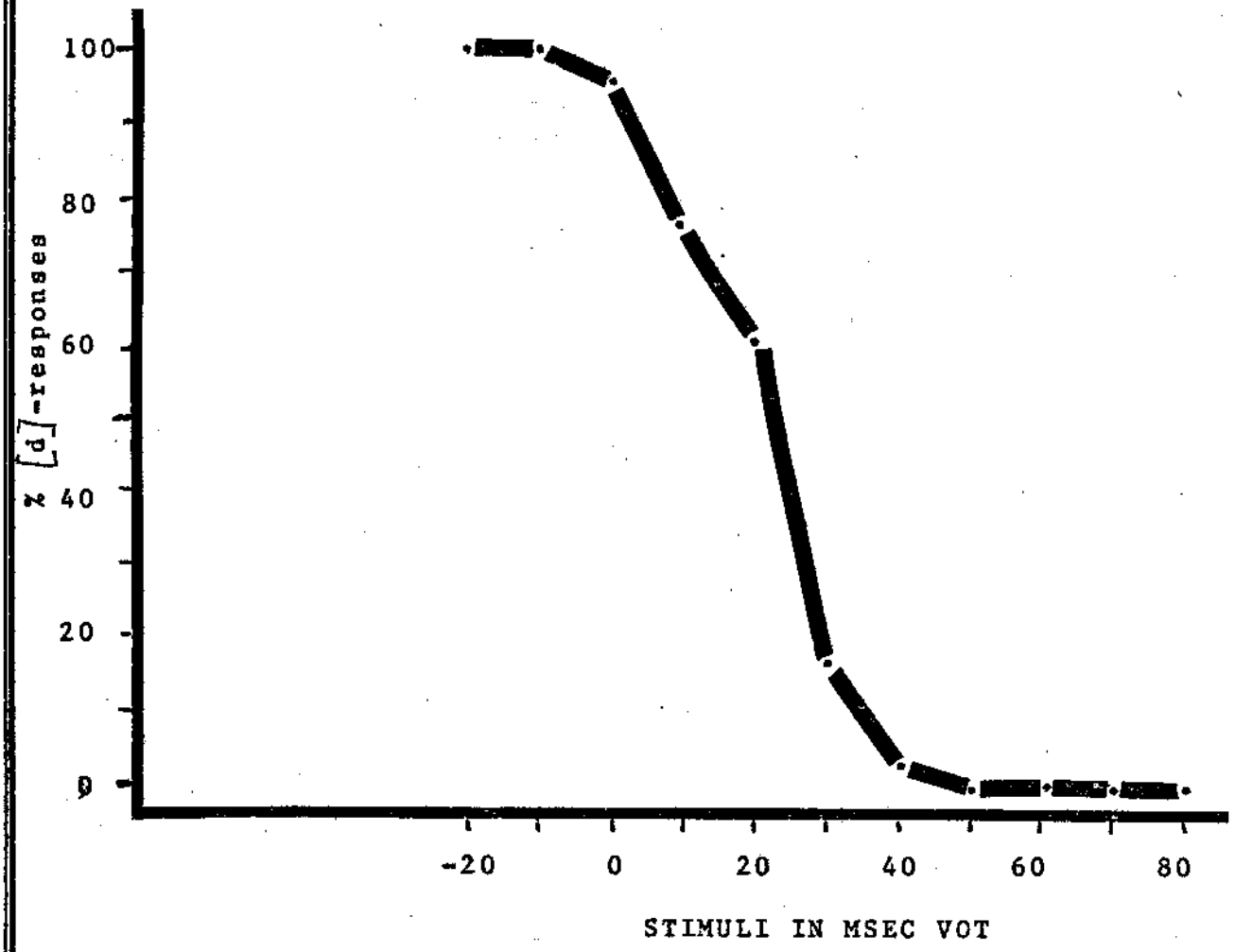

Fig. 2-17 -- Mean identification function for 24 listeners in Wrocław for the synthetic continuum with VOT values from 20 to +80 msec. Individual functions were less steep for this continuum since listeners gradually shifted from lower to higher boundaries.

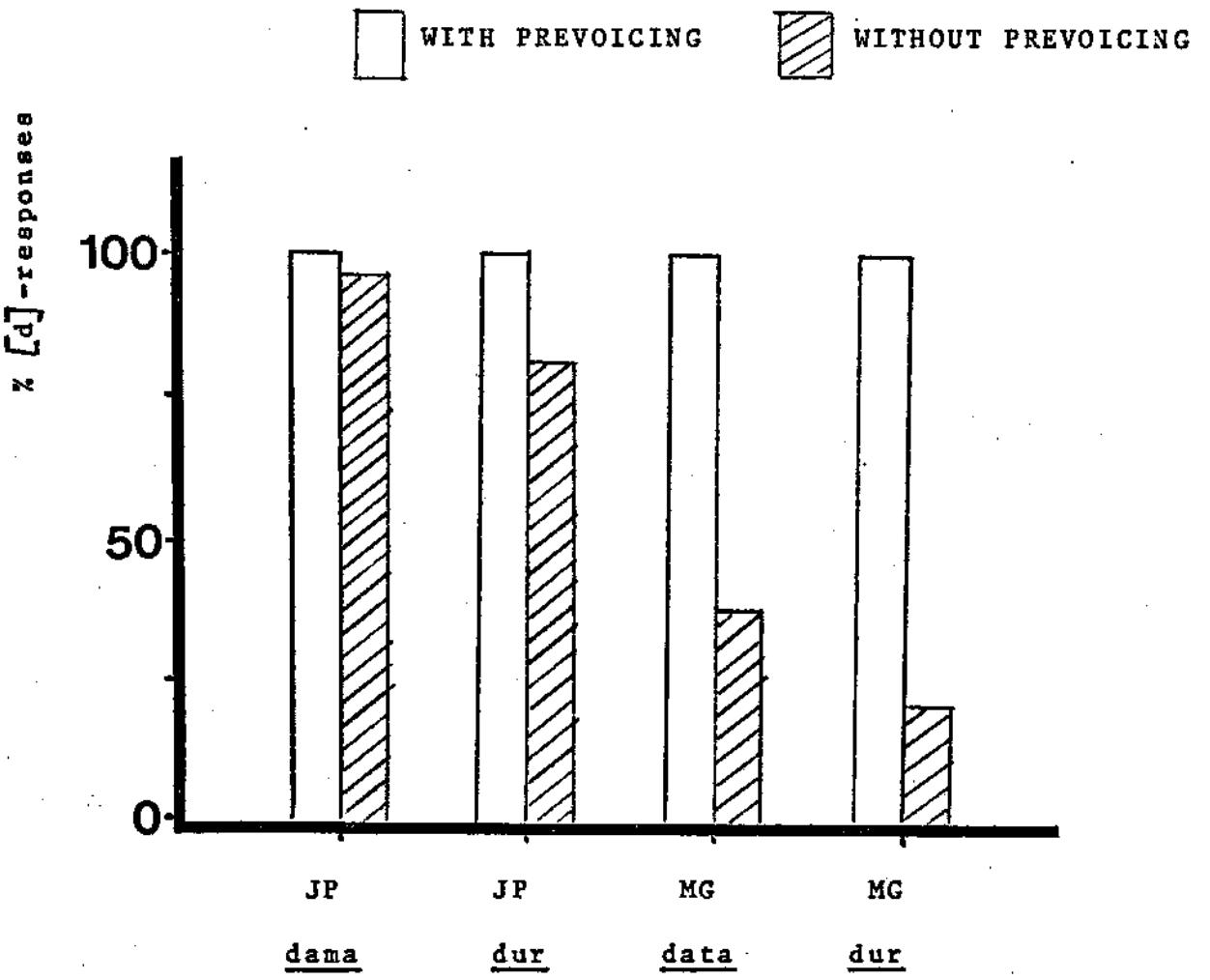

Fig. 2-18 -- Changes in listeners' % [d]-responses to tokens of words begining with [d], with and without prevoicing. The initials indicate the speaker of the original token (with prevoicing).

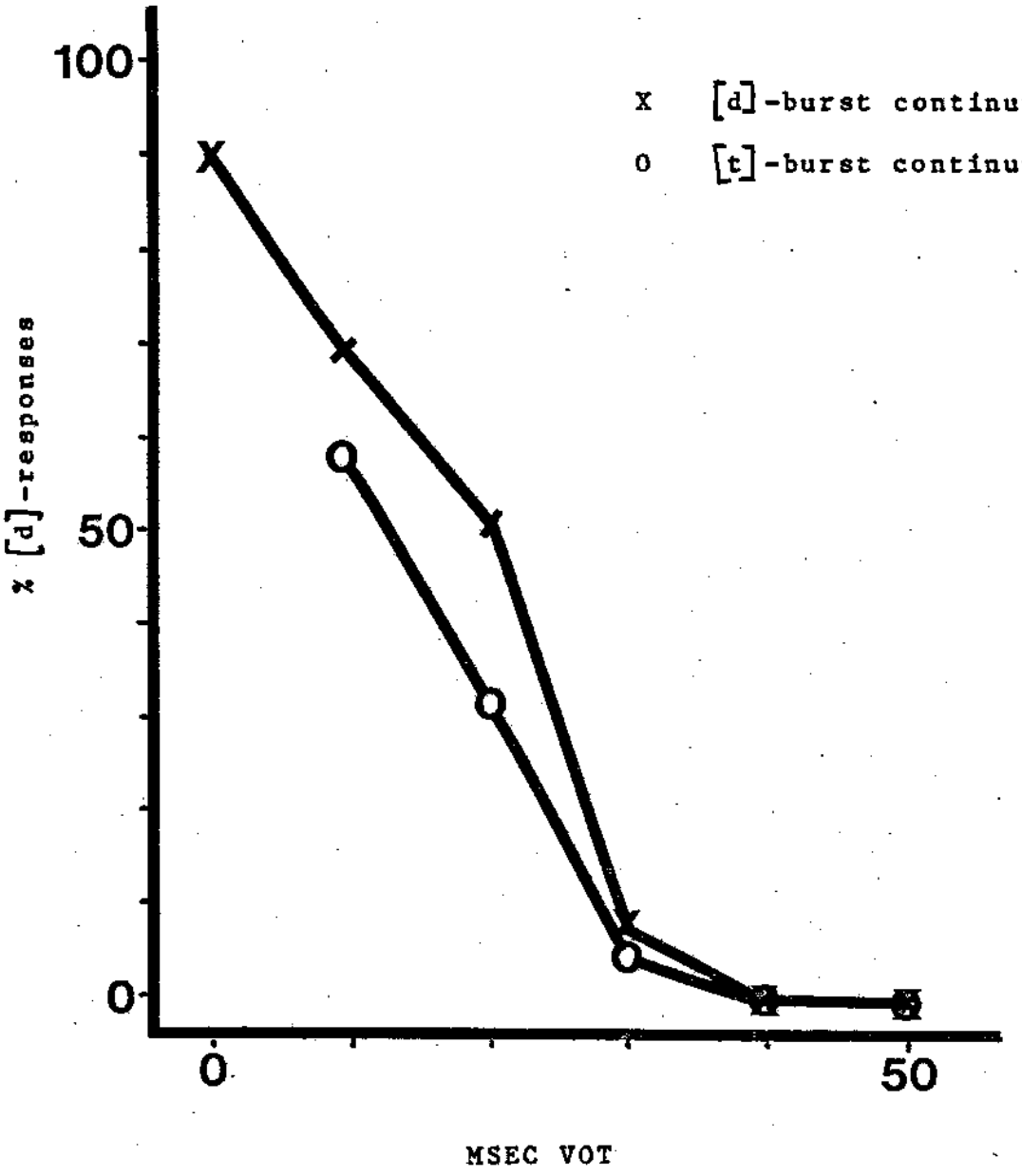

Fig. 2-19 -- Mean identification functions for 24 listeners in Wrocław for two natural-elited VOT continua, one with [t]-bursts and one with [d]-bursts.

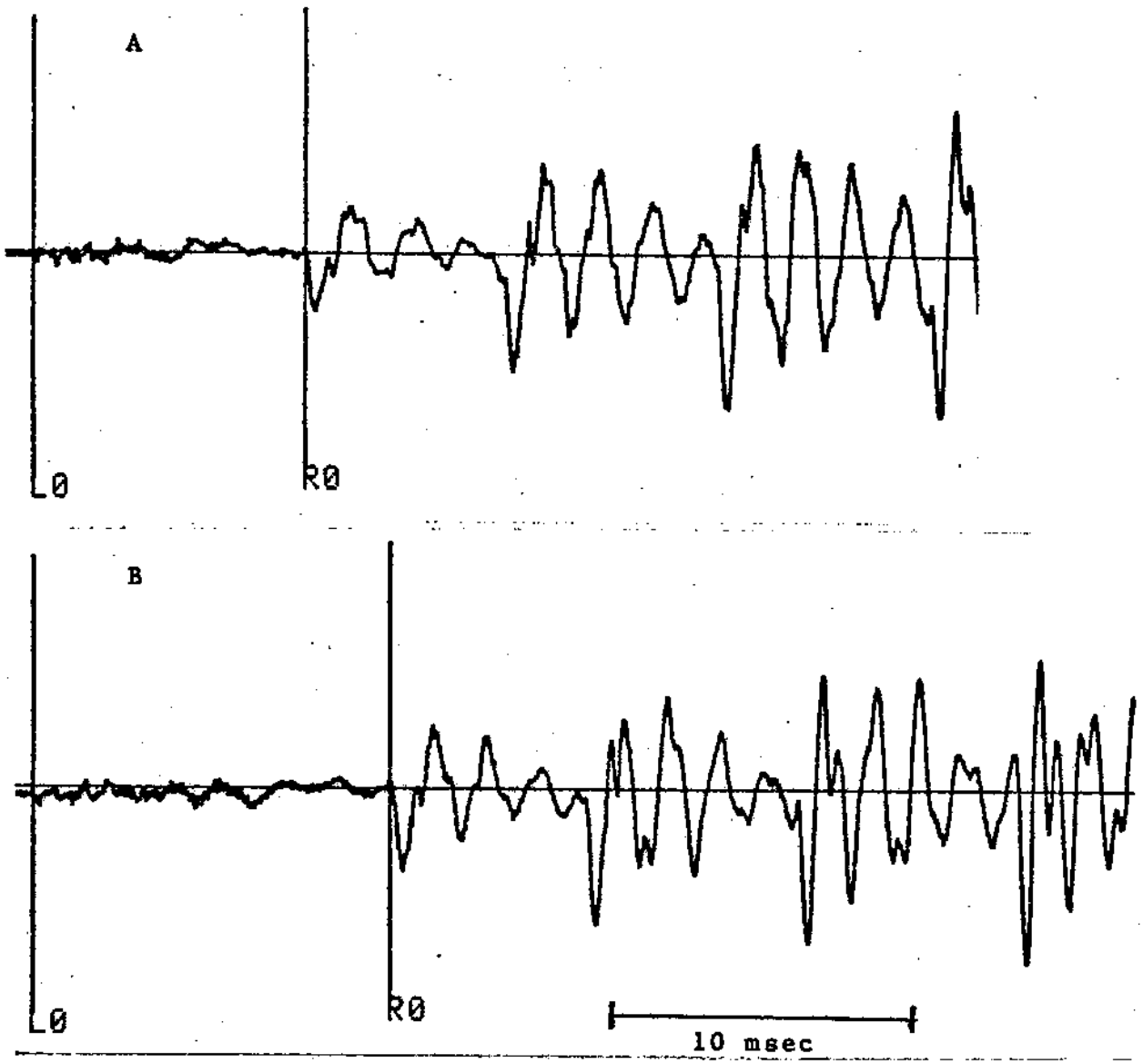

Fig. 2-20a -- Waveform showing token of da read by Wrocław speaker #4, with a VOT value of 9.1 msec as measured between the cursors.

