pairs.

4.2.4. Production Measure III: Preceding Vowel Duration

Durations of vowels preceding voiced or voiceless stops were measured for the same 24 readings of <u>rata</u> - <u>rada</u> from Wrocław. The mean duration of vowels before [t] was 167.4 msec, and of vowels before [d], 169.5 msec. The ratio of these two means is .99. In addition, for each pair of measurements, the ratio of vowel before [t] to vowel before [d] was computed. The mean of these 24 ratios is 1.0. The data for Polish clearly show that vowel duration does not vary systematically according to the voicing of the following consonant. Thus it will not be considered as a potential cue to the medial voicing contrast.

Comparable data for English vowels before voiced and voiceless medial stops have been collected by Sharf (1962) and Klatt (1973). The ratios they obtained were .75 and .79, respectively.[3]

It has been suggested, for example by Delattre (1962) and Chen (1970), that some amount of vowel lengthening occurs universally before voiced obstruents. Previous to this study, at least seven languages have been studied for this effect, and all of them show at least a 10% difference in duration between vowels before voiced consonants and vowels before voiceless consonants. It has been thought that such a difference would be found in <u>all</u> languages, and that it in fact constitutes a phonetic universal.

Our Polish data contradict this claim. The finding was extended by recording speakers of Czech reading contrasting words in that language. Czech, another West Slavic language, is similar to Polish in many ways, but it has phonemic vowel length contrasts. It seemed likely that Czech would not lengthen vowels before voiced consonants, because vowel duration would be reserved for the length contrast. Three native speakers of Czech read several words of the following form:

$$\begin{Bmatrix} p \\ ml \end{Bmatrix} \begin{Bmatrix} a \\ a: \end{Bmatrix} \begin{Bmatrix} t \\ d \end{Bmatrix} \quad V \qquad (C)$$

The number of phonemic short and long vowels was balanced. The mean duration of vowels before [t] was 193.7 msec; before [d], 204.2 msec. The ratio of these two means is .95; the mean of the 24 ratios is .98. Thus there is a slight tendency for vowels to be lengthened before voiced consonants in Czech, at least based on this small amount of data, but the effect is minimal. The difference in durations did not reach statistical significance ($t_{30} = -.37$, $p > .20$). This finding shows that Czech, as well as Polish, does not show the putatively universal vowel lengthening before voiced consonants.

4.2.5. Discussion of Production Data

Two of the three measures made showed systematic variation according to the voicing of a medial stop. Neither, however, was effective in distinguishing [t] from [d] consistently. One measure was VOT, as defined to include closure voicing in running speech for [d]. Notice that in the case of medial stops, this measure must not apply to the occasional voiced [t]- closure, or voiceless [d]-burst. To the extent that a [t] and a [d] can both have voiced closures and voiceless bursts, this measure will show overlap between the two voicing categories. It was suggested that there may be other differences between such tokens, though, such as the nature of the voicing during closure and burst, or the speed of the amplitude rise after stop release.

The other measure that varied systematically with stop voicing was closure duration. Individual speakers produced much longer closures for [t] than for [d], and the means for the group were distinctive for the two consonants. Nonetheless, the distribution of individual tokens for the group showed a fair amount of category overlap. Thus it is not clear how strong a cue closure duration is for the medial voicing contrast. Closure duration for [d] is of course also used as a measure of VOT. Possibly some combination of the closure and voicing parameters would be the most effective cue.

Vowel duration was found not to differentiate a following [t] from [d] in this context. Although this parameter has been thought to be a universal correlate of voicing contrasts, the Polish data does not support such a claim. The possible cue value of vowel duration will not be considered further. Perceptual experiments will focus on closure duration and voicing.

4.3. Perception Data

4.3.1. Methodology

Stimuli were constructed from tokens of rata, rada, tata, and data spoken by MG. The [t] of rata, which is shown in Fig. 4-4a, has about 160 msec of closure, all with low-amplitude voicing, followed by a voiceless burst. Detail of the closure voicing is shown in Fig. 4-10C. The [d] -burst, which is shown in Fig. 4-5d, is voiced throughout. The initial /ra/ syllables of rata and rada were of nearly identical durations.

The initial bursts of MG's tata and data are shown in Ch. 2, Fig. 2-14. This token of data was identified as having initial [d] when prevoicing was present, but initial [t] when prevoicing was removed. The [d]-burst has a voiceless beginning, but some irregular voicing with frication immediately after that.

Various stimuli representing tokens of rata, rada, tata, and data were made. They are described below for each experiment. All stimuli representing tata and data spoken by MG, including those constructed for other experiments, were combined into one test for presentation to listeners. All stimuli representing rata and rada spoken by MG were also combined into a single test. For each test, ten tokens of each stimulus were presented in random order to 24 subjects in Wrocław, Poland. See Ch. 2 for details on the procedure.

The number of [d]-responses (out of 10) to each stimulus was determined for each subject, and a mean percentage for the group was computed. Statistical tests were performed on the number of [d]-responses using repeated-measures analyses of variance.

4.3.2. Experiment I

The object of this experiment was to see whether a medial consonant maintains its voicing quality when the preceding initial syllable is removed, and the final syllable is presented as an initial syllable. In English, this is not the case, as a medial voiceless stop is heard as voiced when it is presented as an initial stop. For example, if the second syllable is excised from rapid, it will be heard as bid. This result indicates that for English a somewhat different set of cues must be relevant for initial and medial stop voicing. There is an allophonic shift in the phonetic voicing categories used for medial stops in English trochees. Cues during the vowel and closure intervals, before the burst, may be as important as cues during and after the burst. Experiment I considers the comparable situation in Polish.

The [ta] of rata and [da] of rada were used as initial syllables of tata and data. (These disyllables were used instead of monosyllabic ta and da so that the stimuli could be combined with other stimuli into a single test.) The final [ta] came from tata or data. Thus there were two initial syllables and two final syllables, allowing four stimuli to be made. The "initial" [ta] and [da] were taken from the burst to the end of the syllable. The final [ta]'s consisted of the voiceless closure plus the rest of the syllable.

Subjects' responses to each stimulus were analyzed with a two-way repeated measures analysis of variance. The means for each condition and factor are given in Table 1, and the results of the analysis in Table 2. As would be expected, the source of the final syllable [ta] had no influence on subjects' responses, either as a main effect or in the interaction. On the other hand, the identity of the "initial" consonant did have a significant effect: a medial [t] in initial position was still heard as [t], and a medial [d] in initial position was heard as [d]. With both consonants, the categorizations were weaker than for the same syllables in medial position.

As reported in Ch. 2, MG's productions of initial [d] from which prevoicing had been removed were heard as [t]. Thus the medial [d], which also had no prevoicing, had more effective cues for voicedness than did the edited initial [d]'s. The reason for this difference may be burst voicing. The [d]-burst from rada, shown in Fig. 4-5d, is entirely voiced, unlike the [d]-bursts from the initial syllables. Bursts tend to be voiced throughout more

TABLE 1

% [d]-responses for 24 subjects, Experiment I

	"initial" [ta]	"initial" [da]	MEAN
final [ta] from <u>tata</u>	11.7	67.5	39.6
final [ta] from <u>data</u>	12.9	65.0	38.9
MEAN	12.3	66.3	

TABLE 2

Results of ANOVA, Experiment I

Source	df	MS	F	p
Initial (ta or da)	1	698.76001	46.12	<.001
error	23	15.15168		
Final (from tata or data)	1	0.09375	< 1	–
error	23	2.22418		
Initial X Final	1	0.84375	< 1	–
error	23	1.97418		

often in medial than in initial position. However, this burst produced only 66% [d]-responses, compared to JP's 80% and better for initial [d] without prevoicing. This result indicates that, while burst voicing can yield a categorization as [d] , some other cue or cues to voicedness are still not optimal in MG's tokens. Data on perception of JP's medial stops in initial position would be useful in this regard.

4.3.3. Experiment II

The object of this experiment is to see whether initial prevoiced [d] and medial voiced [d] are perceptually equivalent in medial position. Fig. 4-8 illustrates phonation in prevoicing and voiced closure. Each figure shows the most common pattern for that context, but both kinds of phonation, as well as intermediate patterns, occur in both contexts. The pattern of phonation in Fig. 4-8b for medial closure resembles nasalization. Fig. 4-9 shows part of the [m] in dama. Voicing during stop occlusion can be effected by opening the velum, resulting in nasalization, but more data would be needed to determine whether nasalization correlates with syllable position. Despite the lack of such data, it seems likely that prevoicing and medial closure voicing would be equivalent in cueing voicedness.

The stimuli were constructed from tokens of rata, rada, and data. The two initial [ra] syllables were isolated once to the beginning of medial closure, and a second time with 60 msec of closure. For rata, the closure was voiced with the low-amplitude voicing sometimes found before medial voiceless stops. For rada, the closure was voiced with high-amplitude voicing. The initial syllable [da] from data was also isolated, once without prevoicing and a second time with 60 msec of prevoicing. The two [ra] syllables without closure were combined with the [da] without prevoicing. Thus all four stimuli contained a [ra] followed by 60 msec of voicing, plus a [da]. The voicing was either prevoicing, strong medial closure voicing, or weak medial closure voicing. The four stimuli are shown in Fig. 4-10.

Subjects' responses to each stimulus were analyzed with a two-way repeated measures analysis of variance. The means for each factor and condition are given in Table 3, and the results of the analysis in Table 4. The means indicate that three of the four stimuli were almost uniformly heard as containing medial [d]. The fourth stimulus, [ra] from <u>rata</u> with 60 msec weak closure voicing plus [da] without prevoicing, was ambiguous.

MG's initial [d] from <u>data</u> requires prevoicing in initial position to be heard as [d]. Prevoicing in medial position also allows this token to be heard as [d]. Medial closure voicing works as well as prevoicing in signaling voicedness, if it is the strong voicing that occurs before medial [d]. The weak voicing that occurs before medial [t] is not sufficient to signal voicedness in that [d], and subjects give slightly more [t]- than [d]-responses. All of the significant effects obtained with the analysis of variance appear to be due to this failure of [t]-closure voicing to produce a voiced percept.

This result, that [t]-closure voicing is not perceptually equivalent to [d]-closure voicing under certain conditions, suggests that weak closure voicing may not have much effect on perception. Thus its exclusion from measurements of VOT may be justified.

In Experiment I, medial syllables were used as initial

TABLE 3

% [d]- responses for 24 subjects, Experiment II

	initial [ra](t)	initial [ra](d)	MEAN
prevoicing	99.6	98.8	99.2
original closure Voicing	42.9	10.0	71.5
MEAN	71.3	99.4	

TABLE 4

Results of ANOVA, Experiment II

Source	df	MS	F	p
Initial Syllable (Ra(t) or Ra(d))	1	189.84363	64.30	.001
error	23	2.95244		
Voicing (Prevoicing or Original Closure)	1	184.26009	61.88	.001
error	23	2.97780		
Initial X Voicing	1	201.26004	67.59	.001
error	23	2.97780		

syllables, and in Experiment II an initial syllable was used medially. In general, the two positions seem to be similar, in that the same cues appear to be relevant and have similar effects. The strength of closure voicing and burst voicing both affect the perception of voiced and voiceless medial consonants.

4.3.4. Experiment III

The object of this experiment is to compare the contributions of strong voicing, weak voicing, and silence during closure to the perception of voicedness in medial stops. A closure interval can be voiced throughout, voiceless throughout, or combine voicing followed by silence in any proportion. For English, there appears to be a hierarchy of cues to medial stop voicelessness (Lisker, 1957; Port, 1977). If the stop is aspirated at all (i.e. VOT about 15 msec), then the percept is voiceless. If not, and if the closure is voiced or short, the percept is voiced. Otherwise the percept is voiceless. This hierarchy of cues is illustrated in Fig. 4-11. Thus a token of medial [b], with a short voiced closure, will still be heard as [b] with the voicing removed from that short closure. If the resulting silent closure is lengthened, however, the percept will be [p], even though the burst and following portion come from a [b]-syllable. Preliminary data of a similar type was gathered for medial [t] and [d] in Polish, using stimuli with various combinations of closure voicing and silence.

Stimuli consisted of (1) [ra] from <u>rata</u> or <u>rada</u>, plus, (2) some amount of closure from <u>rata</u> or <u>rada</u>, and/or (3) some amount of silence, plus (4) [ta] or [da] from <u>tata</u> or <u>data</u>. The original <u>rata</u> token had 160 msec weakly voiced closure (see Fig. 4-4a); the original <u>rada</u> token had 85 msec fully voiced closure (see Fig.

4-5a). The following durations of closure were used in constructing the experimental stimuli: from rata, 0, 60, 110, and 160 msec; from rada, 0, 60, and 85 msec. Several durations of silence were also recorded: 50, 60, 100, 110, 150, 160, 200, 210 msec. Finally, the initial syllables [ta] and [da] (without prevoicing) were isolated from tata and data and used as final syllables. This use of initial syllables in the medial position was intended to allow comparison with data from other experiments. Observation and Exp. I and II suggested that the positional variants are quite similar. However, this particular edited [d] from MG's data apparently contains no cues for voicedness, since with prevoicing removed it was identified as [t]. The medial [d] in MG's rada was identified as [d] when it was presented in initial position without closure voicing, probably because it has a completely voiced burst. Thus it must be borne in mind that the [da] used in constructing these stimuli was probably less likely to contribute to [d]-responses than other [da]'s (initial or medial) that could have been used.

A list of the stimuli constructed is given in Table 5.

The major results of this experiment were the following. First, few stimuli were categorized as containing [d], and those stimuli all had closure voicing from rada. Second, the weak closure voicing from rata has virtually the same effect as does a silent closure. That is, the voicing is either not perceived or is ignored by the listener. Third, all the parameters varied affected the percentage of responses, if not the gross categorization. All effects were larger when the closure interval contained voicing from rada, because of floor effects with silent and [t]-closure.

Table 6 summarizes how listeners responded to the stimuli. Only

TABLE 5

Construction of stimuli, Exp. III. Closures taken from *rata* have weak voicing; closures taken from *rada* have strong voicing. In any one stimulus, the first syllable and the closure *both* come from either *rata* or *rada*.

{ [ra] from *rata* / [ra] from *rada* }

{
closure interval with
total duration of
60, 110, 160, 210 msec,

containing

0, 60, 110 msec medial
closure from *rata*

OR

0, 60 msec medial
closure from *rada*

plus

silence
}

{ [ta] / [da] }

and

[ra] from *rada* + 85 msec medial closure + { ta / da }

TABLE 6

Categorizations of stimuli by 24 listeners, Exp. III.

Key: D, T = [ra] from <u>rada</u> or <u>rata</u>

1st number msec medial closure from token indicated by D or T

2nd number msec silence

da, ta second syllable

% categories	Total Closure Duration (msec) 60	110	160	210
80-100% [d] (strong)	D-60-0-ta D-60-0-da	D-60-50-da		
40-60% [d] (ambiguous)	D-0-60-da T-60-0-da	D-60-50-ta	D-60-100-da	D-60-150-da
60-80% [t] (weak)	T-0-60-da	D-0-110-da		
80-100% [t] (strong)	D-0-60-ta T-60-0-ta T-0-60-ta	D-0-110-ta T-0-110-ta T-0-110-da T-60-50-ta T-60-50-da T-110-0-ta T-110-0-da	D-60-100-ta D-0- 160-ta D-0- 160-da T-0- 160-ta T-0- 160-da T-60-100-ta T-60-100-da T-110-50-ta T-110-50-da	D-60-150-ta D- 0-210-ta D- 0-210-da T- 0-210-ta T- 0-210-da T-60-150-ta T-60-150-da T-110-100-ta T-110-100-da

five stimuli were heard unambiguously as [d], and five were ambiguous. The rest were categorized, either weakly or strongly, as [t]. Since the [d] used in medial position was not a good exemplar of a voiced stop, it is possible that with a "better" [d], the four ambiguous stimuli with final [da] would be heard unambiguously as [d].

