length was mostly 90-100 ms, in a few cases of a 110-120, and in some cases of i or u only 70-80 ms.

The syllables were combined into two tests (A (VC) and B (CVC)), where each syllable was presented once. The order of the syllables was random, and the distance between the syllables was 4-5 sec. The test was given to three groups of students of philology in their first term in the years 1954, 1955 and 1956. There were 94 listeners in all. This means that there were 94 answers to each vowel-consonant combination in test A, and 282 in test B, i.e. 4512 answers in all.

The listeners were told that they were going to hear a list of meaningless syllables containing the vowels i, a, u + unreleased b, d or g in test A, and bdg + the same vowels + unreleased bdg in test B. They were asked to write down what they heard and allowed to put a questionmark if they could not identify the consonant. The syllables had been recorded on a professional tape recorder but were played back on a semi-professional tape recorder via a loudspeaker. The quality was not very good, but as a control the same listeners were asked to identify three additional lists, viz. two CV lists containing ptk or bdg + the vowels i, ε, a, u, and one VC list containing the same vowels + released ptk. There were hardly any mistakes in these lists, and it was also extremely rare that there were mistakes in the initial consonants in test B. The mistakes in the final consonants in test A and B must, therefore, be due to the lack of explosions.

1.2. Results

The results of the two tests are given in Table I and Table II and in graphical form in Fig. 1.

TABLE I

Answers (in percentage) to test 1 A (VC)

Answers:	b	d	g	?		b	d	g	?		b	d	g	?
ib	90	3	0	7	id	24	40	11	25	ig	43	10	35	11
εb	65	23	5	8	εd	11	74	9	7	εg	9	5	78	8
ab	77	2	16	4	ad	25	58	12	5	ag	1	1	91	7
ub	70	4	13	13	ud	4	88	1	7	ug	37	2	46	15

TABLE II

Answers (in percentage) to test 1B (CVC)

Answers:	b	d	g	?		b	d	g	?		b	d	g	?
ib	92	2	1	5	id	55	30	6	9	ig	46	16	26	13
εb	75	12	4	9	εd	31	33	21	12	εg	6	8	76	11
ab	60	7	24	9	ad	19	40	32	9	ag	7	1	86	6
ub	42	7	37	14	ud	2	91	3	5	ug	20	4	58	18

"?" in the tables covers the answers: questionmark, minus sign, or no indication of the consonant in question or of the whole syllable, e.g. bu?, bu÷, bu or nothing for bud.

Some subjects have written ptk finally instead of bdg. As there was, of course, no aspiration at the end, this is simply an influence from orthography. Phonologically there is no distinction between [p] and [b] finally.

On the whole the answers to test A (the VC list) are better than the answers to test B (the CVC list). The percentage of correct answers for all environments taken together are in test 1 A b 76 %, d 65 % and g 63 %, in test 1B b 70 %, d 49 % and g 62 %. The difference is small for b and g, but considerable for d. Particularly the syllables with d-d are heard incorrectly in test 1 B (only 36 % correct answers). One of the purposes of list B was to find out whether there would be perceptual dissimilations in the answers. This seems to be the case for d-d, but there is no clear tendency for the syllables with b-b and g-g. This difference may, of course, be due to chance, but the question deserves further investigation.

b after u is heard better in test A than in test B. In test B it is often heard as g, whereas in test A g is often heard as b in this environment.

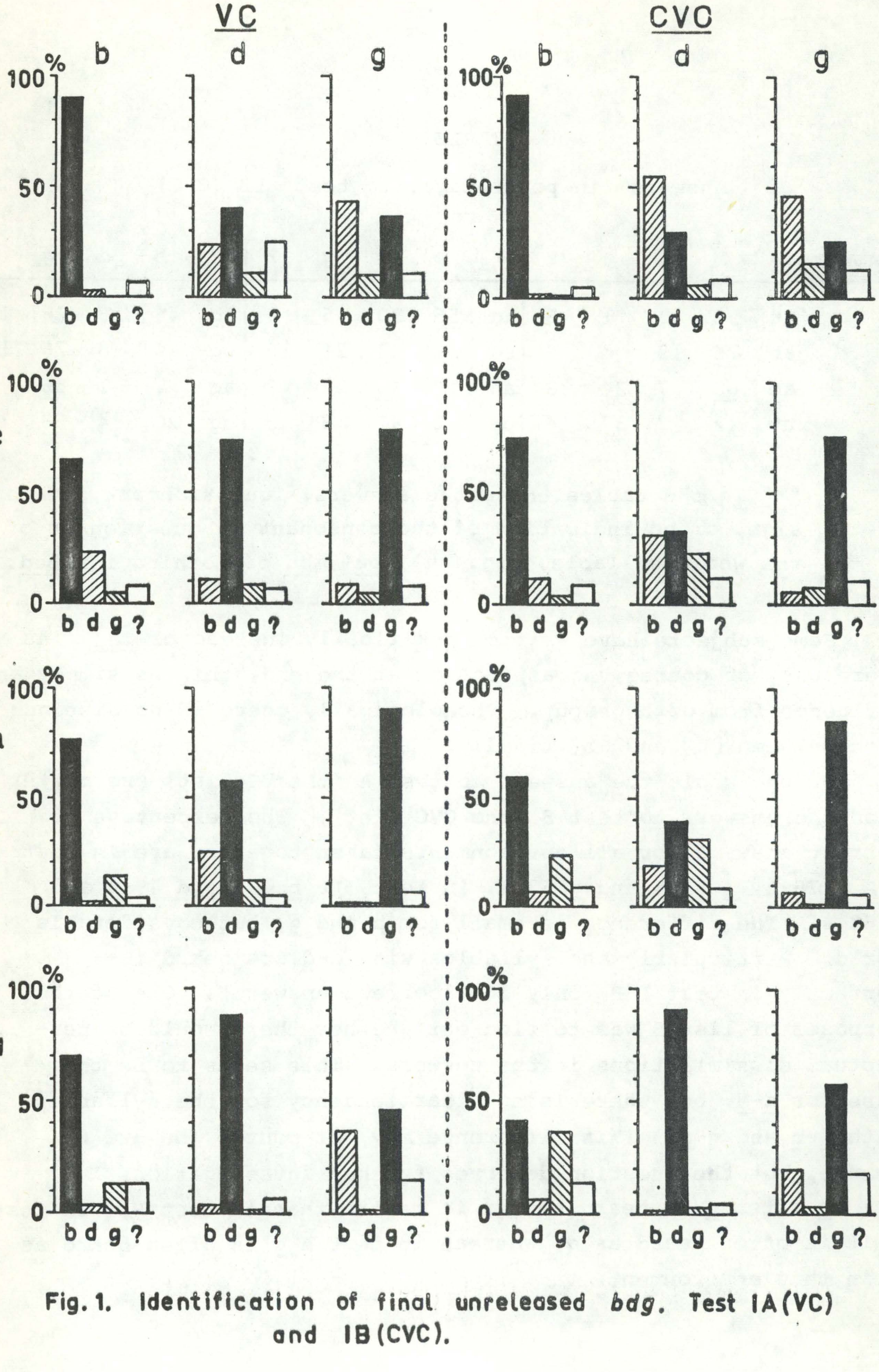


Fig. 1. Identification of final unreleased *bdg*. Test IA (VC) and IB (CVC).

Apart from these differences the answers to test A and B have many features in common. The main tendencies common to both are that b is identified most correctly after i, d after u, and g after ε and a.

Repeated listening has revealed a very slight explosion noise in id and ud in test A, but not in test B. id is heard somewhat better in test A, but there is no difference between the answers to ud in test A and B. This slight explosion noise can therefore not have improved the tendencies found in the answers.

The differences just mentioned between b, d and g can be easily explained when the extent of the transitions is taken into account. The transitions of F2 and F3 are particularly clear in ib and almost non-existent in ub (which has the lowest percentage in test B). Before d, on the other hand, the vowel u has an extensive transition, but i very little, and before g, ε and a have a characteristic converging transition of F2 and F3, whereas the transitions are slight in u and i.

The high number of correct g-answers after a is not astonishing, since the quality of the whole vowel is influenced in this case (a is more retracted). For this reason ε has been included in the test as a better representative of open vowels, but the answers to εg are not much poorer than the answers to ag. a is also somewhat more retracted before b than before d, not so much in my speech as in Copenhagen speech, but clearly visible in the spectrograms.

If the examples with a are kept aside, the percentage of correct answers for g are somewhat lower: 54 and 53 % instead of 63 and 62 in tests A and B.

The confusion between b and g after u is explicable from the fact that they have a similar and very small transition of F2. After i and to a lesser extent after ε, d is often heard as b. This can be explained by the common direction of their F_2 transition. As the extent of this transition is larger in b than in d, it is possible to hear b for d rather than the opposite. Such an explanation can, however, not be given for the cases of g heard as b after i. On the whole, it seems that the labial is often chosen in cases of doubt. It is, in some sense, the most neutral consonant.

2. Experiments 2 A and B (1964)

In 1964 I undertook a new experiment with VC syllables, this time including all the Danish short vowels i y u e ø o ɛ œ ɔ a. Each of the syllables with these vowels followed by unreleased b, d and g were spoken twice by me and by the speaker JR. The tests were run in the same way as in experiment I. My recording (A) was played to two groups, one consisting of 20 students in their first term, and the other of 10 phoneticians. The recording of JR (B) was only played to the 10 phoneticians. The number of correct answers to test A was for the student group b 67%, d 64% and g 48%, which was not quite as good as test 1. For the phoneticians the number of correct answers was higher: b 79%, d 78% and g 58%. In test 2 B (spoken by JR and played to the phoneticians only) the number of correct answers was still higher: b 76%, d 82% and g 88%. The difference is very large for g. One of the reasons why JR's recording gave more correct answers is that he spoke somewhat more slowly than I did. His vowels were 10-20 ms longer, and he had a less steep rise and decay of the intensity. Spectrograms of the syllables with g show more pronounced transitions in some cases.

The answers to the syllables with i ɛ a u can be compared with test 1. In Fig. 2 the percentage of correct answers to the syllables with these four vowels are given for all the tests. The stimuli for tests 1 A and B and test 2 A were

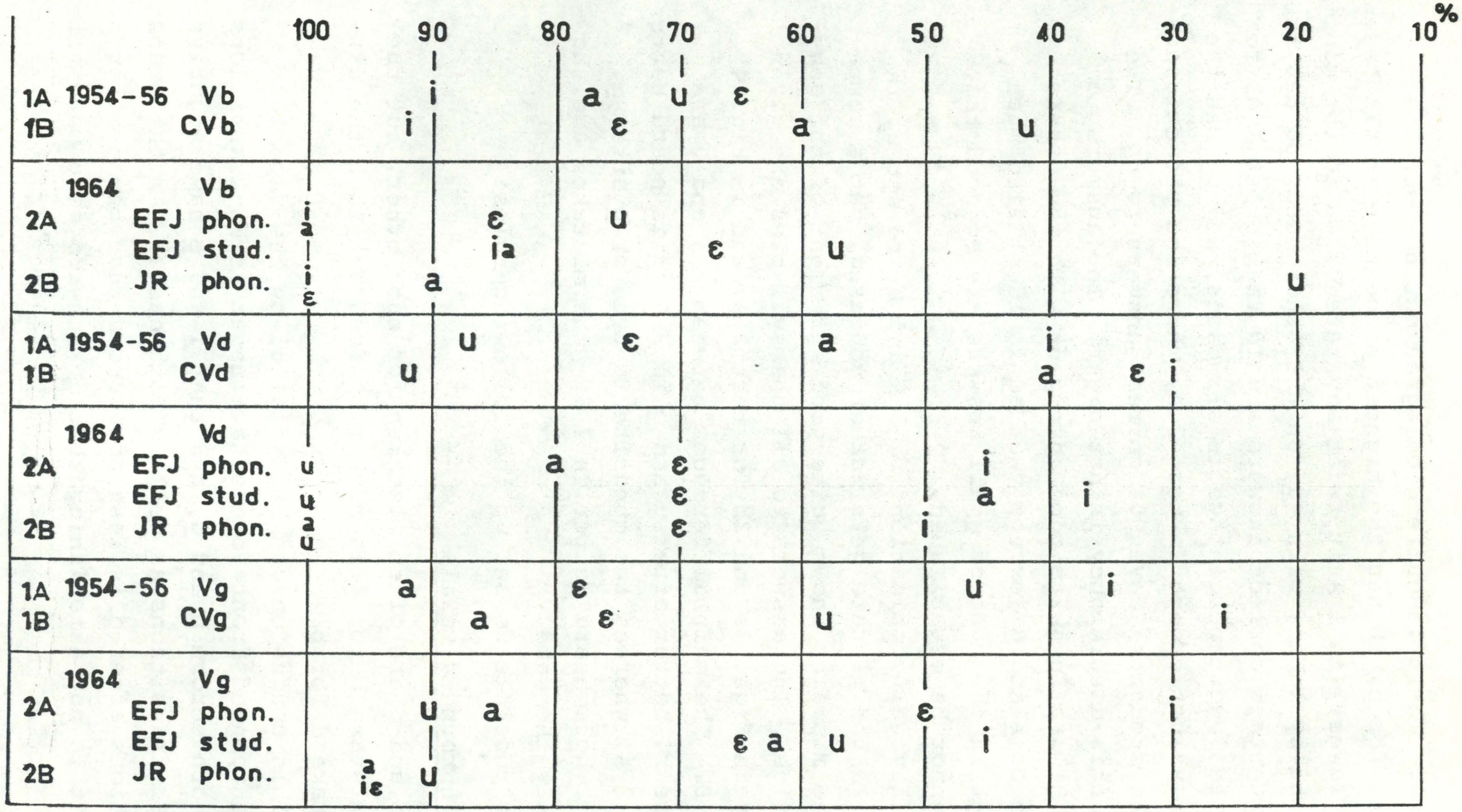


Fig. 2. Percentage of correctly identified unreleased *bdg* in Test 1AB and Test 2AB.

spoken by me. There is quite good agreement between the answers. As for b, ib has the highest percentage in all cases and ub the lowest (in 1 A it is at any rate close to the minimum). In the syllables with -d, on the other hand, ud has the highest percentage correct answers and id the lowest. As for g, ig has the lowest percentage in all cases, and ug has the next lowest values except in test 2 A, where the phoneticians (but not the students) have 90% correct answers to ug. The reason for this discrepancy is not clear. The answers to test 2 B (JR) are in quite good agreement with the others in the case of b and d (except that ad is identified much better) but in the case of g all answers are so good that an order cannot be established.

If all vowel combinations in 2 A and B are set up in this manner, the picture gets rather confused, but a comparison of the mistakes shows quite good agreement of the general tendencies. The answers to all the syllables are given graphically in Figs. 3a and 3b for test 2 A, and in Fig. 4 for test 2 B. The main differences between 2 A and B are that b is heard better after u and o in 2 A, d is heard better after œ in 2 A, whereas the opposite is true of b after œ. g is on the whole heard better in 2 B. The mistakes are, however, roughly the same.

It is possible to set up the following rules:

(1) after unrounded vowels

(1a) b and d are often mistaken for each other, but the mistake b for d is much more frequent than the mistake d for b.

This confusion can be explained by the fact that unrounded front vowels have negative F_2-transitions both before b and d, and that i and e have negative F_3-transitions as well in both cases. As the transitions after d is less pronounced than those after b, it is not astonishing that d is heard as b more often

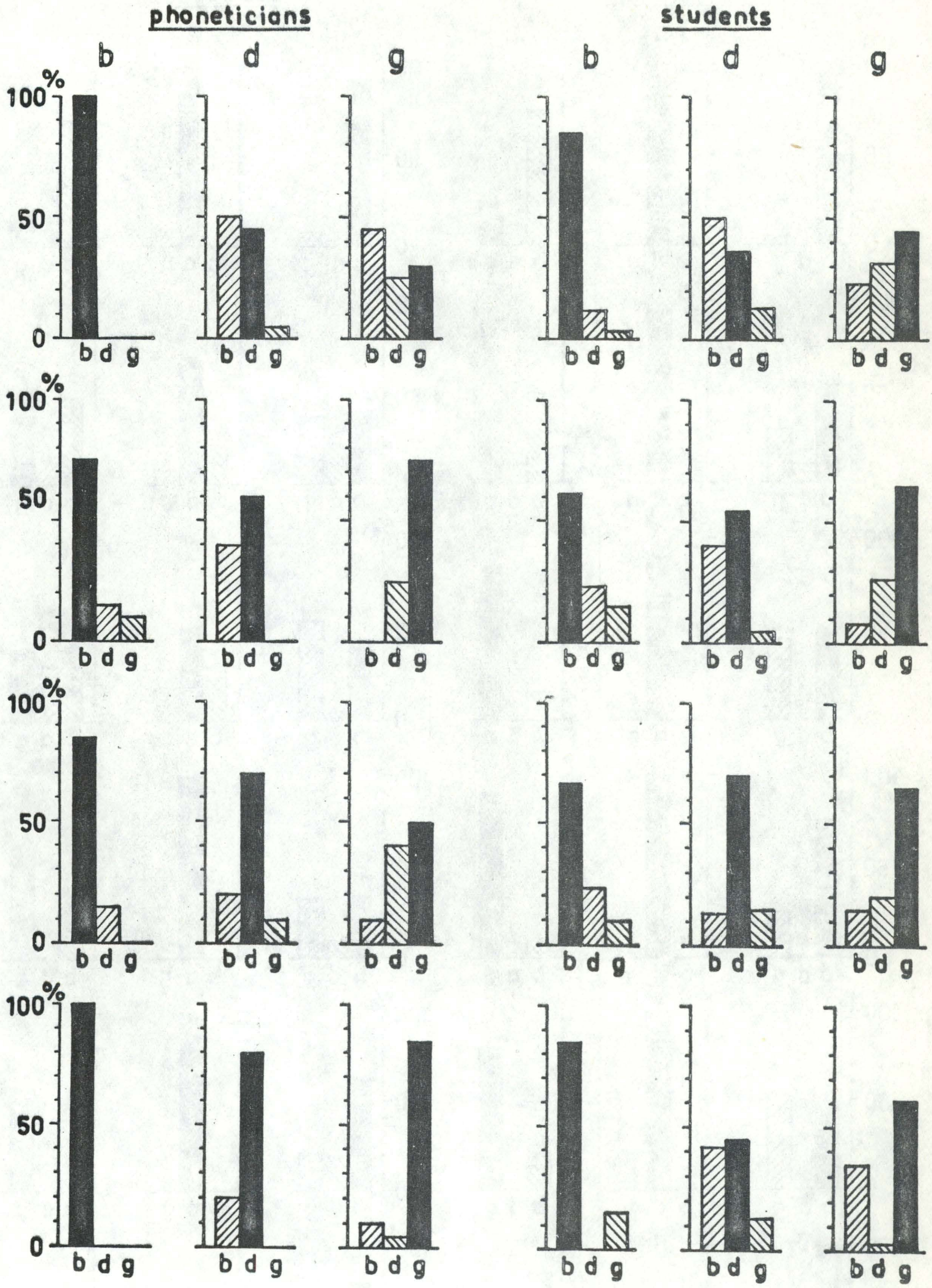


Fig. 3a. Identification of final unreleased *bdg* 1964, spoken by EFJ.
Listeners: 10 phoneticians & 20 students.