Fig. 2-20b -- Waveform showing token of ta read by the same speaker, with a VOT value of 12.1 msec. Note the similarity of the overall patterns of the bursts.

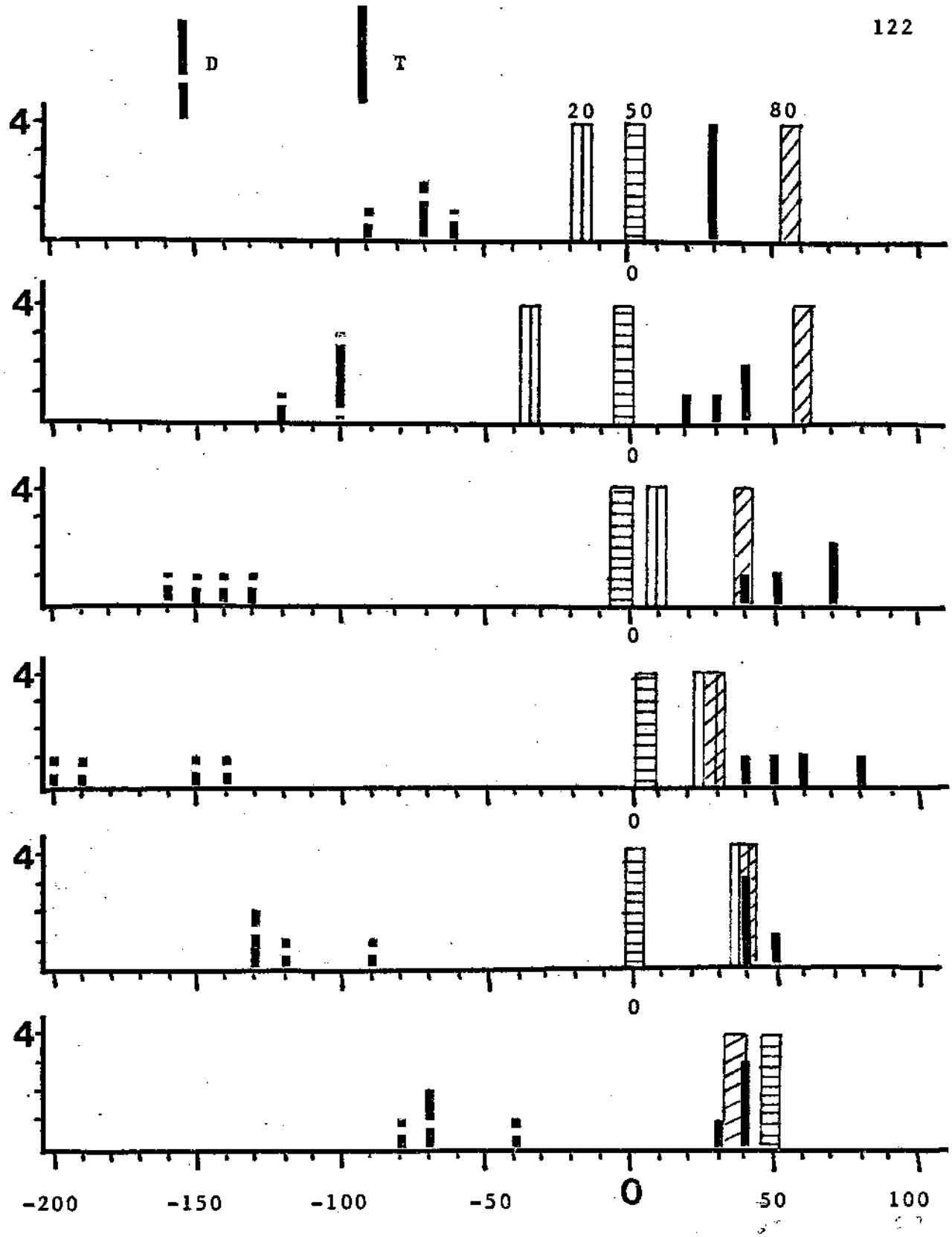

Fig. 2-21 -- Examples of the match between production and perception voicing categories for six subjects from Wrocław. The narrow bars show the number of tokens of [t] and [d] (identified at the top of the page) at each VOT value. The wide bars show the location of the perceptual boundaries for the three synthetic continua, identified in the top graph by the endpoint VOT value.

CHAPTER THREE -- Other Cues to Initial Voicing

3.1. Introduction

It was shown in the preceding chapter that speaker JP's [d]-tokens were heard as [d]'s even when prevoicing was removed. Therefore these tokens must have contained cues for voicedness other than VOT that occur during or after the release burst. The experiments reported in this chapter address the question of whether these cues are in fact located in the burst itself, or after it. It is not the purpose of this chapter to determine which aspects of the burst may act as acoustic cues to the voicing distinction in Polish. Rather, the aim here is simply to see to what extent the burst, taken as a unit, contributes to the perception of voicing in Polish word-initial [t] and [d].

Observation of tokens of Polish initial [t] and [d] indicates that there can be differences in the gross acoustic patterns of the bursts. Fig. 3-1 and 3-2 give examples of the various types of [t]- and [d]-bursts which occur in natural speech. Fig. 3-1a and 3-1b show waveform displays of the bursts of tur and dur read by JP from the list of minimal pairs. The [t]-burst is highly aspirated; the [d]-burst is barely fricated and is superposed on a pitch pulse. This token of dur was judged to contain an initial [d] even when prevoicing was removed. Fig. 3-2a and 3-2b show waveforms of the bursts of another minimal pair, tama and dama, read by a Wrocław subject. The [t]-burst is not especially noisy, but it is much longer than the [d]-burst. Note that the actual onsets of these two bursts are quite similar. These two tokens were not submitted to listeners for labeling. However, it seems likely that more than one acoustic parameter in the burst may help distinguish Polish stop voicing

categories.

One way of testing the influence of the burst on the perception of voicing is simply to interchange [t]- and [d]-bursts. This was done in each of the three experiments reported here. Another way of testing the influence of the burst is to remove it. The perceptual influence of post-burst cues was also tested by interchanging and removing [t]- and [d]-transitions.

A further effort was made to identify cues after the burst by considering whether voicelessness is signaled by a lack of voicing, or by the presence of aspiration. As was noted in the last chapter, in Polish some examples of initial [t] contain aspiration noise in the lag interval; others do not. English voiceless stops overall contain more aspiration than do Polish stops. Furthermore, it has been shown that aspiration is a strong cue for voicelessness in English (Winitz et al., 1975; Repp, 1979). Since aspiration appears to be a less consistent correlate of voiceless stops in Polish, it is not clear how strong a cue it is. Two aspects of voicelessness will be investigated here: first, voicing lag as a cue to voicelessness, and second, aspiration as a cue when it fills that lag interval.

3.2. Methodology

Three sets of stimuli[1] were made for this series of experiments on the PDP-8e computer in the Psychology Dept., using the SPLIT program (Myerow and Millwood, 1978).

3.2.1. Experiment I: VOT Continua

The first set of stimuli was described in detail in 2.2.4.b. It consists of the two VOT continua made by splicing waveform

segments taken from natural tokens of tur and dur spoken by JP. In one continuum the stimuli had a [t]-burst ("[t]-burst continuum") and in the other continuum the stimuli had a [d]-burst ("[d]-burst continuum"). In both continua VOT was varied in approximately 10-msec steps from +10 to +50 msec. This was done by substituting aspiration from the [t]-token for voiced pitch periods in the [d]-token. The exact VOT values that resulted depended on the lengths of the pitch periods in the original token; a complete list of stimuli is given in Table 1. The bursts are illustrated in Figures 3-1a and 3-1b, which show waveforms of the original tokens. The purpose of Exp. I was to compare the effect of the two bursts, [d] and [t], on listeners' VOT boundaries.

3.2.2. Experiment II: Voicing Lag

The second set of stimuli was made from tokens of tama and dama, also spoken by JP. These stimuli were designed to test the effect of a lag interval, both aspirated and silent, vs. the effect of a [t] or [d]-burst, on listeners' voicing judgements. Each token was divided into two or three acoustic segments. These segments were manipulated as follows. First, the burst segments were isolated from tama and dama. The [d]-burst segment included the waveform between burst onset and voicing onset, and was 12.3 msec in duration. A matching segment of burst+aspiration was spliced from tama, also 12.3 msec in duration. An aspiration segment, containing the rest of the [t] from 12.3 msec to voicing onset (at 34.9 msec) was also isolated. An equal duration (22.6 msec) of silence was also recorded. Third, the segments remaining of tama and dama, from voicing onset to the end of the word, formed two available word-endings. These two word-endings, however, were somewhat different, since 22.6 msec of voiceless transitions were removed from the tama-token,

TABLE 1

Complete list of stimuli made for experiments reported in Ch. 3.

Experiment I: VOT Continua

Stimuli with [t]-burst

VOT: 11.7 msec

20.2

29.5

39.1

49.1

Stimuli with [d]-burst

VOT: 0 msec

11.7

20.2

29.5

39.1

49.1

Experiment II: Voicing Lag

Stimulus Components

Burst	22.6 msec Lag	Voiced Part of
T	silent	tama
T	silent	dama
T	aspirated	tama
T	aspirated	dama
D	silent	tama
D	silent	dama
D	aspirated	tama
D	aspirated	dama

TABLE 1 (cont.)

Experiment III: Transitions

Stimulus Compenents		
Burst	Transitions	Steady-State of
T	T	tur
T	T	dur
T	D	tur
T	D	dur
D	T	tur
D	T	dur
D	D	tur
D	D	dur
none	T	tur
none	T	dur
none	D	tur
none	D	dur
T	none	tur
T	none	dur
D	none	tur
D	none	dur

while the dama-word-ending began immediately after the burst. Thus these word-endings, especially the one from dama, comprise almost the entire original token, including all of the voiced transitions.

Some of these stimuli are somewhat unnatural, for example, combinations with a voiced [d]-burst followed by aspiration. One very unnatural combination is that of aspiration followed by the [d]-word. The aspiration segment contains the onset of voiceless transitions, while the [d]-word which follows contains the onset of voiced transitions. Thus stimuli with this combination include two unrelated transition onsets. Such unnatural stimuli were intentionally made to see if cue value would be affected by these discontinuities.

Some of these stimuli seem similar to synthetic ones made by Summerfield and Haggard (1974). Those investigators varied the time separating burst from transitions, as well as VOT. However, since they found that aspiration increased the number of voiced responses, their results (and presumably their stimuli) are not comparable to those of this study.

In total, then, there were three pairs of acoustic segments: [t]- and [d]-bursts, 22.6 msec of silence and aspiration, and word-endings from tama and dama starting at voice onset. These segments were recombined, one from each pair, in all eight possible ways. Table 1 gives a complete list of these stimuli. The original tokens are illustrated in Ch. 2, Fig. 2-13; two recombinations are shown in Fig. 3-3a and 3-3b. Fig. 3-3c shows a spectrogram of the stimulus waveform given in Fig. 3-3b, illustrating the two transition onsets. Fig. 3-3c also indicates that the [d]-burst was not successfully taken at a zero-crossing, since an artifactual click

3.2.3. Experiment II: Transitions

The third set of stimuli was made from the tokens of tur and dur spoken by JP. (These are the same tokens used for the natural VOT continua.) These stimuli were designed to compare the relative contributions of the burst and transitions to the perceptions of voicing. Again, three acoustic segments were manipulated in constructing the stimuli. First, the [t]- and [d]-bursts (7.2 and 7.3 msec) were isolated. Second, the transitions, from end of burst to steady-state vowel, were isolated. Spectrograms of the original tokens were made, and from them it was determined that by 118 msec from stimulus onset, all transitions were completed in both tokens. Therefore 110.3 msec of [t]-transitions and 110.2 msec of [d]-transitions were isolated.[2] These transition portions contain all voice-onset information except the burst frication. Third, the rest of tur and dur from steady-state vowel to the end of the word was available. These portions differed in pitch contour and duration: dur, which was read in a list after tur, had a lower fundamental frequency and was about 25 msec longer.

From these three pairs of acoustic segments, the eight possible three-way combinations were formed. In addition, eight two-way combinations were formed. The bursts alone were combined with the steady-states with no temporal separation of these two segments.[3] The transitions alone were also combined with the steady-states. Thus there were four stimuli containing the various combinations of burst+steady-state, and four stimuli containing the various combinations of transitions+steady-state, as well as the eight stimuli containing combinations of burst+transitions+

steady-state. A complete list of these stimuli is given in Table 1. The original tokens are shown in Chapter 2, Fig. 2-13. Fig.3-4a and 3-4b show stimuli with interchanged bursts; Fig. 3-4c shows a spectrogram corresponding to 3-4b in which the connection of the transition segment to a steady-state segment can be seen. Fig.3-4d and 3-4e illustrate stimuli without bursts; the transitions show fairly abrupt onsets. Fig. 3-4f is a spectrogram of a stimulus without transitions.

3.2.4. Test Tapes

All of the stimuli were recorded onto a cassette, with stimuli representing the same minimal-pair contrast combined into one test. In each test, the original [t] and [d] tokens as well as the [d] -token with prevoicing removed were also included.[4] Ten tokens of each stimulus were randomized and recorded in blocks, with an ISI of 2 sec and an IBI of 10 sec. There were 10 blocks per test, and each block contained a randomization of all the stimuli.

The stimuli made by combining the bursts from dur and tur directly with the steady-state vowels, with no intervening transitions or silence, sounded somewhat strange. Therefore the 10 tokens of each of these were recorded as a separate test, at the end of the entire session.

All the other stimuli made from tur and dur -- the two VOT continua, the transition+steady-state stimuli, and the burst+transition+steady-state stimuli -- were randomized and recorded as a single test. There were 25 stimuli altogether. The stimuli made from tama and dama by manipulating the lag interval formed another test, with a total of 11 stimuli.

The cassettes were sent to Wrocław, Poland, where 24 listeners

identified the tokens as tama or dama and as tur or dur in a forced-choice task.

3.3. Results

For each perceptual test, the number of [d]-responses to each stimulus for each subject was determined.

3.3.1. Experiment I: VOT Continua

The first set of results to be considered are those for the two VOT continua in which the burst was taken either from a token of tur or dur. Category boundaries were computed for each continuum for each subject by Probit Analysis as described in the preceding chapter. When a subject had only [t]-responses, the category boundary was arbitrarily (and conservatively) set at 1 msec below the endpoint of the continuum. If both boundary values for a subject fell beyond the continuum endpoints, that subject was not included in the analysis. Twenty-one subjects' boundaries were used; the mean values of these boundaries are 16.6 msec for the [t]-burst continuum and 19.9 msec for the [d]-burst continuum. This difference is significant ($t_{20} = -5.21$, $p < .001$). The mean percentages of [d]-responses for the 24 listeners are given in Table 2.

3.3.2. Experiment II: Voicing Lag

The second set of results to be considered are those for the stimuli in which the burst and lag were manipulated. There were three segments of the stimuli that were interchanged: the burst, the lag, and the word-ending. The % [d]-responses to each stimulus, given in Table 3, was analyzed with a three-way repeated-measures analysis of variance, where the three factors corresponded to the three pairs of acoustic segments.