At extreme stimulus values, perception is quite consistent. With 160 msec of [t]-voiced or silent closure, any stimulus is labeled as [t] 90% or more. With 60 or 85 msec of completely-voiced closure (the only values tested), a stimulus is labeled as [d] almost uniformly. The sources of the initial and final syllables have significant effects on the percept, but the differences produced are very small. All of the more interesting effects are obtained at shorter closure durations or when closure voicing is present in some of the closure interval. This pattern is consistent with that found in other chapters, in that some combinations are potentially ambiguous and allow minor cues some role in determining the percept.

For the stimuli without extreme values of closure duration and consistent voicing, the listeners' responses can be approximately accounted for by assigning cue-weightings as follows. Closure from <u>rata</u> is treated like silence. An initial [ra] from <u>rata</u> is equal to 50 msec of silent closure as a cue for [t]. A final [ta] is equal to 100 msec of silent closure. Each 50-msec of silent closure is weighted cumulatively, but <u>voiced</u> closure from <u>rada</u> is <u>subtracted</u> from the voiceless closure. With these weightings, all the stimuli with strong [t]-categorizations have a "[t]-score" of 3 or more. The stimulus with a strong [d]-

categorization has a "[t]-score" of 0. The stimuli with ambiguous and weak [t]-categorizations have "[t]-scores" of 1 or 2.

Consider the stimuli with none of the original closure, but instead with artificial silent closures of 60, 110, 160, and 210 msec. For the two longer closures, stimuli are all labeled as [t] well over 90%, but the sources of the initial and final syllables do have an effect on the exact percentage of responses. These effects are larger for the two shorter closures. Even though the closure contains no voicing, the stimulus with an initial syllable (to the point of closure) from rada, 60 msec silence, plus final [da] was perceived as rada 49%. The same 60 msec silent closure, but with [ra] from rata and final [ta], was perceived as rada only 3%. The effect of the initial syllable is particularly interesting, since no closure is included from the original token. For 60 msec silent closure, the initial syllable gave an 11% difference in [d]-responses, collapsed over final syllables. This difference is not due to differences in the durations of the vowels or syllables, which are almost identical. It may be due to the final irregular pitch periods in the [ra] from rata, or to fundamental frequency differences.

The effects of silence and closure voicing in the potentially ambiguous stimuli can be seen in Table 7. For stimuli with initial [ra] from rata, the voiced closure is essentially equivalent to silence. Lengthening closure from 60 to 110 msec reduces the % [d]-responses, but lengthening it even further to 160 msec does not further reduce the % [d]-responses. For stimuli with initial [ra] from rada, the 60 msec of closure is voiced, and this segment is crucial to the perception of medial [d]. Even when 100 msec of

TABLE 7

% [d]- responses for various combinations of original voiced closure and silence, collapsed over final syllable, Experiment III

Closure	Initial Syllable		
	ra(ta)	ra(da)	MEAN
60 msec closure	26	97	61
60 msec silence	21	32	26
MEAN	24	65	44
60 msec clos + 50 msec sil	7	62	34
110 msec sil	7	12	9
MEAN	7	37	22
60 msec clos + 100 msec sil	7	38	23
160 msec sil	2	6	4
MEAN	5	22	14

TABLE 8

% [d] - responses to stimuli made from initial [ra] from rada plus 60 msec voiced closure, Experiment III

	Final Syllable	
Closure Silence	[ta]	[da]
0 msec	98	96
50	41	82
100	19	58
150	8	45

silence is added to the 60 msec of closure voicing, the % [d]-responses collapsed over final syllable (38%) is greater than that for 60 msec silent closure (32%), although this difference is of course quite small. With 50 msec silence added to the 60 msec closure, a stimulus is still categorized as [d] 62%, collapsed over final syllable (vs. 97% [d]-responses with no silence).

The final syllable has its largest effect for just these stimuli, those made with voiced closure from rada and some silence inserted in the closure interval. This effect is shown in Table 8. Here the final syllable nearly produces true category shifts. In fact, these stimuli with [ra] from rada, 60 msec closure voicing, 50 or 100 msec silence, and [da] as the final syllable are the only stimuli with any silence during closure that are categorized as [d]. When the closure contains no voicing, or 150 msec silence, the categorization is [t]. Note, however, that when a closure of 60 msec is entirely voiced, it overrides a final syllable [ta] to produce an almost uniform categorization as [d].

By collapsing 50 and 60 msec silence, 100 and 110 msec silence, and 150 and 160 msec silence, it was possible to perform a four-way analysis of variance for some of the stimuli used in this experiment. Completely crossed factors of (original) closure and silence are found only for those stimuli with no closure or 60 msec closure, and with 50, 100, 150 msec inserted silence. The other two factors tested were initial syllable ([ra](t) or [ra](d)) and final syllable [ta] or [da]. The means for each factor are given in Table 9, and the results of the analysis are given in Table 10. Note that none of the % [d]-responses represents a categorization as [d]. Only two stimuli with silence during closure were categorized as

TABLE 9

Mean % [d] - responses for each of 4 factors in ANOVA, Experiment III

Initial	ra (ta)	7 %
	ra (da)	29
Closure	0 msec	13
	60 msec	23
Silence	50 msec	30
	100 msec	16
	150 msec	9
Final	ta (ta)	8
	da (ta)	28

TABLE 10

Results of 4-way repeated measures ANOVA on Initial Syllable (from rata or rada); Closure (0 or 60 msec of original); Silence (50,100 or 150 msec inserted); Final Syllable(initial syllable of tata or data), Experiment III

Source	df	MS	F	p
INITIAL	1	669.51392	53.02	.001
error	23	12.62792		
CLOSURE	1	145.00110	38.28	.001
error	23	3.78796		
SILENCE	2	239.41089	50.06	.001
error	46	4.78283		
FINAL	1	606.38867	60.71	.001
error	23	9.98844		
IN X CL	1	322.50024	52.19	.001
error	23	6.17926		
IN X SIL	2	48.31760	27.26	.001
error	46	1.77242		
IN X FIN	1	118.26494	44.09	.001
error	23	2.68228		
CL X SIL	2	4.07465	1.18	.315
error	46	3.44240		
CL X FIN	1	2.91840	2.89	.102
error	23	1.00898		
SIL X FIN	2	41.26540	20.98	.001
error	46	1.96671		
IN X CL X SIL	2	17.35582	11.78	.001
error	46	1.47365		
IN X CL X FIN	1	74.39012	38.68	.001
error	23	1.92323		
IN X SIL X FIN	2	1.00520	<1	-
error	46	2.75339		
CL X SIL X FIN	2	14.07455	5.31	.008
error	46	2.64892		
R X C X S X F	2	0.41145	<1	-
error	46	2.13428		

[d], and one was marginal enough to be considered ambiguous. Both had 60 msec voiced closure from rada and final [da]; they had inserted silences of 50 and 100 msec. (Of course, similar stimuli with no inserted silence were also categorized as [d], but they could not be included in this analysis of variance.)

The results of the analysis indicate that all four main effects and most of the interactions were significant. Since all the stimuli included in this analysis contained some closure silence, there was a preponderance of [t]-responses. Most of the interactions obtained here seem to involve floor effects with the % [t]-responses. That is, listeners heard so many [t]'s that the effects for initial and final syllables, as well as lengthened silent closure, were weaker simply because one cannot hear more than 100% of the tokens as [t]. Larger differences were found at lower silence durations, with initial [ra] from rada or final [da], and with the 60 msec of original closure voicing from rada. Some of the significant interactions are illustrated in the graphs in Fig. 4-12.

To summarize the results of Exp. III: the most important factor in determining the voicing feature of a Polish medial stop seems to be the presence of at least some strong voicing during the medial closure interval. A moderate amount of silence may follow the closure voicing, and the percept will still be voiced. With moderate amounts of silence, the influence of the initial and final syllables increases, but with extreme amounts of silence the percept is unambiguously voiceless regardless of these other parameters. The relevant parameter in the initial syllable may be the nature of voicing offset; in the final syllable, it may be burst voicing. They assert themselves most at the intermediate, and ambiguous, values of the closure cues. There is an interesting trading relation operating among these cues. The interacting cues are initial syllable (voicing offset), closure voicing, closure silence, overall closure duration, and burst voicing in the medial stop. Of these, the closure cues interact among themselves at a more primary level, and then the combination of these interacts with the more minor cues.

4.3.5. Discussion of Perception Data

Three preliminary experiments on the perception of medial apical stops in Polish were carried out. In Exp. I, final syllables with medial stops were put in initial position, while in Exp. II, an initial [d] was used in medial position both with and without prevoicing. In Exp. III initial [d] and [t] were used in medial position with various combinations of closure voicing and silence. A major result of Exp. II and III was that the low-amplitude ("weak") closure voicing found before medial [t] has virtually the same perceptual effect as absolute silence, that is, it does not cause any increase in [d]-responses. Exp. II also showed that prevoicing and closure voicing are equally effective in producing [d]-responses.

Another result seems to be that there is a "continuum" of burst voicing which influences perception of stop voicing. A voiceless [t]-burst with some aspiration is a cue for stop voicelessness, while a voiced-throughout [d]-burst is a cue for stop voicedness. Partially-voiced [d]-bursts, that is, bursts with voiceless beginnings but some internal noisy voicing, which are common, may be less effective cues for stop voicedness. However, unlike English, burst voicelessness and aspiration do not seem to be overriding cues to stop voicelessness. At most, they produce a 40% response shift when combined with some closure silence.

A third result is that few of the stimulus combinations constructed resulted in [d]-categorizations. Only a short voiced closure interval was an overriding cue to stop voicelessness. The effect of a long voiced closure was not investigated, but a long (160 msec) closure interval filled with weak voicing and/or silence is an overriding cue to stop voicelessness. A short silent closure, or a medium-length closure combining voicing and silence, is a cue to stop voicedness that interacts with other cues in determining the percept. The weightings of all the cues involved, including those found in the initial syllable before closure and in the final syllable including the medial stop burst, remain to be determined.

4.4. General Discussion

In this chapter, acoustic analysis and perceptual experiments were used to identify some correlates of and cues to the Polish medial [t] - [d] contrast. The contrast seems to be acoustically quite similar to the initial contrast in running speech, and the perceptual experiments in fact used initial tokens in the medial position. The most relevant parameters appear to be closure duration and voicing, which interact. The closure voicing that is sometimes seen with medial [t] has no perceptual effect, and presumably can be discounted. The pattern identified for [d] consists of closure voicing which continues through the burst, or which dies down somewhat before the voiceless or partly-voiced burst, followed quickly by am amplitude rise for the formants. Perceptually, voiced closure is the best cue for [d]. If the closure is all or part voiceless, then the burst must be voiced to produce a voiced percept.

Preliminary results on the critical closure durations for perception agree with the production data. Acoustic analysis showed category overlap for closures between 90 and 140 msec long. Perceptual experiments showed that shorter and longer durations were indeed unambiguous (although the effect of long entirely-voiced closures was not tested). At the intermediate duration tested (110 msec), the presence of at least some closure voicing, plus a fully-voiced burst, was critical for a [d]-percept. Inspection of the production data indicates first, that the few voiceless [d]-closures were all quite short, and second, that the [d]-closure intervals that were longer than 90 msec all contained about 90 to 115 msec of voicing. However, most tokens with some silent closure also have voiceless medial bursts, which is surprising given the perceptual finding that voiced bursts are very influential with ambiguous closure combinations.

The production and perception results are enough in agreement to outline the general medial voicing contrast. However, more work is needed to fill in the details of this contrast. Still unclear is the proportion of closure voicing and silence that is produced, and that is tolerated perceptually. Also not clear is how burst voicing is perceived -- that is, how voiced a burst has to be to have cue value, and how it interacts with closure voicing.

FOOTNOTES

[1] Recall that there is no final voicing contrast in Polish.

[2] In a few instances the [d]-burst is voiceless, and a VOT measure equal to the burst duration can be made. Because this is a relatively uncommon pattern, such measurements were not made.

[3] English has a rule of medial flapping, so these data are for labials and velars.

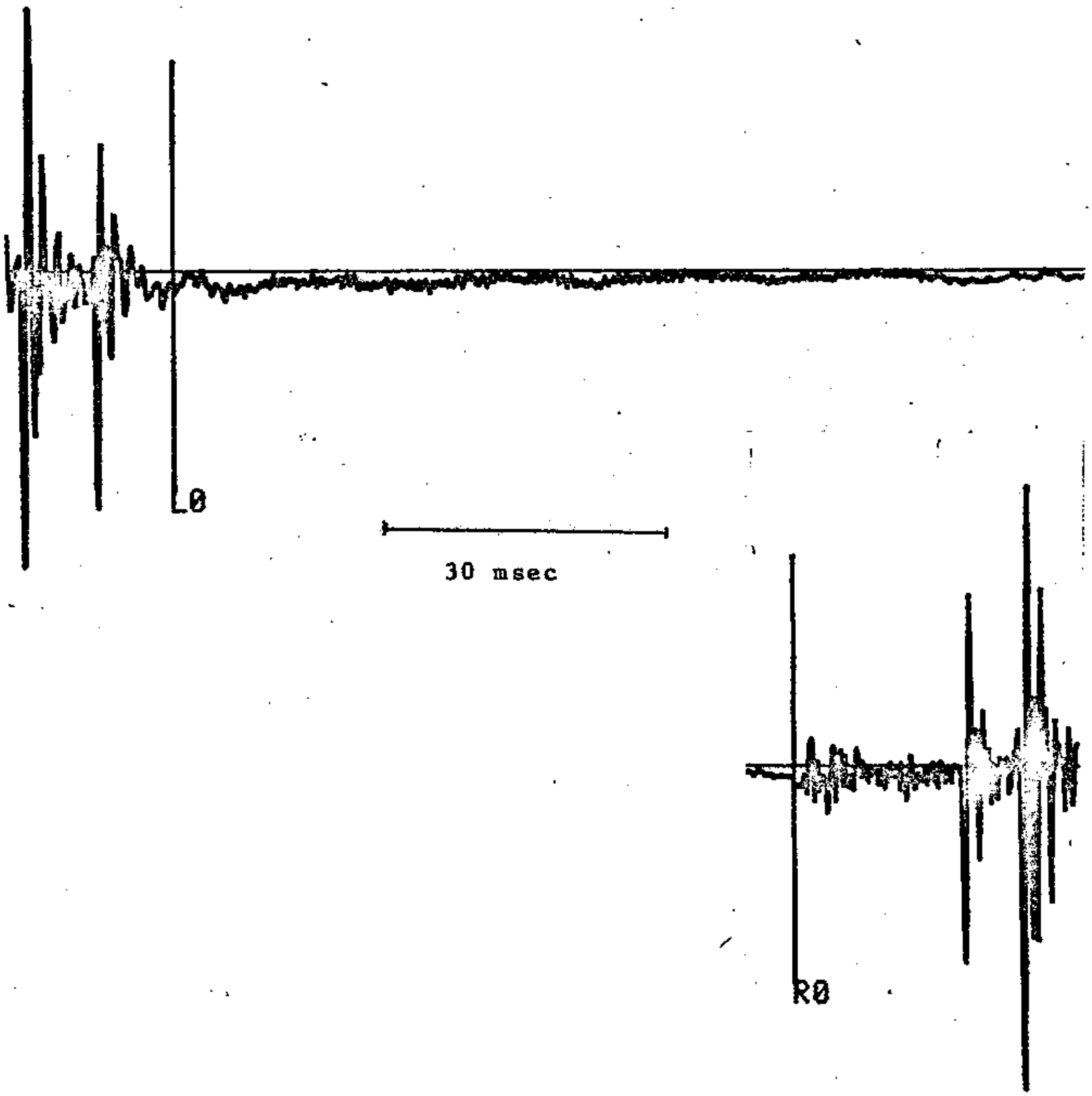


Fig. 4-1a -- Waveform showing closure and release of medial [t]. The left cursor is set at closure onset, and the right cursor is set at the burst. The display is continuous across the two lines. The closure duration as measured between the cursors is 105 msec. Note that the time scale has been compressed in this display. The token is <u>rata</u> read by Wrocław speaker #3.