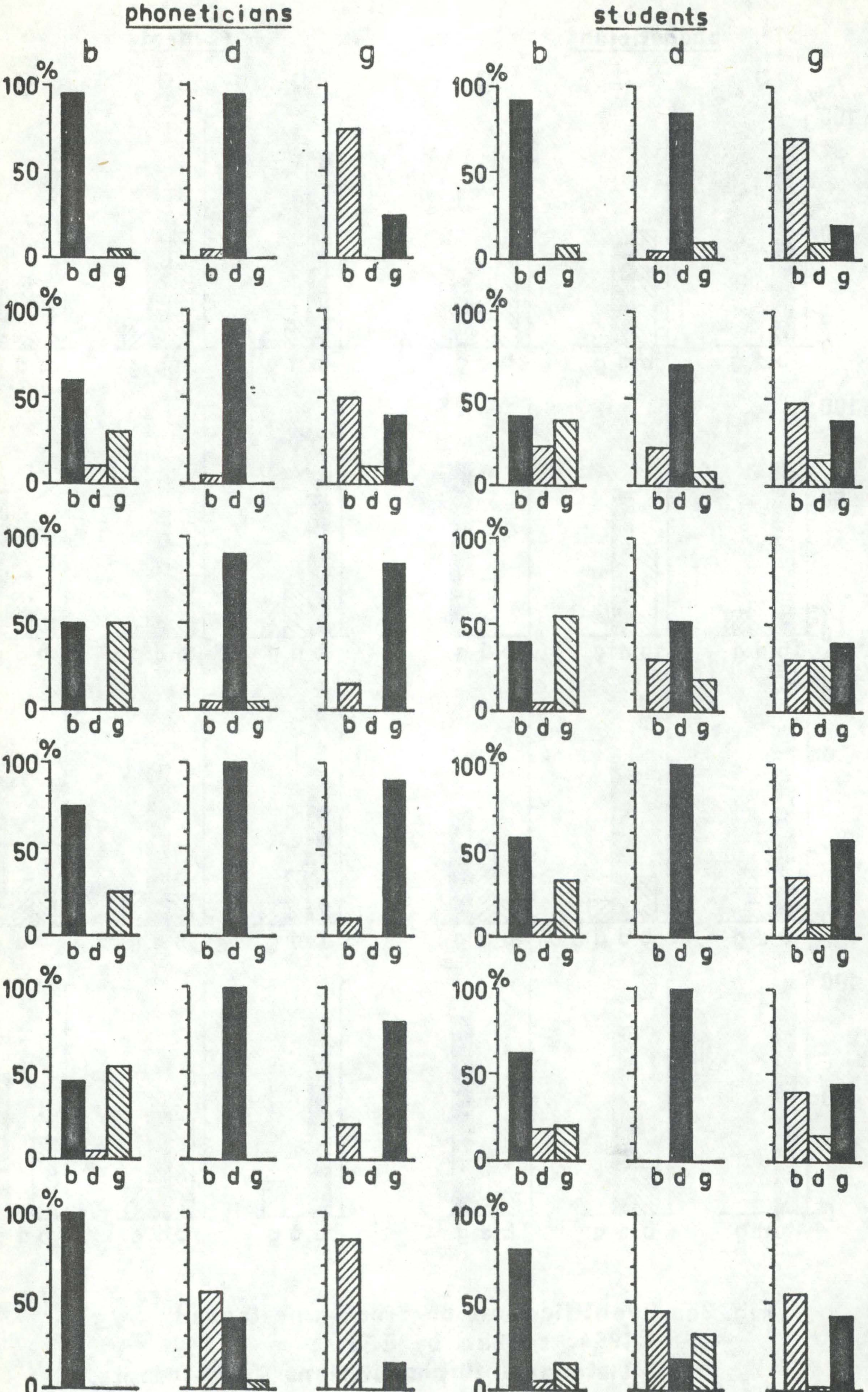


Fig. 3b. Continuation of 3a.

than the opposite. A further reason is general tendency,mentioned above, to choose labial in case of doubt.

(1b) g is heard either as b or d (it seems to be heard more often as d after i and a, and more often as b after e and ε, but these differences may be due to chance. They are difficult to explain).

(2) after rounded vowels

(2a) b and g are often confounded. The mistakes go both ways, except after y and ɔ where g heard as b is much more common than the opposite.
It is easy to understand that b and g are confounded after rounded back vowels, since they have almost identical transitions (straight or slightly positive F_2-transitions, straight or negative F_3-transitions). It is not quite as evident why they are confounded after rounded front vowels; but it may be explained by the fact that they have similar transitions of F_3, and that both may have negative F_2-transitions, although this is more stable in combination with labials than in combination with velars.

(2b) d is often heard correctly after rounded vowels, if not, it is heard as b.
It is easy to understand that d is heard correctly after rounded back vowels (u and o), but more difficult to understand why it is often heard as b after rounded front vowels and after ɔ. Here again, the main reason may be that the labial is the neutral consonant, most often heard when there is no reason to hear others. The transitions may be fairly level in rounded front vowels in combination with d.

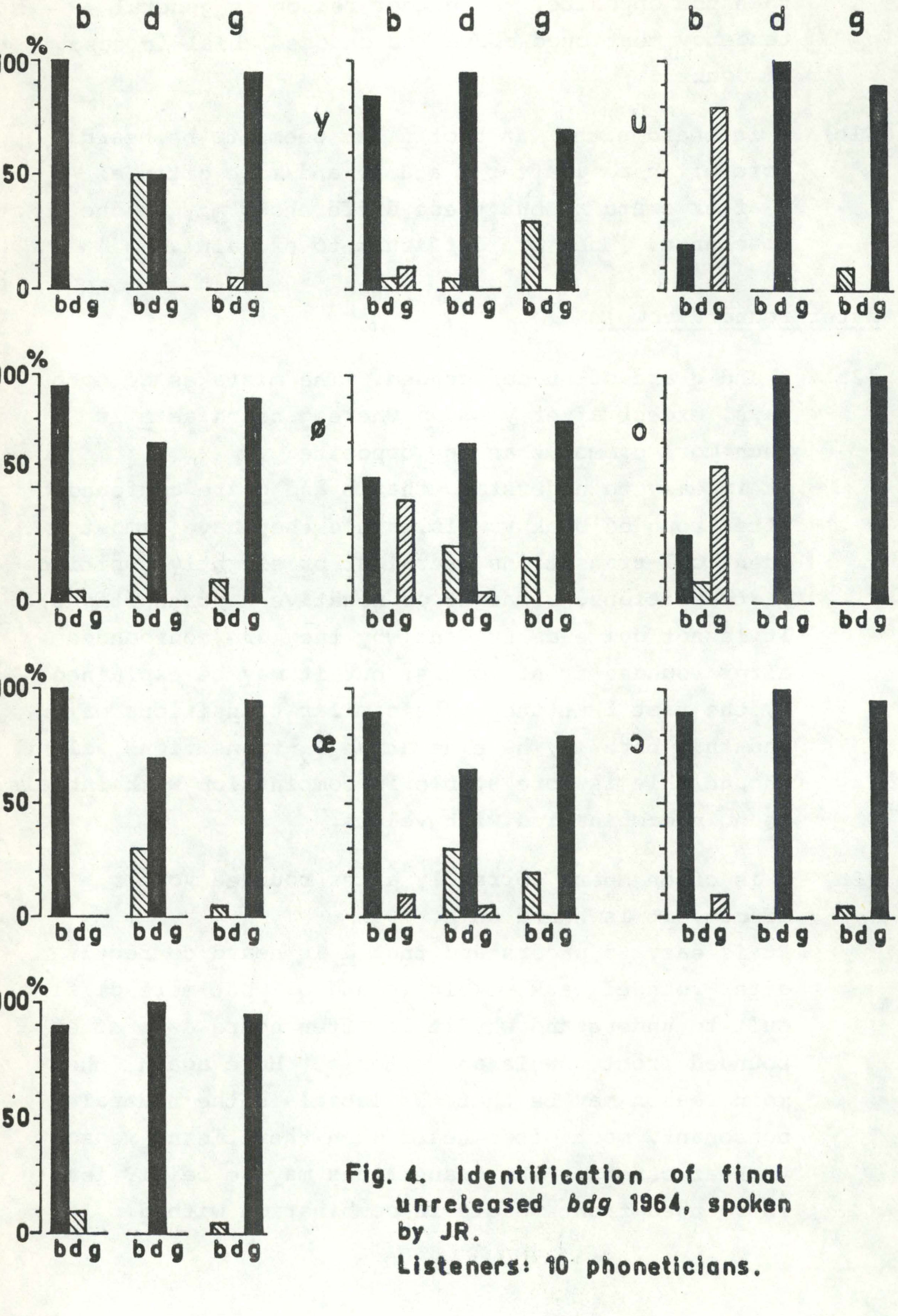


Fig. 4. Identification of final unreleased *bdg* 1964, spoken by JR.
Listeners: 10 phoneticians.

The tendencies found in test 2 are in agreement with the results of test 1, and they are for the main part confirmed by experiments with removal of initial explosions (see the article on "Tape cutting experiments with Danish stop consonants" later in this volume).

3. Comparison with other investigations

Since the experiments described here were started, a good deal of perceptual tests with unreleased stop consonants have been carried out, e.g. Halle-Hughes-Radley (1957), Householder (1956), Malécot (1958), Wang (1959), (all based on American English), Andresen (1960) (British English), Malécot-Lindheimer (1966) (French) and Kurt Johansen (1969) (Swedish).

The main result of the present investigation, i.e. that unreleased stop consonants are more easily identified when the transitions are extensive (e.g. ib, ud, ag) than when they are small (e.g. id, ig, ug), appears as a tendency in many of the above mentioned studies. In Halle-Hughes-Radley's test labial stops are identified better after i than after u, alveolar stops better after u than after i, and velars have a particularly high degree of correct identification after I and ʌ (which have pronounced formant transitions), and a low degree of correct identification after iː and uː. Andresen finds that p and k are identified better after I than after ɔ, t better after ɔ than after I, and Malécot finds the same for ɛ and ɔ, although the answers to ɔd and ɔt are only clear in the test with French listeners. Householder's results are not very clear. Wang-Fillmore examined consonant perception in noise and found that initial bilabials are identified with much higher correctness before i than before u, alveolars better before u than before i, and velars very poorly both before i and u, whereas all have relatively high percentages of correct identification before a and ɛ.

Both Wang-Fillmore and Halle et al. mention that an extensive transition may be more important for perception than a small transition. This has been contended by Delattre (1958), who maintains that the decisive thing is that the transition points to the locus of the consonant, and this is true of most transitions, with the exception of the transitions of rounded back vowels in combination with velars; in this combination the explosion is necessary for the identification. Delattre tries to interpret Householder's and Halle's results in accordance with this view, but he can only do so by leaving out i as "a special case". Malécot's result can be quoted in support of Delattre (ɔ + velar was difficult to recognize without release), but his material is too restricted. The present investigation demonstrates (i) that velars can be recognized without release after u o ɔ (cp. test 2 B) and (ii) that other combinations may be more difficult (e.g. id, ig). It is thus not sufficient that the transition points to the locus.

As for the mistakes made in identification, rule 1a (confusion of labial and alveolar after unrounded vowels) is on the whole supported by the results of Householder, Halle et al. and Malécot. Rule 1b (g is heard either as b or d) is not really a rule but rather lack of a rule, but the same irregularity can be seen in other investigations. Rule 2a (b and g are often confounded after rounded vowels) is only partly confirmed; velars are often heard as labials in Halle's and Householder's material, but not often vice versa.

Rule 2b (d is often heard correctly after rounded vowels; if not, it is heard as b) is true of long vowels in Householder's material, after short vowels t is heard either as p or k.

It can be concluded that whereas the absolute number of mistakes is evidently dependent on the precision with which the speaker articulates, the relative number in different combinations is subjected to some more general tendencies which are not in agreement with Delattre's hypothesis, and that there are also certain common traits in the type of mistakes made.

References

Andresen, N.S. 1960: "On the Perception of Unreleased Voiceless Plosives in English", Language and Speech 3, p. 109-121.

Delattre, P. 1958: "Unreleased Velar Plosives after Back-rounded Vowels", JASA 30, p. 581-582.

Halle, M., Hughes, G.W. and Radley, J.P.A. 1957: "Acoustic Properties of Stop Consonants", JASA 29, p. 107-116.

Householder, F.W. 1956: "Unreleased ptk in American English", For Roman Jakobson, p. 235-244.

Johansen, K. 1969: "Transitions as Place Cues for Voiced Stop Consonants: Direction or Extent?", Studia Linguistica XXIII 11, p. 69-81.

Malécot, A. 1958: "The Role of Releases in the Identification of Released Final Stops", Language 34, p. 370-380.

Malécot, A. and Lindheimer, E. 1966: "The Contribution of Releases to the Identification of Final Stops in French", Studia Linguistica XX, 2, p. 99-109.

Wang, W. S-Y. 1959: "Transition and Release as Perceptual Cues for Final Plosives", Journal of Speech and Hearing Research 2, p. 66-73.

Wang, W. S-Y. and Fillmore, Ch.J. 1959: "Intrinsic Cues and Consonant Perception", Journal of Speech and Hearing Research 4, p. 130-136.

TAPE CUTTING EXPERIMENTS WITH DANISH STOP CONSONANTS IN INITIAL POSITION

Eli Fischer-Jørgensen

1. Introduction

The experiments described in the present paper were inspired by the perceptual tests with stop consonants carried out by the Haskins group in the early fifties (e.g. Liberman et al. 1952, Cooper et al. 1952, Liberman et al. 1954) and by the tape cutting experiments of Carol Schatz (1954). I decided to work with natural speech, not only because we had at that time no synthesizer in Denmark, but also because it might be of interest as a necessary supplement to the work with synthetic speech. Most of the experiments were carried out in 1954-55, but they were supplemented at some points in 1972 (particularly concerning removal of explosions and transitions of bdg). The purpose was to find out what part of the acoustic sequence, bursts or aspirations and transitions, was the most important for the perception, and I wanted to use at least three different vowels, because I had a suspicion that the results might differ for different vowels. Some of the results were mentioned very briefly in a paper on the commutation test (1956), but a more detailed report has not been given until now. In the meantime others - and particularly the Haskins group - have worked with similar questions, but the results of the present experiments may still be of interest, not only because of the specific information they give about a language of the Danish type, but also because they throw some light on more general problems, particularly the importance of vowel transitions and some aspects of the locus theory.

2. Material and listeners

2.1. The material

The material consisted of Danish words with initial stop or fricative consonant followed by either a, i or u, viz. the words [panə, tanə, kanə, kalə, banə, danə, galə, bas, das, gas, pi:lə, ti:lə, ki:lə, kilə, bilə, dilə, gi̯lə, pu:ə, tu:ə, ku:ə, bu:ə, du:ə, gu:lə, - fanən, sanə, hanə, halə, fi:lə, si:lə, hi:lə, hilə, fu:ə, su:ə, hu:ə, hu:lə]. They are all existing Danish words, with the exception of [hi:lə]; two cases ([hanə] and [ti:lə]) are names.

The tests with fricatives will only be mentioned here in so far as they form part of the investigation of stop consonants. Only initial consonants were investigated, since Danish ptk and bdg are only distinguished in syllable initial position.

The words were spoken by me on tape several times, so that typical examples could be chosen for the tests on the basis of spectrograms. Two or three examples of each word were used in the test.

Fig. 1 shows schematic spectrograms of the stop consonants used in the test including the transition part of the following vowel. It appears from the figure that Danish ptk are strongly aspirated, and that t is affricated (more so before close than before open vowels, and particularly before i). It is possible to distinguish burst (in the following called explosion) from aspiration in p and k, and often explosion, fricative phase (with high frequency noise) and aspiration phase in t, e.g. in ta. But the two latter segments were not separated in the experiments, and they will be combined under the designation "aspiration" in this report. It may be difficult to separate the explosion of t from the following fricative phase, but in my pronunciation it is possible. The explosion is somewhat lower in frequency than the fricative phase and very weak. It should be borne in mind that

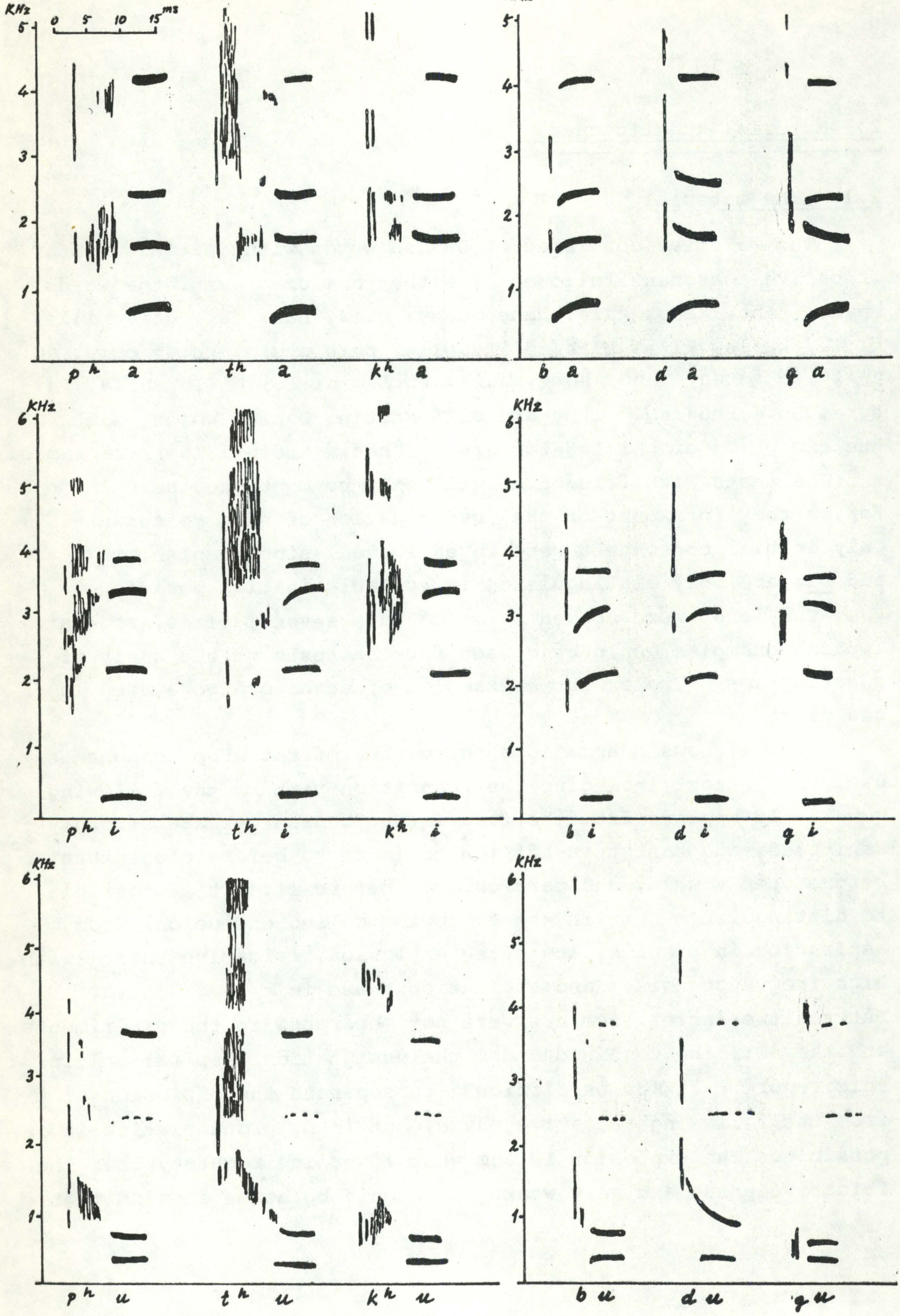


Fig. 1. Schematic spectrograms of Danish sound sequences used in cutting and splicing experiments.

in languages where t is unaffricated, a segmentation into a relatively short explosion noise (corresponding in frequency to the fricative phase of Danish t) and an aspiration phase (or nothing, if the consonant is unaspirated) will be more natural, and the removal of the explosion will not give the same result in these languages as in Danish. The duration of the aspiration in the words with ptk utilized in the experiment was (in ms) pa 60-80, ta 70-85, ka 60-70, pi 50-75, ti 75-90, ki 60-70, pu 55-70, tu 70-75, ku 55-65. The duration of the open interval in the words with bdg was (in ms) ba 5-10, da 10, ga 10-20, bi 10, di 10-15, gi 15-20, bu 10-15, du 10-15, gu 10-20. Exchanges of explosions may have involved small changes in the open intervals of bdg, of 5-10 ms, but this has hardly influenced the identification of the place of articulation, and the differences were too small to influence the distinction between ptk and bdg.