TABLE 2

Mean % [d]-responses for 24 listeners to stimuli in Exp. I, VOT Continua

VOT	Burst [t]	Burst [d]
0	-	91
12	58	74
20	31	55
30	4	7
39	0	0
49	0	0

TABLE 3

Mean % [d]-responses for 24 listeners to stimuli in Exp. II, Voicing Lag

Burst	Lag	Voiced Part of	% [d]
T	silent	tama	0
T	silent	dama	19
T	aspirated	tama	0
T	aspirated	dama	8
D	silent	tama	0
D	silent	dama	28
D	aspirated	tama	0
D	aspirated	dama	16

The mean % [d]-responses for each factor is given in Table 4. The results of the analysis of variance are given in Table 5.

Fig. 3-5 illustrates the results given in Table 4: that the effects of aspirated vs. silent lag and [t]- vs. [d]-burst are obtained only when the voiced portion of the word (word-ending) is taken from dama, which consisted of the entire token except for the burst. In just these cases, a [d]-burst and a silent lag contribute to an increase in the number of [d]-responses. Otherwise the percept is [t] uniformly.

3.3.3. Experiment III: Transitions

The third set of results to be considered are those for the stimuli in which bursts, transitions, and steady-states were interchanged. Table 6 gives the mean % [d]-responses for each of the 16 stimuli. The results from the combinations of all three stimulus segments were analyzed with a three-way repeated-measures analysis of variance, with the three factors corresponding to the three pairs of acoustic segments. The mean % [d]-responses for each factor is given in Table 7, and illustrated in Fig.3-6 . The results of the analysis of variance are given in Table 8.

Fig. 3-6 shows that the effects of burst and steady-state are obtained only when the transitions are voiced. When the transitions are voiceless, responses are almost uniformly [t]. When the transitions are voiced, a [d]-word steady-state and a [d]-burst contribute to a larger number of voiced responses.

The fourth set of stimuli tested were combinations of burst+ steady-state alone. The mean % of [d]-responses to these factors are given in Table 9. The data were analyzed with a two-way repeated-measures analysis of variance; results are given in Table

TABLE 4

Mean % [d]-responses for 24 listeners to burst, lag, and voiced word-ending segments of stimuli in Exp.II, Voicing Lag

Stimulus Segments		% [d]
Burst	T	7
	D	11
Lag	silent	12
	aspirated	6
Voiced Part of	tama	0
	dama	18

TABLE 5

Results of analysis of variance for Exp. II, Voicing Lag

Source	df	MS	F	p
Burst (t or d)	1	10.08333	13.32	.001
error	23	0.75724		
Lag (aspir. or silent)	1	16.33327	13.10	.001
error	23	1.24637		
Voiced Word Ending (from *tama* or *dama*)	1	146.99944	16.22	.001
error	23	9.06521		
Burst X Lag	1	0.08333	.24	.627
error	23	0.34420		
Burst X Word	1	8.33333	11.86	.002
error	23	0.70290		
Lag X Word	1	16.33327	13.58	.001
error	23	1.20289		
Burst X Lag X Word	1	0.08333	.24	.627
error	23	0.34420		

TABLE 6

Mean % [d]-responses for 24 listeners to stimuli in Exp. III, Transitions

Burst	Transitions	Steady-State of	%[d]
T	T	tur	4
T	T	dur	5
T	D	tur	3
T	D	dur	60
D	T	tur	5
D	T	dur	5
D	D	tur	80
D	D	dur	91
none	T	tur	5
none	T	dur	5
none	D	tur	82
none	D	dur	83
T	none	tur	34
T	none	dur	30
D	none	tur	60
D	none	dur	42

TABLE 7

Mean % [d]-responses for 24 listeners to burst, transition, and steady-state segments of stimuli in Exp. III, Transitions

Stimulus Segments		% [d]
Burst	T	18
	D	46
Transitions	T	5
	D	59
Steady-State of	tur	23
	dur	40

TABLE 8

Results of analysis of variance for Exp. III, Transitions, stimuli with both bursts and transitions

Source	df	MS	F	p
Burst (t or d)	1	360.25366	217.36	.001
error	23	1.65738		
Transitions (t or d)	1	1392.12842	218.94	.001
error	23	6.35847		
Steady-State (from tur or dur)	1	138.37938	87.21	.001
error	23	1.58672		
Burst X Transitions	1	344.00366	172.49	.001
error	23	1.99434		
Burst X Steady-State	1	66.50471	34.87	.001
error	23	1.90737		
Transitions X Steady-State	1	135.00439	90.34	.001
error	23	1.49433		
Burst X Transition X Steady	1	59.62968	35.86	.001
error	23	6.66281		

TABLE 9

Mean % [d]-responses for 24 listeners to burst and steady-state segments of transition-less stimuli in Exp. III, Transitions

Stimulus Segments		% [d]
Burst	T	32
	D	51
Steady-State of	ɾur	47
	dur	36

10. The burst factor but not the steady-state factor was significant: a [d]-burst gave more [d]-responses. The interaction of burst and steady-state was significant in that there was more of an effect for the vowel with [d]-bursts. This effect, however, went in the opposite direction from what would be expected. A [d]-burst +[t]-word vowel combination resulted in more [d]-responses than the [d]-burst+[d]-word vowel combination. This result is illustrated in Fig. 3-7.

The final set of stimuli tested were combinations of transitions+steady-state alone. The mean % of [d]-responses to these factors are given in Table 11. The results of a two-way repeated-measures analysis of variance are given in Table 12. The transitions had a significant effect on the number of [d]-responses, while the steady-state did not. This result is illustrated in Fig. 3-8. A complete shift in response categories was produced.

To summarize the results for these burst and transition manipulations:

1) Both burst and transitions, alone or together, always have a significant influence on the number of responses in each voicing category. However, it must be emphasized that many of these significant shifts in responses do not involve a change in the categorization itself. Often only the strength of the categorization is affected. The operational definition of "categorization" understood here is that roughly 50-60% responses with one category label is ambiguous, 60-80% is a weak categorization, and 80-100% is a strong categorization. A "category shift" is a change in the % responses of at least 40-50% which involves a change from one label to the other. This point will be discussed further in the next section.

TABLE 10

Results of analysis of variance for Exp. III, Transitions, stimuli with bursts and no transitions

Source	df	MS	F	P
Burst (t or d)	1	86.26030	32.53	.001
error	23	2.65171		
Steady-State (from tur or dur)	1	31.51009	2.90	.102
error	23	10.85824		
Burst X Steady-State	1	12.76039	7.83	.010
error	23	1.62998		

TABLE 11

Mean % [d]-responses for 24 listeners to transition and steady-state segments of burst-less stimuli in Exp. III, Transitions

Stimulus Segments		% [d]
Transitions	T	5
	D	83
Steady-State of	tur	44
	dur	44

TABLE 12

Results of analysis of variance for Exp. III, Transitions, stimuli with transitions and no bursts

Source	df	MS	F	p
Transitions (t or d)	1	1449.26001	226.77	.001
error	23	6.39085		
Steady-State (from *tur* or *dur*)	1	0.01041	< 1	-
error	23	1.05389		
Transitions X Steady-St.	1	0.09375	< 1	-
error	23	1.22418		

2) The transitions+steady-state stimuli showed no significant effect of the steady-state vowel. The burst+steady-state stimuli did show a significant effect of the steady-state vowel, but in the wrong direction. The burst+transitions+steady-state stimuli showed a significant interaction with the steady-state vowel that is somewhat complex. If the transitions were voiced, a [d]-word steady-state gave more [d]-responses. This influence was greater if the burst was voiceless. Thus the effect of steady-state appears to be somewhat idiosyncratic with each set of stimuli.

3.4. Discussion

3.4.1. Experiment I: VOT Continua

In the first set of stimuli, 5 lag values of VOT were combined with [d]- and [t]-bursts. With the [d]-burst, listeners needed a greater voice onset time to shift their responses to the [t] category: 18 msec VOT with the [d]-burst vs. 15 msec VOT with the [t]-burst. In this sense, whatever cues to voicedness are present in the [d]-burst, they are "worth" about 3 msec of aspirated VOT lag. That is, there is a trading relation that holds here between burst-cues and VOT: one offsets the other.

It is only at VOT values of 12 and 20 msec, which straddle the boundary values averaged for 21 listeners, that burst cues have any effect. At a VOT of 12 msec, both the [t]-burst and [d]-burst stimuli are heard as [d], but the categorization of the [t]-burst stimuli is marginal (58%) for the group of 24 listeners. The VOT value of 20 msec is clearly ambiguous, as the burst changes the overall categorization, although neither categorization is particularly strong. (31% with the [t]-burst vs. 55% with the [d]-burst).

At higher VOT values the percept is uniformly [t], regardless of the burst. Two [d]-burst stimuli show the greatest percentage change in responses across a single step: from 55% [d]-responses at 20 msec VOT to 7% [d]-responses at 30 msec VOT. The [t]-burst stimuli show a more gradual shift in responses. See Fig. 2-19 for graphs of the group labeling functions.

There are at least two ways that aspiration could act to change the percept in the [d]-burst stimuli. In these stimuli, systematically-varied durations of aspiration are substituted for equal durations of voiced transitions. We know that for JP, a [d]-token with prevoicing removed is still heard as [d]. Therefore some cue(s) to voicedness are present somewhere in the signal after the point of release. The cue(s) could be in the burst itself, after the burst, or both. If the cues are in the burst, they are present even when the aspiration is inserted after the burst in making the VOT continuum. The aspiration, as a cue for voicelessness, competes with the cues for voicedness in the burst, perhaps by masking them. The two sets of cues are weighed against each other perceptually, and presumably some amount of aspiration could outweigh the burst-cues and give a [t]-percept. On the other hand, if the cues are located immediately after the burst, and the burst itself is irrelevant, then the aspiration replaces these cues because as aspiration is inserted, the voiced transitions are deleted. Presumably some duration of aspiration would entirely replace the other cues, and only at shorter durations would both sets of cues be present in the signal.

The results of Experiment I on VOT continua suggest that cues for voicedness are located both during and after the burst.

Interchanging bursts produced a reliable shift in the category boundary, indicating that a [d]-burst has some cue value for voicedness. However, the shift was small enough (3 msec VOT mean for 21 subjects) that it seems likely that other cues are found after the burst as well. The first 10+ msec after the 7-msec burst are probably more crucial to the percept than is the burst itself. For most listeners, when the VOT is less than about 10 msec or greater than about 25 msec, the percept is clearly [d] or [t] respectively, regardless of the burst. That is, only stimuli which are ambiguous or equivocal with respect to VOT show the trading relation with the burst. In the intermediate VOT region, the burst offsets the effect of a small amount of aspiration. At no VOT value represented here does the burst have a strong enough effect to completely shift the categorization of an otherwise clear stimulus.

3.4.2. Experiment II: Voicing Lag

Further evidence that burst information is outweighed by a moderated VOT lag comes from the experiment in which the bursts, a 22.6 msec voiceless lag interval, and the voiced portions of the tokens were interchanged. Both a [d]-burst and a silent lag interval contributed to an increase in [d]-responses, but only when the voiced portions of the stimuli were taken from the [d]-token. It is not difficult to see why this should be so. Recall that stimuli made with the voiced portion from the [d]-token were extremely unnatural stimuli, because the voiced portion began immediately after the burst frication, at the onset of transitions. However, all these stimuli also contained a lag interval of 22.6 msec, between the burst and these voiced transitions. The voiced portion from

the [t]-token, on the other hand, began 22.6 msec after the burst (34.9 msec from the beginning of the burst), and the transitions were already well underway. Thus there was a higher F1 starting frequency in the voiced transitions from the [t]-token, and they corresponded with the lag interval which was manipulated. When the voiced portion from the [t]-token was used in a stimulus (with the voiced transitions that matched the lag interval), listeners uniformly gave [t]-responses, regardless of the other manipulations On the other hand, when the voiced portion came from the [d]-token and the lag was inserted artificially and unnaturally between the burst and the voiced transitions, the stimuli were still labeled as [t], but the categorization was not as strong. That is, as far as the overall categorization is concerned, the presence of a lag interval was the strongest cue, but other factors influenced the strength of the categorization. Only when the transitions were inappropriate for the voicing lag did the burst and lag-contents have an effect. In just this case, a [d]-burst increased the number of [d]-responses, and a silent (unaspirated) lag also increased the number of [d]-responses. These two contributions appear to be additive. Figure 3-5 and Table 3 show that the difference in % [d]-responses due to burst and aspiration (19.5%) is approximately equal to the sum of the difference due to the burst alone (8%) and the difference due to the aspiration alone (11%).

These stimuli were all weighted towards a [t]-percept in that they had a 22.6-msec voiceless interval after the burst.[5] Thus it is not surprising that <u>all</u> stimuli were categorized as [t] for at least 70% of the total responses. The significant effects

discussed here have no effect on the gross overall categorization of the stimuli; no stimuli are actually ambiguous with respect to voicing. The 22.6 msec of lag, whether silent or aspirated, is clearly the strongest cue in the stimuli. However, having voiced transitions that did not correspond to the lag interval, but instead looked as if the lag interval were not present, made the stimuli more ambiguous and allowed two other cues to affect the percept. The first is the burst itself, and the second is the "contents" of the lag interval. Both a [d]-burst and silence during the lag interval increased the number of [d]-responses, apparently in an additive way.

Regardless of the stimulus configuration surrounding it, the 22.6-msec lag interval is a powerful one that gives a [t]-percept for at least 70% of the tokens. When the lag interval cue is supported by voiced transitions that correspond, then the percentage of [t]-responses increases to 100%, regardless of the burst or the lag "contents". When the lag interval is contradicted by voiced transitions that do not correspond, then the secondary cues have their (additive) effects. Only in these equivocal stimuli do the minor cues assert themselves. In the comparatively natural configuration, they are completely overridden.

The stimuli used in Exp. I on VOT continua were made by replacing voiced transitions with aspirated ones. The stimuli in Exp. II on voicing lag, however, were made by inserting aspirated transitions before the voiced ones. Thus when the voiced transitions came from the [d]-token, there were two transition-onsets, one voiceless and one voiced. The 22.6-msec inserted lag followed a [d]-burst of 12.3 msec. In Exp. I, the various lag values

followed a [d]-burst of only 7.3 msec, so the same amount of lag added to the two bursts would give VOT values that differ by 5 msec. Thus the Exp. II VOT of 34.9 msec is roughly equivalent to the Exp. I VOT of 29.5 msec. This Exp. I stimulus was categorized as [d] 7% with a [d]-burst. The corresponding stimulus from Exp. II ([d]-burst+aspirated lag+[d]-word ending) was categorized as [d]for 16% of the tokens. The 9% difference is probably due to the mismatch of the voiceless and voiced transitions.

3.4.3. Experiment III: Transitions

The results of this experiment, in which transition information was manipulated, also indicate that burst information was less important than information that followed the burst. In this experiment, the burst, 112 msec of transitions, and steady-state vowels were isolated and recombined. When transitions alone were combined with the steady-states, the percept was determined by the transitions. When bursts alone were combined with the steady-states (which is a very unnatural combination, especially with no temporal separation), then the burst had a significant effect on the number of [d]-responses, but the tendency was nevertheless to label all these anomalous stimuli as [t]. A [d]-burst increased the number of [d]-responses somewhat without affecting the overall categorization. That is, voiced transitions alone were a sufficient cue for voicedness, but a voiced burst alone was not, although it did have some effect on listeners' responses.