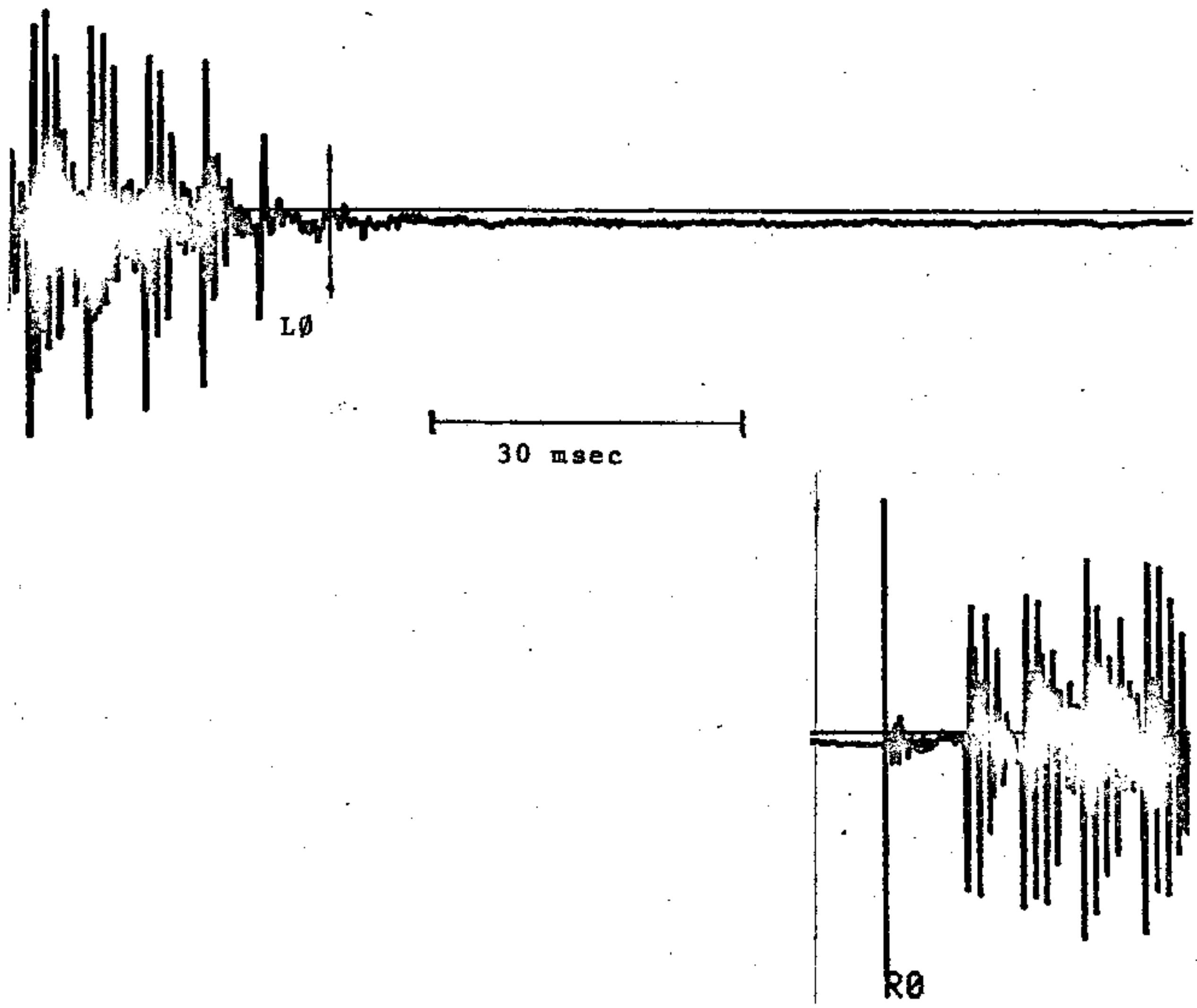


Fig. 4-1b -- Waveform showimg closure and release of medial [t]. The left cursor is set at closure onset, and the right curser is set at the burst. The display is continuous across the two lines. The closure duration is 122 msec. Note that the time scale has been compressed in this display. The token is <u>rata</u> read by Wrocław speaker #5.

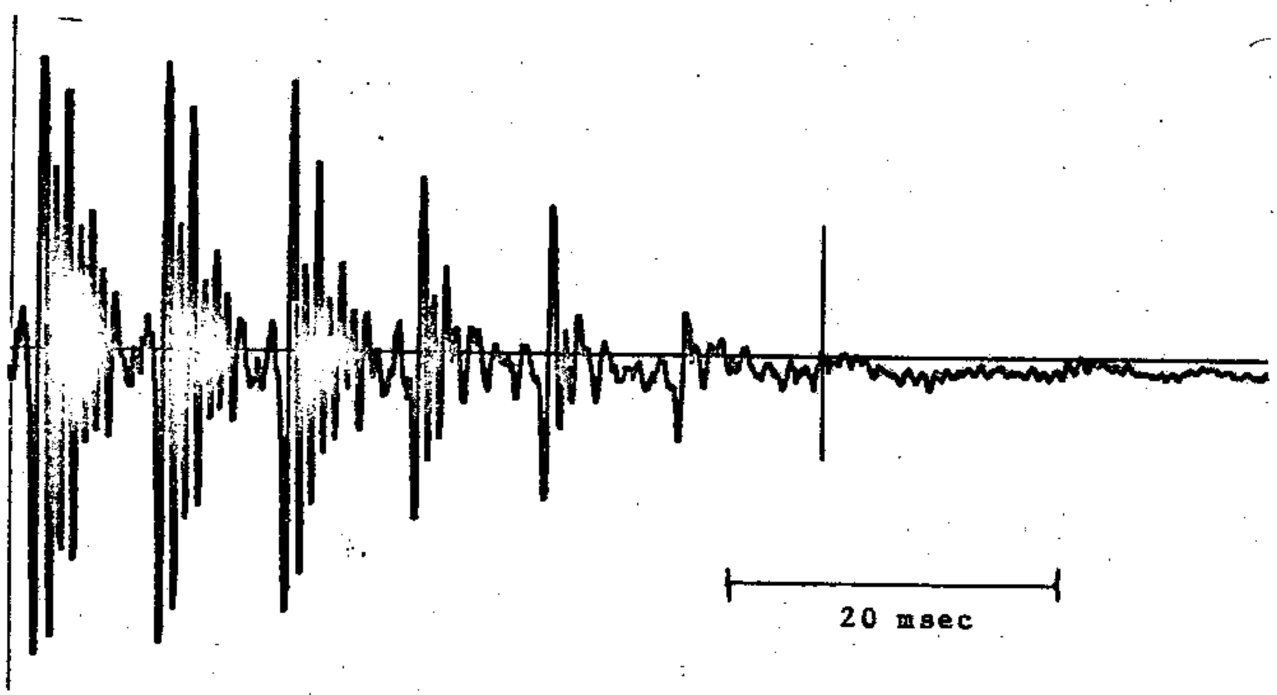


Fig. 4-1c -- Waveform showing detail of closure onset for medial [t]. The token is <u>rata</u> read by Wrocław speaker #4.

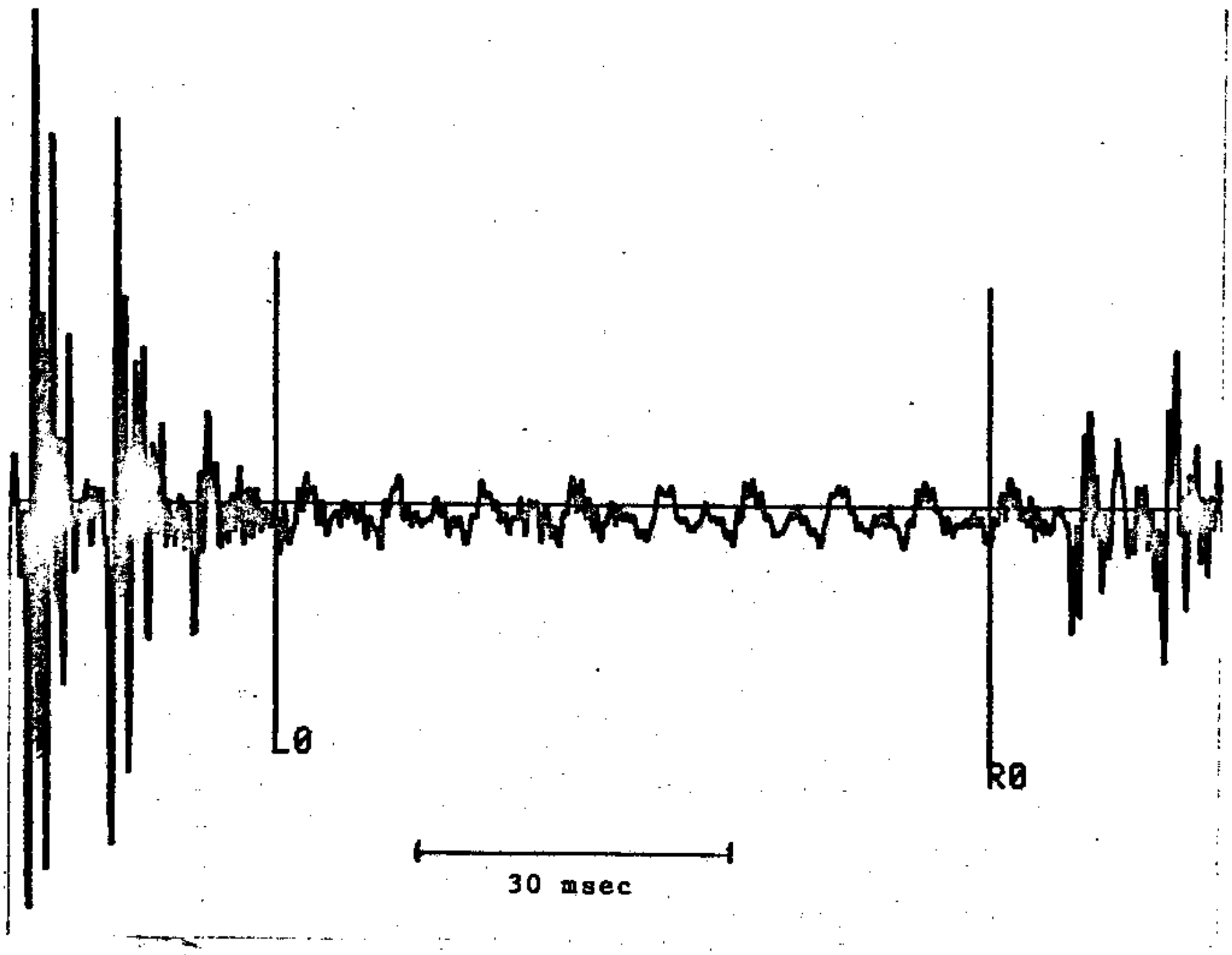


Fig. 4-2a -- Waveform showing closure and release for medial [d]. The left cursor is set at closure onset, and the right curr r is set at the burst. The closure duration is 70 msec. Note that the closure is voiced throughout, and that the time scale has been compressed in this display. The token is <u>rada</u> read by Wrocław speaker #7.

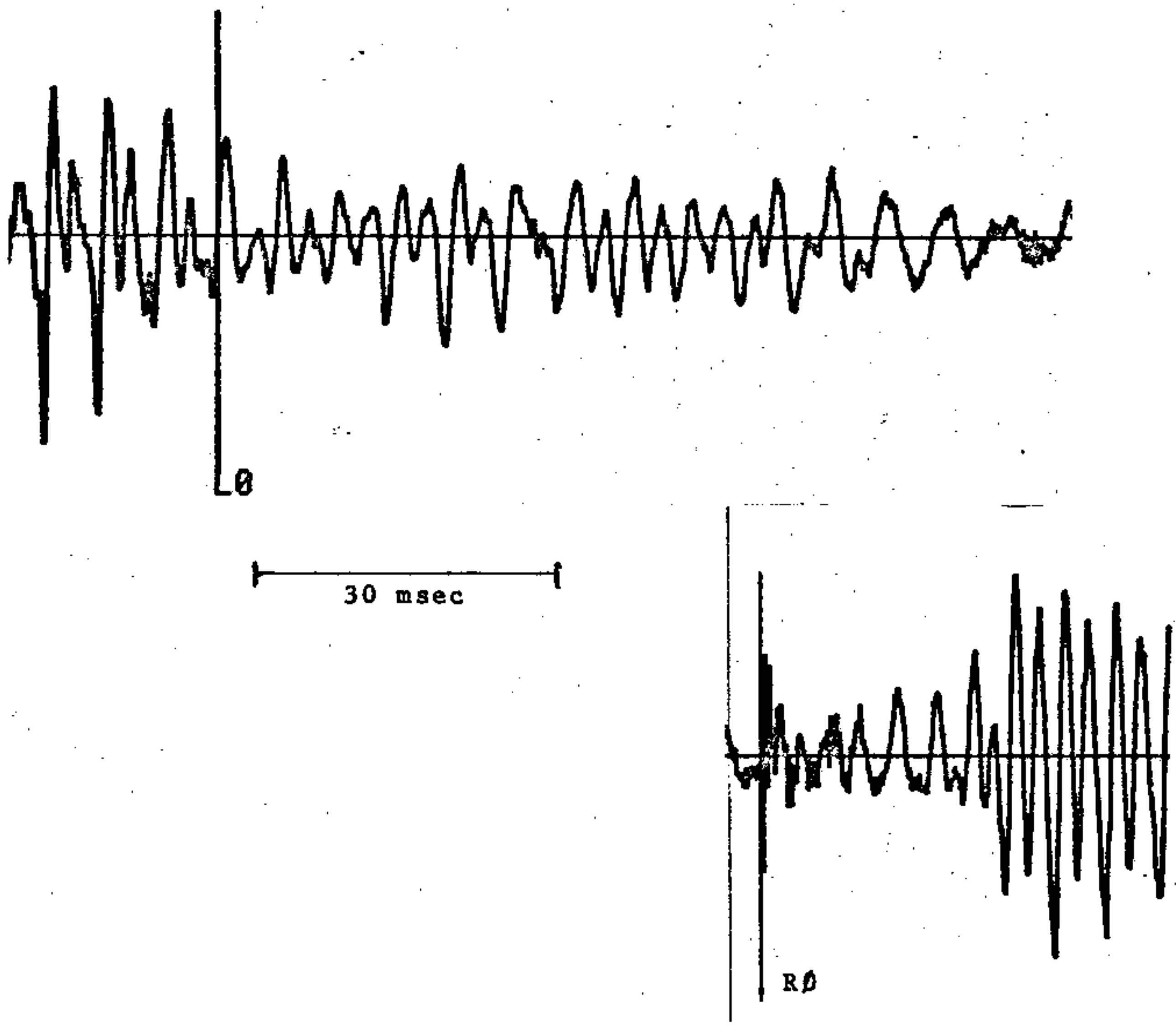


Fig. 4-2b -- Waveform showing closure and release of medial [d]. The left cursor is set at closure onset, and the right cursor is set at the burst. The closure duration is 82 msec. Note that the closure is voiced throughtout, and that the time scale has been compressed in this display. The token is *rody* read by MG. The display is continuous across the two lines.