There is very little transition to be seen after ptk because most of the movements of the speech organs from the consonant to the vowel take place during the aspiration phase (this is clearly seen, e.g. in tu). After the strongly affricated t in ti there is, however, a pronounced transition of F_3, because the constriction of the vocal tract continues during the fricative phase, so that the movement to the vowel starts later than after p or k. Danish i is very fronted, so that the front cavity may become responsible for F_3, and this formant shows often more extensive transitions than F_2. The velars are more strongly influenced by the place of articulation of the following vowel than in English; for this reason there are hardly any vowel transitions in i and u after velars.

The fact that bdg are voiceless simplifies the cutting experiments.

The experiments consisted in removal of explosions, aspirations and transitions, and exchange of these segments, but always before the same vowel (in this respect the tests differ from those of Carol Schatz). All cutting and splicing was done by hand in 1954. For the supplementary test in 1972, some cutting was done by means of an electronic segmentator, but the splicing had to be done by hand anyway. Some cuttings done by hand were repeated by means of the segmentator. The two methods gave the same results. In all cases the cutting was sharp. Spectrograms were taken before and after each cutting and splicing.

The test played in 1954-55 consisted of 500 different items, the supplementary test 1972 contained 92 items. There were, however, only 1-3 examples of each single phenomenon (e.g. b-explosion removed before i), and the results must therefore be taken with some reservation. It would also have been preferable to have more than one speaker. But 592 words and 21 (20) listeners gave 12,340 answers, and without access to a computer the processing was rather time consuming.

2.2. Listeners and testing procedure

Each word was repeated three times on the test tape with half a second between the repetitions, and there was a pause of five seconds before the next test word. The tape was played over a loudspeaker in 1955 to a group af 21 listeners and over headphones in 1972 to a group of 20 listeners. They were all phoneticians or dialectologists. The listeners were asked to identify the words as existing (or at least possible) Danish words, and to make a note if they found the pronunciation abnormal. It has not been possible to find any differences in the answers indicating that the frequency of the word or even its existence played a role. On the contrary, although the subjects were asked to use normal Danish spelling, they sometimes made spelling mistakes showing that they had heard a Danish phoneme sequence and written it down in a way which was possible according to Danish spelling rules, but not the one used for the word in question.

In a certain number of cases they indicated that ptk were unaspirated. Less sophisticated listeners would probably have written bdg instead (cp. the paper on the identification of foreign stops by Danish listeners in this volume). But the answers given in this test have the advantage of giving some indication of cues that may be used in distinguishing voiceless bdg from unaspirated ptk.

Unfortunately there was some low-frequency noise on the tape on which the test words were recorded in 1955, but none on the tape used for intervals between the words. This had the result that the listeners sometimes heard an h or an f (particularly before u) where an initial consonant had been cut off, whereas the listeners to the supplementary test in 1972 more often heard ? or nothing.

3. Results concerning the distinction between ptk and bdg

In this section we are only concerned with the distinction between the two categories ptk and bdg, and not with possible mistakes in place of articulation within these two categories.

3.1. Identification of bdg after removal of explosions and/or transitions

In Table I the percentual identifications of bdg before i, u and a are indicated for four conditions: (1) unchanged, (2) both explosion and transition removed, (3) explosion removed, and (4) transition removed. The same is given in graphical form in Fig. 2. There were 6-9 different examples in each case, and as there were 20 or 21 listeners, the possible number of answers reach from 120 to 189. The actual numbers are indicated in the table.

TABLE I [1]

Identification (in %) of bdg, unchanged and after removal of transition and/or explosion

	i				a				u			
	N	ptk	bdg	x	N	ptk	bdg	x	N	ptk	bdg	x
1.+ex.+tr. BVV	144	-	99	1	186	1	98	1	186	-	97	3
2.-ex.-tr. (BV)V	180	1	14	85	140	-	1	99	183	4	25	71
3.-ex.+tr. (B)VV	144	1	48	51	186	3	86	11	186	2	56	42
4.+ex.-tr. B(V)V	120	-	97	3	183	2	42	56	120	-	91	9

(1) bdg are identified almost 100 % correctly in the unchanged words. Most exceptions are due to one example of [bu:ə]with very weak explosion, which has sometimes been heard as [fu:ə] because of the noise on the tape, and to one example of [dilə], sometimes heard as [gilə].

(2) If both explosions and transitions are removed, no stop consonant is heard before a, and only a small number before the other vowels. 25 % before u is however somewhat surprising. The consonant heard here is always b, and most examples belong to three individual words from the supplementary test, where the cutting has produced a slight click.

---

1) Here and in the following tables and figures the following abbreviations are used: B = bdg, P = ptk, V = vowel, V = vowel transition, H = aspiration, x = 0,?, h or f, () = removed, / = splicing between sounds from different words.

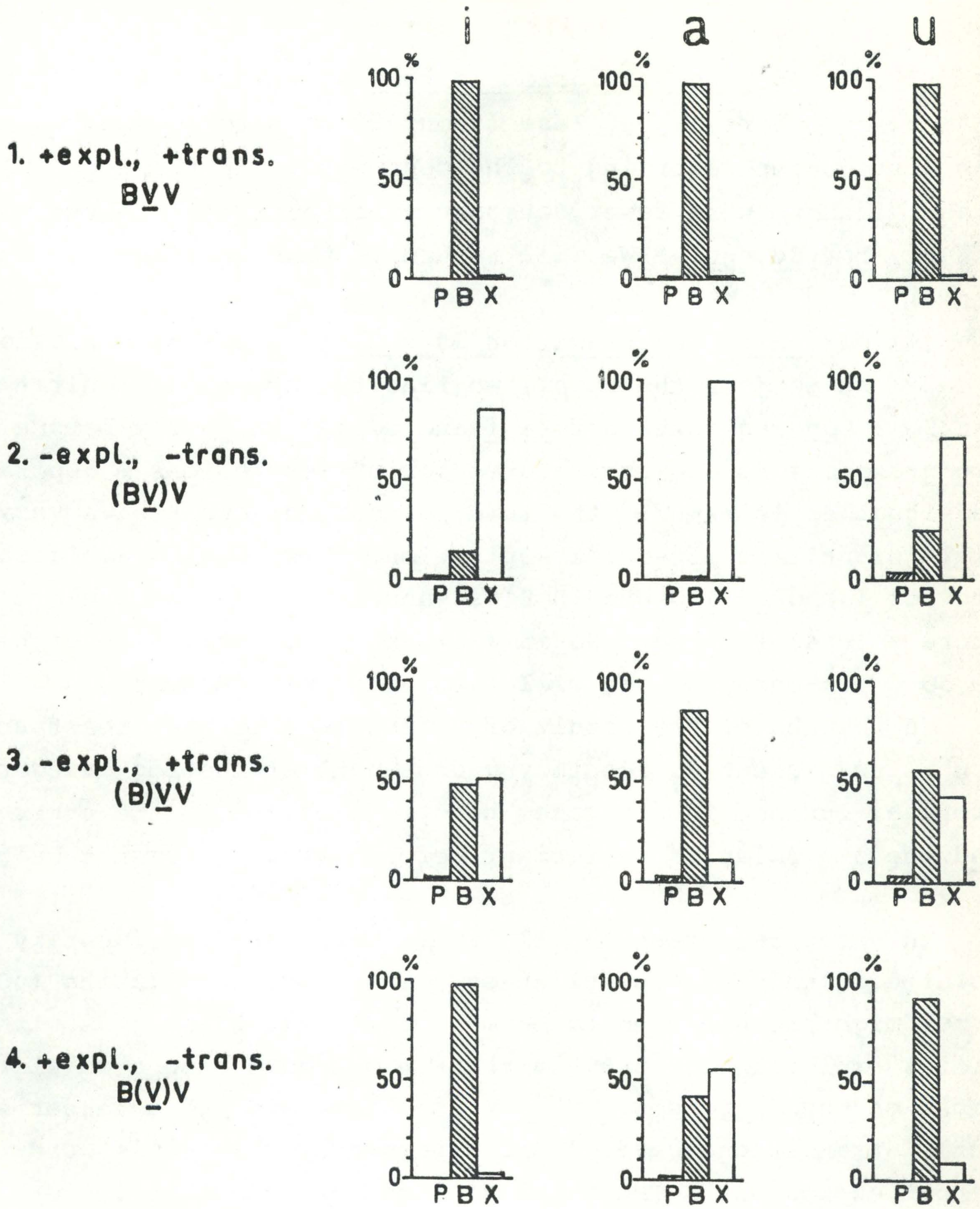


Fig. 2. Identification of *bdg* after removal of explosion and/or transition.

P = *ptk*-explosion
B = *bdg*-explosion
H = aspiration
V = vowel
$\underline{V}$ = vowel transition
X = other answers (0, *h*, *f*)
( ) = removed

(3) If <u>only the explosion is removed</u>, there are still 86 % <u>bdg</u>-answers before <u>a</u>, but less (about 50 %) before <u>i</u> and <u>u</u>, with some differences according to the consonant. In the words with <u>di</u> and <u>gi</u> there were fewer consonant-responses than in the words with <u>bi</u>, <u>bu</u>, <u>du</u>, <u>gu</u>. (We will return to this question in section 4).

(4) If <u>only the transition is removed</u>, and the explosion moved so as to join the vowel, we find the opposite result: more than 90 % <u>bdg</u>-responses before <u>i</u> and <u>u</u>, and only 42 % before <u>a</u>. If <u>bdg</u>-explosions are substituted for <u>ptk</u>-explosions + aspiration the situation is almost the same, since the vowels have very little transition after the aspiration. The results are also the same for <u>i</u> and <u>u</u> (100 and 90 % <u>bdg</u>-answers). The percentage <u>bdg</u> before <u>a</u> is 43 % (almost as in (3)), but instead of zero there are 50 % <u>ptk</u>-answers. We will return to this question in 3.7.

On the whole, the result of these tests is that the transition is sufficient to identify a consonant of the <u>bdg</u>-category before <u>a</u>, but not in all cases before <u>i</u> and <u>u</u>. On the other hand, the explosion is sufficient before <u>i</u> and <u>u</u>, but not before <u>a</u>.

In a separate test the listeners were asked to identify isolated transitions as syllables. These short syllables (40 - 80 ms) might be expected to be more difficult than words with the initial explosion cut out. <u>a</u>-transitions gave also a smaller number of <u>bdg</u>-responses, but <u>u</u>-transitions and (to a lesser extent) <u>i</u>-transitions gave more <u>bdg</u>-responses than whole words without explosions. The <u>u</u>-transitions were somewhat longer than the others.

The results are given in Table I a.

TABLE I a

Identification of stop consonant on the basis of isolated bdg-transitions

i				a				u			
N	ptk	bdg	x	N	ptk	bdg	x	N	ptk	bdg	x
61	-	63	37	81	-	79	21	61	-	61	39

3.2. Identification of ptk after removal of explosions and/or aspirations

In table II and in Fig.3 the perceptual identification of ptk before i, a and u is indicated for four conditions: (1) unchanged, (2) both explosion and aspiration removed, (3) explosion removed, and (4) aspiration removed. The same is given in graphical form in Fig.3. There were from 3 to 9 different examples of each case, and 21 listeners, which gives from 63 to 189 answers, as indicated in the table.

TABLE II

Identification (in %) of ptk, unchanged and after removal of explosion and/or aspiration

	i				a				u			
	N	ptk	bdg	x	N	ptk	bdg	x	N	ptk	bdg	x
1.+ex.+asp. PHV	126	99	-	1	189	98	-	2	126	100	-	-
2.-ex.-asp. (PH)V	63	-	21	79	63	61	17	22	63	2	65	33
3.-ex.+asp. (P)HV	63	100	-	-	126	98	-	2	189	74	7	19
4.+ex.-asp. P(H)V	63	-	97	3	63	27	70	3	63	-	100	-

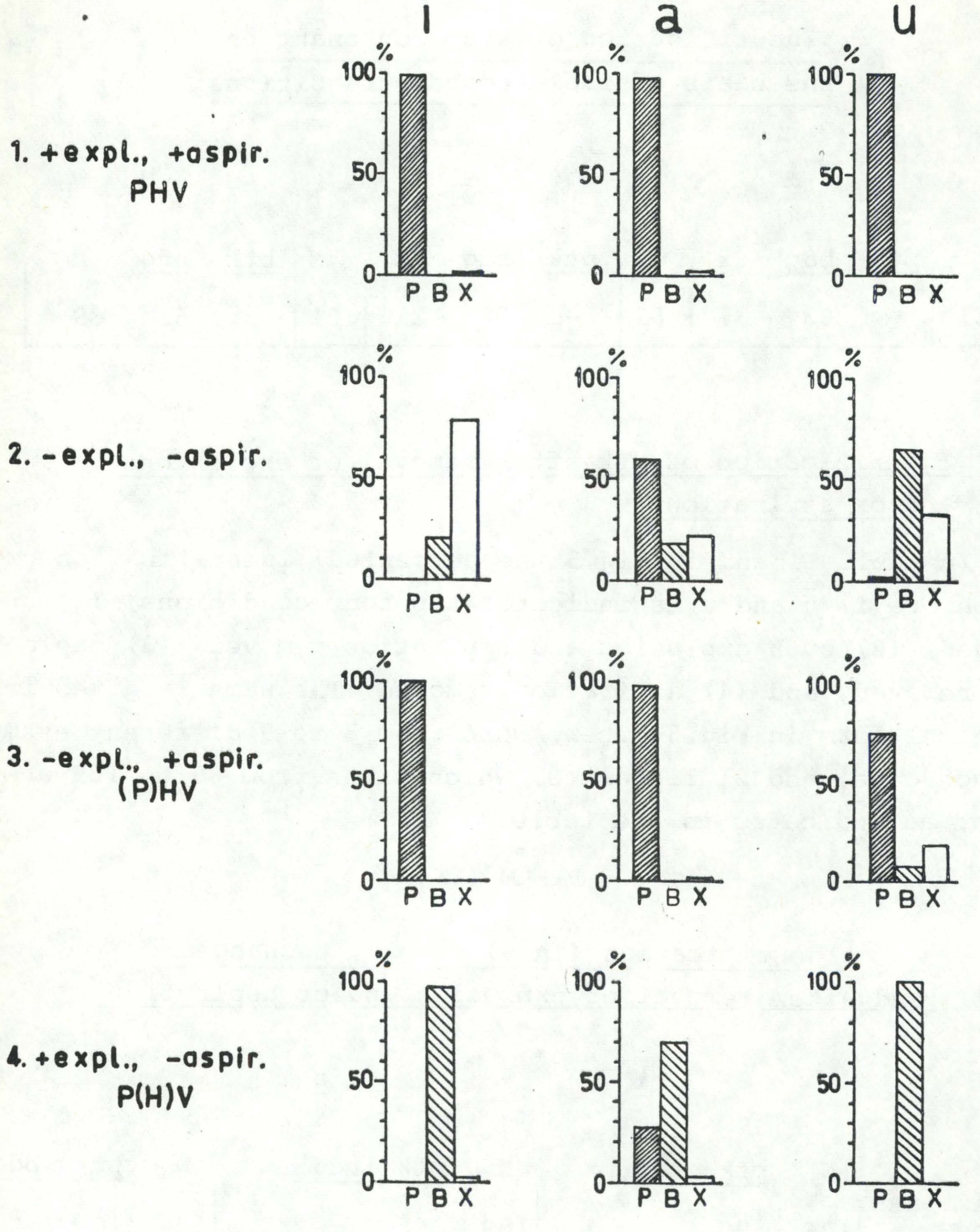


Fig. 3. Identification of *ptk*, unchanged and after removal of explosion and/or aspiration.

(1) Unchanged words are heard correctly in almost 100 % of the cases. The very few exceptions concern the word [kanə], sometimes heard as [panə].

When both explosion and aspiration are present, the start of the following vowel is not very important. Substitution of ph, th and kh for bdg + transition has been tried before a only. The result was 93 % ptk. Substitution of ph, th, kh for bdg-explosions (i.e. with the transitions retained) before i and u gave almost the same result (97 %), when one example is kept aside, namely kh for g in galde, which gives only 52 % k. (This k had a rather weak aspiration, and ga had a strong transition). ph, th, kh can also be identified in isolation.

(2) When both explosion and aspiration are removed, the result is in most cases zero before i, as might be expected, but there are 65 % bdg-answers before u (almost exclusively b), and before a there are 61 % p-answers and 17 % b-answers. An abrupt vowel start without pronounced transition is often interpreted as labial + vowel.

The labial seems to be the most neutral of the stops. (See also the paper on unreleased stop consonants in this volume). But a labial before an i involves an extensive transition of F_2 or F_3, and it is understandable that no consonant is heard when such transitions are absent. There are also more b-answers before u than before i when both explosions and transitions are removed from words with bdg (see 3.1), though the difference is less pronounced in this case. The problem of p-answers before a will be treated in 3.7.

(3) If only the explosions are removed, the words are still heard as starting with ptk in 90 % of the cases. And if one example of pu with a very weak aspiration is kept apart, the percentage ptk-answers is 97. This means that the aspiration is sufficient for the identification of the ptk-category. (ka is heard as pa, however).

(4) If <u>only the aspiration is removed</u>, and the explosion moved so as to join the vowel, the normal response is <u>bdg</u>, with a few exceptions before <u>a</u>.

The conclusion is that the aspiration is very important, whereas the explosion without aspiration is not sufficient to evoke <u>ptk</u>-responses.

3.3. Pause or aspiration noise?

The question may be raised whether a pause between explosion and vowel start is sufficient for the identification of <u>ptk</u>, or whether aspiration noise is necessary.

In order to investigate this question, the aspiration was removed in some cases and the explosion placed at a distance of 60-70 ms from the start of the vowel. The results are given in table III.

TABLE III

Identification (in %) of
ptk-explosion + pause + vowel

	N	ptk	bdg	x
i	63	38	29	33
a	63	65	28	7
u	63	33	41	26
aver.		45	33	22

A comparison with table II, 4 shows that the introduction of a pause of 60-70 ms between explosion and vowel start increases the <u>ptk</u>-responses considerably, but there is also an increase in zero-responses, and a majority of ptk-answers is only found before a. Before i and u the answer depends on the consonant-vowel combination. pi, ki, ku have 62 % ptk-answers, pu, tu, ti only 10 %. This means that, on the whole, the poorest ptk-identifications are found with consonants having a weak explosion and requiring a strong aspiration in normal speech (tu, ti), whereas consonants with stronger explosions and a normally weak aspiration (ku, ka) are identified more easily as ptk under these conditions. In ki the strong explosion may be the only cause for the good identification.

The general conclusion is that a simple pause is not sufficient for the identification of Danish ptk.

Before u a distance of 30 ms was also tried out. Although 60-70 ms should be the optimal distance, there were, on the average, almost as many ptk-answers when the distance was 30 ms (36 %). This apparent similarity is, however, due to an increase of pt-answers and a decrease in k-answers accompanied by an increase of g-answers. The explanation is probably that if weak explosions (as in pu, tu) are removed far from the vowel, they do not fuse with it to form a syllabic unit, and sometimes they are not heard at all.

3.4. Aspiration noise replaced by fricative noise from f,s,h

In order to see how critical the type of noise is, an experiment was carried out consisting in replacing the aspiration noise by approximately 50 ms fricative noise cut out of s, f, h taken from the same vowel environment; f from fi was thus substituted for the aspiration of pi, ti, ki etc.

The results are given in table IV.

TABLE IV

f, s, h substituted for aspiration
in ptk (answers in %)

	N	ptk	bdg	affr.	x
f	63	49	9	13	29
s	63	99	-	1	-
h	63	80	13	-	7

It can be seen from the table that s-noise evokes the highest number of ptk-answers, f-noise the lowest number. It should be remembered that in this section we are not concerned with the place of articulation, only with the difference between ptk and bdg. As a matter of fact s-noise gives t-responses in 92 % of all cases. - Without a preceding explosion an abruptly starting s-noise of 40-70 ms is heard 100 % as t. The explosion is thus of no importance here.