When bursts and transitions were both present in a stimulus, voiceless transitions were seen to be a strong cue for voicelessness, regardless of the burst or steady-state. When the burst and transitions were both voiced, the percentage of tokens labeled

[d]was high. Only when the transitions were voiced but the burst voiceless was the stimulus potentially ambiguous. Such a stimulus may be similar to a [t]-token with a low (7 msec) VOT. In this stimulus only, the steady-state portion played a major role in determining the percept. If the steady-state portion came from the [t]-word, then the percept was consistently [t]. The steady-state portion of a [d]-word, combined with voiced transitions, produced a [d]-percept somewhat less often than voiced transitions did without a burst (60% vs. 82%). This difference of 22% [d]-responses may be seen as the opposing effect of the [t]-burst. These particular stimuli, then, were the only ones where either burst or steady-state had any particularly large influence on the percept.

The burst+transitions+steady-state stimuli are similar to the extreme VOT stimuli used in Experiment I, with some differences in how the original tokens are divided into acoustic segments. Figure 3-9 illustrates these differences. In both experiments the same 7-msec bursts are interchanged; in Experiment III the entire 42-msec lag interval is included in the [t]-transitions. It is interesting and encouraging that the percentage of [d]-responses is nearly identical for similar stimuli, indicating that listeners are consistent in their responses. For example, the [t]-burst VOT-continuum stimulus with a 12 msec VOT is similar to the [t]-burst +[d]-transitions+[d]-steady state, since the burst is 7 msec in duration. The % [d]-responses to the first stimulus was 58%, and to the second, 60%.

To summarize the results of this third experiment, both the burst and the vocalic information can influence the strength of

the listeners' categorizations, and sometimes produce category shifts. However, in most stimuli the crucial information seems to be contained in the portion following the burst frication. The 42-msec voiceless period, after either burst, is a very strong cue to voicelessness. Stimuli with voiced transitions, either alone or with a [d]-burst, are perceived as [d]as often as is the [d]-token without prevoicing (80%). With a [t]-burst and the same voiced transitions, the percentage of [d]-responses decreases to 60%. Comparing [t]- and[d]- burst pairs in this experiment indicates that overall the burst <u>can</u> produce a 10-30% change in responses, but that it is usually overridden by transition (VOT) cues.

Voiceless transitions are an unambiguous cue for voicelessness which is to be expected, since the [t]-token has a VOT of about 50 msec. This lag value overrides all other stimulus parameters. When the transitions are voiced, then the stimulus is <u>potentially</u> ambiguous. No burst, or a [d]-burst, produces a voiced percept. Stimuli with the equivocal [t]- burst + voiced-transition combination show a significant effect of the steady-state vowel. When there are <u>no</u> transitions, then the burst has a significant but minor effect on the responses. That is, in just those cases where the stimulus is equivocal or anomalous with regard to the transition (VOT) information, then the secondary cues play some role in the percept.

3.5. General Discussion and Conclusions

It was originally hypothesized that some cue or cues in the burst might account for the perception of voicedness in [d]- tokens from which prevoicing (VOT) information was removed. The stimuli

which still produced [d]-responses had completely voiced bursts, and the stimuli which did not had partly voiced,or voiceless,noisy bursts, but the number of stimuli involved was small enough that no conclusions could be drawn about the cues responsible. In this chapter various experiments were carried out to determine how important the burst is in voicing judgements when prevoicing is removed.

The first experiment reported, on VOT continua, indicated that simply interchanging bursts does not completely shift voicing responses, but rather that there is a trading relation between the burst cues and VOT. The VOT boundary with a voiceless burst lies at about 15 msec; with a voiced burst, it lies at about 18 msec VOT. At the same VOT value, interchanging bursts can change the percentage of responses up to 25%. Extreme VOT values (less than 10 msec or greater than 25 msec) override the burst cues. Experiments II and III corroborated the finding that a VOT value of 35 or 50 msec will be the strongest cue in a stimulus. Stimuli that are made with ambiguous or impossible combinations of acoustic segments show the influence of secondary cues such as bursts.

Stimuli with a voiced burst at stimulus onset, optionally followed by a voiceless interval and a second voicing onset within 20 msec of the first, are perceived as [d]. On the other hand, stimuli with a voiceless burst at stimulus onset, and stimuli with a voiced burst followed by a voiceless interval and a second voicing onset more than 20 msec after the first, are perceived as [t]. The generalization may be that a fairly brief voiceless interval between two voiced segments is not perceived. A voiceless interval may be perceived either when it is longer than a certain critical relative

duration, or when it is in stimulus-initial position, which makes it particularly salient. Aspiration may also make a voiceless interval more salient. That is, there may be a psychoacoustic explanation (perhaps based on a masking phenomenon) for the trading relation between VOT and the burst, as well as for the results with editing of prevoicing.

In an unedited [d]-token with a voiceless or partly-voiceless burst, there are two voiced portions, the prevoicing and the vocalic segment. The briefness of the voiceless burst that intervenes may make it difficult to distinguish that there are two separate voiced intervals. When the prevoicing is removed, the voiceless burst is then at stimulus onset, and this position may make it quite salient, despite its briefness. If the burst is completely voiced, of course, the stimulus onset is always voiced, with or without prevoicing. However, the unnaturalness of a stimulus with a voiced burst but no prevoicing may weaken the categorization slightly.

There may also be a hierarchy of burst "voicedness". The bursts which are superposed onto pitch pulses are "more voiced" that those with weak pulses within the burst. MG's bursts are probably less voiced in this way than JP's. MG's [d]-bursts are more prominent in the waveform, because of higher relative amplitude and more frication. The very beginning of the burst is voiceless, followed by a dense concentration of irregular and noisy pulses. In contrast, JP's [d]-bursts are completely voiced, with less frication, and resemble the surrounding signal more.

The results of these experiments also support a distinction between primary and secondary cues. Primary cues are those that

have the strongest weightings, and which can effect a complete reversal in categorization. Secondary cues are those whose weightings are weak enough that they can change only the strength of a categorization. The secondary cues identified here usually change the percent responses by 20-30%. In the most natural stimulus configurations the primary cues override the secondary ones, if those conflict, and the categorization is strong. In ambiguous configurations the secondary cues are able to assert themselves (for example, a voiceless burst followed directly by voicing onset allows the steady-state to have cue value). In impossible configurations the same is true (for example, in the burst + steady-state combination the burst has one of its largest effects). An unnatural stimulus confuses the listener, and he seems to abandon his hierarchy of cue weightings (primary and secondary). Instead he uses any acoustic information available, regardless of the normal weightings. The strength of the categorization is also reduced; labeling becomes more variable as cues are treated more equally. Responses may be less tied to the linguistic system and more tied to the psychophysical system under these conditions.

FOOTNOTES

[1]Two additional sets of stimuli were made from tokens spoken by MG. Results for tests using these stimuli are not reported, because all tokens were heard as [t]. It was reported in the preceding chapter that when prevoicing is edited from MG's tokens of word-initial [d], listeners consistently hear [t]. Prevoicing was never retained in stimuli made by editing [d]-tokens. Consistent with the results reported earlier, all of MG's [d]-tokens without prevoicing were heard as [t]. MG apparently has only cues for voicelessness from the point of release burst on, and the effect of various stimulus manipulations cannot be assessed.

[2]The exact point at which the cut was made was by necessity a zero-crossing in the waveform.

[3]Combining a burst with a steady-state is in itself quite unnatural, but leaving no time interval between them is even more unnatural. It was done in order to see how effective, if at all, the burst would be in a "worst possible" stimulus.

[4]These stimuli are identical to some of the recombined items. For example, the original tur is identical to the stimulus made from the [t]-burst+ [t]-transitions + [t]-steady-state.

[5]Lack of prevoicing by itself was not a cue for voicelessness, since the mean % [d]-responses to the [d]-token without prevoicing was 96%. However, lack of prevoicing may interact with and enhance the lag interval.

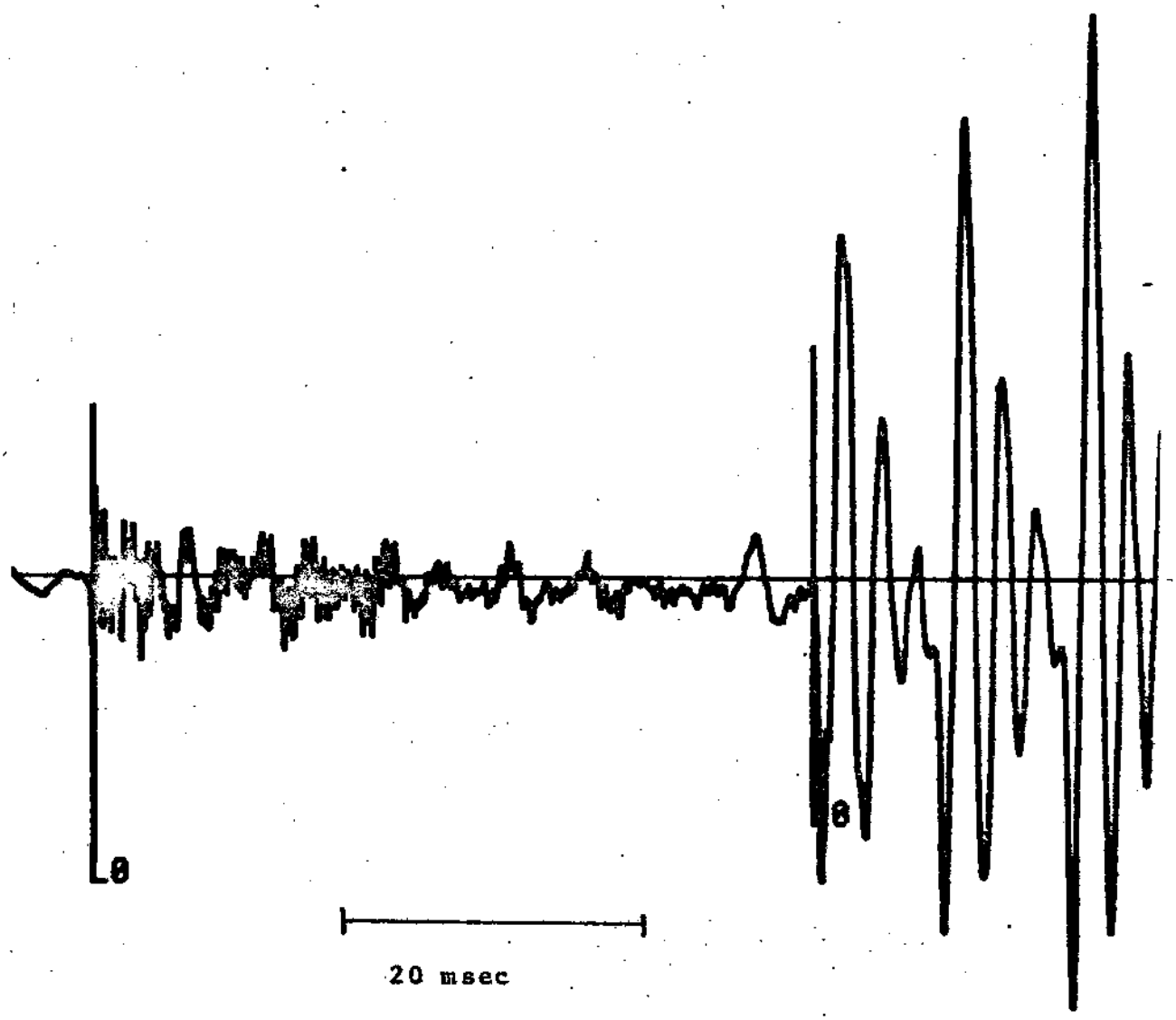

Fig. 3-1a -- Waveform showing token of tur read by speaker JP, with a VOT value of 48.7 msec as measured between the cursors.

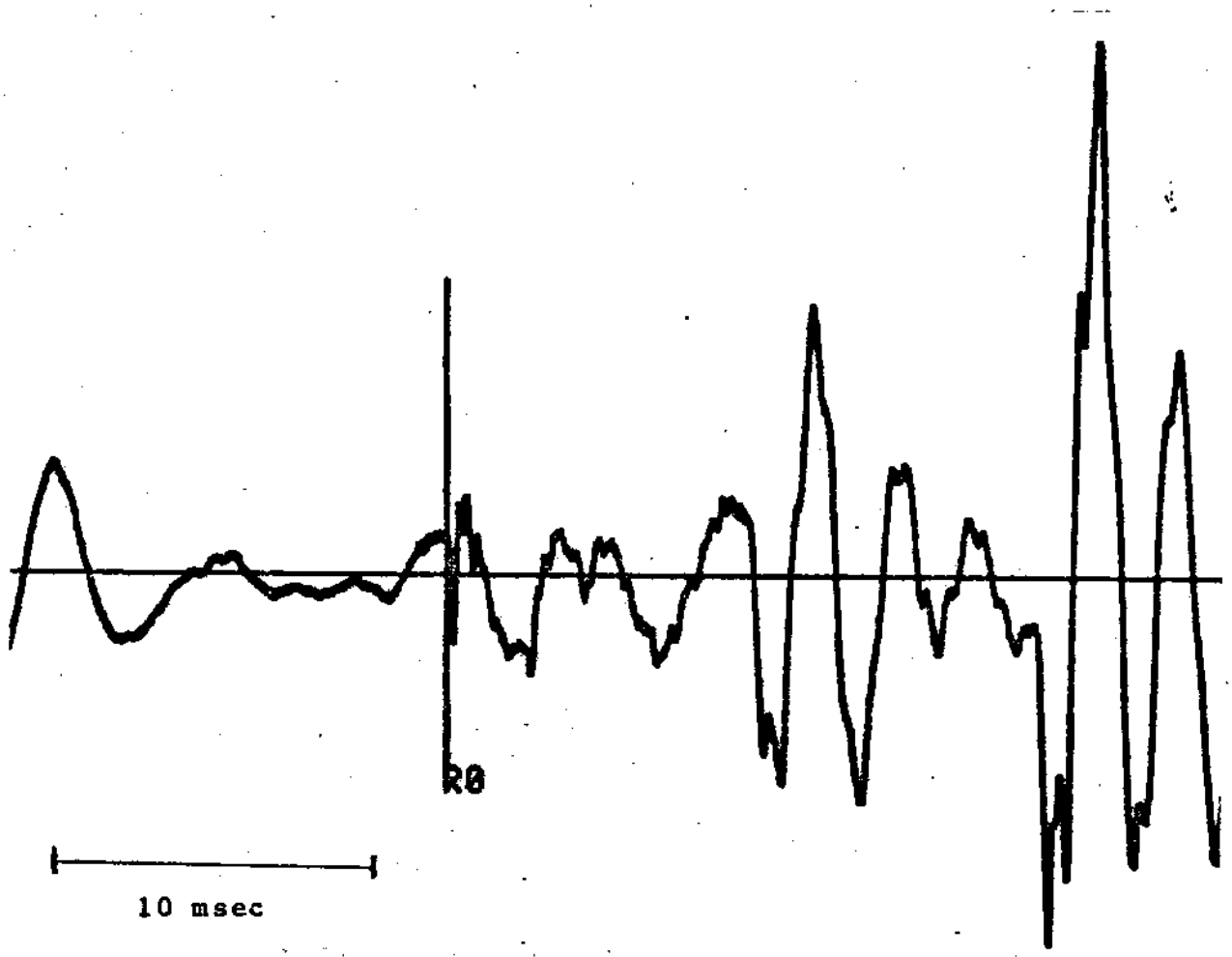

Fig. 3-1b -- Waveform showing token of dur read by speaker JP. The cursor is set at the burst, which is voiced through from prevoicing. This token was still identified as [d] without prevoicing.