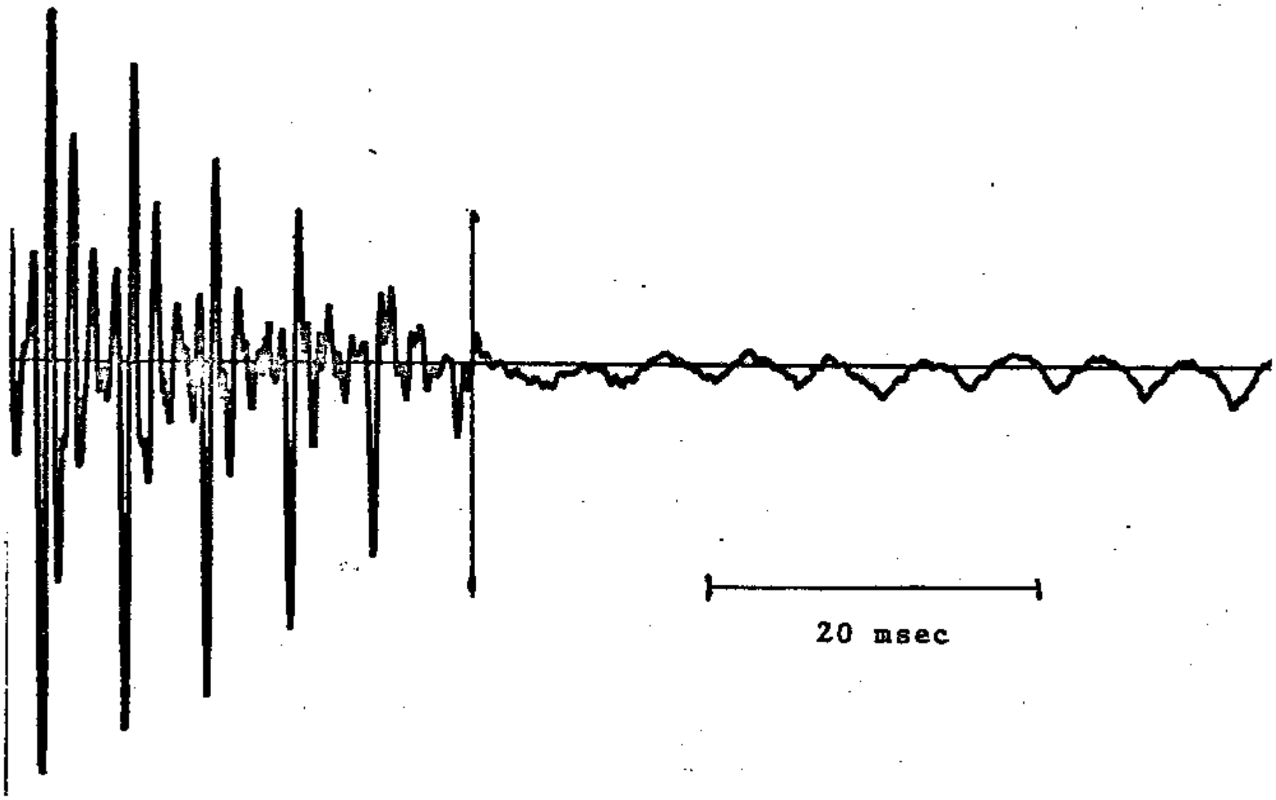


Fig. 4-2c -- Waveform showing detail of closure onset and voicing. The token is rada read by Wrocław speaker #1.

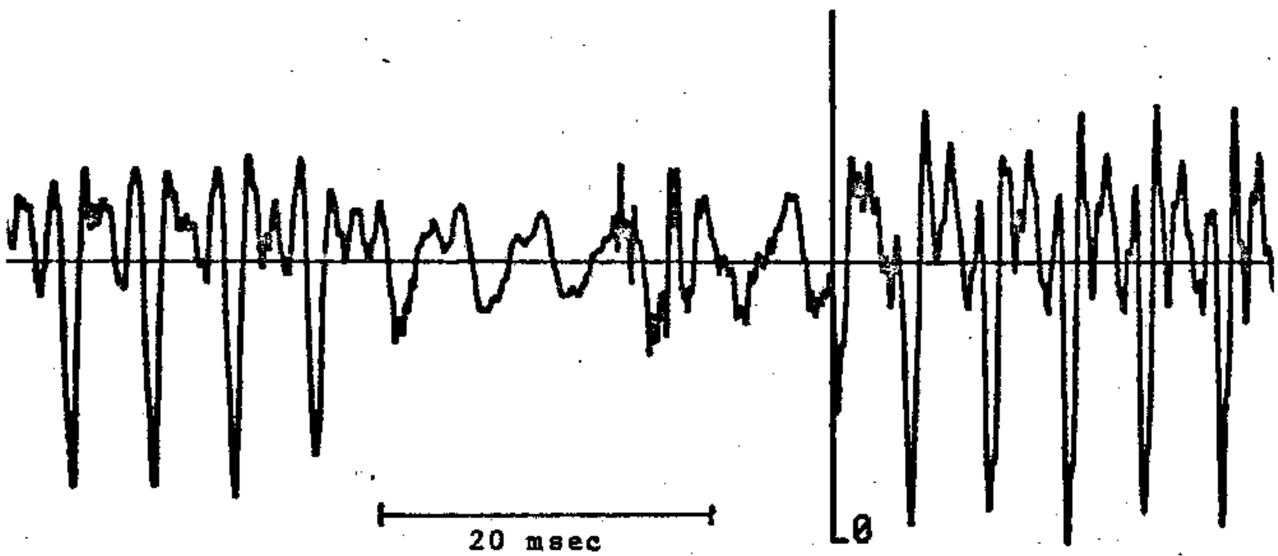


Fig. 4-3 -- Waveform showing point from which vowel duration was measured following [r]. The display shows the transition from one tap of the trill to the next, which then continues directly into the following vowel.

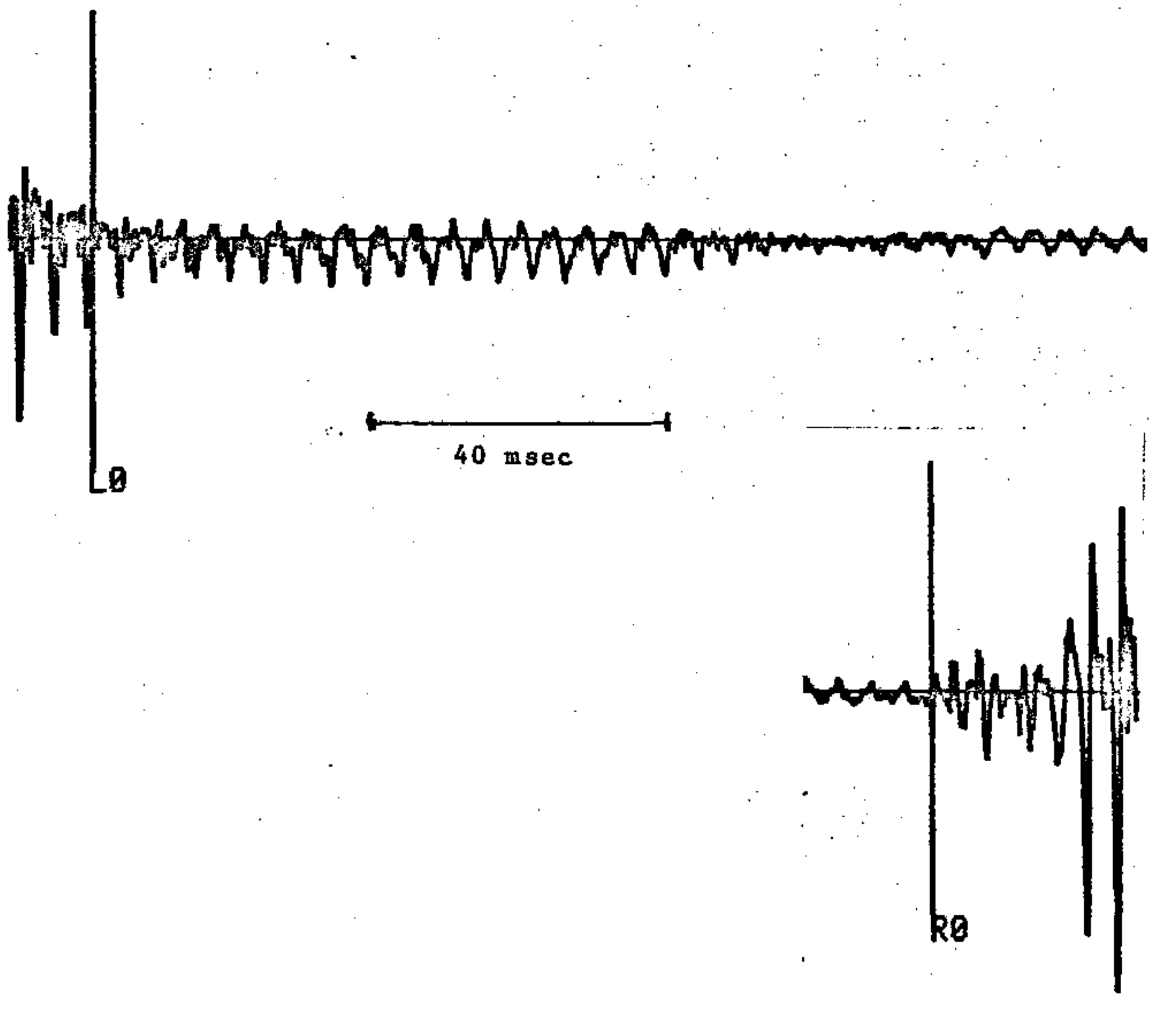


Fig. 4-4 -- Waveform showing closure and release for medial [t]. In this token the closure is voiced. The closure duration is 164 msec. Note that the time scale in this display has been greatly compressed. The token is <u>rata</u> read by speaker MG.

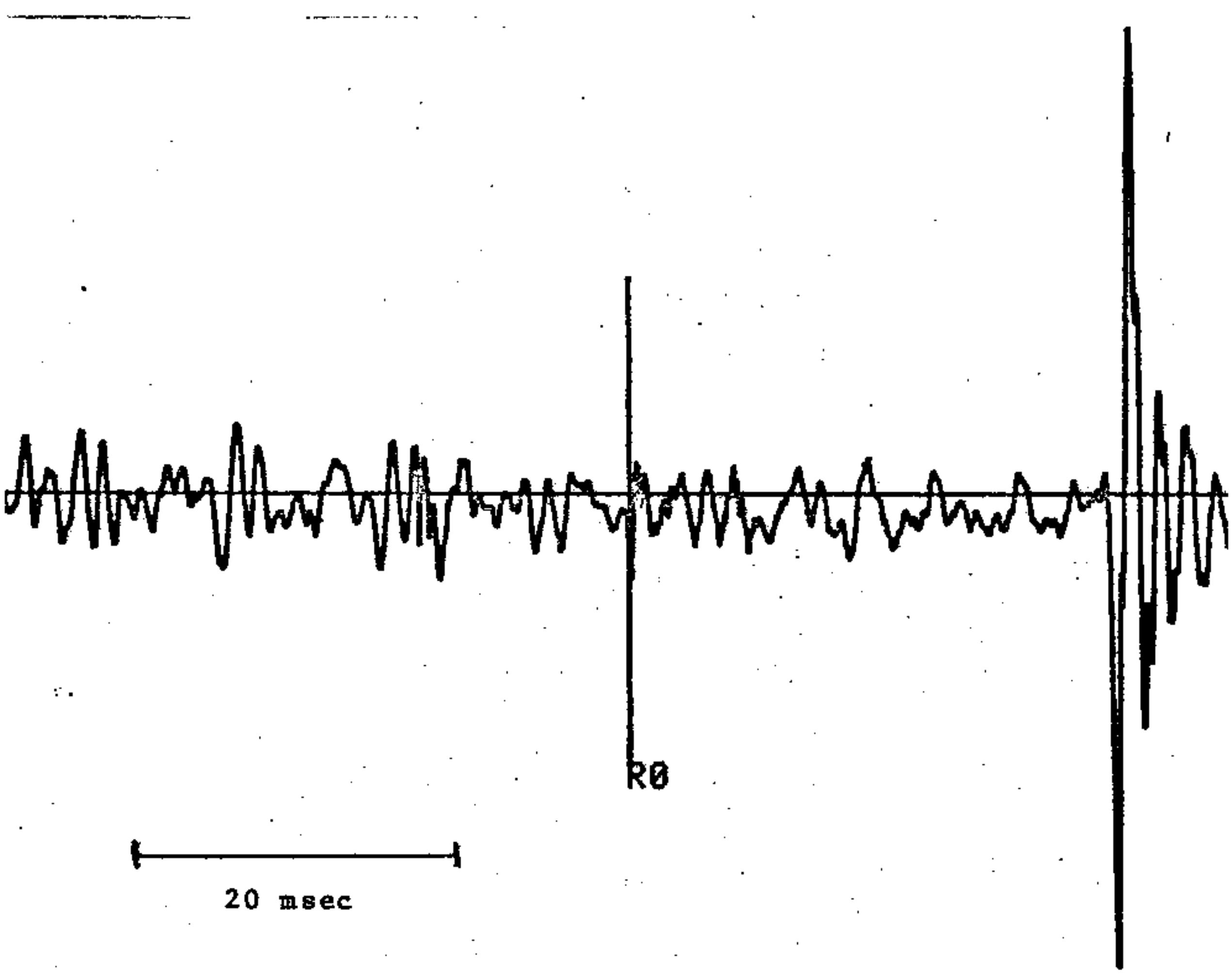


Fig. 4-4b -- Waveform showing detail of medial [t]-burst and comparatively long lag VOT. The cursor is set at the burst. The hard-copy shown here does not indicate the burst frication as well as does the original display. The token is rata read by speaker JP.