As for the effect of f-noise it depends chiefly on the following vowel. There are 81 % ptk-answers before u and only 34 % ptk-answers before a and i. The lower part of the f-noise is apparently related to the aspiration noise before u. Differences according to preceding consonant are less pronounced. The highest percentage is reached after k (68), the lowest after t (30), probably because a stronger explosion is needed, when the noise of the aspiration is not quite appropriate. The extremes are k/f/u (95 % k) and t/f/i (0 % ptk). p/f/u gives only 71 % pu (the other answers are fu and pfu). Abruptly starting f-noise without explosion gave f in almost all cases. The noise is apparently not strong enough to give the impression of a stop consonant until just before the vowel (25 ms f gives b), h substituted for aspiration noise gives ptk-responses in most cases, the exceptions are k/h/i and t/h/i, which are heard as gi, because they require a stronger noise.

An abruptly starting h without preceding explosion may be heard as p before a, but before i and u it is rarely heard as a stop.

3.5. The importance of the duration of the aspiration

A certain number of test words, both words with initial ptk and words with initial fricatives, had been cut off at different distances from the vowel start; but the number of examples is not sufficient and the cutting points not sufficiently parallel to allow precise conclusions. Shortening of the aspiration with retained explosion has not been tried. If my tests had been conducted after the appearance of the papers of Lisker and Abramson on VOT-value, I would have given more consideration to this problem. Moreover, cutting of natural sounds will give less clear results than synthetic sounds, because the aspiration noise is not homogenous in the time domain. The noise has, for instance, very often increasing intensity in pa and (rather suddenly) decreasing intensity in ta. And if the cutting point lies after the fricative phase of ta, the result is pa.

As mentioned above, only short durations of f-noise give stop-answers. Durations of 75,70 and 50 ms give f-answers, 15, 20 and 25 ms give b-answers for fi and fa, but f for fu. (Here the noise of the tape may have played a role). - Cuttings of h give very irregular results, from which no conclusions can be drawn.

The fricative s and the stop consonants give relatively clear results, but the number of examples are too small. In table V the answers of the majority for each consonant and duration is given. Underlining indicates a majority of more than

90%. - The answer f for pu is due to an example with very weak aspiration.

TABLE V

Answers to p, t, k and s cut off at different distances from the vowel start

	10	15	20	25	30	35	40	45	50	55	60	65	70 ms
pi				g									p
pu	f					f			f	p			
pa							p		p		p		
ti				d					t				t
tu	d				f				t		t		t
ta							t		t		t	t	
ki					g/o				k		k		
ku	b/f					p		k		k	k		
ka						k	p/k	k					
si			d						t				
su		d					t				t	t	
sa			d				t						

In spite of the restricted number of examples a pretty clear dividing line appears between 30 and 35 ms for g/k and between 25 and 35 ms for b/p and d/t. This is in complete agreement with the results from identification of foreign stop consonants by Danish listeners (see the report in this volume). But although a dividing line can be found for the majority of answers, there are still ptk-answers at much shorter VOT-values (see 3.7.).

3.6. Influence of explosion type (ptk or bdg)

The explosions of ptk and bdg are not very different. The only relatively stable difference seems to be that the explosion of k is somewhat stronger than that of g and that the explosion of t is somewhat weaker and lower in frequency than that of d.

Thus it is not astonishing that it makes no great difference to the identifications whether the explosions belong to one or the other type. Other cues are much stronger and override the small differences which may exist between the explosions.

This appears from a great number of identifications, which are combined in table VI. The examples are ordered in comparable pairs, where the only difference lies in the explosions. Only ptk- and bdg-answers are included, not zero, h etc.

TABLE VI

Comparison between answers (in %)
to stimuli with ptk- and bdg-explosions

		N	ptk	bdg
1a	BVV	516	-	98
1b	P/(B)VV	189	4	96
2a	PHV	441	99	-
2b	B/(P)HV	189	98	-
3a	P(H)V	189	7	90
3b	B/(PH)V	189	17	78
4a	P/f/V	189	49	9
4b	B/f/V	231	55	11
5a	P/s/V	189	99	-
5b	B/s/V	189	92	-
6a	P/h/V	189	80	16
6b	B/h/V	231	8o	11

It is evident from the responses listed in table VI that the type of explosion has been without any influence. The same appears from three other cases (all with 126 answers), viz.(1) substitution of aspiration for transition in bdg-words, which gives 99% ptk-answers, (2) substitution of transition for aspiration in ptk-words, which gives 100% bdg-answers, and (3) aspiration placed before the transition in bdg-words, which gives 98% ptk-answers.

In all these cases the difference between the explosions has been overridden by the strong cue: aspiration/lack of aspiration, or there has been fricative noise.

If there is only a pause between the explosion and the vowel start, the type of explosion has some effect. As mentioned in 3.3., some of the stimuli consisted in explosion noise + pause + vowel. If ptk-explosions are replaced by bdg-explosions in this context, there is a small decrease of ptk-answers. It has, however, only been tried before the vowel u. The results are given in table VII.

TABLE VII

Substitution of bdg for ptk before pause with aspiration removed. (Answers in %)

		N	ptk	bdg	x
1. 60-70ms	P-(H)u	63	33	41	26
pause	B-(PH)u	63	19	37	44
2. 30-40ms	P-(Hu)	63	37	49	14
pause	B-(PH)u	63	33	33	33

The difference is appreciable only with the longer pause, and here it is only due to the difference between g and k. In the case of the shorter pause, both p- and k-explosions give some more ptk-answers than b- and g-explosions, but for d/t the opposite is the case.

If the explosions in words with bdg are moved the same distances away from the vowel,there are more bdg-answers, but this depends primarily on the character of the vowel start (see 3.7.).

3.7. Importance of the vowel start

Although stimuli with a short open interval are normally heard as bdg, there is a good number of cases with ptk-answers. They are somewhat more common with velars than with alveolars and labials. This may be due to the stronger explosion of velars, since cases without explosion show the same number of ptk-answers for alveolars and velars.Much more evident is a difference according to the nature of the following vowel: ptk-answers are almost exclusively found in words with a and hardly ever in words with i and u. Finally ptk-answers are only found in cases where an aspiration has been cut off, not in cases where the vowel originated from a bdg-word. In table VIII and in Fig.4 the different types of stimuli without aspiration are indicated with the percentages of ptk- and bdg-answers (only the stimuli with a are considered, since the others show none or at most 3% ptk-answers).

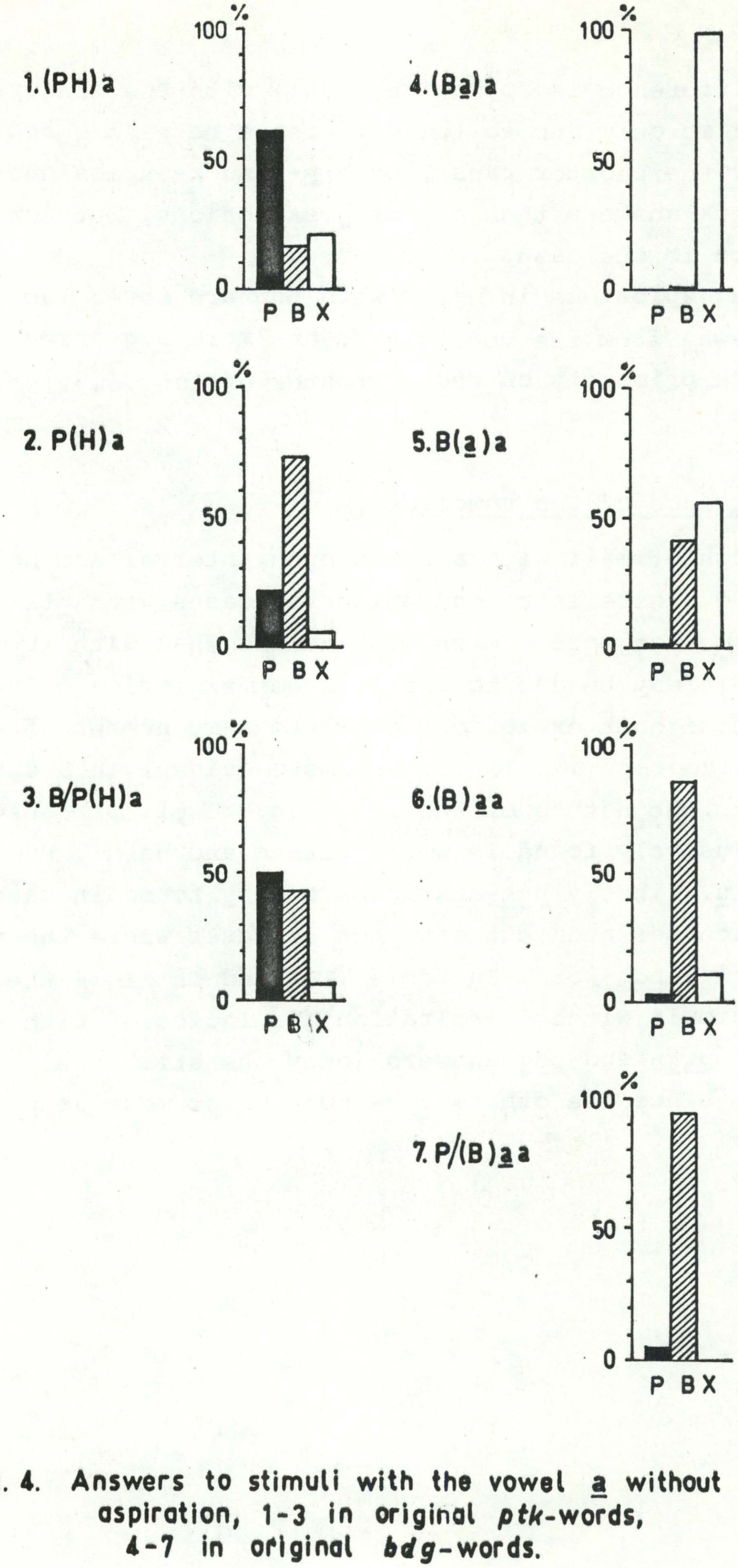


Fig. 4. Answers to stimuli with the vowel <u>a</u> without aspiration, 1-3 in original *ptk*-words, 4-7 in original *bdg*-words.

TABLE VIII

Answers (in %) to stimuli with the vowel a without aspiration. 1-3 original ptk-words, 4-7 original bdg-words.

	N	ptk	bdg	x
1. (PH)a	63	61	17	21
2. P(H)a	63	22	73	6
3. B/(PH)a	63	50	43	7
4. (Ba)a	140	-	1	99
5. B(a)a	183	2	42	56
6. (B)aa	186	3	86	11
7. P/(B)aa	186	5	95	-

It appears from table VIII that only the cases where an aspiration has been removed show a considerable number of ptk-answers, whereas the number of ptk-answers is insignificant if the vowel has followed a bdg-stop. A comparison between e.g. 3 and 7 shows that the explosion is irrelevant (there are 50% ptk-answers after a bdg-explosion, and only 5 after a ptk-explosion). The duration of the open interval is at most 35 ms. A closer inspection of the words with velars shows the intervals 20, 25 and 35 ms in the words with a majority of k-answers, and 10, 25, 30 and 35 in words with g-answers. Thus, the difference in the responses cannot be due to differences in open interval. Moreover, in (1) there is no explosion, and consequently no interval.

Two other explanations are possible: (a) The presence of transitions favours bdg-answers, whereas lack of transitions or rudimentary transitions favour ptk-answers. This would also explain why only the vowel a is involved. F_1 of a has a more pronounced transition than F_1 in i and u, and these may be necessary for the identification of bdg. However, this explanation is hardly sufficient.

It is true that (1), (2) and (3), which have many ptk-answers, are characterized by very rudimentary transitions because of the original aspiration, and that, on the other hand, (6) and (7), which show a majority of bdg-answers have full transitions. But (2) has a majority of bdg-answers in spite of its very slight transitions, and in (4) and (5), where the transitions have been removed altogether, there are no ptk-answers, but (in 5) some bdg-answers.

(b) The ptk-answers in (1-3) must therefore be due to a different cue in the vowel start. An inspection of spectrograms revealed that the vowel start after aspirated stops, particularly in the vowel a, is more irregular than after unaspirated stops. There are often one or two vibrations at a very low frequency before the start of the first formant, and at the same time there may be noise at the frequency of F_2-F_4, and these formants may continue to be somewhat noisy after the start of the first formant. This can be explained physiologically by the fact that the vocal chords are wide open at the explosion and may start to vibrate before they have reached complete closure. The early start may be more frequent before open vowels than before close vowels because the free passage in open vowels will cause the pressure above the vocal chords to drop more quickly. In Danish bdg, on the other hand, the vocal chords are close together already at the explosion, and a breathy start should not be expected (see Frøkjær et al. 1971).

The hypothesis that this vowel start is of importance for the identification is corroborated by a closer inspection of (1) and (2) and of individual examples of (2). At first sight it seems paradoxical that there are 61% ptk-responses to stimuli where both explosion and aspiration have been removed, and only 27% ptk-responses to stimuli where the explosion has been retained. But the cuts were slightly different in the two cases. In the latter case the cuts were closer to the vowels, in the former the cuts were placed a little earlier, and the irregular start with low frequency vibrations was retained. Moreover, besides the examples with the cut close to the vowel, experiment (2) contained two other examples of the same words where the cuts were placed 15 ms earlier in ta and 10 ms earlier in ka. This addition of 15 and 10 ms changed the responses drastically. The ptk-answers for ta were changed from 19 to 43% and for ka from 52 to 95%. The same words were cut 20 ms later with the result that the ptk-answers decreased to 14% both for ta and ka. (Only the words with the middle position of the cuts were included in the averages). The spectrograms (Fig.5) of the three different cuts show that it is not a question of vowel transitions. There are no transitions in the examples (c) heard as bdg by the majority of the listeners. - The very start of the vowel thus seems to be crucial.

In about half of the cases of ptk-answers to stimuli with removed aspiration, the subjects designated the consonants as unaspirated ptk, not normal Danish ptk-sounds. An inspection of some spectrograms, however, did not reveal a similar start of the unaspirated French and Dutch stops. Probably the glottis is less open in unaspirated stops than in aspirated stops, and a breathy start of the vowel is less probable, cp. that Slis and Damsté (1967) have found a wider glottis opening in voiceless fricatives than in voiceless stops in Dutch, whereas no such difference is seen in Danish.

The stimuli with pause between the explosion and the vowel are of interest in this connection. Unfortunately only words with u and i were used both with ptk- and bdg-stops and substitutions of bdg-explosions for ptk-explosions was only tried with u. But a comparison of individual words confirms the influence of the vowel start for these vowels, although less clearly than in the case of a. The answers are given in table IX. The alveolars are left out because the very weak t-explosion is only able to evoke a few ptk-responses under these conditions.

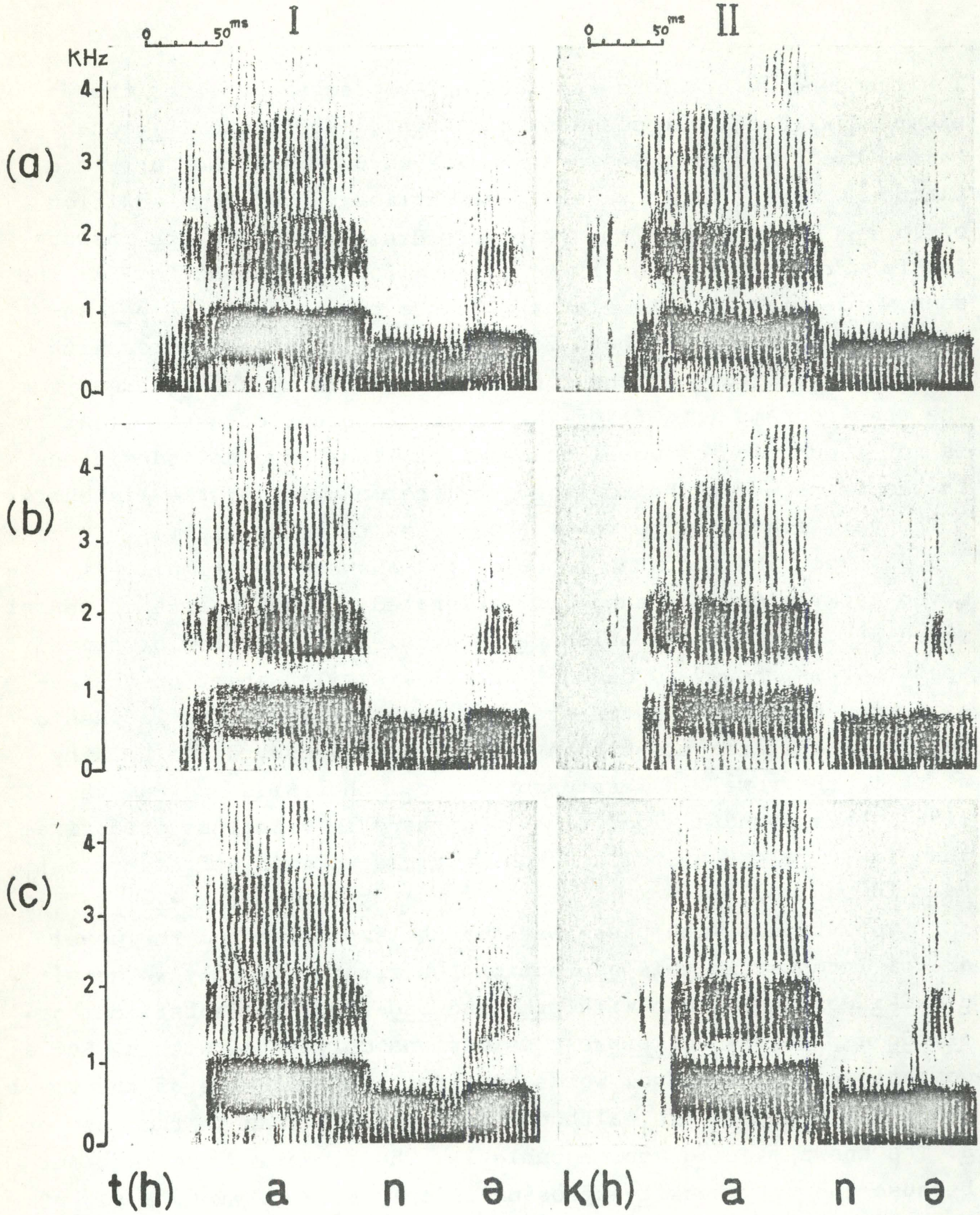


Fig. 5. Spectrograms of [t(h)anə] and [k(h)anə] with aspiration cut out (female voice, played at half speed, with expanded scale):
(a) cut at start of vibrations: I, 43% *ptk*, 48% *bdg*: II, 95% *ptk*, 5% *bdg*.
(b) 15 and 10 ms more cut off: I, 19% ″, 81% ″ ; II, 52% ″ , 48% ″ .
(c) further 20 ms cut off: I, 14% ″, 81% ″ ; II, 14% ″ , 57% ″ .

TABLE IX

Answers (in %) to different types of words with pause (60-70 ms) between explosion and vowel start (N=21)

	ptk	bdg	x		ptk	bdg	x
p-(h)i	71	5	24	k-(h)i	43	38	19
p-(b)i	52	48	-	k-(g)i	-	95	5
b-i	43	57	-	g-i	19	76	5
p-(h)u	9	48	43	k-(h)u	71	29	-
b-(ph)u	14	29	57	g-(kh)u	38	38	24
b-u	-	81	19	g-u	29	62	9

The answers to the words with i show that stimuli with original aspiration have a relatively high percentage of p- and k-answers, and words without original aspiration have a relatively high percentage of bdg-answers, whereas the difference in explosion type is almost irrelevant.

The words with u show that a high percentage of bdg-answers is found only in words without original aspiration, and that the explosion type has some effect for the velars (k is stronger than g).