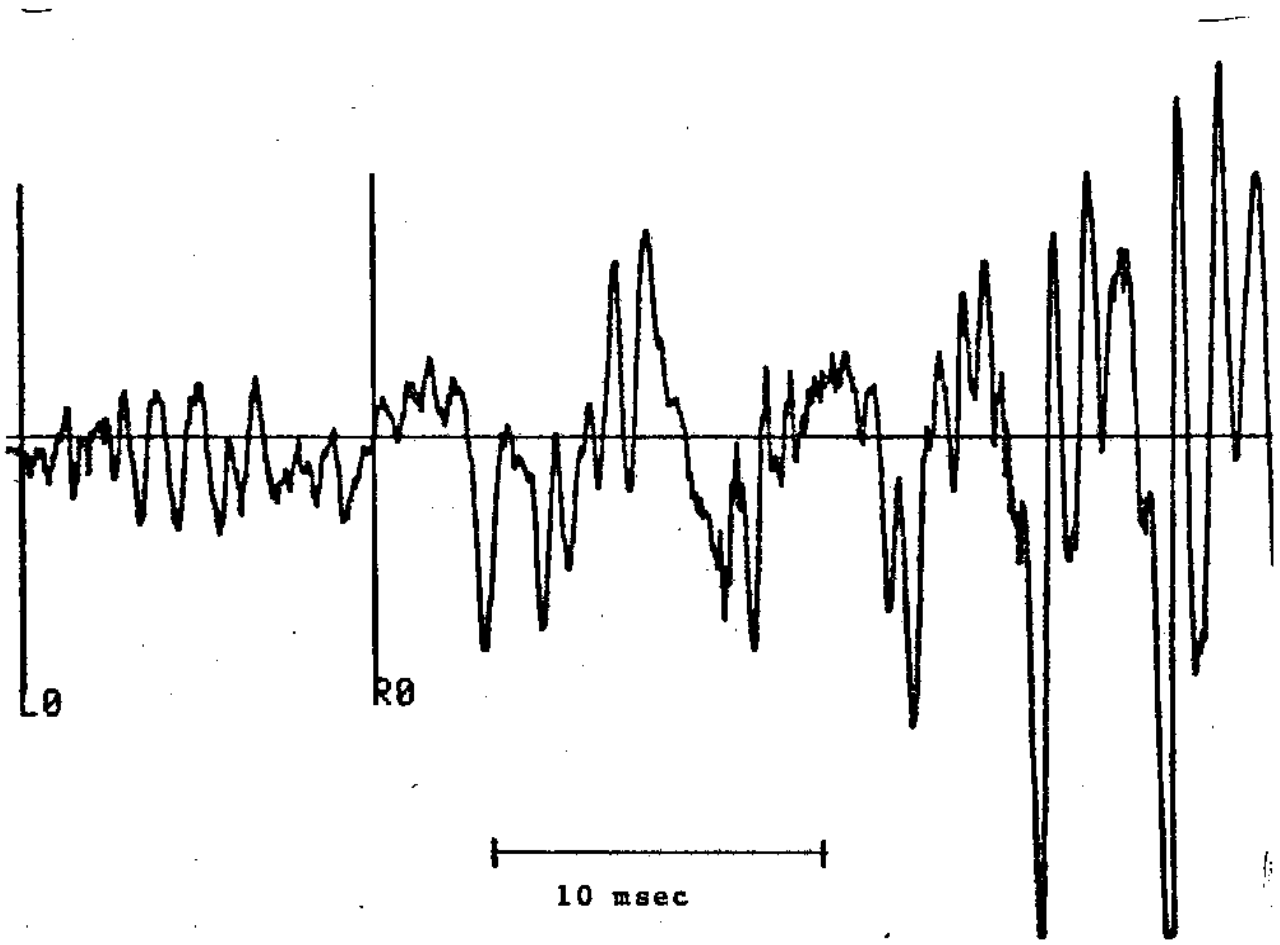

Fig. 3-2a -- Waveform showing token of tama read by speaker #16 with a VOT value of 10.8 msec. The burst is relatively long but not particularly noisy.

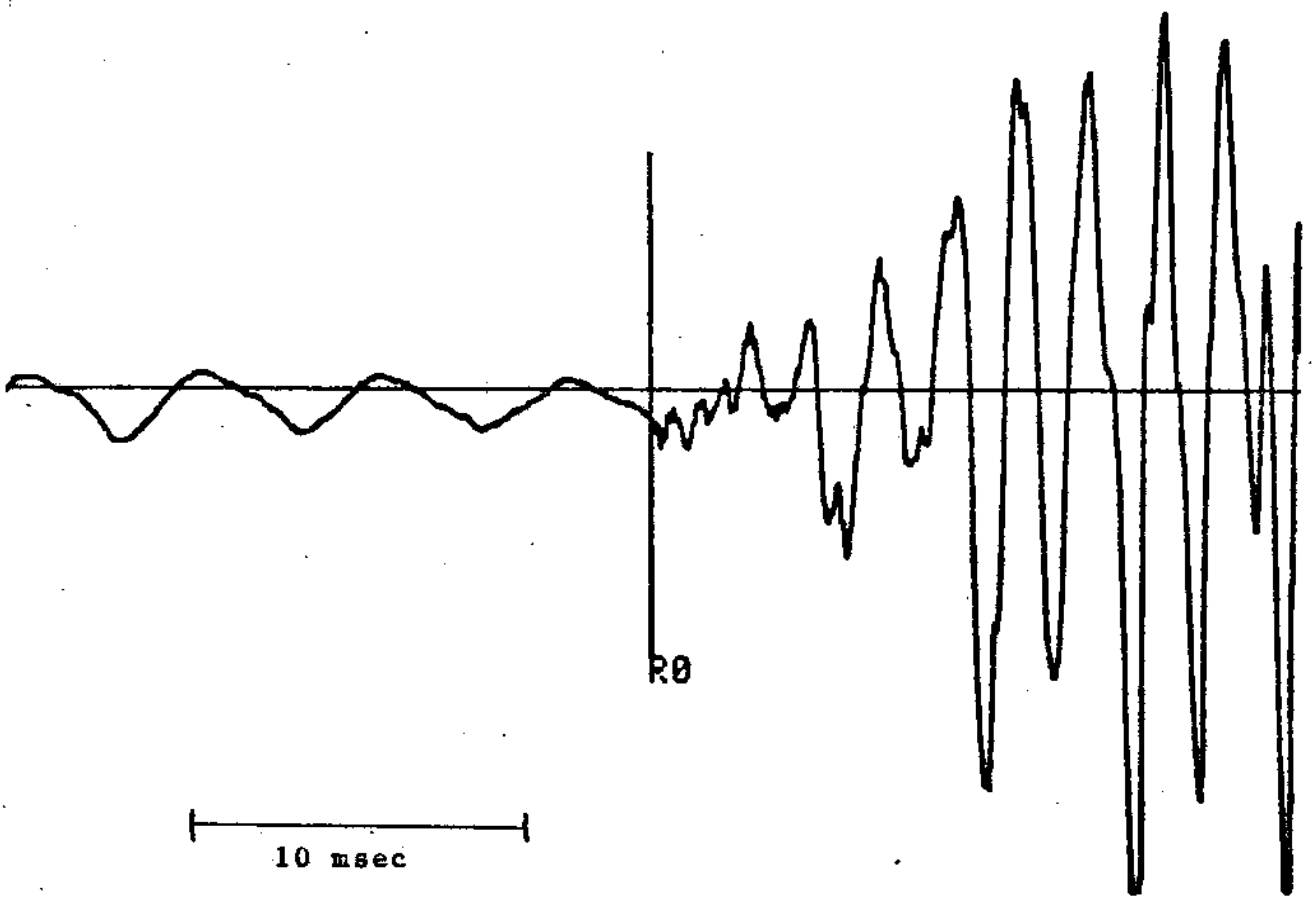

Fig. 3-2b -- Waveform showing token of dama read by speaker #16. The burst, located at the cursor, is very short and essentially voiced through from prevoicing. Note that its onset is much like that of the [t]-burst shown in Fig. 3-2a.

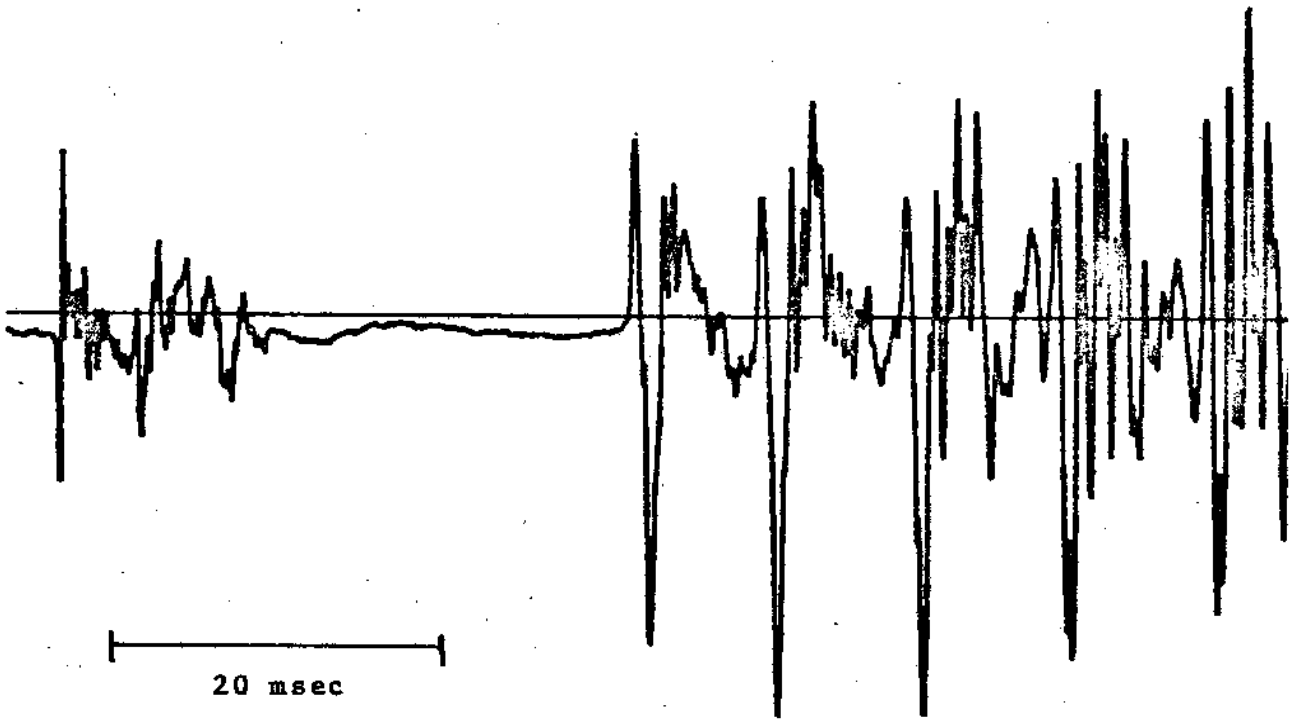

Fig. 3-3a -- Waveform showing natural-edited stimulus with [t]-burst from tama, inserted silence, and voiced transitions and vowel from tama.

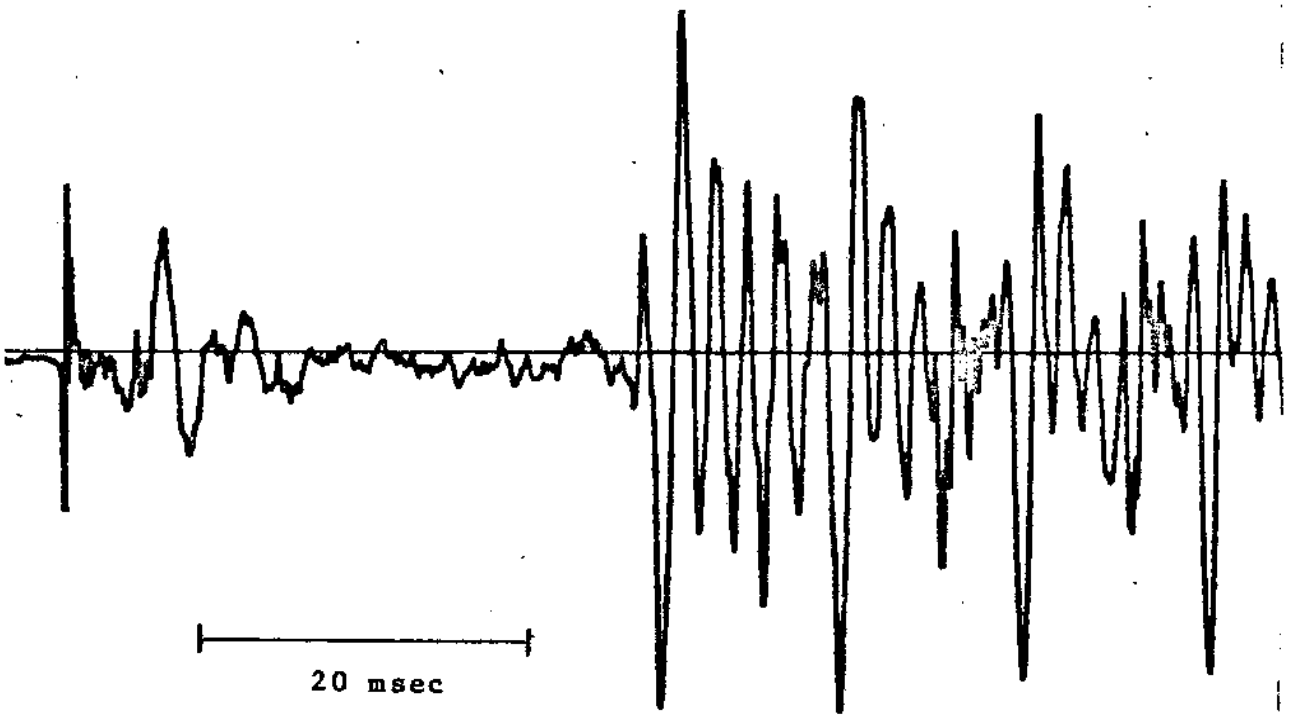

Fig. 3-3b -- Waveform showing natural-edited stimulus with [d]-burst from dama, aspiration from tama, and voiced transitions and vowel from dama.