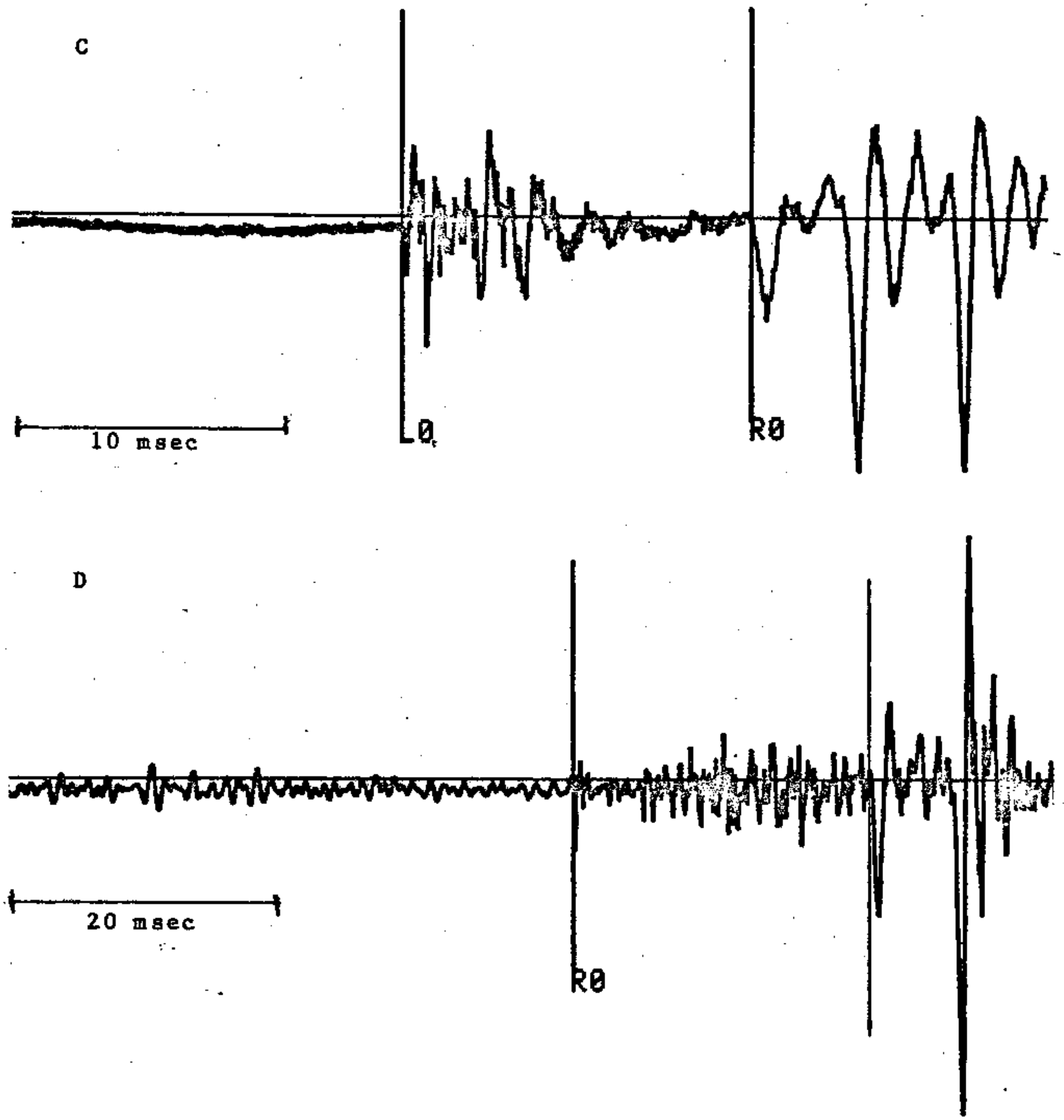


Fig. 4-4c -- Waveform showing detail of voice-onset and burst in medial [t]. The VOT is 13.2 msec. The token is <u>rata</u> read by Wrocław speaker #1.

Fig. 4-4d -- Waveform showing burst and voice-onset of medial [t]. The VOT is 23 msec. The token is <u>rata</u> read by Wrocław speaker #8.

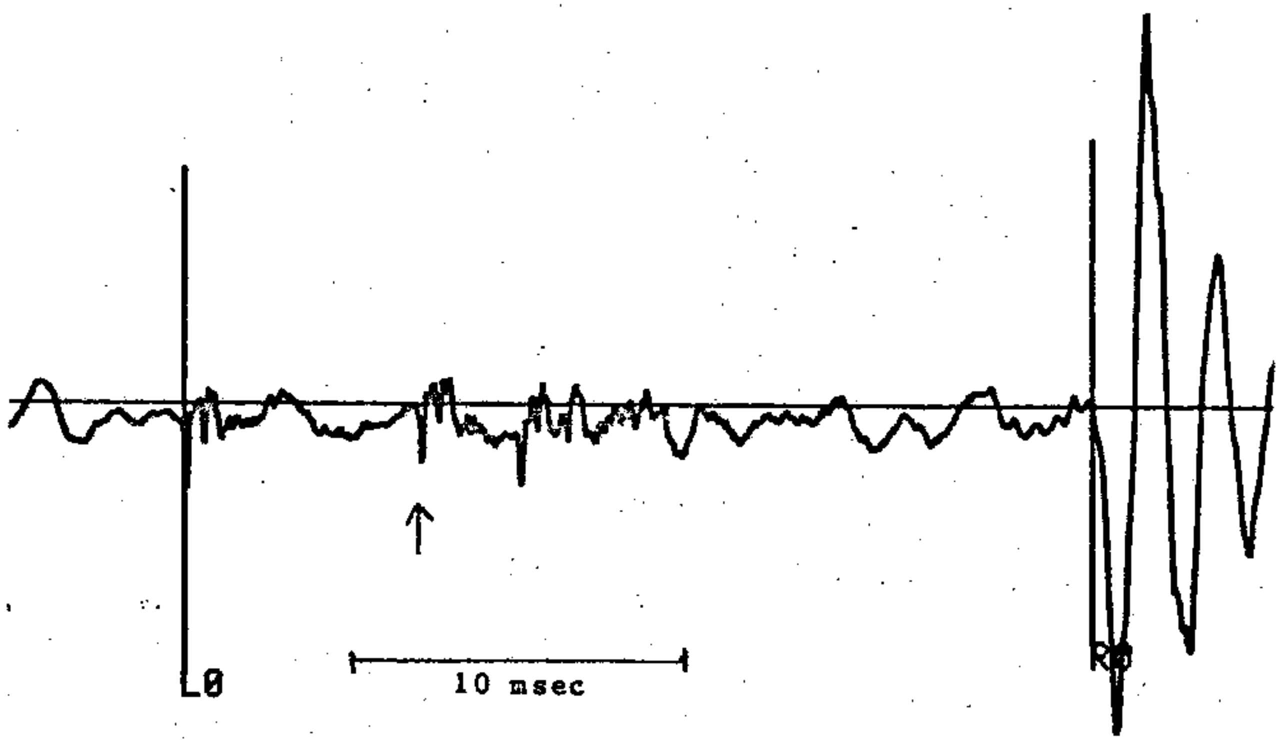
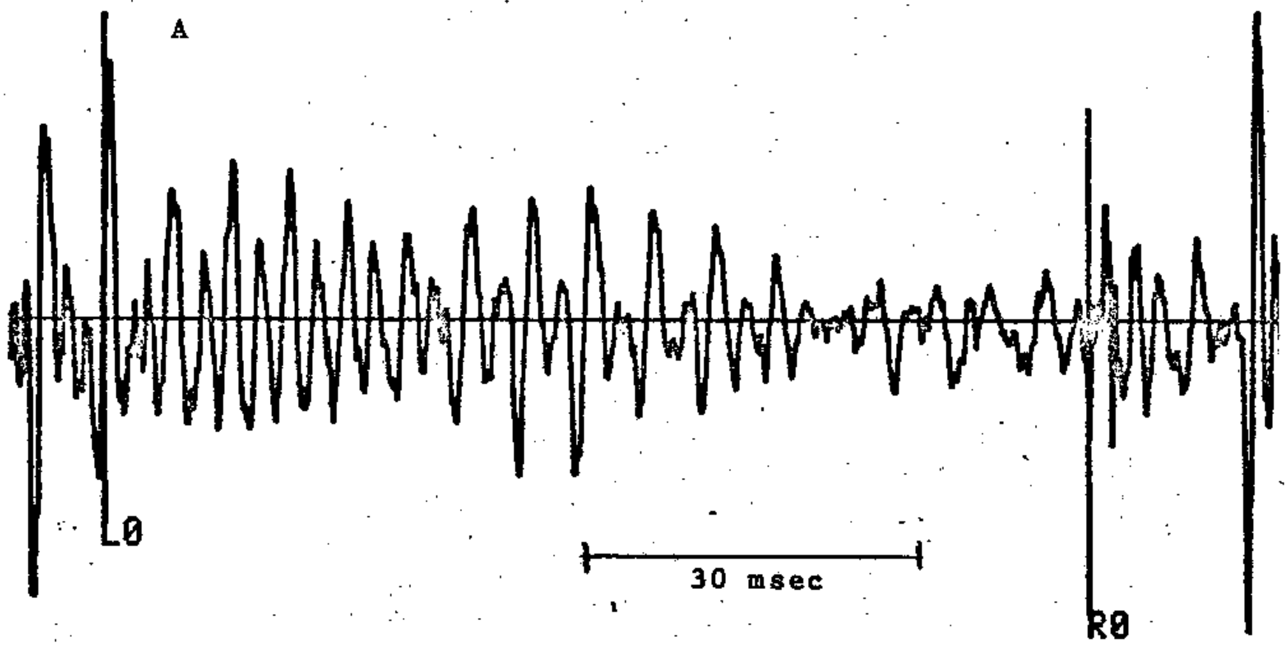


Fig. 4-4e -- Waveform showing burst and voice-onset of medial [t] in sentence context. This token has a double burst, indicated by the cursor and the arrow. The VOT is 28 msec. Some of the preceding voiced closure can also be seen. The token is gazeta read by speaker JP.



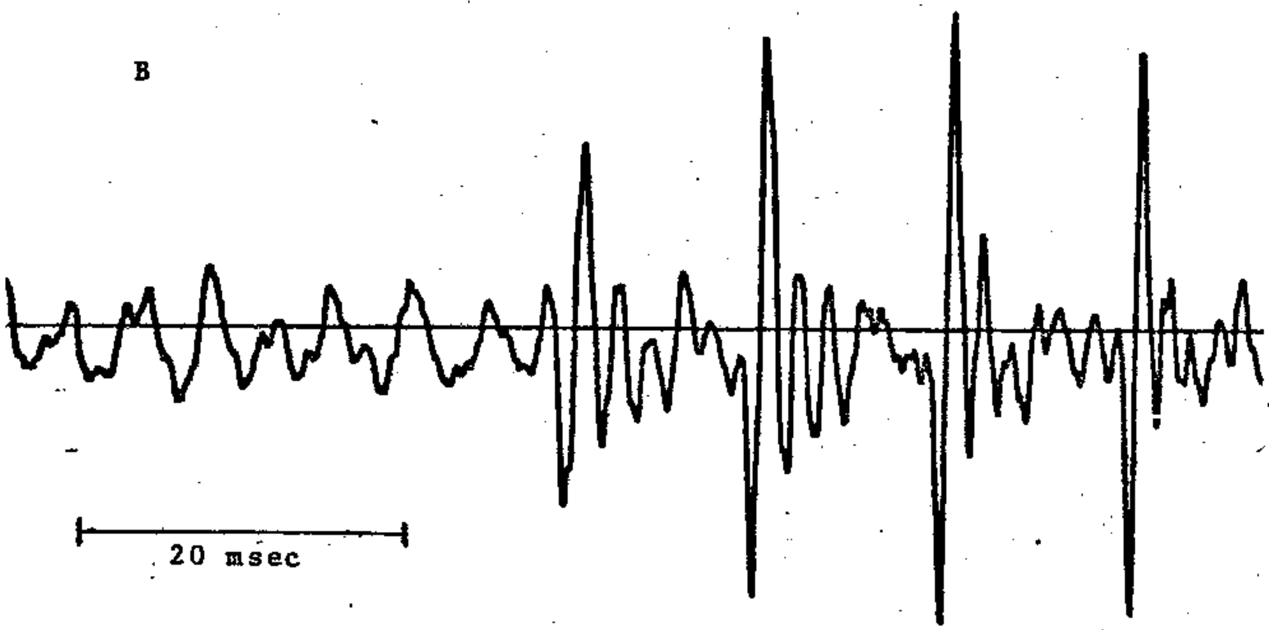
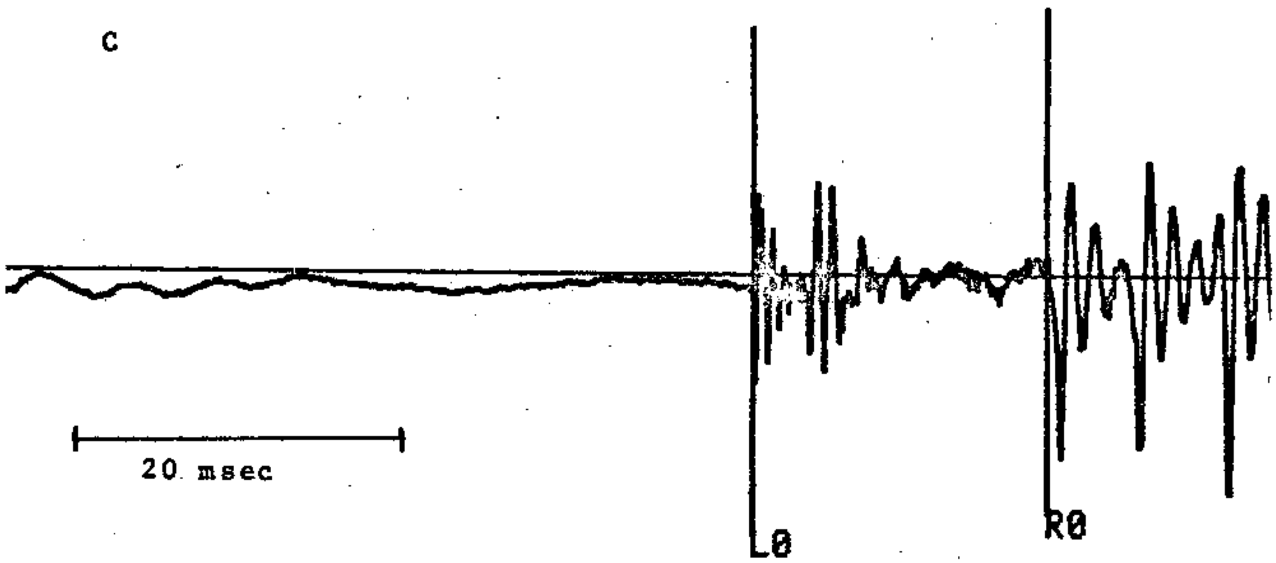


Fig. 4-5a -- Waveform showing voiced closure and burst in medial [d]. The closure voicing has relatively high amplitude. Note that the time scale has been compressed in this display. The closure duration is 92 msec. The token is <u>rada</u> read by speaker MG.

Fig. 4-5b -- Waveform showing end of closure and begining of amplitude increase for following vowel with no visible burst. The token is <u>rada</u> read by speaker JP.