4. Identification of place of articulation

4.1. Identification of place of articulation in bdg

4.1.1. Identification of unchanged bdg

In 3.1. it was mentioned that the identification of unchanged bdg-words was almost 100% correct, as far as the category of stops was concerned. As for the place of articulation, it is 100% correctly identified for b and g (there are a few f-answers for b), for d it is correctly identified in 92% of the cases. The mistakes for d are: 2% p before a, 6% b before u, and 17% g before i. The number of answers were 165 for b and d, and 186 for g.

4.1.2. Identification of bdg-words after removal or exchange of both explosion and transition

A. It was mentioned in 3.1. that when both explosion and transition are removed the most common answer was zero or ?. There were, however, a certain number of bdg-answers. The place of articulation heard in these cases is indicated in table X.[1]

---

1) Here, and in the following tables, a few ptk-answers are included in the bdg-answers, since only the place of articulation is relevant. x covers 0, ?, and sometimes h and f. Since the f-answers were caused by noise on the tape they are not counted as correct answers for labial place of articulation.

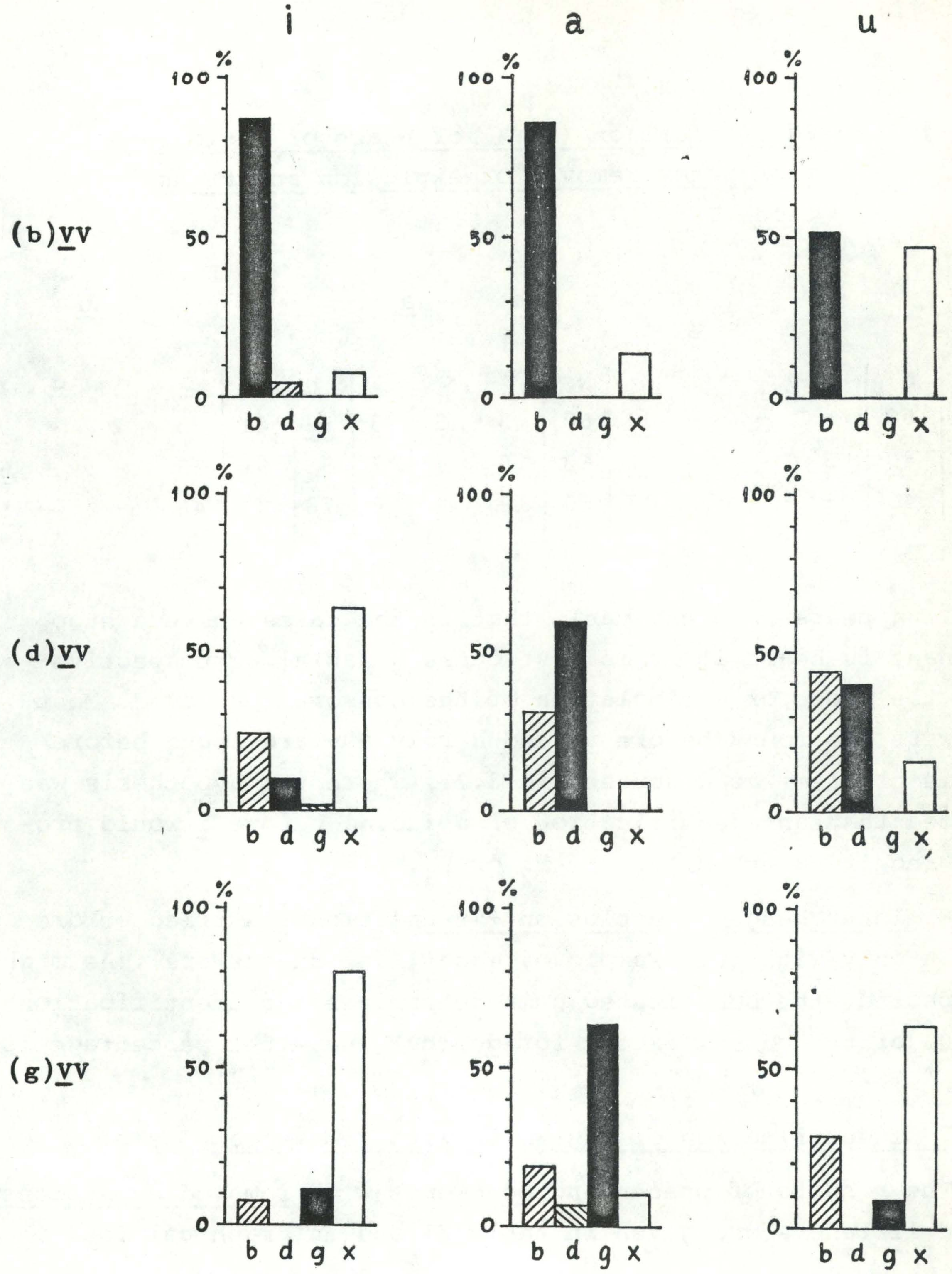


Fig. 6. **Identification of bdg + iau after removal of the explosion.**
V = transition, X = 0 (h, f)

TABLE X

Identification (in %) of place of articulation after removal of explosion and transition

	i					a					u				
	N	b	d	g	x	N	b	d	g	x	N	b	d	g	x
(bV)V	60	-	-	-	100	60	3	2	3	92	41	3	2	-	95
(dV)V	40	3	-	-	97	60	8	4	5	83	61	33	-	-	67
(gV)V	4o	-	-	-	100	60	13	4	5	78	81	48	-	-	52

It appears from the table that in the cases where a stop consonant is heard it is in most cases a labial, irrespectively of the place of articulation of the consonant removed. Most b-answers are found before u, and hardly any are found before i. This problem was discussed in 3.2., where the hypothesis was advanced that an identification of a labial before i would probably require a transition of F_2 or F_3.

B. Interchange of explosion + transition was tried before i and u only, with one example of each (i.e. 21 answers). As might be expected, the substituted unit determines the identification in 100% of the cases, except for du (bu) where the percentage is 95.

4.1.3. Identification of bdg-words with removed explosion

The results of presenting bdg-words with removed explosions to the listeners are given in table XI and in graphical form in Fig.6

TABLE XI

Identification (in %) of bdg-words after removal of the explosion

	i					a					u				
	N	b	d	g	x	N	b	d	g	x	N	b	d	g	x
(b)VV	62	87	5	-	8	42	86	-	-	14	62	52	-	1	47
(d)VV	62	24	10	-	64	82	31	60	-	9	62	44	40	-	16
(g)VV	61	8	-	12	80	62	19	7	64	10	62	29	-	7	64

The table and Fig.6 show that when the vowel is a, it is in most cases possible to remove the explosion and still get correct answers. Before i only b is heard correctly, before u b and d are recognized in about 50% of the cases, whereas g cannot be identified.

If the transition is cut out of words and presented alone, it is still possible to hear stops in many cases. The results are given in table XII.

TABLE XII

Identification (in %) of transitions cut out of bdg-words (N=40)

	i				a				u			
	b	d	g	x	b	d	g	x	b	d	g	x
(b)V(V)	90	-	-	10	70	-	-	30	82	-	5	13
(d)V(V)	42	-	3	55	15	35	3	47	25	60	-	15
(g)V(V)	7	3	43	47	12	3	45	40	45	-	20	35

A comparison with table XI shows certain deviations. da and ga are perceived less correctly, bu, du, and gi better than in words. This I cannot explain. Perhaps short transitions are heard better in isolation.

On the other hand the agreement is quite good between the test with removal of explosions in words and the test with unreleased bdg (see the report in this volume). In table XIII the two tests are compared. The percentages given for unreleased bdg are averages of the VC and the CVC test.

TABLE XIII

Comparison between correct identifications of words with removed initial bdg and words with unreleased final bdg

(b)i	87	i(b)	91	(d)i	10	i(d)	35	(g)i	12	i(g)	45
(b)a	86	a(b)	69	(d)a	60	a(d)	57	(g)a	64	a(g)	89
(b)u	52	u(b)	56	(d)u	40	u(d)	90	(g)u	7	u(g)	51

On the whole, the identification is better in final than in initial position. This is not surprising since the open interval of 10-30 ms in initial bdg cuts off part of the transition. Both finally and initially b is identified more correctly in combination with i and a than with u, d more correctly in combination with u and a than with i, and g more correctly in combination with a than with i and u.

As for mistakes, the number of zero-answers is greatest in the test with initial bdg. In most of the remaining mistakes a labial was heard. This was also the most common mistake in the test with final unreleased stops, but in some cases b was heard as g after a rounded vowel.

4.1.4. Identification of bdg-words after removal of the transition

The result of removing the transition in bdg-words is seen in table XIV and in Fig. 7.

TABLE XIV

Identification of bdg-words after removal of the transition

	i					a					u				
	N	b	d	g	x	N	b	d	g	x	N	b	d	g	x
b(V)V	40	70	8	10	12	61	18	-	-	82	40	82	-	-	18
d(V)V	40	-	80	20	-	61	23	21	-	56	40	60	35	-	5
g(V)V	40	13	5	82	-	61	25	-	44	31	40	5	-	90	5

A comparison between Figs. 6 and 7 shows striking differences between the results of removing the explosion and of removing the transition. In the first case all consonants were identified pretty well before a, in the second case they are identified most poorly before a. The opposite is true of i and u, except that bi was also identified well without explosion, and du is not heard better with explosion than with transition. The percentages of correct identifications for the different vowels with removed explosions are: a 70%, i 36%, and u 34%, with removed transitions a 24%, i 77%, u 69%.

A glance at the schematic spectrograms in Fig.8 will give an explanation to these differences. In ba, da, and ga the transitions are clearly different, and the explosions very similar. Therefore the transitions are more important for the perception,

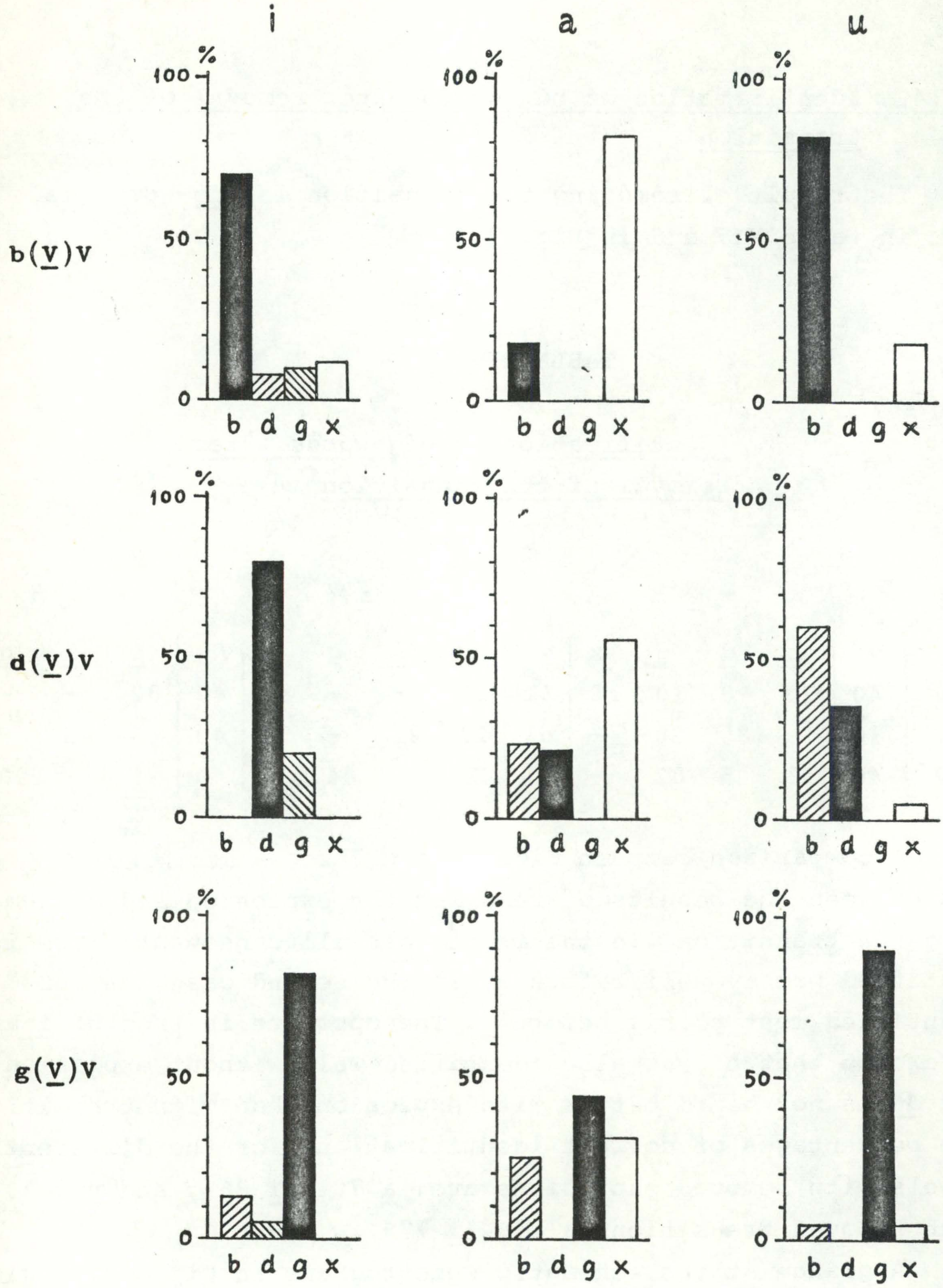


Fig. 7. **Identification of bdg + iau after removal of the transition.**
V = transition, X = O (h, f)

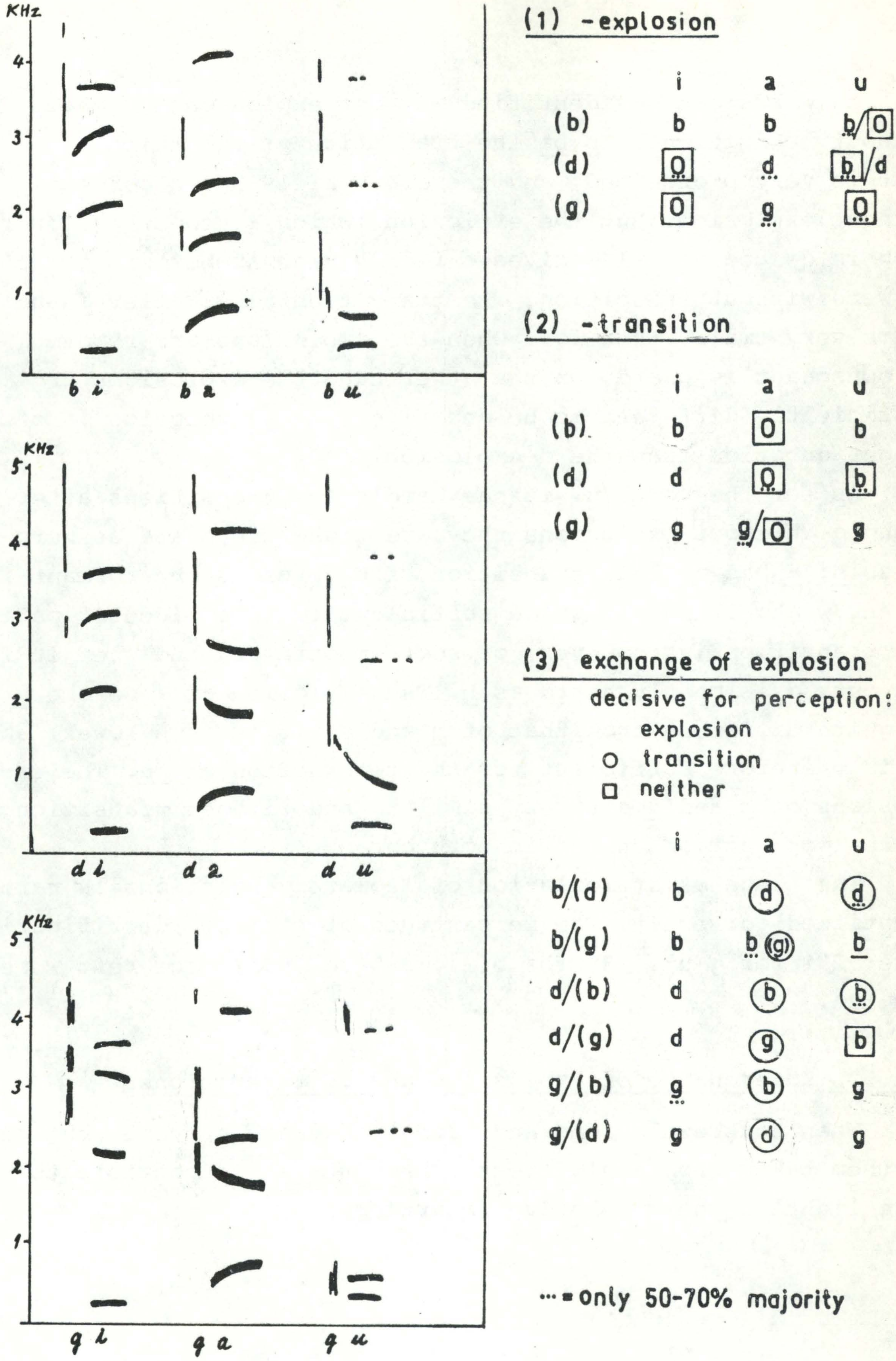


Fig. 8. Schematic spectrograms of *bdg + iau* together with a survey of the most frequent answers to stimuli with (1) removed explosions, (2) removed transitions, and (3) interchanged explosions.

and the words can be identified without explosions but not without transitions. In bi the transition of the third formant is very pronounced, and therefore bi is heard correctly without explosion. But the explosion (which is somewhat lower than in di and gi) also gives a fairly high number of correct answers without transition. The transitions in i after d and g are very small, therefore when the explosions are removed no consonant is heard. On the other hand the explosions are sufficiently different to be decisive (the g-explosion is of longer duration than the d-explosion).

As for the vowel u, it has hardly any transitions after b and g. Without explosions they are therefore heard as zero or as b; d has a clear transition of F_2, but as the formant is rather weak, it is not quite sufficient to allow identification. The transition is, however, of some importance, and when it is cut out, d is mostly heard as b. The explosion of g before u is quite different from that of b and d (it is much lower) and it is therefore sufficient for the recognition of gu. The explosions of b and d are very similar, and without transition b is the most common answer in both cases.

The place of articulation of isolated explosions is rarely identified correctly. The percentages of correct identifications were: 29% for g(u), 33% for b(i), 30% for d(i), the rest were below 12%.

4.1.5. Interchange of explosions and of transitions

When both explosions and transitions are present, but one of them taken from a different consonant, it is possible to throw more light on their relative importance.

A. Interchange of explosions

The result of interchanging explosions in bdg-words is given in table XV and in graphical form in Fig.9.

TABLE XV

Identification (in %) of bdg-words after interchange of explosions (N=21 for i and a, 42 for u)

	i				a				u			
	b	d	g	x	b	d	g	x	b	d	g	x
b(d)	71	24	5	-	5	95	-	-	33	64	-	2
b(g)	80	10	10	-	57	5	38	-	86	-	-	14
d(b)	-	100	-	-	100	-	-	-	68	10	-	21
d(g)	-	76	24	-	-	19	81	-	95	5	-	-
g(b)	29	10	57	5	76	10	14	-	-	-	100	-
g(d)	-	10	90	-	-	81	19	-	-	29	71	-

It is possible to summarize the results in a few simple rules:

(1) Before i the explosion is decisive

(2) Before a the transition is decisive, except for b(g)

(3) Before u the results are different according to the individual combinations.