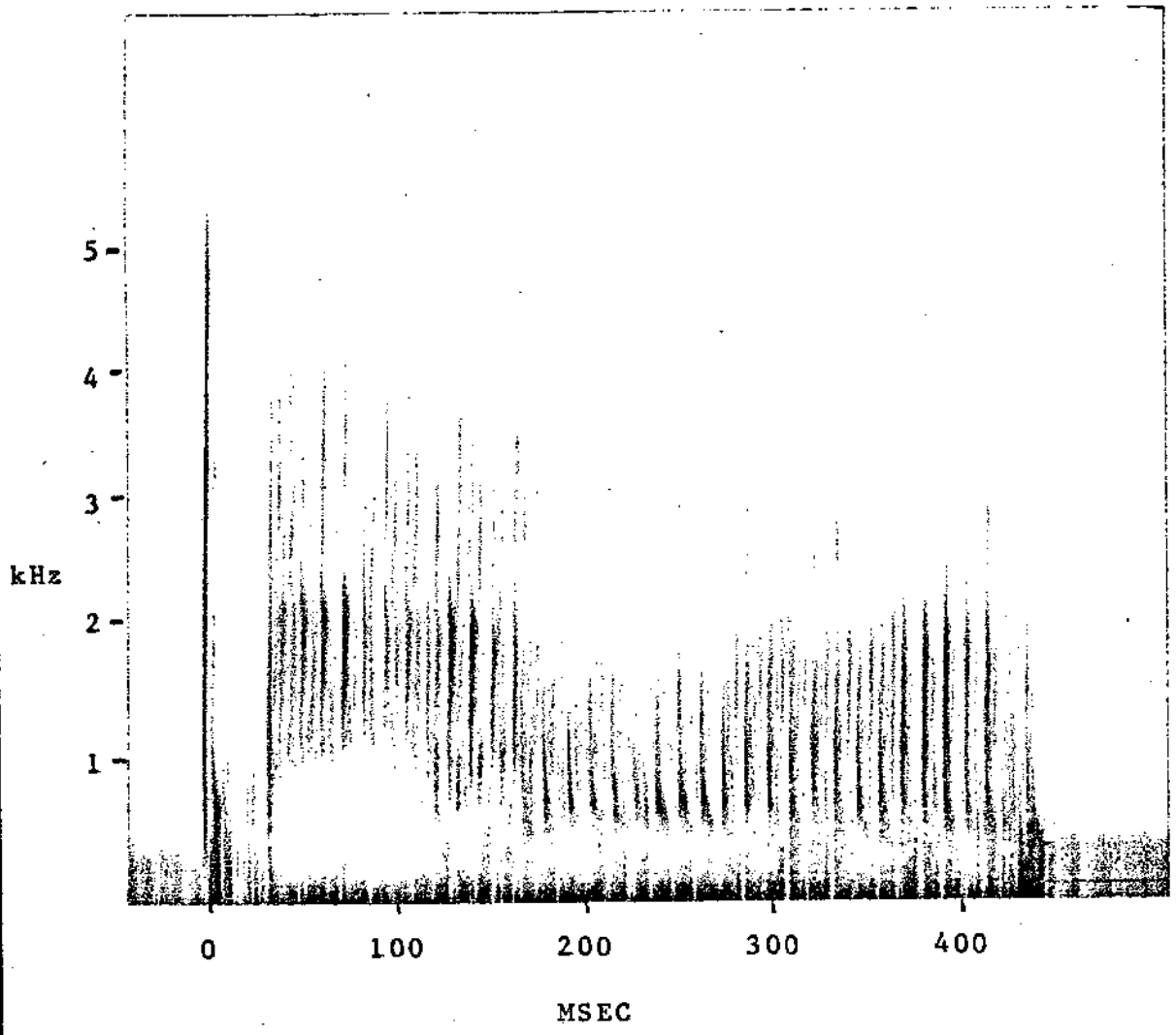

Fig. 3-3c -- Spectrogram of natural-edited stimulus shown in Fig. 3-3b where the two transition onsets, voiceless and voiced, can be seen.

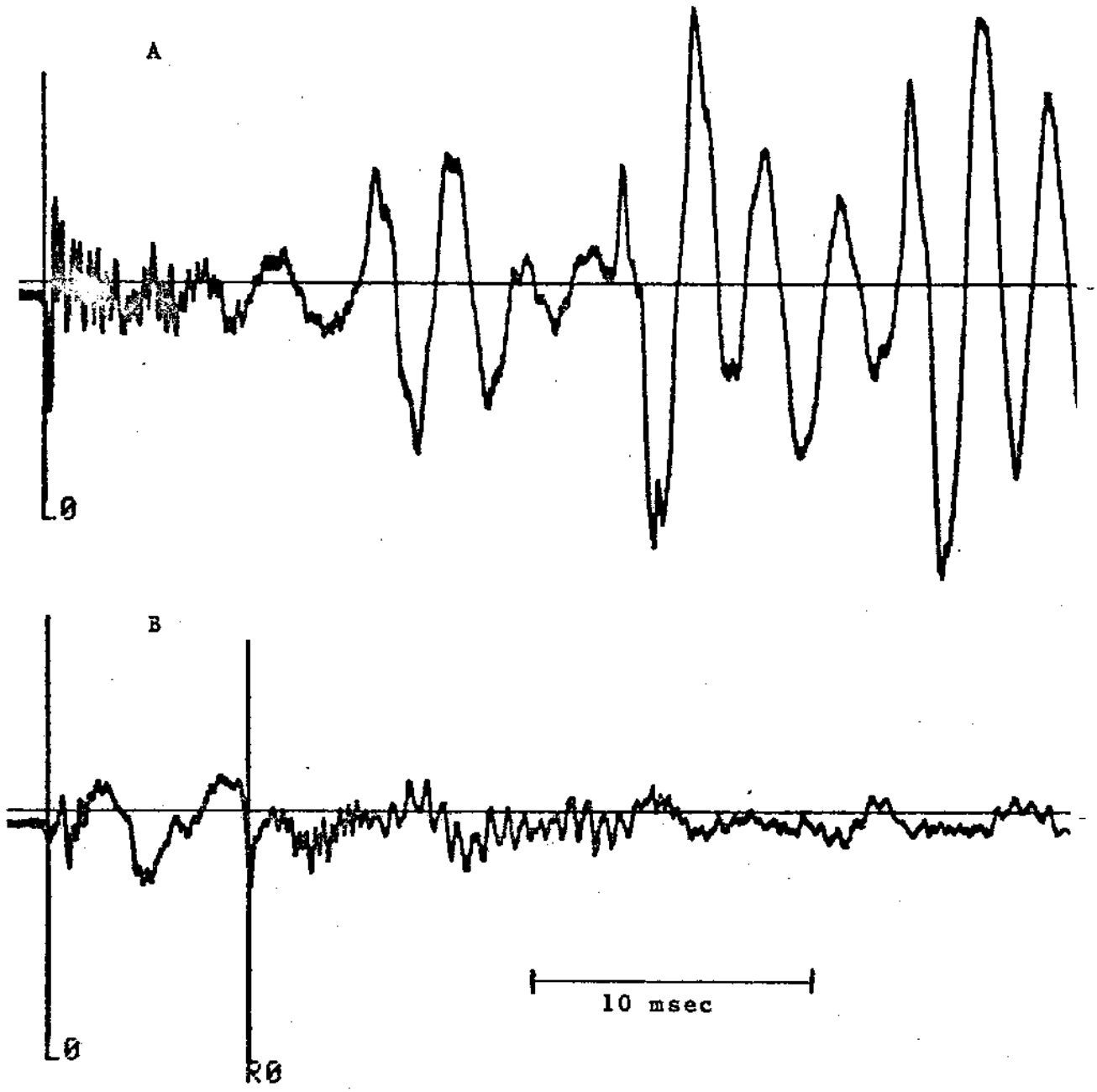

Fig. 3-4a -- Waveform showing natural-edited stimulus containing [t]-burst and [d]-transitions.

Fig. 3-4b -- Waveform showing natural-edited stimulus containing [d]-burst and t -transitions. The burst duration measured between the cursors is 7.3 msec.

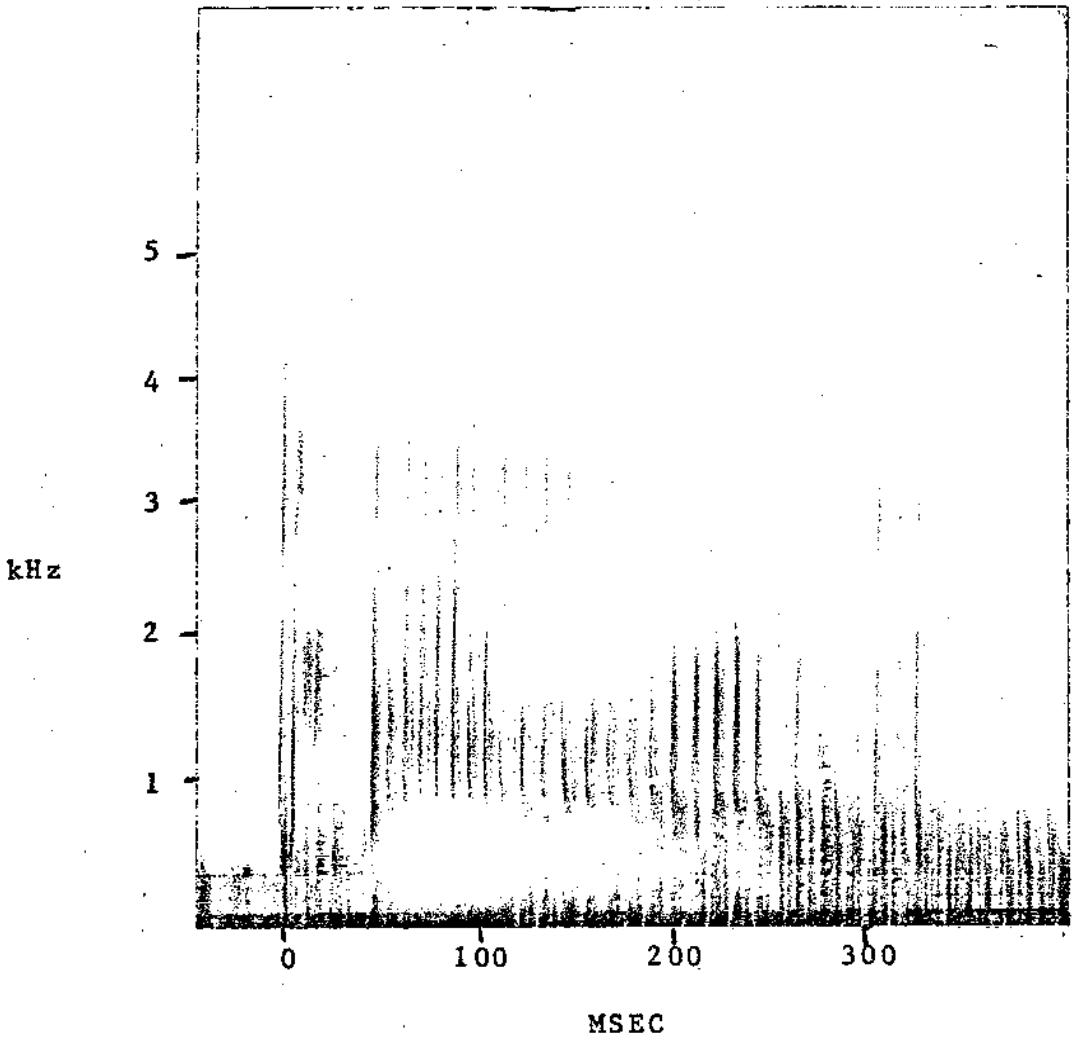

Fig. 3-4c -- Spectrogram of stimulus shown in Fig. 3-4b, showing the match of transitions (from tama) and steady-state vowel (from dama).

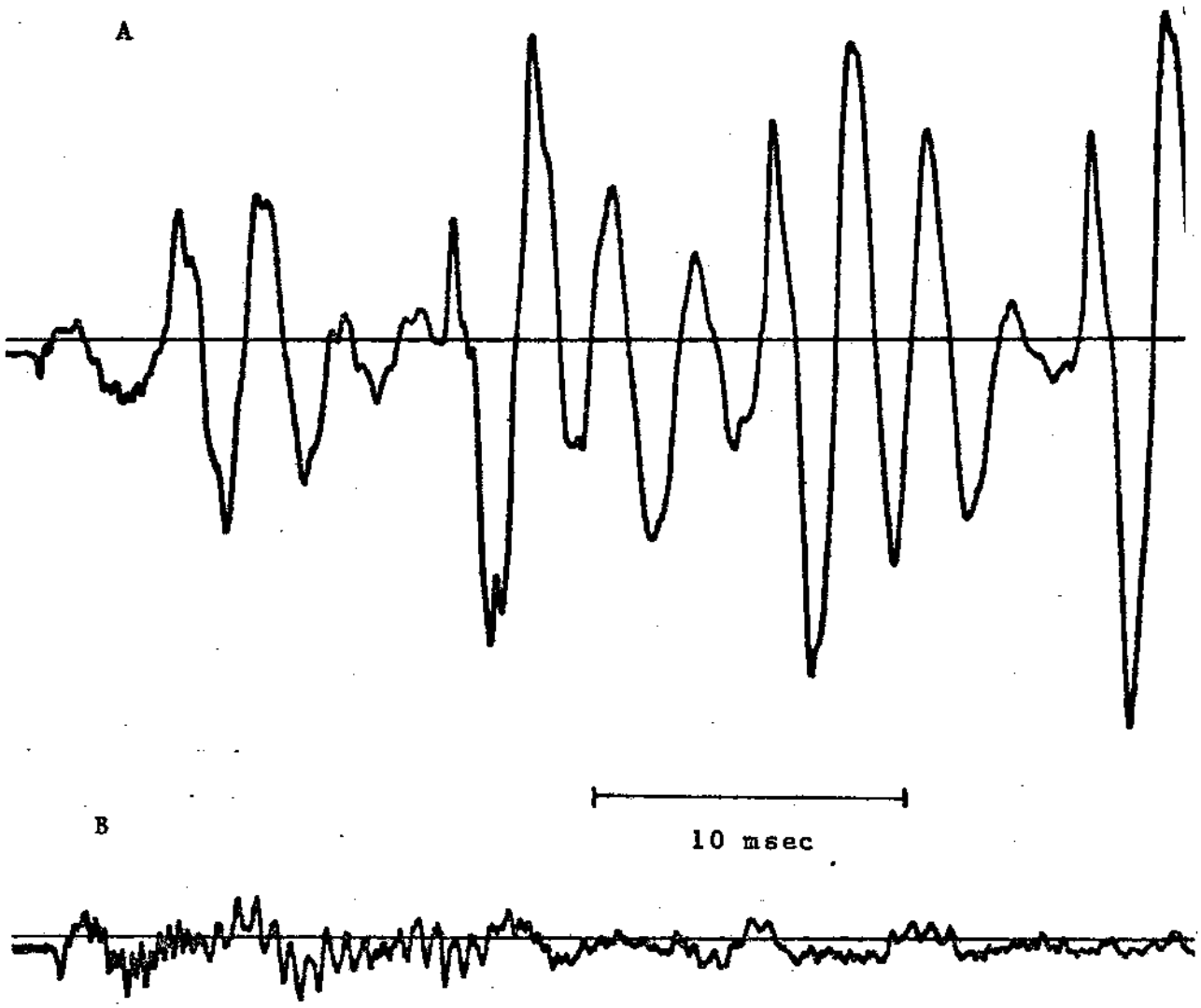

Fig. 3-4d -- Waveform of natural-edited stimulus with no burst, showing onset of transitions (from dama).

Fig. 3-4e -- Waveform of natural-edited stimulus with no burst, showing onset of transitions (from tama).