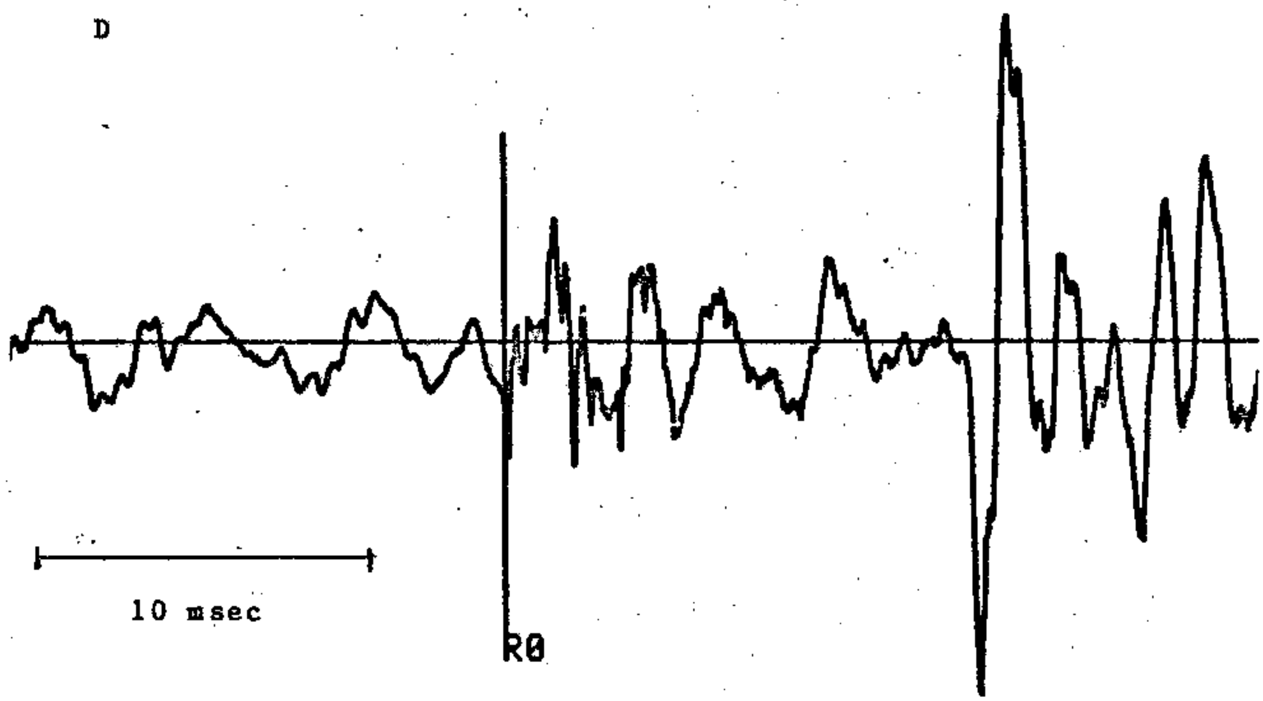
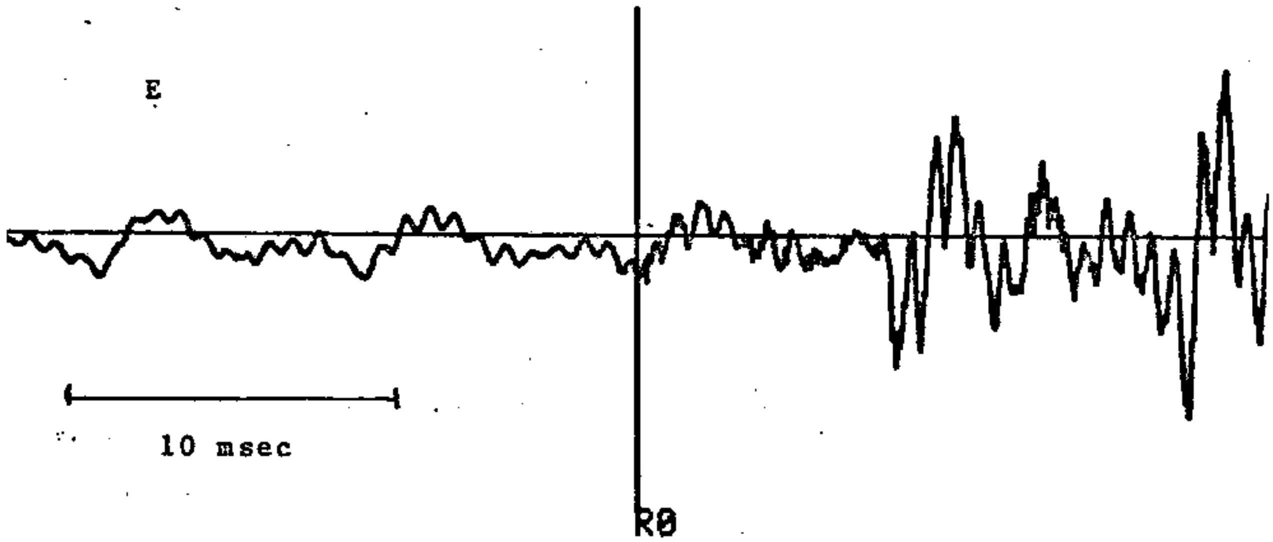


Fig. 4-5c -- Waveform showing closure silence followed by voiceless burst in medial [d]. The "second" VOT that could be measured is shown between the cursors; it is 18.2 msec. The token is <u>rada</u> read by Wrocław speaker #1.

Fig. 4-5d -- Waveform showing voiced burst of medial [d]. The token is <u>rada</u> read by speaker MG.



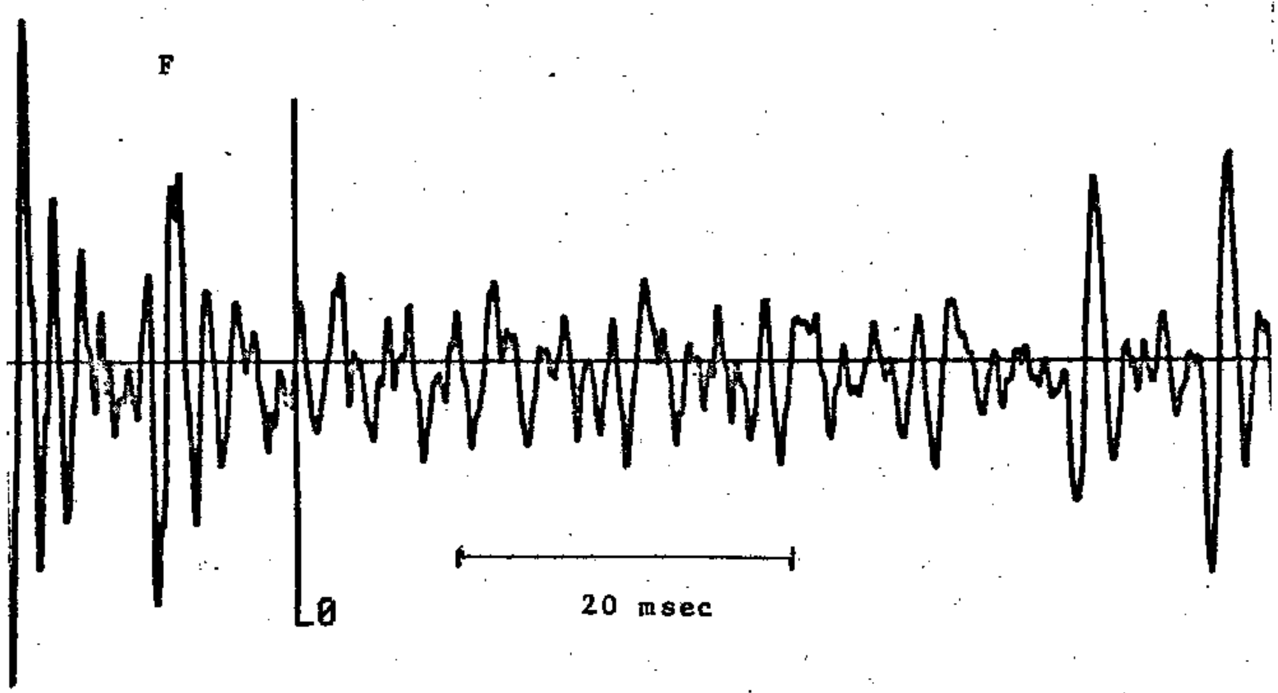


Fig. 4-5e -- Waveform showing fully-voiced burst of medial [d]. The burst frication is obscured on the hard-copy display. The token is <u>rada</u> read by Wrocław speaker #7.

Fig. 4-5f -- Waveform showing closure and following vowel of medial [d] in sentence context. The cursor is set at closure onset. No burst is visible. The token is <u>jade</u> read by speaker JP.

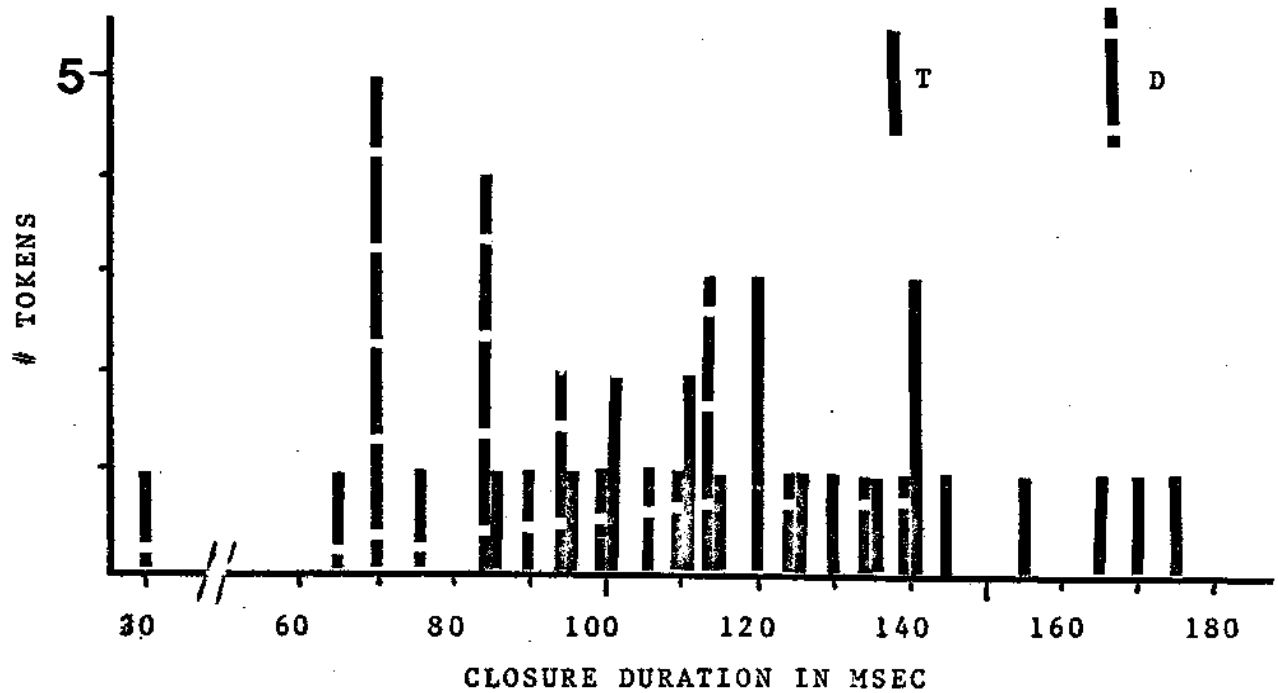


Fig. 4-6 -- Distribution of [t]- and [d]-closures for 24 pairs of rata and rada read by speakers in Wrocław.

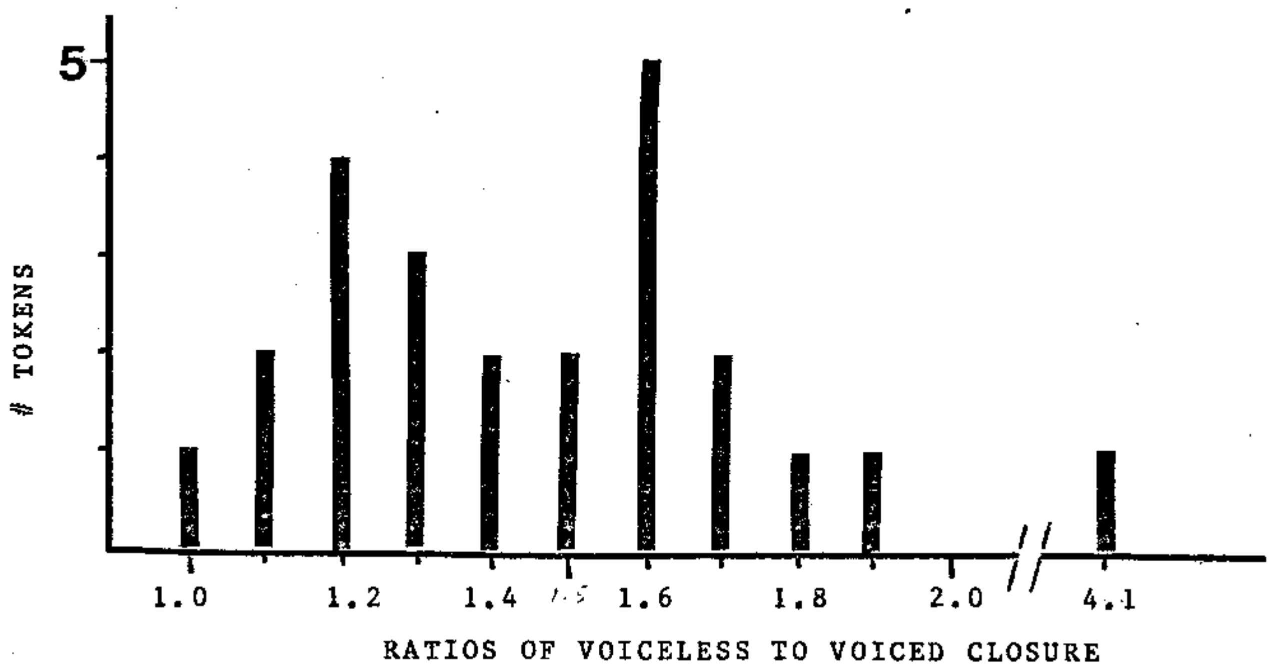


Fig. 4-7 -- Distribution of closure duration ratios for 24 pairs of rata and rada read by speakers in Wrocław.