These rules are given in schematic form in table XVI

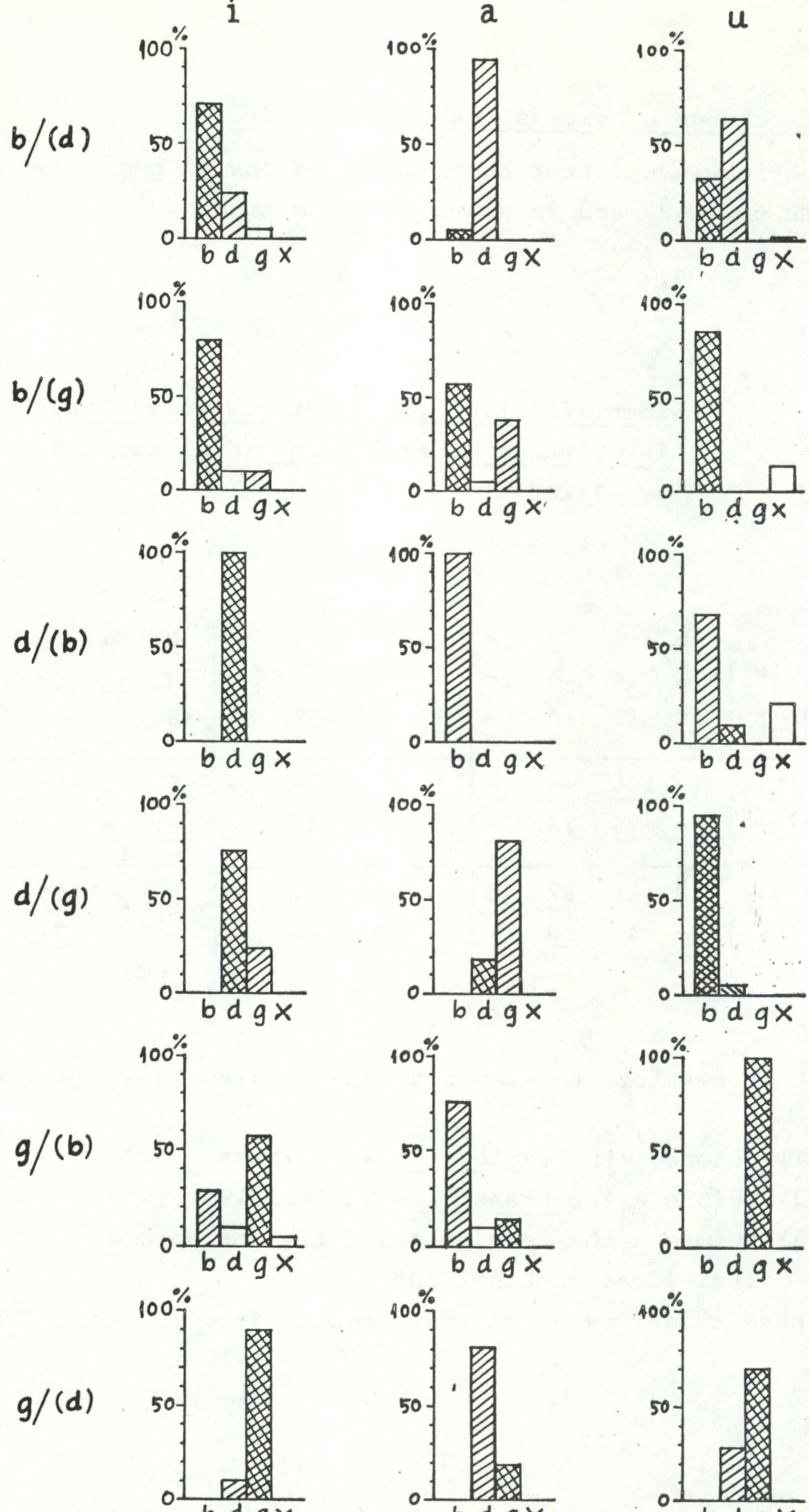


Fig. 9. Identification of *bdg* after interchange of explosions.
▩ = explosion, ▨ = transition.

TABLE XVI

The relative importance of explosions and transitions when explosions are interchanged in bdg-words (+=decisive, - not decisive)

	i		a		u	
	expl.	trans.	expl.	trans.	expl.	trans.
b/(d)	+	-	-	+	-	+
b/(g)	+	-	+	-	+	-
d/(b)	+	-	-	+	-	+
d/(g)	+	-	-	+	-	-
g/(b)	+	-	-	+	+	-
g/(d)	+	-	-	+	+	-

The schematic spectrograms in Fig.8 can again be used to explain the results.

As for the vowel i, di and gi have very small transitions and therefore the explosion must be decisive. This is also in agreement with the results from removing explosion or transition (Figs.6 and 7). One might have expected the strong transition in bi to have influenced the perception of d/(b) and g/(b) before i, but the explosions of d and g have probably been too high and too strong to allow interpretation as b.

The vowel a has different transitions after b, d, and g, whereas the explosions are not very different. Therefore the transitions determine the responses. This is in agreement with the results of removing explosions or transitions (Figs.6-7), where the transitions alone were found to be sufficient, but not the explosions. The exception b/(g) heard as b is somewhat astonishing.

As far as u is concerned, g has a characteristic explosion which must be decisive for g/(b) and g/(d), although the transition after d diminishes the majority for g somewhat. In b/(g) the explosion must also be decisive, because the explosions of b and g are very different and the transitions very similar. On the other hand, when b and d are interchanged, the transitions must be decisive because they are rather different, whereas the explosions are very similar. Finally, when a d-explosion is placed before a gu-transition, the result is b. This is not as surprising as it may look at first sight. The d-explosion is similar to the b-explosion. Thus it could be either d or b (but not g). The g-transition is similar to that of b, but different from the d-transition. Thus it could be g or b (but not d). The remaining possibility is b.

B. Interchange of transitions

An interchange of transitions should give the same result as an interchange of explosions, but the operation is somewhat more complicated and requires very precise splicing. It was only carried out for the vowel u, and it gave the same result as the interchange of explosions (table XV), but with less pronounced majorities and a small deviation for b/(d) which had the same number of b- and d-answers. The percentages are given in table XVII.

TABLE XVII

Identification (in %) of bdg before u after interchange of transitions (N=21)

	b	d	g	x
b/(d)	38	38	5	19
b/(g)	67	-	10	24
d/(b)	52	48	-	-
d/(g)	71	19	-	-
g/(b)	-	-	100	-
g/(d)	-	10	90	-

4.2. Identification of place of articulation in ptk

ptk are more complicated than bdg, because the aspiration covers most of the transition time, so that there is very little vowel transition, and because of the affrication of t. The answers therefore cannot be compared directly with those given to bdg-words, and the results apply more particularly to the Danish language than the results for bdg which can probably be generalized to other languages as well.

4.2.1. Identification of unchanged ptk

There were 21 unchanged ptk-words in the test, and thus 441 answers. The words with t were heard 100 % correctly, those with p 99 % (there was one k-answer for pa). For the k-words the percentage was 91. There were 8 % p-answers (all for ka) and 1 % zero (for ku).

4.2.2. Identification of ptk-words after removal or exchange of both explosion and aspiration

A. When both explosion and aspiration are removed there are still a good number of ptk-answers before a and of bdg-answers before u (see 3.2.). But the place of articulation is identified only in some examples of (th)i in which the opening movement starts rather late because of the strong affrication and which, therefore, have vowel transitions in F_3. Otherwise almost all stops heard are identified as labials. The results for the single consonant-vowel-combinations are seen in table XVIII.

TABLE XVIII

Identification (in %) of ptk-words with both explosion and aspiration removed (N=21)

	i				a				u			
	lab.	alv.	vel.	x	lab.	alv.	vel.	x	lab.	alv.	vel.	x
(ph)V	5	-	-	95	76	-	-	24	62	-	-	38
(th)V	14	19	-	67	86	-	5	9	43	-	9	48
(kh)V	19	-	5	76	67	-	-	33	81	-	5	14

Since labials are heard just as frequently when the removed consonant was alveolar or velar as when it was labial, it would be misleading to say that the labials were correctly identified before a and u. If the vowel starts abruptly and there are some rudimentary transitions indicating a preceding consonant, but no clear indications of the place of articulation, the normal answer is apparently "labial". The exception before i may, as proposed above (3.2. and 4.1.2.), be due to the fact that labials before i require extensive transitions of F_2 or F_3.

On the other hand, the isolated segments consisting of explosion + aspiration can on the whole be identified as far as the labials and alveolars are concerned, whereas the velars are badly identified. This was only tried before the vowels i and a. The percentage correct answers were

ph(i) 67%, th(i) 90%, kh(i) 33%

ph(a) 100%, th(a) 90%, kh(a) 24%

kh(i) was often heard as t, kh(a) as indefinite noise. This k had a relatively weak aspiration.

B. Interchange of explosion and aspiration as a whole gives the result that the substituted segment determines the identification. This is true in 96-100 % of the cases with the exception of kh/(th)i which gave 81 % k, and 19 % p.

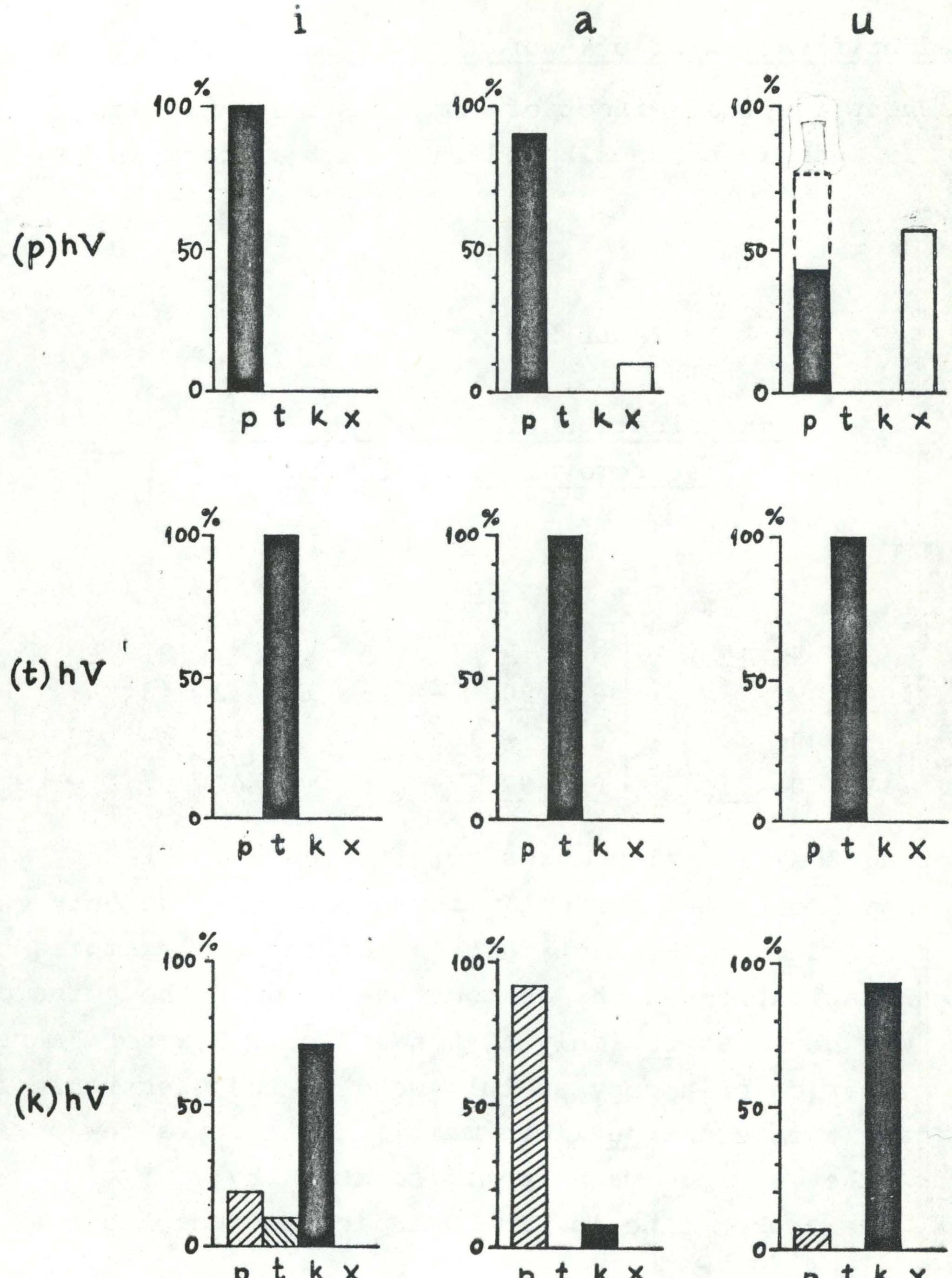


Fig. 10. Identification of *ptk* + *iau* after removal of explosion.

4.2.3. Identification of ptk-words with removed explosion

A. The perceptual effect of removing the explosion in ptk-words is seen in table XIX and in graphical form in Fig. 10.

TABLE XIX

Identification (in %) of ptk-words after removal of explosion

	i					a					u				
	N	p	t	k	x	N	p	t	k	x	N	p	t	k	x
(p)HV	21	100	-	-	-	63	90	-	-	10	42	43	-	-	57
(t)HV	21	-	100	-	-	63	-	100	-	-	42	-	100	-	-
(k)HV	21	19	10	71	-	63	92	-	8	-	42	7	-	93	-

Removing the explosion makes no difference at all for t, and p is also identified correctly in most cases. The only exception: pu is due to the fact that one of the two examples had a very weak aspiration which, in combination with the noise on the tape, was heard as fu (only 10 % heard pu). However, a renewed segmentation of a copy of this word on the electronic segmentator gave the result pu (informal listening by a few persons). The other example was identified as pu by 76 % of the listeners. The broken line in Fig.10 indicates that the result was probably not typical.

k is identified correctly in most cases before i and u, but it is heard as p before a. This answer, which is identical for all three examples, can be explained by the fact that the aspirations of p and k are very similar, whereas the explosions are different. The explosion of k is normally stronger, of longer duration (it often contains two maxima at a short distance) and somewhat more concentrated in frequency than that of p. When this characteristic explosion is lacking and only the aspiration is left, the listeners hear a p (see also Fig.12, where the main results of the tests with ptk can be compared with schematic spectrograms).

B. Interchange of aspirations (with removed explosions) was tried with p-, t- and k-aspirations (one example of each before i a and u). The aspirations were interchanged before the same vowel; thus there were 18 different items. In practically all cases the identification was determined by the substituted segment (100 % for p and t, 95 % for k). The pu-explosion was identical with the one which gave 76 % pu-answers before the interchanges. When substituted for (t)h and (k)h, it was heard 100 % as pu.

The aspiration is thus sufficient for the identification of ptk in all cases except for ka, and it does not matter if the following short vowel transitions are taken from other consonants.

4.2.4. Identification of ptk-words after removal of the aspiration

If the aspiration is removed and the explosion moved so as to join the vowel, most listeners hear bdg (with some ptk-answers before a, cp.3.2.). The same result is obtained when ptk-explosions are replaced by bdg-explosions (see 3.6. and 3.7.). In table XX and Fig.11 the answers are distributed according to place of articulation.

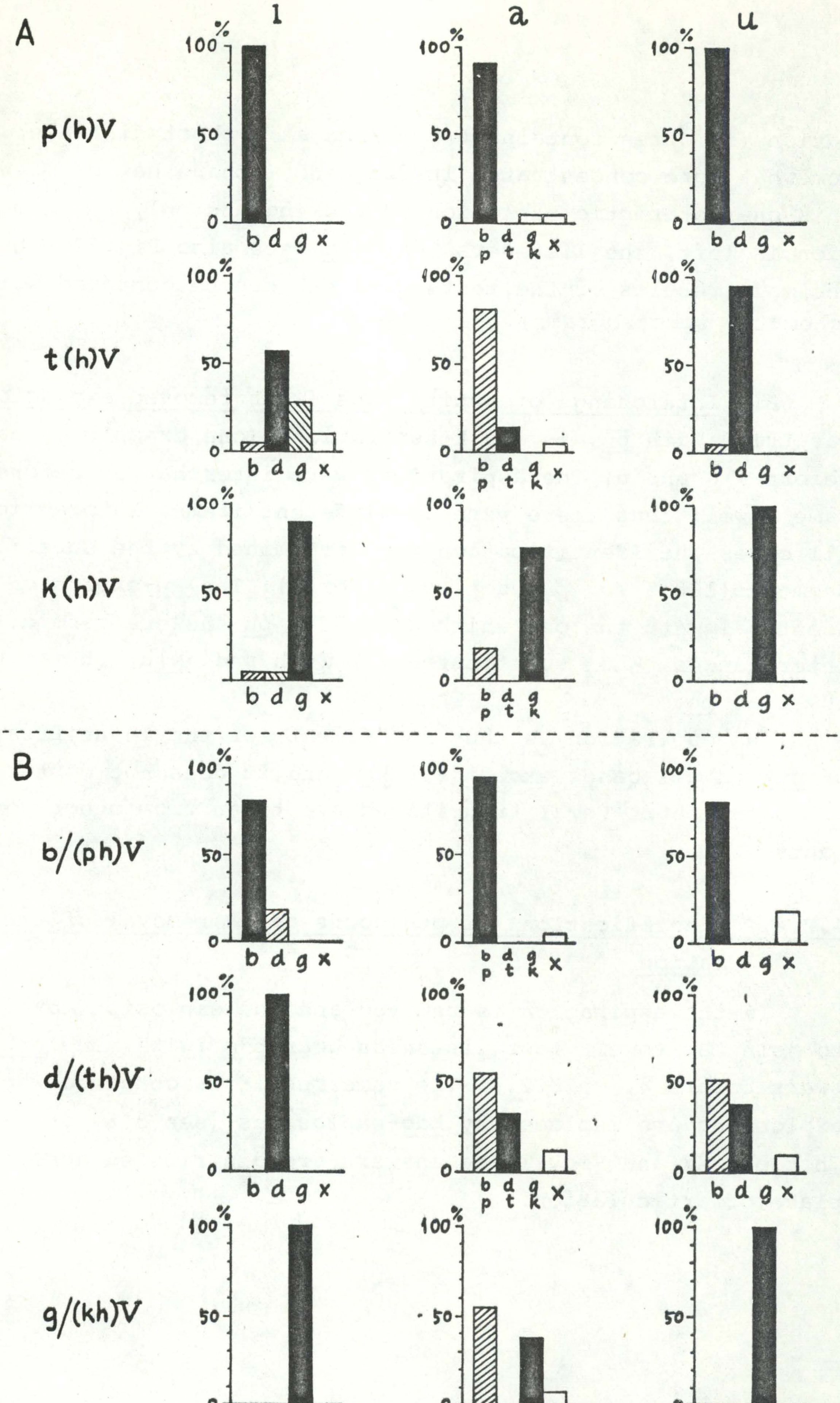
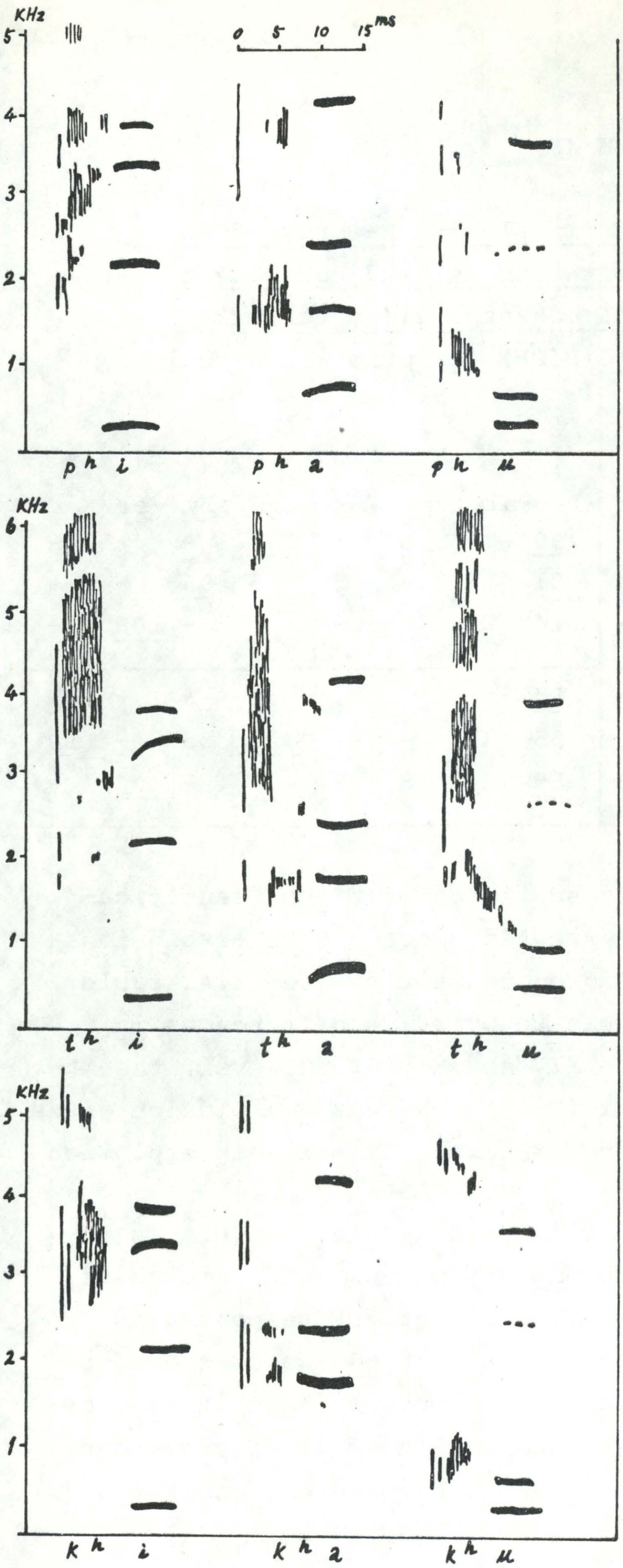


Fig. 11. Identification of place of articulation after removal of aspiration in *ptk* (A) and after subsequent replacement of *ptk*-explosions by *bdg*-explosions (B).