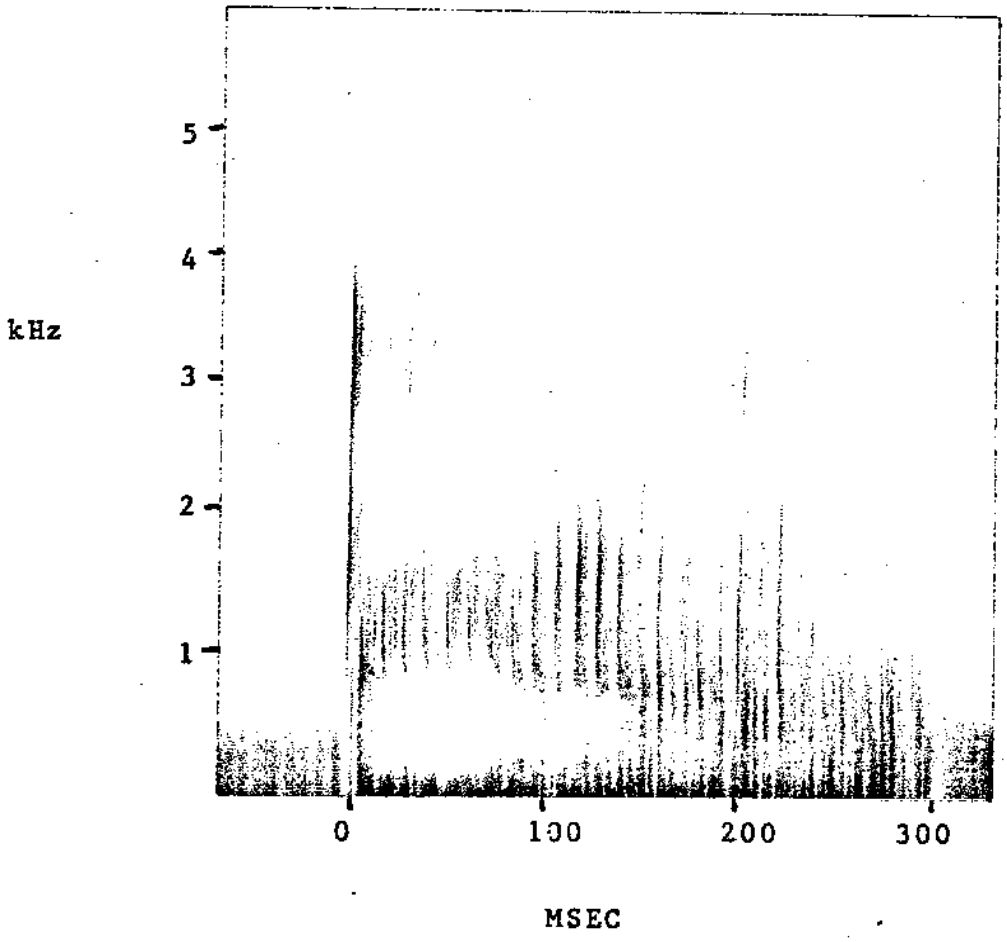

Fig. 3-4f -- Spectrogram showing stimulus with no transitions, but with burst joined directly to steady-state vowel (from tama).

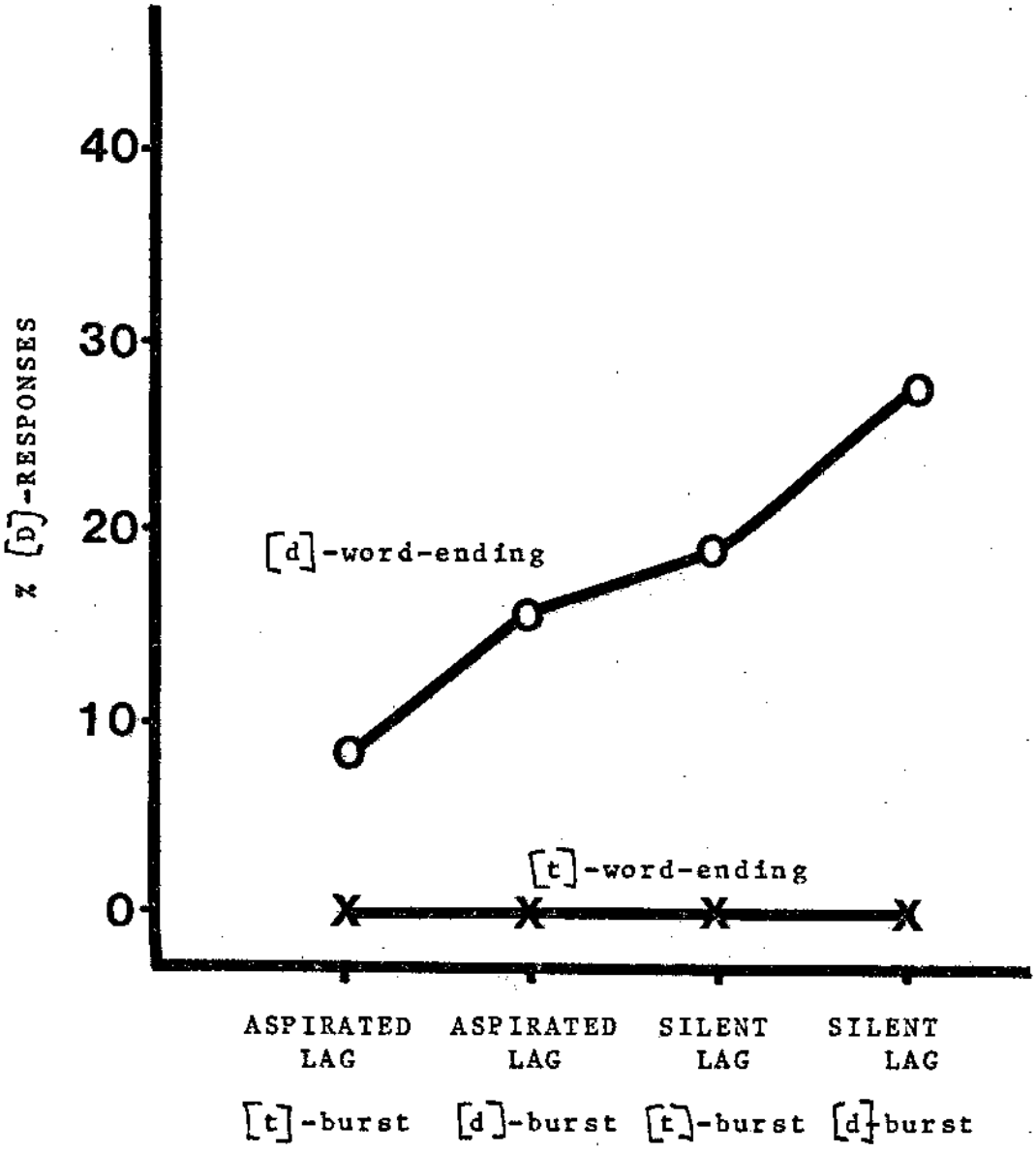

Fig. 3-5 -- Results of Exp. II, Voicing Lag, given as %[d]-responses for 24 subjects for each of eight stimuli.

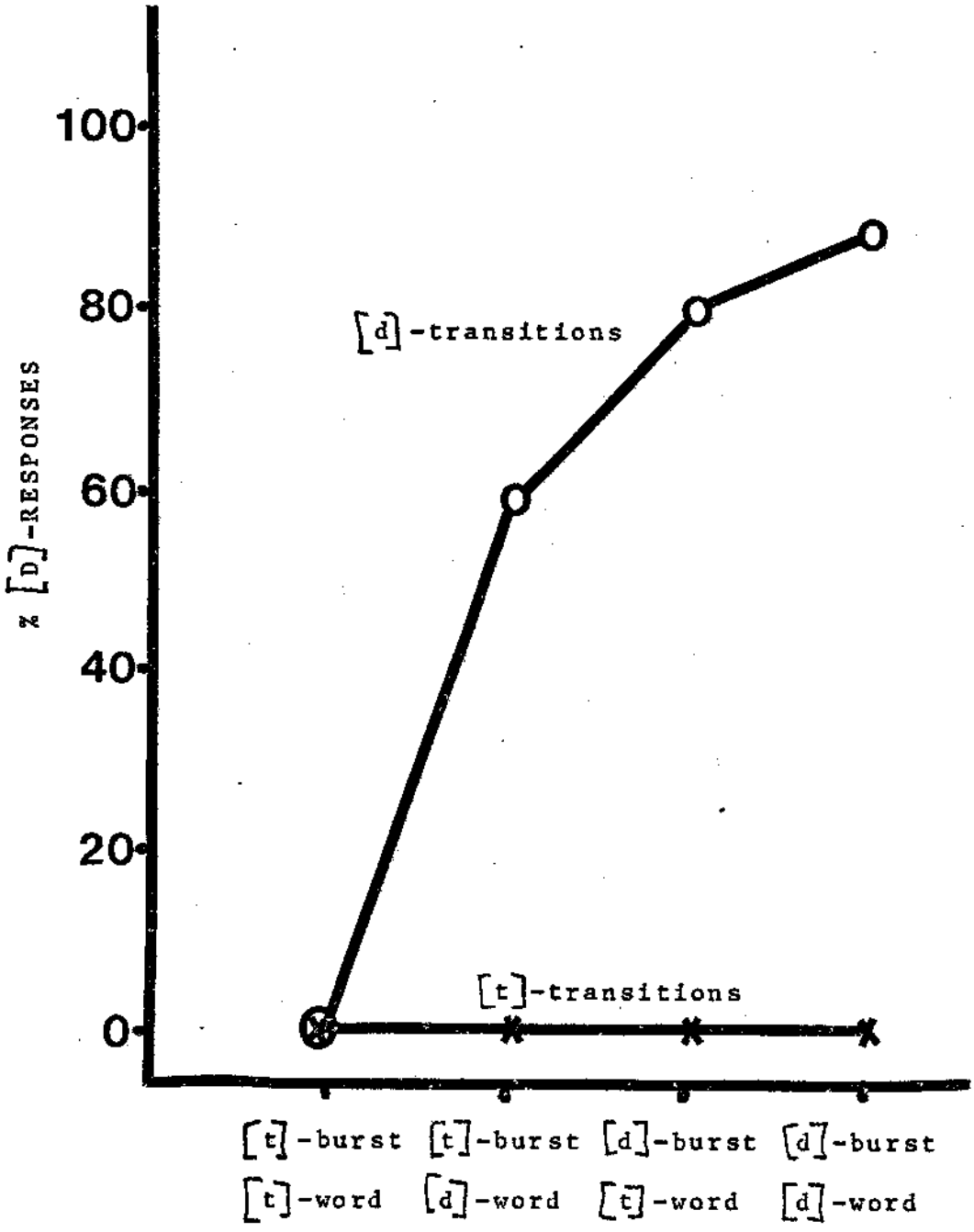

Fig. 3-6 -- Results of Exp. III, Transitions, given as % [d]-responses for 24 subjects for each of eight stimuli.

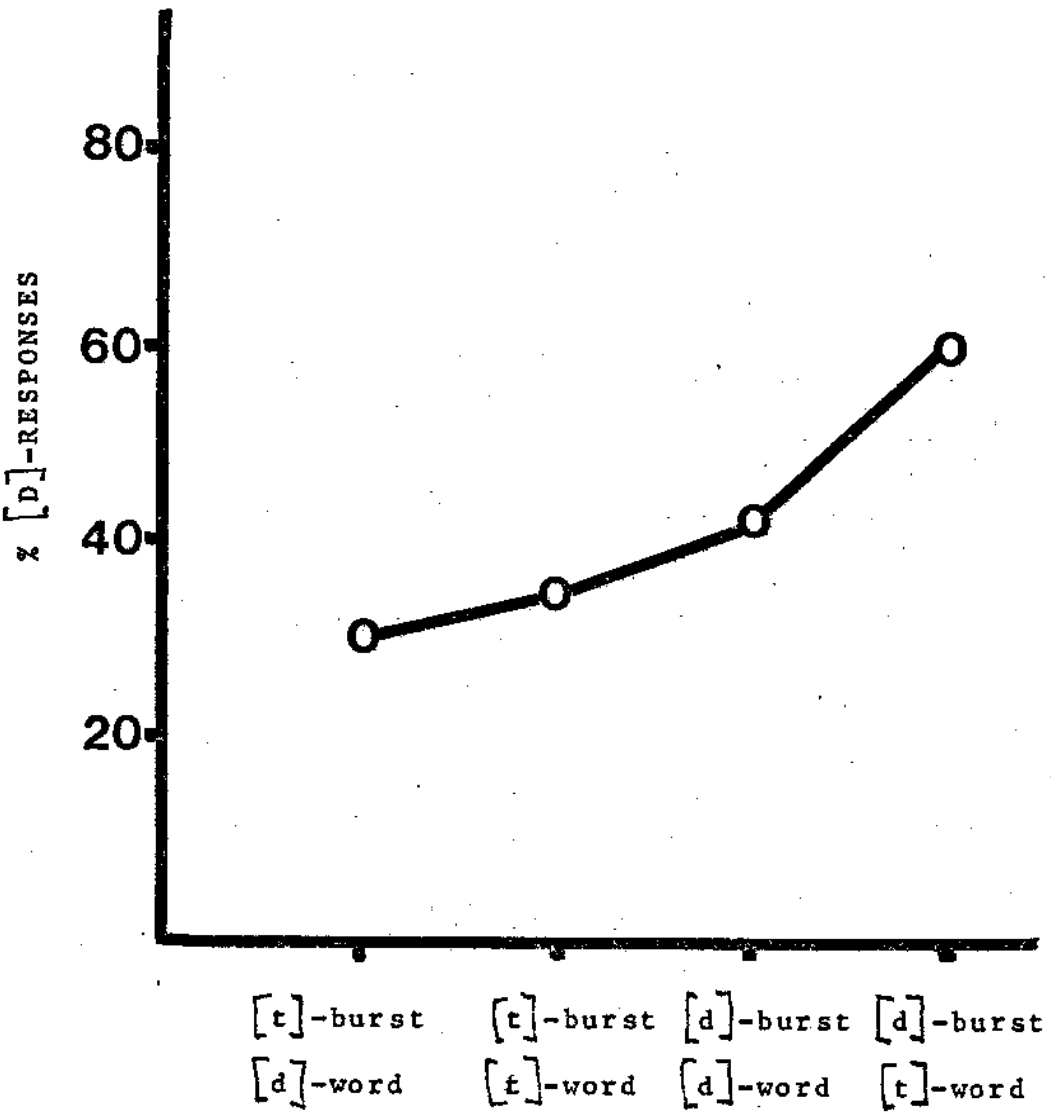

Fig. 3-7 -- Results of Exp. III, Transitions, for transitionless stimuli, for 24 subjects, given as % [d]-responses for each of four stimuli.