Fig. 4-8a

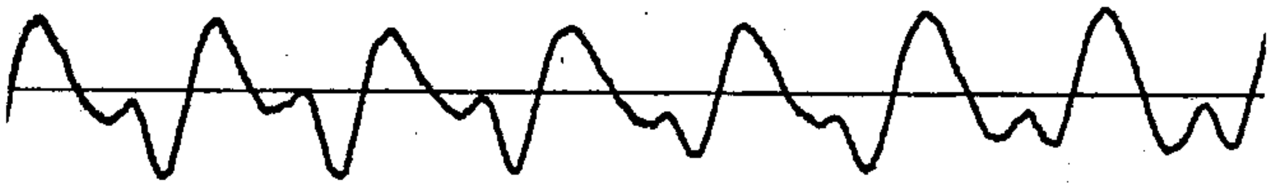


Fig. 4-8b

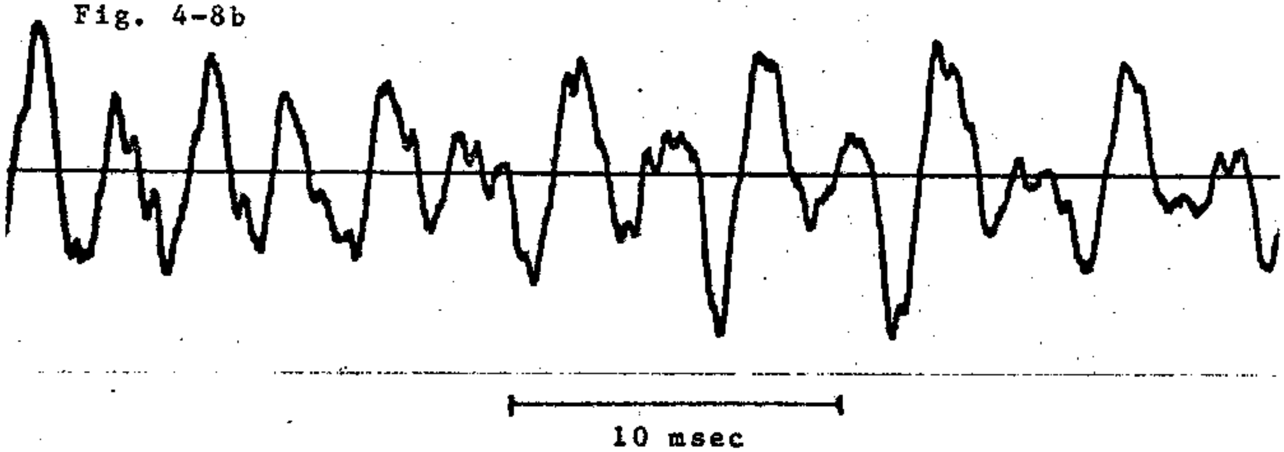


Fig. 4-9

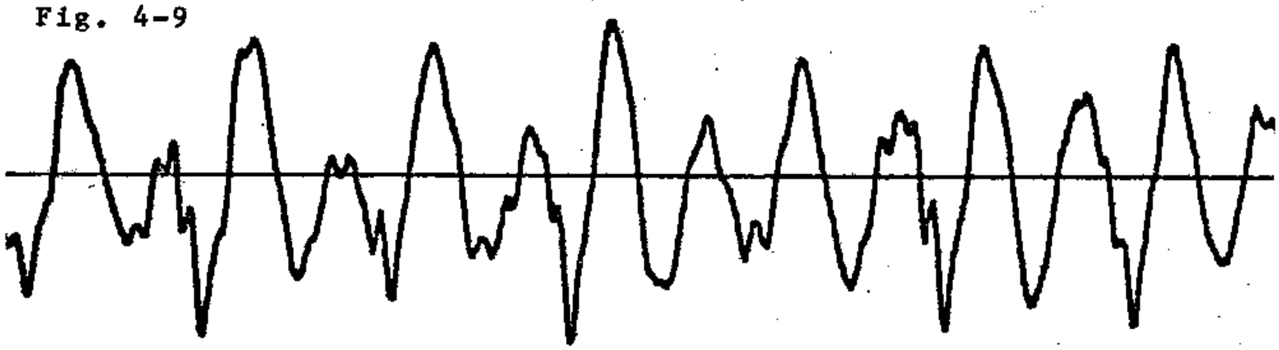


Fig. 4-8a -- Waveform showing prevoicing from initial [d] in data read by speaker MG.

Fig. 4-8b -- Waveform showing voiced closure from medial [d] in <u>rada</u> read by speaker MG.

Fig. 4-9 -- Waveform showing nasal consonant [m] from token of <u>dama</u> read by speaker MG.

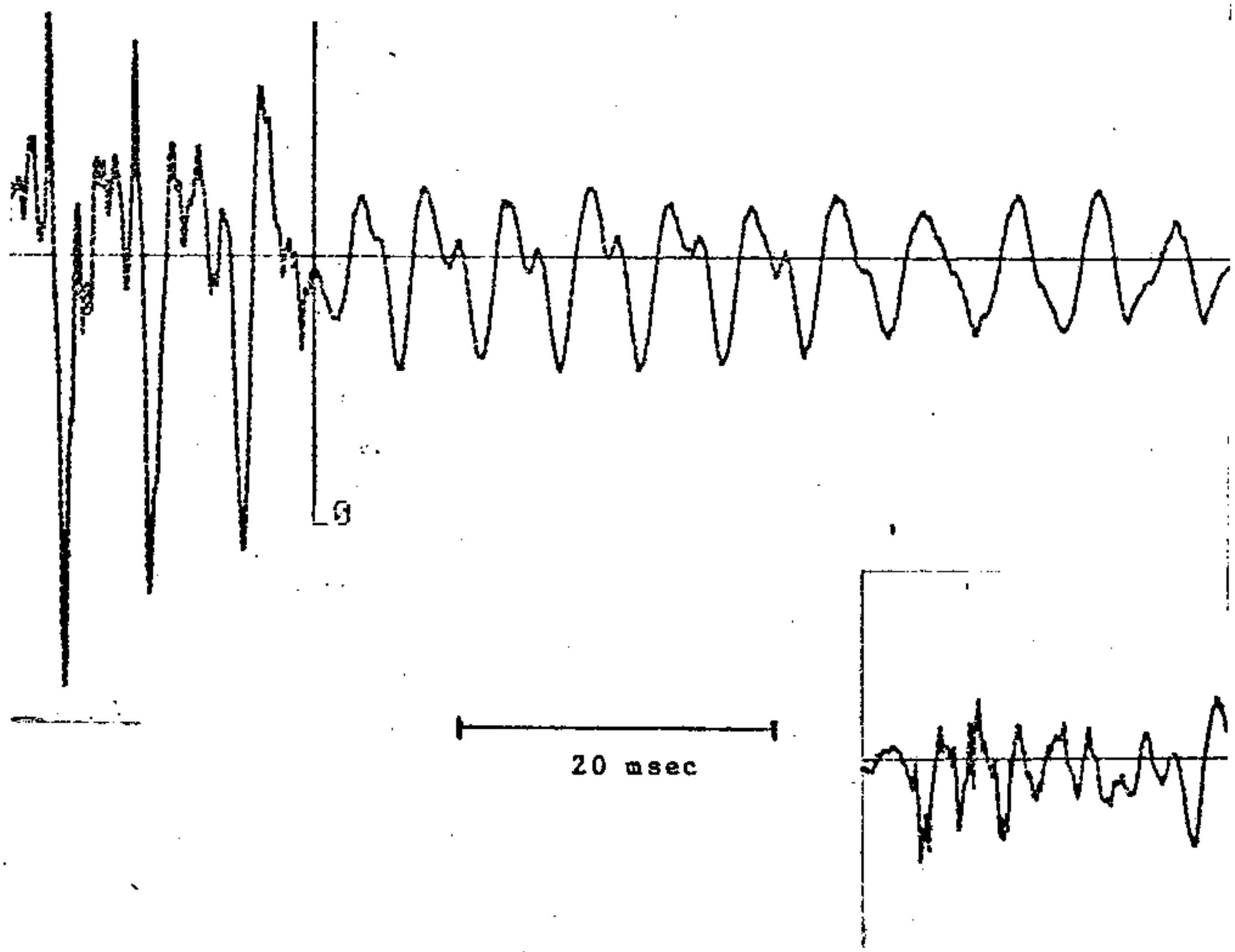


Fig. 4-10a -- Waveform showing natural-edited stimulus with first syllable from rada plus 60 msec prevoicing plus [da] from data as the second syllable. The cursor is set at closure onset and the display is continuous across the two lines. The speaker of the original tokens is JP.

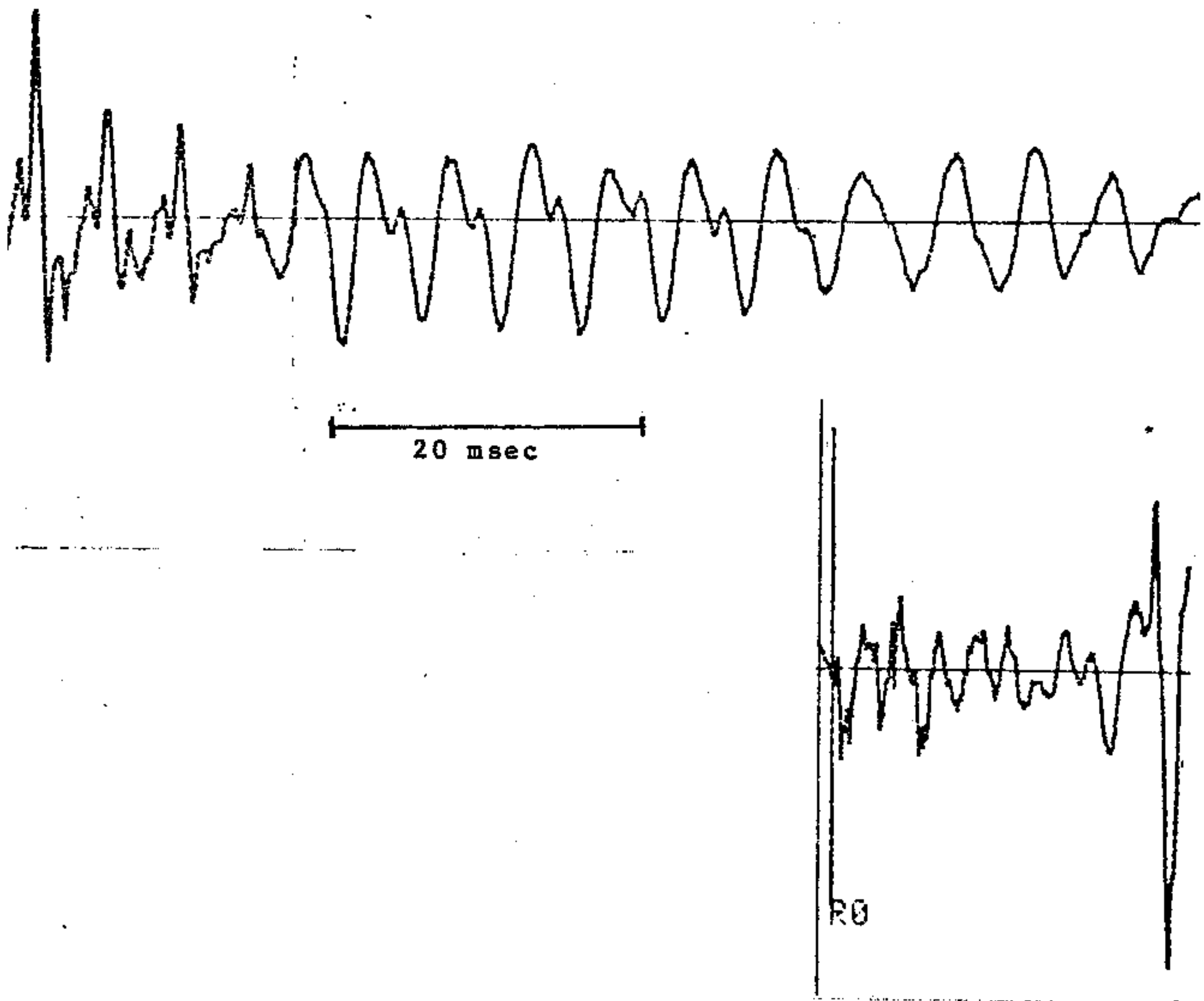


Fig. 4-10b -- Waveform showing natural-edited stimlus with first syllable from rata plus 60 msec prevoicing plus [da] from data as the second syllable. The cursor is set at the burst and the display is continuous across the two lines. The speaker of the original token is JP.

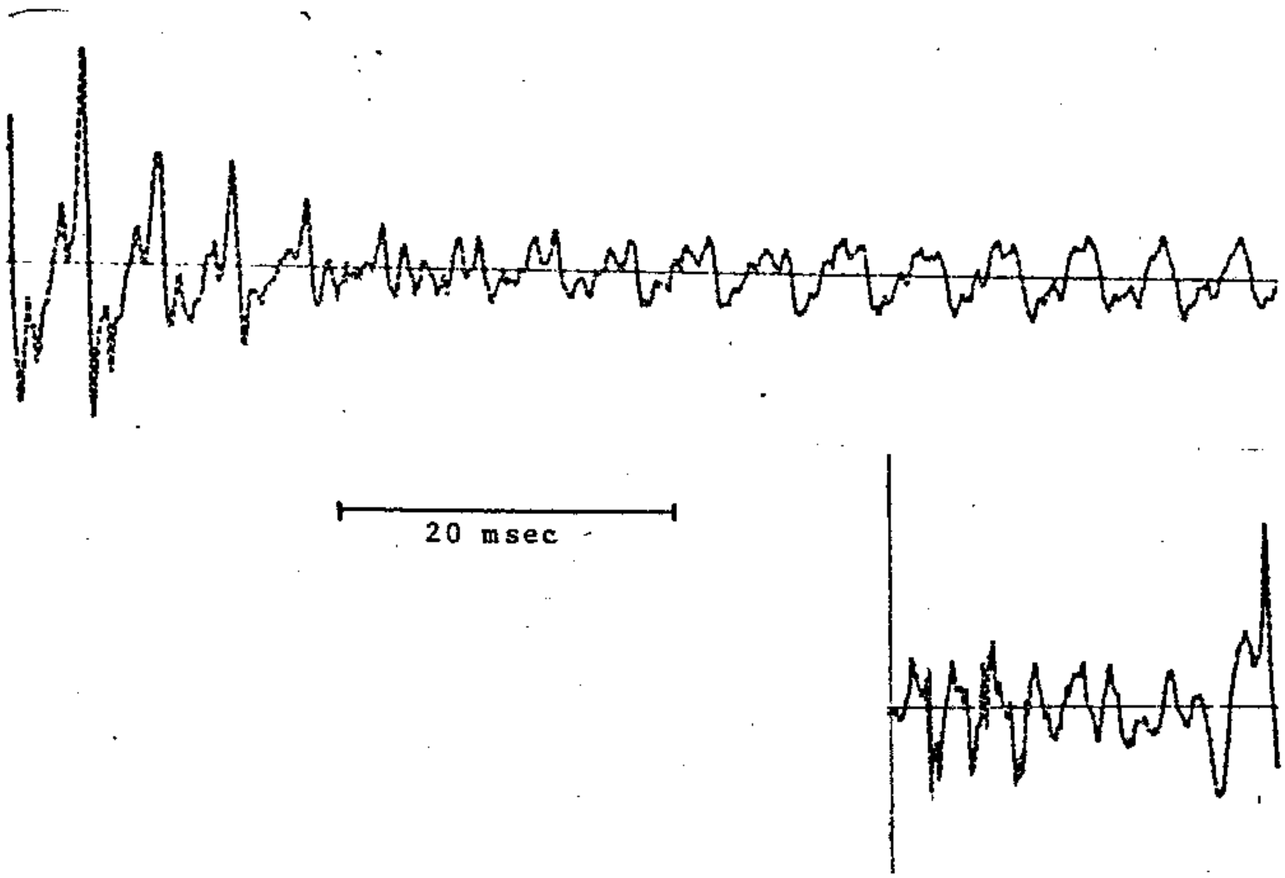


Fig. 4-10c -- Waveform showing natural-edited stimulus with first syllable from rata plus 60 msec (voiced) closure from rata plus [da] from data as the second syllable. The display is continuous across the two lines. The speaker of the original tokens is JP.