(1) –explosion

	i	a	u
(p)h	p	p	(p)
(t)h	t	t	t
(k)h	k	□p	k

(2) –aspiration

	i	a	u
p(h)	b	b	b
t(h)	d... □(g)	□b	d
k(h)	g	g	g

(3) exchange of explosion

decisive for perception:
explosion
○ aspiration
□ neither

	i	a	u
p/(t)h	○t	○t	○t
k/(t)h	○t	○t...	○t
p/(k)h	p	p	p
t/(k)h	○k	□p	□p
t/(p)h	○p	○p	○p... (t)
k/(p)h	k	k	k

... = only 50 70% majority

Fig. 12. Schematic spectrograms of *ptk+iau* together with a survey of the most frequent answers to stimuli with (1) removed explosions, (2) removed aspirations and (3) interchanged explosions.

TABLE XX

Identification (in %) of ptk-words after removal of aspiration (A) and subsequent substitution of bdg-explosions for ptk-explosions (B).(B(PH) 42, otherwise N=21)

	i				a				u			
	lab.	alv.	vel.	x	lab.	alv.	vel.	x	lab.	alv.	vel.	x
p(h)V	100	-	-	-	90	-	5	5	100	-	-	-
t(h)V	5	57	28	10	81	14	-	5	5	95	-	-
k(h)V	5	5	90	-	24	-	76	-	-	-	100	-
b(ph)V	81	19	-	-	95	-	-	5	81	-	-	19
d(th)V	-	100	-	-	55	33	-	12	52	38	-	10
g(kh)V	-	-	100	-	55	-	38	7	-	-	100	-

Before i the explosion is sufficient for the identification. This is what should be expected since the bdg-explosions were found sufficient with removed transition (see 4.4. table XIII and Fig.7)and were even able to override differences in transitions (see 4.1.5., table XIV and Fig.9). However, the majority is small for t before i (57 %), whereas d-explosion gives 100 % d. This difference can be explained by the weak explosion of t, particularly before i.

Before a most explosions are heard as labial. This is in accordance with the answers to bdg-explosions with removal of transitions in so far as the number of correct responses is approximately the same, but in the case of bdg-words the most common answer was zero, whereas in the present case it is p or b. This can be explained by the presence of rudimentary transitions and the particular vowel start after aspiration (see 3.7.).

Only the k-explosion is sufficiently strong to cause a majority of k-answers (see also Fig.12). The d-explosion is somewhat more efficient than the t-explosion because it is stronger.

Before u labial and velar explosions are sufficient. This is what could be expected on the basis of the experiments with bdg-words (table XIV and Fig.7). But in this case the t-explosion is superior to the d-explosion (the t in question had a relatively strong explosion, and the d-explosion contained energy at relatively low frequencies).

Before u the ptk-explosions were interchanged. But they were still decisive for the identification in 90-100 % of the answers. This also shows that the rudimentary vowel transitions are of very little importance.

4.2.5. Interchange of explosions and of aspirations

In the preceding sections we have seen that the aspiration alone (including affrication) is sufficient to allow identification of ptk, with the exception of ka, and that the explosion alone is sufficient to allow identification of the place of articulation except for ta and partly ti (bdg-explosions were not sufficient for da, du and ga).

Now the question is what happens when explosion and aspiration are in conflict. This question was investigated by interchanging explosions (this was done before i, a, and u) and aspirations (before i and u only).

The results of these two experiments are given in table XXI. Fig.13 contains a graphical representation of the former experiment (XXI,A).

TABLE XXI

A. Interchange of explosions in ptk (N=21)

		i				a				u			
expl.	asp.	p	t	k	x	p	t	k	x	p	t	k	x
p	(t)h	-	100	-	-	-	95	-	5	-	100	-	-
k	(t)h	-	95	-	5	-	57	14	29	-	86	5	10
p	(k)h	95	-	5	-	95	-	5	-	90	-	5	-
t	(k)h	10	-	90	-	100	-	-	-	100	-	-	-
t	(p)h	95	-	-	5	100	-	-	-	62	38	-	-
k	(p)h	19	-	81	-	14	-	86	-	-	-	100	-

B. Interchange of aspirations (N=21)

		p	t	k	x		p	t	k	x
p	(t)h	10	90	-	-		-	100	-	-
k	(t)h	-	100	-	-		-	86	10	5
p	(k)h	86	-	14	-		95	-	-	5
t	(k)h	10	-	90	-		71	24	-	5
t	(p)h	95	5	-	-		43	57	-	5
k	(p)h	71	-	24	-		-	-	95	5

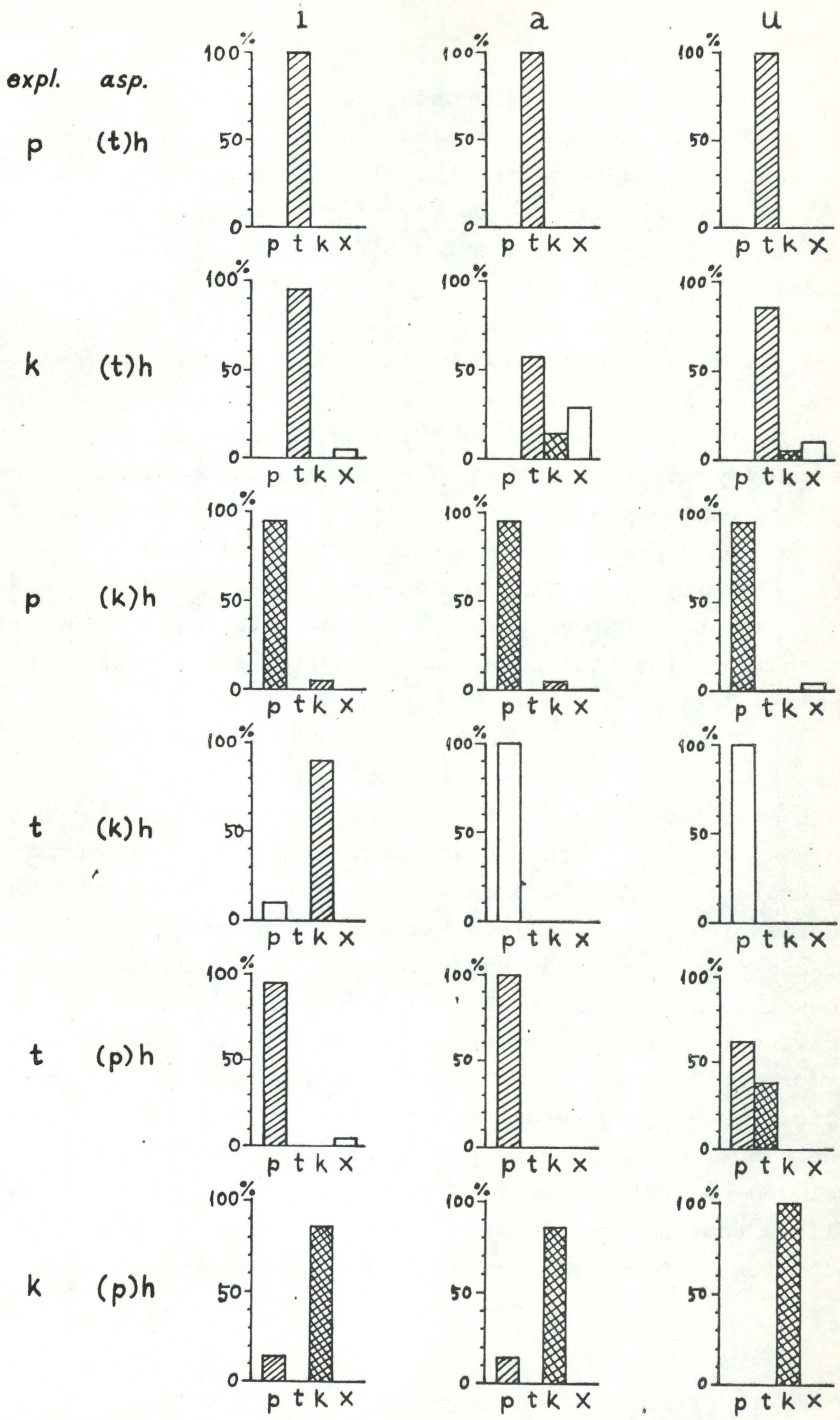


Fig. 13. Identification of *ptk* after interchange of explosions.
▩ = explosion, ▨ = transition

The answers are - with one exception - in good agreement in the two experiments, and the results can be described in relatively simple terms (cp. also the schematic spectrograms in Fig.12): (1) The characteristic affrication is necessary and sufficient for the identification of t.When it is present, the listeners hear a t irrespectively of the preceding explosion. The majority is large except for k-explosion + t-aspiration before a (57 %). Here the k-explosion is so strong that some listeners (19 %) hear a k and others (24 % of the 29 listed under x) hear the consonant cluster kt. When the t-affrication is absent, listeners fail to identify the t, i.e. the explosion is not sufficient. The only exception is t-explosion + p-aspiration + u in experiment B, which has a small majority for t (in experiment A there are 33 % t-answers. The t-explosion before u was relatively strong, and the p-aspiration looks like the later (low) part of the t-aspiration (see also Fig. 12)).

(2) In the other cases, i.e. in the examples with p- or k-aspiration, the explosion is decisive in the sense that p- and t-explosions give p, and k-explosion gives k. The only exception is t-explosion + k-aspiration before i, which gives k. The explanation of these answers is that p- and k-aspirations are relatively similar (see Fig.12), therefore the distinction between p and k must depend mainly on the explosions. k-explosions are characteristically different from both p- and t-explosions by being longer and stronger and (particularly before u) more concentrated. This explosion is necessary for the identification of k. p- and t-explosions are more similar to each other and might both give p or t. Since identification as t is excluded because the characteristic t-affrication is absent, it follows that not only must p-explosion + k-aspiration give p-responses, but so must t-explosion + p- and k-aspirations. The

exception t-explosion + k-aspiration before i heard as ki can be explained by the fact that pi requires either the explosion or the aspiration to begin at a rather low frequency.

The only evident case of conflict between the two experiments (A and B) is k-explosion + p-aspiration + i. In A there is a clear majority for k, which is in accordance with the rules given above. In B, however, there is a clear majority for p. The spectrograms look very similar, and when I listened to them again, I found that they also sounded alike;both of them gave the impression of a sound somewhere between p and k. In such a case the environment in the test may have played a role. The example heard as ki came (with one word between) after a clear example of pi, the example heard as pi came (with one word between) after a clear gi. But such a context effect in consonant perception with another word in between, is not very probable. Another explanation may be sought in the intensity level. The example heard as ki had a somewhat higher level than the other. Some informal experiments with changes of intensity level showed that this might have an effect on the identification as k or p.- Reactions to stimuli with two conflicting cues are, on the whole, subject to variations.

The importance of the affrication phase for the identification of t, also appears from the fact that a t-explosion put before h + vowel does not give t, but either a labial (before a and u) or a velar (before i). (p- or and k-explosions before h are heard as p and k respectively in almost all cases). Similarly, if explosion and affrication is cut off in ta 4o ms before the vowel start (i.e. after the affrication phase), the listeners hear a labial.

5. Summary of results

5.1. Summary of the results concerning the difference between the ptk- and the bdg-category

The primary cue for distinguishing between Danish ptk and bdg is the presence or absence of aspiration. ptk normally cannot be identified without aspiration, whereas the explosion noise is of very little importance (3.2.). It does not matter whether the explosion is a ptk- or a bdg-explosion (3.6.) or whether it is there at all (3.1.).The aspiration phase must contain noise; a simple pause is not sufficient (3.3.). The duration of the aspiration noise must exceed 3o ms to give a majority of ptk-answers, and a high percentage of ptk-answers is not reached until about 45-50 ms (3.4.).

A secondary cue is the vowel start, at any rate for the vowel a (3.7.). After an aspiration there may be some weak low frequency vibrations accompanied by somewhat noisy second and third formants. This breathy start may give a relatively high percentage of ptk-answers even at very short distances from the explosion (e.g. 2o ms). These ptk-sounds are, however, often described as "unaspirated ptk", i.e. not as normal Danish ptk-sounds,but different from bdg.

Lack of aspiration is insufficient for the identification of bdg. This requires the presence of explosion and/or vowel transitions (3.1.). The relative importance of explosion and transition depends on the vowel. Before i and u the explosions are necessary and sufficient, and the transitions constitute a secondary cue. (Transitions alone give only about 50 % bdg-answers). Before a the explosion is not sufficient and of minor importance.Here transitions alone give 86 % identification. The rudimentary transitions found after ptk may give some improvement compared to absence of transition (3.2.), but they are unable to give a majority of bdg-answers.

The aspiration-cue of ptk is much stronger than the transition-cue of bdg. If both are combined, the result is ptk (3.2.).

5.2. Summary of the results concerning place of articulation

The contributions of explosions, aspirations and transitions to the identification of place of articulation when each of them is the only present cue is indicated in schematic form in table XXII.

TABLE XXII

Contribution of explosions, transitions and aspirations to the identification of place of articulation (+= more than 70 % identification, ±=50-70 % identification, -=less than 5o % identification)

		i			a			u		
bdg	expl.	+	+	+	-	-	-	+	-	+
	trans.	+	-	-	+	±	±	±	-	-
ptk	expl.	+	±	+	+	-	+	+	+	+
	asp.	+	+	+	+	+	-	+	+	+

In the case of ptk both explosions and aspirations were found to be sufficient in most cases (but without aspiration the sounds were heard as bdg). The only exceptions were ka without explosion heard as pa, ta without aspiration heard as pa, and ti without aspiration which was often heard as gi.

For the identification of bdg the explosions were found to be sufficient before i and u (with the exception of du) but not before a. The somewhat lower efficiency of explosions in bdg-words than in ptk-words is probably partly due to the fact that g normally has a somewhat weaker explosion than k, and that b before a has a particularly weak explosion; an additional reason is that the explosions in ptk-words are probably supported by rudimentary vowel transitions. Why the d-explosion before u should be inferior to the t-explosion, is not quite clear. It may not have been typical (it contained rather low frequency noise).

The vowel transitions in bdg constitute a weaker cue than the aspirations in ptk, also for the place of articulation. They are, however, important in bi and in ba da ga (but not quite sufficient in da and ga). In bu the identification without explosion is just above 5o %, in du just below 50 %.

Figs. 8 and 12 give a survey of a somewhat different kind. Here the most frequent answers to stimuli with removed explosions or transitions (aspirations) are given together with the answers to exchange of explosions. Schematic spectrograms have been added as a visual support to the explanations.

Detailed explanations have been given in the preceding sections, see particularly 4.1.4. and 4.1.5. for bdg and 4.2.3. - 4.2.5. for ptk. Most identifications can be explained when it is assumed that extensive transitions and strong explosions (and aspirations) are more efficient than small transitions and weak explosions, and that transitions, aspirations, and explosions of a given consonant which differ characteristically in frequency or intensity from those of other consonants before the same vowel are more efficient than those which differ little from one consonant to the other before the same vowel.

Thus velar explosions differ from labial and alveolar explosions in being stronger, of longer duration, and with a concentration of energy close to F_2 of the vowel. They are therefore relatively strong cues; an exception is the explosion of *ga* which is less concentrated and more similar to that of *d*. Similarly the strong affrication of *t* makes it easily recognizable from *p* and *k*, which have mutually similar aspirations. The relatively pronounced and mutually different transitions of *ba da ga* constitute a better cue than the relatively similar explosions. As for *i* and *u*, on the other hand, only *bi* and *du* have extensive transitions. Therefore the transition is sufficient in *bi* and of some importance in *du*, though not sufficient in the latter case because the formant in question is weak.

The answer *b* for *d*-explosion + *g*-transition before *u* and the answer *p* for *t*-explosion + *k*-aspiration before *a* and *u* can also be explained in this way. These stimuli lack the characteristic velar explosion and the alveolar transition (or affrication), and the remaining possibility is thus a labial. On the whole, there is a tendency to hear labials when no other strong cues are present.

5.3. The reliability of the results

I have not found it worth while to test all the results mentioned in the preceding sections for significance. As each table contains information both about the percentual identifications and about the number of responses upon which the calculations are based, the reader will be able to judge about the reliability for himself. Moreover, in the cases with several examples the significance is often quite evident. This is particularly true of the responses to the categories ptk and bdg. In other cases,where sometimes only one example has been used (e.g. in some tests about place of articulation),a significance test might be misleading. If there is a high degree of unaminity among the 21 listeners, the identification may well be statistically significant in the sense that this unaminity cannot be due to chance. It might, however, be due to some atypical characteristics of the example in question, which does not allow of any generalisation. I have therefore found it more important to point to consistencies among different tests and to find explanations of the responses based on the typical features of the sounds in question. If such consistencies are found and such explanations can be given, the probability that the results can be generalized and that they tell something about the perceptual cues, is relatively high. Examples of consistencies are the dominance of the t-affrication before all vowels irrespectively of preceding explosions, or the weakness of the explosion as a cue for bdg before a, both when the transitions are removed, and when explosions are interchanged, also the importance of the transitions in the same cases. Examples of plausible explanations are the specific character of velar explosions, the similar and strong affrication noise after t before all vowels, and the extensive transitions in ba da ga bi.

6. Comparison with the results of other investigations

6.1. The distinction between ptk and bdg

In the literature on stop consonants there has been a certain tendency to talk about the various acoustic characteristics and perceptual cues without reference to any specific language. But ptk and bdg are labels covering rather different sound types, and the different cues and their relative importance must be described for different languages or language types separately. The confusion is considerably increased by the use of the terms "voiced" and "voiceless" in the same sense as bdg and ptk, i.e. as vague labels for a variety of sound types. bdg may not be voiced in the narrower sense of the word; they are voiceless in Danish, and may be voiceless initially for instance in English and German. The results found for Danish stops cannot be directly compared with the results of tests with French and Dutch stops, which are of quite a different type, but they can partly be compared with the results for English, since in both languages the main difference between ptk and bdg initially is one of aspiration. The aspiration is, however, on the whole somewhat longer in Danish, and the strong affrication of Danish t, involving a short and weak explosion followed by a long fricative phase gives some specific results which cannot be generalized to languages without affrication, (with a different segmentation, for instance the first 20 ms counted as explosion, the explosion phase would have been more like that of English t, but the "aspiration" phase would have been different anyway).