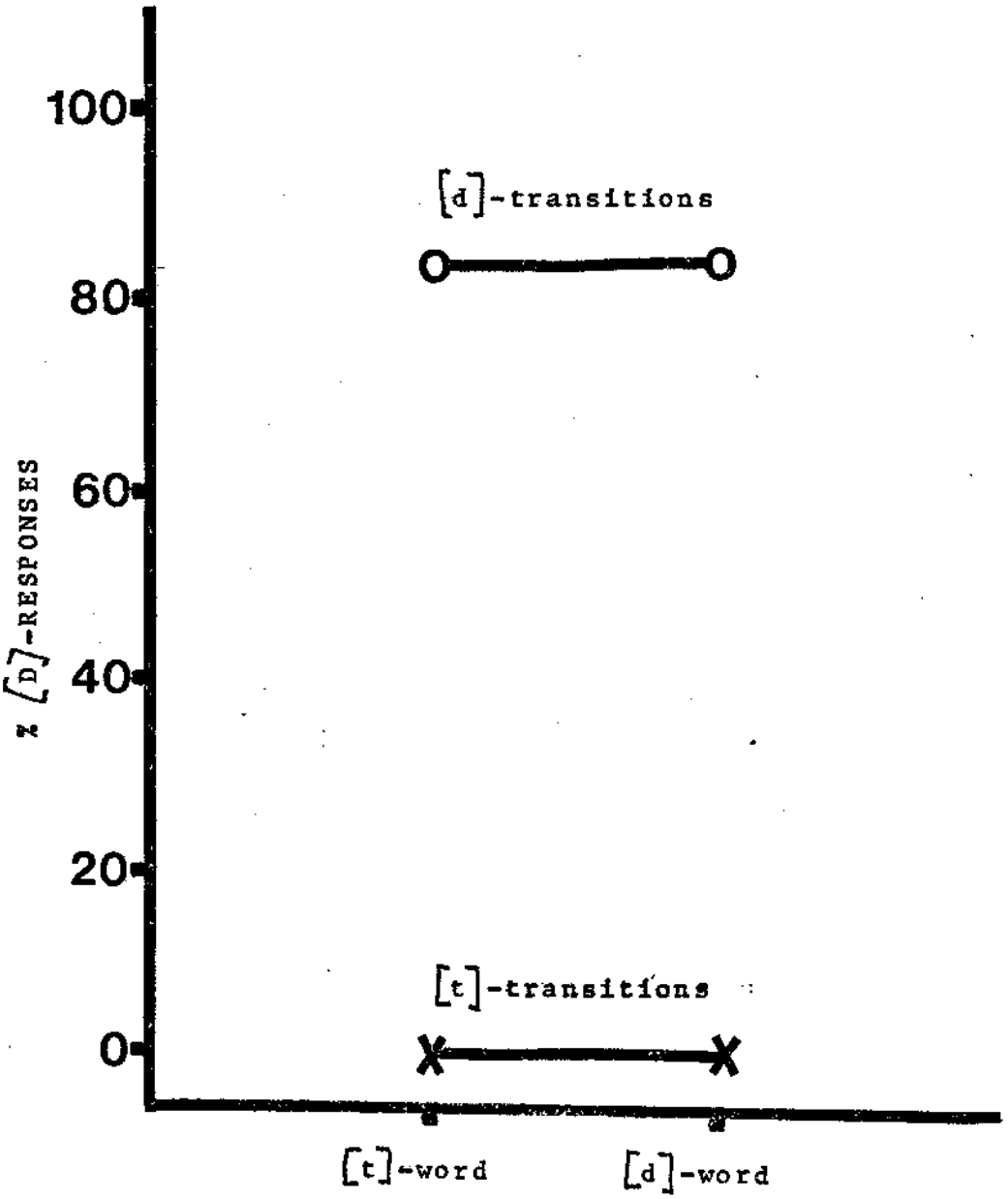

Fig. 3-8 -- Results of Exp. III, Transitions, for stimuli with transitions and steady-states (words) but no bursts, for 24 subjects, given as % [d]-responses to each of four stimuli.

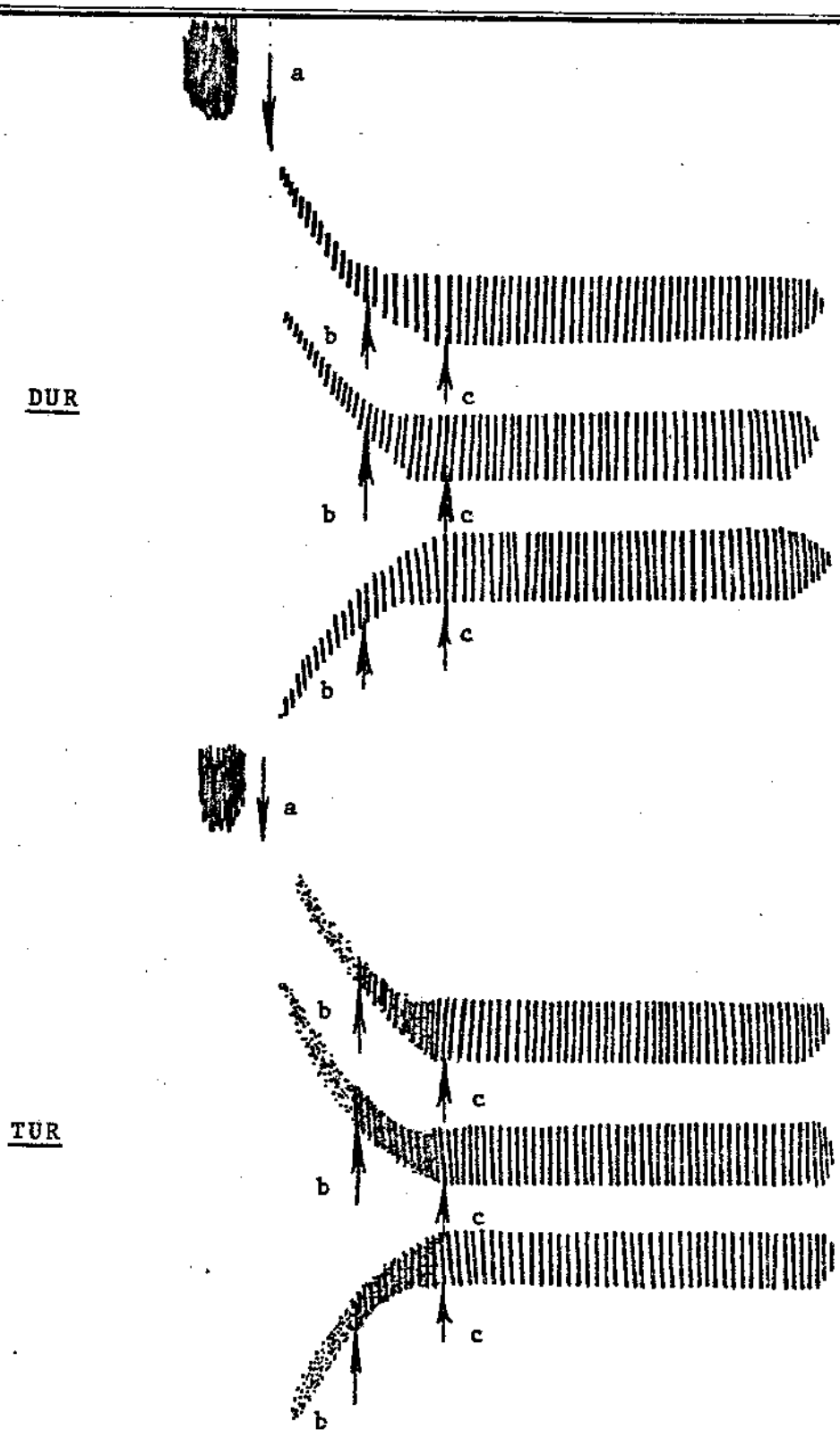

Fig. 3-9 -- Schematic stimuli showing how the stimuli used in Exp. I and III differ. Both sets use bursts, but Exp. I stimuli use aspiration (between a and b in bottom schematic) plus voiced transitions (from b on), while Exp. III stimuli use transitions (from a to c) plus steady-state vowels (from c on).

CHAPTER FOUR -- Medial-Stop Voicing Contrast

4.1. Introduction

One of the goals of a unified description of voicing contrasts should be to account for variation in the acoustic correlates of and cues to the contrast across phonetic environments. In Ch. 2 the differential effects of speaking style (e.g. word lists or running speech) were considered. In this chapter, the effect of position in the syllable or word on the voicing contrast is addressed. Work on English voicing has focused on word-initial position, with less work done on medial and final contrasts. It has not been clear how to incorporate the results for the non-initial contrasts into the account developed for initial contrasts. Therefore it seems important to gather at least preliminary data on a Polish medial contrast to determine how it differs from the corresponding initial one.[1] Data on both production and perception is reported in this chapter.

Medial voicing contrasts can be described in two ways, and the descriptions can then be tested perceptually.The first is to apply the same descriptive mechanism that was developed in Ch. 2 for initial stops in running speech, the modified VOT measure based on closure voicing. The second is to consider those parameters that have been found to correlate with the English contrast, such as closure and preceding vowel duration. An important feature of the English medial contrast is that the closure interval for voiced stops is often voiced throughout, and in Ch. 2 it was seen that this is true as well of Polish word-initial [d] in running speech. The English medial contrast and Polish initial and medial contrasts are generally described as that of "fully voiced" - "voiceless

unaspirated". Therefore some acoustic similarities between the English and Polish medial contrasts may be expected.

4.2. Production Data

4.2.1. Methodology

All observations on the production of Polish medial [t] and [d] are derived from the same corpus of recorded speech described in Ch. 2. In particular, the minimal pair rata - rada, which was recorded by 24 speakers in Wrocław, Poland and by three speakers in Providence, contributes most of the data presented here. Also available are a few tokens of medial stops that occur in the sentences read by five speakers in Łódź, Poland and three speakers in Providence (they occurred in the words bilety, gazeta, jadę, ktoredy) and the additional minimal pair roty - rody read by the three speakers in Providence.

All measurement and observation of these medial stop tokens was made from oscillographic displays of the PDP-11/34 computer in the Brown Univ. Linguistics Dept. Phonetics Lab.

Examples of medial [t] and [d] are shown in Fig. 4-1 and 4-2. The "a" and "b" figures show entire closure intervals; note that the horizontal (time) scale is decreased from that of other figures. The "c" figures give more detail of the onset of closure. Landmark points for measurement of closure duration are shown in these figures; in general the procedure is the same as that described in Ch. 2 for initial stops in running speech. Positive VOT measurements for [t] were also made according to the criteria described in Ch. 2.

The duration of the preceding vowel was measured to the

beginning of closure. Deciding where to mark the "beginning" of the vowel is usually problematical. Here the same procedure was used for the beginning of the vowel as for the end of it: a change in the spectral pattern, particularly peaks and dips in the waveform, after the initial consonant. Polish /r/ is trilled, and each tap is clearly distinguishable acoustically. For the minimal pairs beginning with /r/, the beginning of the vowel was set after the last tap. An example is shown in Fig. 4-3.

4.2.2. Production Measure I: Voicing

As for initial stops, medial stop voicing can be seen to have two characteristics. First, a straightforward VOT measurement can be made for virtually all tokens of [t] and for some tokens of [d]. Second, for [d], voicing can sometimes be observed continuing from closure through the burst and into the syllabic nucleus without a break.

Several examples of waveform displays of medial [t] and [d] are shown in Fig. 4-4 and 4-5. It can be seen that closure for [t] can be voiced (especially in running speech), or very noisy, or silent. Bursts for both [d] and [t] can be either voiced or voiceless. The closure voicing for [d] can have an unusually high amplitude relative to the surrounding signal--almost as high as the vowels--or it can die down immediately before the burst, although this is rare. The [d]-burst can be quite weak, or it can be missing altogether.

From the 24 readings in Wrocław, VOT was measured (from the burst) for 80 tokens of medial [t] in the words tata, data, and rata. (There are extra tokens because some readers repeated the pairs.) The mean VOT for these medial-[t]-tokens was +20 msec, about

the same value as was obtained for initial [t] in the minimal pair readings. The VOT values that occur above about +25 msec are mostly due to multiple bursts, rather than to long lag. For medial [d], as for initial [d] in running speech, a true measure of VOT is often impossible, since voicing continues through the burst from closure.[2] The voiced closure duration measure was used exclusively for [d]. Closure voicing for medial [d] is sometimes found with such high relative amplitude that the burst is effectively obscured. In cases where the burst could not be located, an estimate like that used in Ch. 2 was made. The "VOT" measure then for [d] is equal to the closure duration, and the mean value was -92 msec. Values in sentence contexts appear to be much lower.

Since there is always some voicelessness after a [t]-burst, only positive VOT measures were made, even when the closure was voiced. However, even with a voiced [t]-closure, there are still differences between [t] and [d]. First, the voicing during [t]-closure has a very low amplitude and is often somewhat irregular over the course of the whole interval. It can even resemble the type of voicing that can be found during a voiced burst: many shorter irregular pulses and sub-pulses with some noise, compared to the more sinusoidal closure voicing. Secondly, a [t]-burst is never entirely voiced throughout. Sometimes, after voiced closure, there is a double burst, with the first burst being voiced and the second being voiceless. Thirdly, there is a distinction in the time between the burst and the amplitude build-up for the vowel after [t] and [d]. After a partly-voiced [t]-burst, there is often a fairly long "lag" interval containing very low-amplitude voicing, before the amplitude increase for the vowel. This voicing looks like weak,

noisy closure-voicing. The VOT value as it has been defined here will be quite low in such cases, but a measure from the burst to the amplitude rise would probably be highly contrastive. The amplitude rise in a [d]-token almost always begins immediately after the burst.

4.2.3. Production Measure II: Closure Duration

Closure duration was measured for each of the 24 readings of the rata - rada minimal pair from Wrocław. The distribution of measurements is shown in Fig. 4-6. The mean duration for [t] is 130.1 msec, and for [d], 91.5 msec. The difference between the [t] and [d] closure durations was statistically significant (t_{23}= 8.81, $p < .001$). It can be seen in Fig. 4-6, however, that, across speakers, the closure durations are not distributed into two separate groups, even in these minimal pairs. While 22 of the 24 speakers show a sizable difference, there is no single criterial value separating the two classes across subjects. Durations less than 90 msec are almost uniformly [d], and greater than 140 are [t], but durations from 90 to 140 msec correspond to either [t] or [d].

Inspection of some of the sentence tokens for speakers DB and MG indicates, however, that there is good category separation for [t] and [d], with overall lower values than those found for minimal pairs. Medial [t] in gazety and bilety have closure durations clustering at about 100 msec; medial [d] in jade and Ktoredy have shorter values of 50 to 70 msec.

Closure duration in English trochee minimal pairs does not show the category overlap the Polish pairs do. Lisker (1957) collected data on medial labial stop closures, and found

essentially no overlap. The [b] durations ranged from 65 to 90 msec, with a mean of 75 msec, and the [p] durations ranged from 90 to 140 msec, with a mean of 120 msec.

A ratio of the [t]-closure durations to the [d]-closure durations is usually calculated as a measure of differentiation. There are two ways of doing this, which give different results. One way is simply to take the mean [t]-duration and the mean [d]-duration (130.1 and 91.5 msec in the Polish case) and form their ratio. For the Polish data, that ratio is 1.42; for Lisker (1957)'s labial data, it is 1.6. Port (1977), using carrier sentences for his pairs, obtained a ratio of 1.35 for English medial labials in trochees. Another way of calculating ratios is to calculate the ratio of each pair, and then average all the ratios. For the Polish data, this ratio is 1.53. Note that the second method, averaging many ratios, gives a larger distinction than does taking the ratio of two means. Fig. 4-7 shows the distribution of the ratios of the [t]- to [d]-closures for the 24 speakers. Only one speaker did not have a longer [t]- than [d]- closure.

Thus the mean closure durations for [t]and [d], and the ratio of those means, show a distinction in this parameter along the voicing dimension. However, the frequency distribution of the measurements does not show good category separation, and this general overlap makes it possible that closure duration is not as strong a cue for voicing in Polish as it is in English. There is a substantial range of durations that may be perceptually ambiguous, at least in the minimal pairs condition. In sentences there may actually be less overlap. If this is so, it is the reverse of the usual trend for distinctions to be enhanced in isolated minimal