The importance of the aspiration cue for the identification of Danish ptk is in good agreement with the results of perceptual tests with English listeners. Aspiration is practically the same as "F_1 cut back" (Liberman et al. 1958) or "voicing lag", which has found to be the most important cue for the distinction between bdg and ptk in English (e.g. Lisker and Abramson 1964). Only it should be emphasized that pure voicing lag is not sufficient; there must be noise at the frequencies of the higher formants in the interval, as is also the case in the experiments by Lisker and Abramson.

There is also good agreement between the crossing point found in the experiment with English listeners and the approximate crossing found in the present experiment and in the experiment with identification of foreign stops. - I do not know of any experiments with breathy start of the vowel.

6.1. Identification of place of articulation

The Haskins group have made extensive investigations by means of synthetic stimuli in order to throw light on the importance of explosions and transitions in stop-consonant perception. They have been able to synthesize stimuli heard as labial, alveolar and velar stops both by means of explosions alone (Liberman et al. 1952, Cooper et al. 1952) and transitions alone (Liberman et al. 1954).

There is, however, particularly in the cases with explosion alone, a good deal of overlapping between the answers. The result of the present experiment, i.e. that the explosions are more important before i and u and the transitions in the case of a, does not appear from the experiments of the Haskins group. But a comparison between synthetic and natural sounds is not quite straightforward. The experiment with synthetic explosions may, for instance, have given relatively many velar responses because the noise employed as a stop cue was of relatively long duration (15 ms) and of concentrated frequency, and therefore more like a velar explosion than like an alveolar or labial explosion. The experiment with synthetic formant transitions shows better discrimination for a than for i and u in the section with F_1 transition and bdg-answers, but not in the section with ptk-answers. di gives a particularly bad result (like in our experiment), but it is improved by the addition of F_3 (Harris et al. 1958 and Hoffman 1958). Also ga gives an unsatisfactory result, but it is improved by adding a third formant (cp. the results Harris et al. for gæ).

On the whole, the transitions in natural Danish words do not seem as efficient as the synthetic transitions, and there is a certain contradiction between our results and the formulation of the locus theory found in the later writings of P. Delattre (e.g. 1958 and 1969). According to Delattre the locus is the frequency toward which the formant transitions must point in order to contribute maximally to the perception of a given place of articulation as found in experiments with synthetic speech. In natural speech the formant transitions do not point to the locus in the case of velars before rounded vowels because the coarticulation, involving rounding already in the consonants, lowers the resonance. Therefore explosions are needed in this case, but not in others. Delattre critisizes Halle and Householder who have assumed that an extensive transition contributes more to perception, and states that the extension of the transition is irrelevant; it may be straight. What matters is that it points to the locus. Delattre's own experiments seem to corroborate this assumption (Delattre 1969), but the results obtained with Danish stops confirm the hypothesis of Halle and Householder (cp. also Wang and Fillmore 1961): that it does matter whether the transition is extensive or not. And this seems quite natural. If the transition is straight, the listener cannot know whether it points to a consonant locus or to nothing. It must have a certain extension in order to be an efficient cue. The results of the experiments with Danish stops show, for instance, that the explosion is not only necessary in the case of velars in combination with rounded vowels, but also in di, gi, bu, and du. In di and gi the transitions alone give only about 10 % d- and g-answers, in bu and du around 50 %. Even in da and ga the percentage reaches only 60 and 64 % without explosion. But the more extensive the transitions are the more they contribute to the identification.

The importance of explosion was also emphasized in Malécot's experiments with English and French final stops (Malécot 1958, Malécot and Lindheimer 1966), but transitions alone gave also a relatively high percentage of identification (after the vowels ɛ and ɔ).

The importance of the initial explosions before i and u found in the present investigation has also some interest for the general theory of speech perception. On several occasions (e.g. 1969) Liberman has emphasized that speech is characterized by parallel transmission and overlapping cues (since the transitions contain information about both consonants and vowels) and that the characteristic feature of speech perception is therefore the decoding of such overlapping cues. He makes an exception for fricatives (interchange of Danish f and s before i, a and u also shows that the fricative noise is alone decisive here), but he might have made an exception for some combinations of stops and vowels too. The examples most frequently quoted by Liberman are bag and the syllables di and du. bag is in fact a good example, but it is remarkable that the transition in spoken Danish di seems to be of no importance for perception. It cannot be denied that overlapping cues are very often found in speech, and it remains the merit of the Haskins group to have demonstrated this fact, but its role in the perception of natural speech may have been somewhat overemphasized.

The finding that the most common answer is labial when no pronounced cues are present has been confirmed by Öhman (1961) and by Wajskop (1971). Wajskop finds that the stimuli without transitions or explosions are most often heard as labials, but if transitions are present the most common mistake is t. It is valid for Wajskop's material, where velars are often heard as dentals. It is, however, an observation which can hardly be generalized. t is also a common mistake in my experiments with final stops after unrounded vowels, but not in other cases (see the paper on final unreleased stop consonants in this volume).

References

Cooper, F.S., P.C.Delattre, A.M.Liberman, J.M.Borst and L.J.Gerstman 1952: "Some Experiments in the Perception of Synthetic Speech Sounds", JASA 24, p. 597-606.

Delattre, P.C. 1958: "Unreleased Velar Plosives after Back-rounded Vowels", JASA 30, p.581-582.

Delattre, P.C. 1969: "Coarticulation and the Locus Theory", Studia Linguistica XXIII, p.1-25.

Fischer-Jørgensen, Eli 1954: "Acoustic Analysis of Stop Consonants", Miscellanea Phonetica II, p. 42-59.

Fischer-Jørgensen, Eli 1956: "The Commutation Test and its Application to Phonemic Analysis", For Roman Jakobson, p. 140-151.

Fischer-Jørgensen, Eli 1969: "Voicing, Tenseness and Aspiration in Stop Consonants, with Special Reference to French and Danish", ARIPUC 3, p. 63-114.

Frøkjær-Jensen, B., C. Ludvigsen and J.Rischel 1971: "A Glottographic Study of Some Danish Consonants" Form and Substance, p. 123-138.

Harris, K.S. and H.S.Hoffman, A.M. Liberman, P.C. Delattre and F.S.Cooper 1958: "Effect of Third-Formant Transitions on the Perception of the Voiced Stop Consonants", JASA 30, p. 122-126.

Hoffman, H.S. 1958: "Study of Some Cues in the Perception of the Voiced Stop Consonants", JASA 30, p. 1035-1041.

Liberman, A.M., P.C.Delattre and F.S.Cooper 1952: "The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants", The Am. Journal of Psych. LXV, p. 497-516.

Liberman, A.M., P.C.Delattre, F.S.Cooper and L.J.Gerstman 1954: "The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants", Psychological Monographs 68, p.1-13.

Liberman, A.M., P.C.Delattre and F.S.Cooper 1958: "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position", Language and Speech 1. p. 153-167.

Liberman, A.M. 1969: "The Grammars of Speech and Language", S.R.Haskins 19/20, p. 79-125.

Lisker, L. and A.S.Abramson 1964: "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements", Word 20, p. 384-422.

Lisker, L. and A.S.Abramson 1970: "The Voicing Dimension: Some Experiments in Comparative Phonetics", 6. Int. Cong. of Phon. Sciences 1967 1, p.563-567.

Malécot, A. 1958: "The Role of Releases in the Identification of Released Final Stops", Language 34, p. 370-380.

Malécot, A. and E.Lindheimer 1966: "The Contribution of Releases to the Identification of Final Stops in French", Studia Linguistica XX, p. 99-109.

Öhman, S. 1961: "Relative Importance of Sound Segments for the Identification of Swedish Stops in VC and CV Syllables", Stockholm QPSR 3, p. 6-14.

Slis, I.H. and P.H.Damsté
1967: "Transilluminations of the Glottis during Voiced and Voiceless Consonants", Ipo Annual Progress Report 2, p. 103-109.

Schatz, C.D. 1954: "The Role of Context in the Perception of Stops", Language 30, p. 47-56.

Wajskop, M. 1971: "Identification des occlusives intervocaliques en Français", mimeographed.

Wang, W.S-Y. and Ch.J.
Fillmore 1961: "Intrinsic Cues and Consonant Perception", JSHR 4, p. 130-136.

IDENTIFICATION OF INITIAL STOP CONSONANTS IN SYLLABLES PLAYED AT DOUBLE AND HALF SPEED

Eli Fischer-Jørgensen

1. Stimuli and listeners

The stimuli used in the present investigation consisted in 18 CV-syllables with initial ptkbdg + the vowels i, a or u, spoken by me on tape. They were played in random order, once at double speed and once at half speed, to four groups of students of philology (two in 1954 and two in 1955) over a loudspeaker in a normal classroom. The listeners were asked to write down what they heard. 52 students listened to the syllables presented at double speed and 54 to those presented at half speed. The answers in 1954 and 1955 differed at two points. The listeners in 1954 made more mistakes than the listeners in 1955 for the vowel i at double speed (it was often heard as u), but less mistakes for p at double speed (it was less often heard as t). Probably the loudspeaker used in 1954 had stronger attenuation at high frequencies. But on the whole, the answers were the same in the two cases. There were 312 responses to each vowel and 156 responses to each consonant at double speed, and 324 and 162 respectively at half speed.

2. Identification of syllables played at double speed

2.1. Identification of the vowels

When the syllables were played at double speed the vowels were on the whole correctly identified (except that the listeners in 1954 heard i as u in 32 % of the cases, probably because formants 3 and 4 were raised too much). u has most of the energy concentrated in F1 at 250 Hz, but the doubling to

500 Hz did not change the identification. There were no o-answers and only a few y-answers. It is at first sight surprising that a was heard correctly when its first formant of about 900 Hz and its second formant of 1800 Hz were shifted to 1800 and 3600, respectively. The most probable explanation is that under these conditions the former F_1 functions as F_2, and the former F_2 as F_3, whereas a former subformant at about 300 takes over the function of F_1.

2.2. Identification of the consonants

The consonant identifications at double speed are given in table I and in Fig. 1 A. In the latter case ptk and bdg have been combined under the designation PTK.[1]

TABLE I

Identification (in %) of ptk and bdg played at double speed (N = 52). Answers horizontally.

	i				a				u			
	p	t	k	O	p	t	k	O	p	t	k	O
p	24	63	14	-	36	31	31	3	20	73	2	5
t	2	93	5	-	2	88	7	3	-	100	-	-
k	10	61	27	2	2	49	46	3	2	61	32	5
	b	d	g	O	b	d	g	O	b	d	g	O
b	30	56	2	12	46	46	-	8	22	68	-	10
d	27	64	2	8	29	68	3	-	-	95	5	-
g	2	85	9	5	19	58	22	2	-	58	36	7

1) Thus capital P is here used to cover p and b, not, as in the paper on tape cutting experiments, ptk.

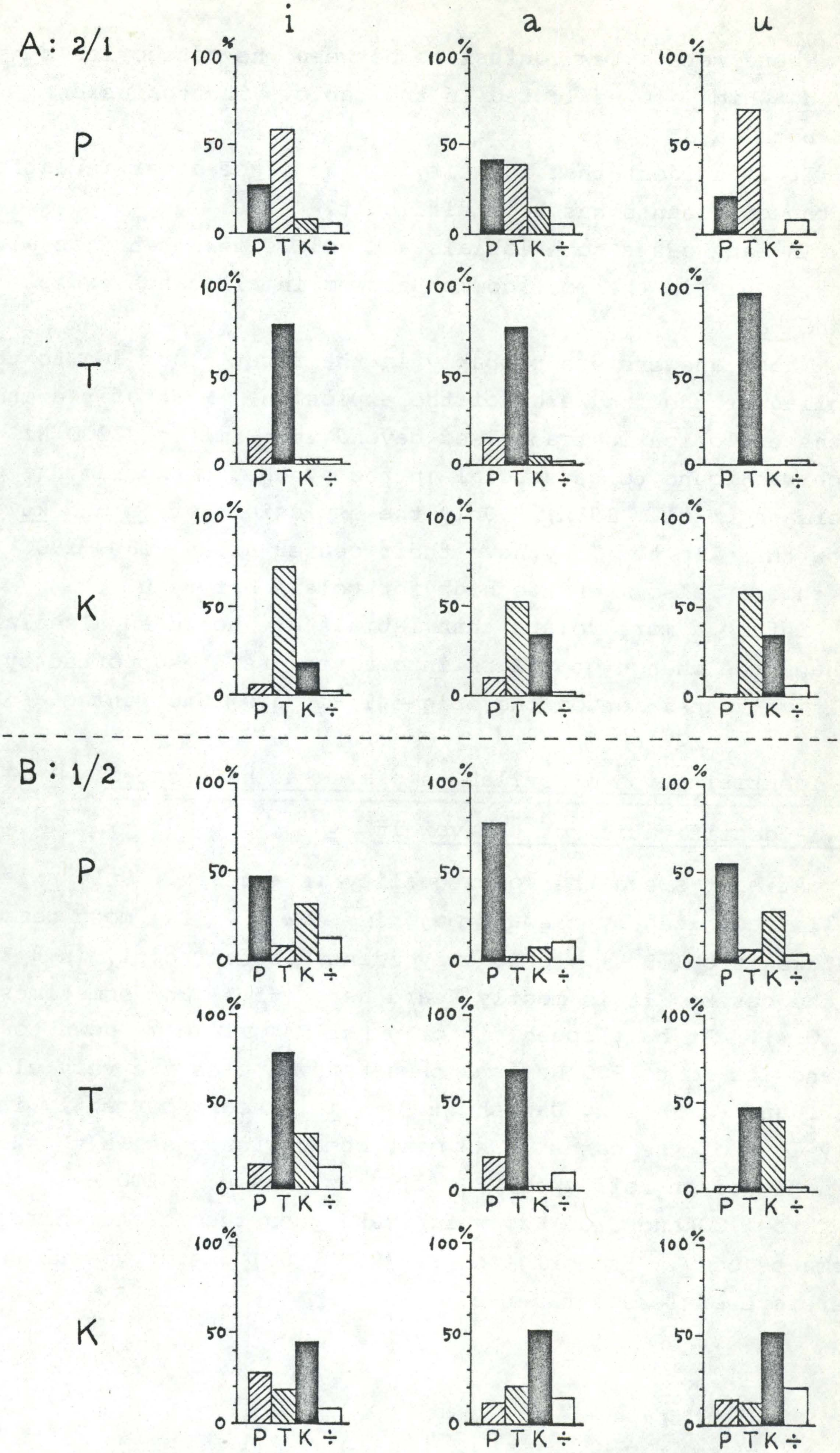


Fig. 1. Identification of labial, alveolar and velar stops (symbolized by P, T and K), played at double speed (A) and half speed (B).

There were a few confusions between the categories ptk and bdg which are neglected in the table. The confusions went both ways.

It is evident that identifying the place of articulation for the consonants has been difficult.

In many cases both labials and velars were heard as alveolars. This is the most common answer in all cases except for pa and ba.

These answers are probably in the first place due to the doubling of the frequency of the explosion. Most of the energy of the explosion is transposed beyond the limit of 3000 Hz which was found to be crucial in the synthetic experiments of Libermann et al. (1952). Only the explosions of gu and ku are below this limit. They have their center around 1800-1900 Hz, but this is also much too high for velars before u.

Before a more velars than labials are heard as alveolars. Perhaps the change of velars into alveolars is supported by the positive transition of the original F_2, now functioning as F_3.

3. Identification of syllables played at half speed

3.1. Identification of the vowels

At half speed the vowel quality is altered. Only u is still recognized by a small majority (52 %). The most common mistake is o (19 %). a is only identified correctly in 8 % of the cases. It is mostly heard as ɔ (65 %) and sometimes as o (10 %). At half speed a's original F_1 will come down to 450 Hz and its F_2 to 900 Hz, and these frequencies are very close to F_1 and F_2 in long Danish ɔ:. - i is heard correctly in only 10 % of the cases. The most common response is y (62 %). When played at half speed F_3 (3400 Hz) and F_4 (4200 Hz) come down to 1700 and 2100 Hz, which make good second and third formants of y. The original F_2 (2200 Hz) was very weak and therefore of less influence.

3.2. Identification of the consonants

The consonant identifications at half speed are given in table II and in Fig. 1 B. In the latter case ptk and bdg have been combined under the designation PTK.

TABLE II

Identification (in %) of ptk and bdg played at half speed (N = 54). Answers horizontally

	i (y)				a (ɔ)				u			
	p	t	k	O	p	t	k	O	p	t	k	O
p	31	13	41	15	69	2	17	13	56	5	26	13
t	2	98	-	-	-	93	-	7	-	59	33	7
k	24	11	56	9	24	11	56	9	20	6	56	18
	b	d	g	O	b	d	g	O	b	d	g	O
b	63	4	22	11	87	4	-	9	54	7	30	9
d	26	57	6	11	39	44	4	13	6	37	46	11
g	31	26	35	7	-	31	48	20	7	20	48	24

At half speed there are some confusions between the categories ptk and bdg, which have been neglected in the table. There are very few bdg-responses to ptk-stimuli, but the opposite case is somewhat more common. The percentages are, however, negligible except for di 11 %, gu 26 %, and gi 44 %. This can be explained by the doubling of the open interval, which for di, gu, gi becomes 40, 50 and 70 ms respectively. The others are shorter.

The place of articulation is identified somewhat better at half speed than at double speed. The highest percentages of answers correspond to the correct consonants except for pi and du, but the highest percentages are often below 50. The general uncertainty also shows up in the higher number of zero-answers as compared to double speed.

In almost 100 % of the cases t in ti and ta are identified correctly because of the high frication noise which in ti reaches above 3000 Hz even at half speed, and in ta at least above 2000 Hz. tu still has some noise concentrated about 3000 Hz, but the lower concentration around 1400 Hz reminds of k before u (1200 Hz).

The characteristic transition in the aspiration of pi (py) gets reduced at half speed, so that it looks very much like ky, in which the aspiration often starts below F_2 of the vowel. Before the other vowels p and k are heard correctly by a small majority. The most common mistake is mutual confusion between p and k.

d is identified correctly less often than t, because it lacks the high fricative noise. It is often heard as b before i (y) and a (ɔ), (b is, on the whole, a common mistake, and the transitions of d look more like those of b than like those of g), but before u the majority has heard g, probably because the frequency of the explosion comes down in the neighbourhood of the g-explosion at half speed (cf. that t is often heard as k before u).

b is relatively correctly identified. Before i (y) and u it is heard as g in a number of cases. This never happens before a (ɔ), nor is g before a (ɔ) heard as b, probably because the transitions are too different.

4. Summary

The identification scores show a greater number of mistakes for the consonants at double speed than at half speed. This can perhaps be explained by the fact that the change in absolute frequency is larger in the former case (a noise of 1000 Hz goes up to 2000 Hz at double speed, but only down to 500 Hz at half speed). For the vowels, however, doubling gives less distortion. But this is because a happens to get formants at the right frequencies, and u and i do not get into the areas of any other vowels. On the contrary, when the speed is halved, i and a get formants corresponding to those of other Danish vowels (y and ɔ).

The perceptual results of doubling and halving the speed for a and i are in good agreement with the results of Chiba and Kajiyama (1941), who, however, consider the preserved relative distances between formants to be a sufficient explanation of the preservation of a at double speed.

The most common mistakes for the consonants are an increase of d- and t-responses at double speed, and confusion between labials and velars at half speed. This indicates that the explosions (and aspirations) play a great role. This is particularly true of ptk, where the transitions are of minor importance. In bdg the transitions are also of some influence, particularly in combinations with a.

The general conclusion of this small experiment, as also of the extensive experiment with tape cutting, is thus that explosions should not be neglected. They play a fairly important role for perception.

References

Chiba, T. and M. Kajiyama 1941: The Vowel.

Liberman, A.M., P.C. Delattre and F.S. Cooper 1952: "The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants", The Am. Journal of Psych. LXV, p. 497-516.