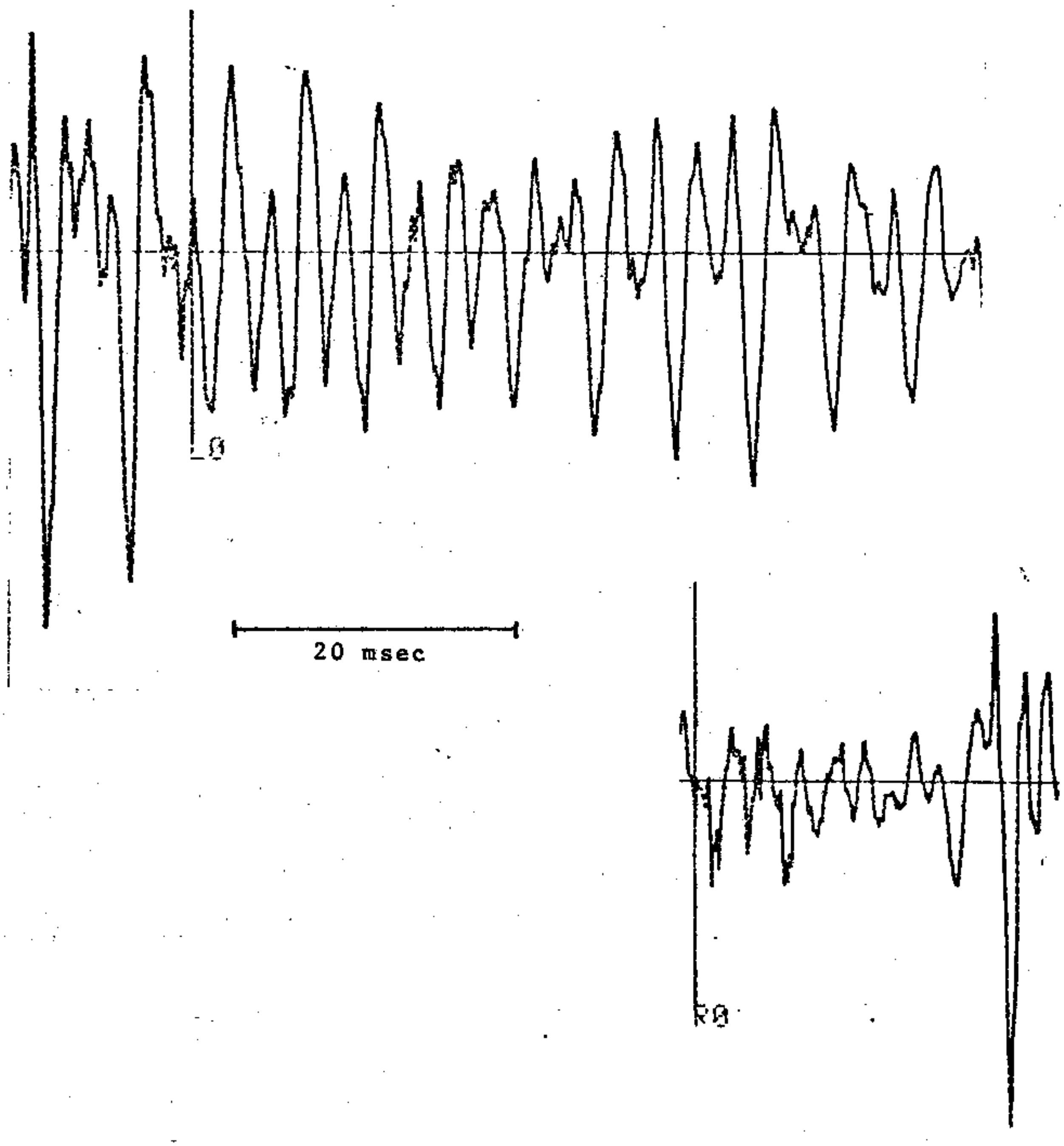


Fig. 4-10d -- Waveform showing natural-edited stimulus with first syllable from rada plus 60 msec of voiced closure from rada, shown between cursors, plus [da] from data as the second syllable. The display is continuous across the two lines. The speaker of the original tokens is JP.

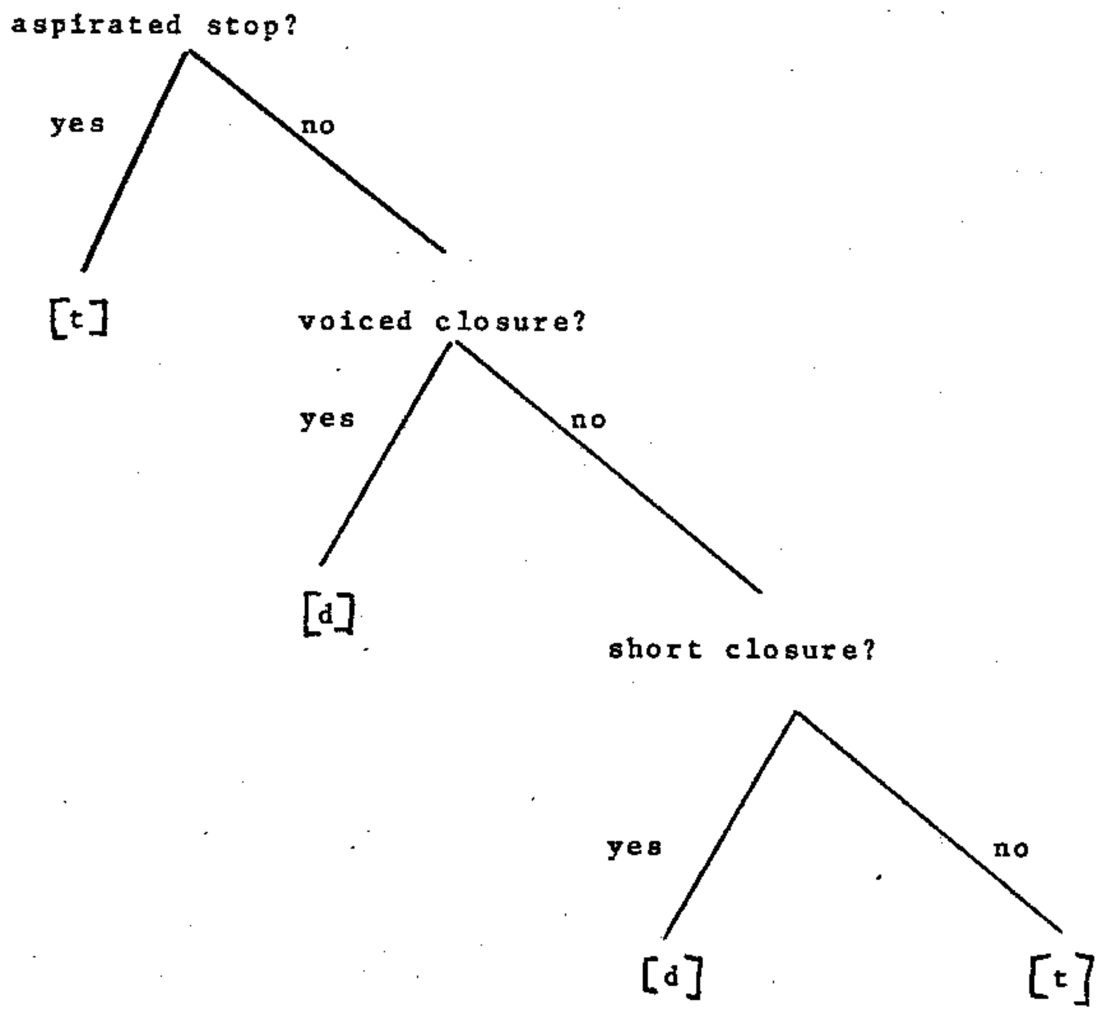


Fig. 4-11 -- Proposed hierarchy of some cues to medial voicing in English stops.

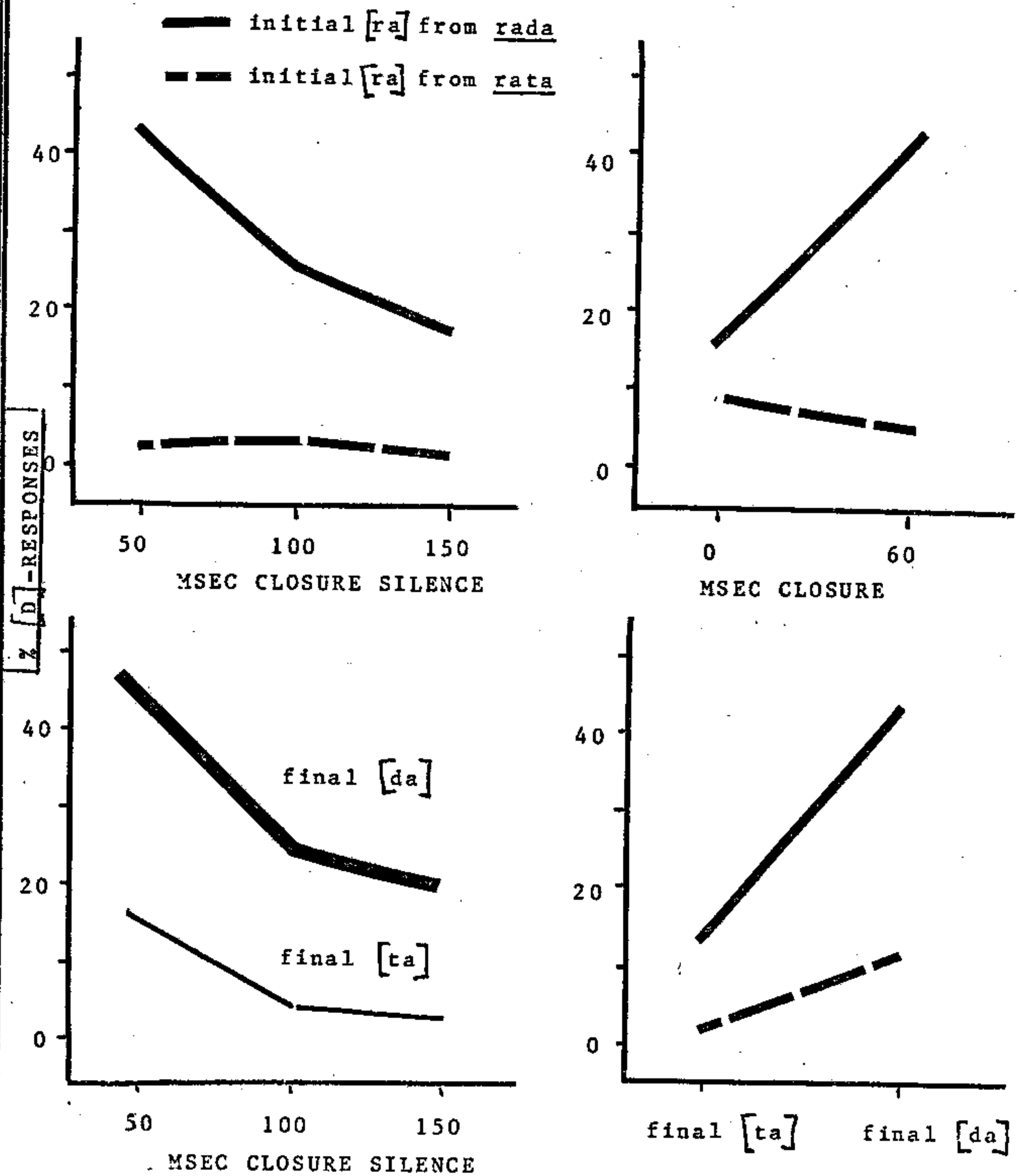


Fig. 4-12 -- Examples of interactions among two factors in experiment on perception of medial [t] and [d].

CHAPTER FIVE -- Summary and Conclusions

5.1. Summary and Discussion of Experimental Findings

The experiments described in Ch. 2 - 4 were designed to determine the importance of VOT as a cue to voicing in Polish stop consonants. The overall result of these experiments was that VOT is the best dimension for representing the Polish voicing contrast when compared to, e.g., burst parameters. VOT is strongly correlated with <u>production</u> categories, and two acoustic aspects, closure voicing and voicing lag, of VOT act as primary cues to the <u>perception</u> of voicing. It was found that the two voicing categories are differentiated in production by the presence or absence of voicing during closure. Voiced stops consistently have some voicing during closure; voiceless stops do not, but rather have a moderate interval of voicing lag after the release burst. Correspondingly, closure voicing and voicing lag were each found to be an overriding cue for one of the voicing categories. Thus the production and perception data are complementary.

The production data can be summarized as follows. VOT values were measured for Polish [t] and [d] in initial and medial position in both minimal pairs and running speech. For initial [d] in running speech, and for medial [d] , the duration of closure before stop release was used as a measure of VOT. Given these measures, the production contrast was found to be generally straightforward. In all positions and speech styles, the mean VOT value for [t] is about +20 msec (voicing lag). For initial [d] in minimal pairs, the mean VOT value is -120 msec (voicing lead). For medial [d] in minimal pairs, the mean VOT value is -90 msec. For initial [d] in a running speech context, the mean VOT value is about -50 msec ;

preliminary observation of medial [d] tokens in running speech indicates that about the same value is found in that context.

For medial stops, two other parameters were measured, closure duration and preceding vowel duration. The second of these was found not to correlate with stop voicing, although it has been thought that vowels are universally lengthened before voiced consonants. The first parameter does correlate with voicing categories, but there is a fair amount of category overlap. This is due to the occurrence of a silent interval between closure voicing and the burst in some tokens of [d] , resulting in increased total durations extending into the [t]-category.

The perception results can be summarized as follows. For initial stop voicing, synthetic VOT continua covering three ranges of VOT values were used, plus natural speech continua covering a fourth range. Listeners responded systematically to these continua, but the placement of the category boundary is subject to strong range effects. A continuum with more long lag stimuli and fewer prevoiced stimuli results in a higher category boundary. In contrast to these results for Polish listeners, American listeners maintain a constant boundary across various test continua. This difference suggests that VOT is not as stable a perceptual dimension for Polish as for English. Rather, in Polish the relation of the voicing categories to VOT is relative, reflecting the particular stimulus set used in an experiment. In an experimental situation, Polish listeners require a range of VOT values that corresponds approximately to the range of values produced in natural speech. The Polish mean category boundaries varied from about +5 msec VOT to +20 msec VOT, compared to +35 msec VOT for the Americans. The lower of the Polish boundaries, obtained with continua representative of natural production values, are well-matched to the production categories. The higher Polish boundary does not match the production categories, but falls at the modal VOT values for [t]. Thus a good match between production and perception categories is obtained only with an appropriate range of VOT in the test continuum.

In natural tokens of [d], there is always some closure voicing. This suggests that prevoicing may be a necessary cue for the perception of voicedness in initial stops. However, experiments showed that this is true for only some tokens of [d]. Other tokens have additional cues for voicedness during or after the burst. The chief additional cue seems to be a fully-voiced burst. Burst voicing is a secondary cue, in that a voiceless burst in a [d]-token is completely overridden by closure voicing. In the same way, a fully-voiced burst can be completely overridden by a moderate voicing lag following after it. However, if neither closure voicing nor a voicing lag is present in a stimulus, then the secondary cue asserts itself, and a fully-voiced burst results in a voiced percept. The results of these perceptual experiments suggest that there may be degrees of voicing in [d]-bursts, which vary in their strength as cues. Burst cues affect listeners' responses only at ambiguous VOT values, which in these stimuli correspond to +10 to +20 msec VOT.

The situation with medial stops seems somewhat more complicated. Voicing during closure is still the strongest cue for voicedness, and a long silent closure gives a voiceless percept. There is a trading relation between closure voicing and silence, and total closure duration, in that voiced closures tend to be short, and voiceless ones tend to be long. Long voiced closures were not tested; however, a short voiceless closure gives an ambiguous percept. (It is possible that with a fully-voiced medial [d]-burst the percept would be voiced, even without closure voicing.) At intermediate closure durations, roughly corresponding to the overlapping durations found in production, closure silence offsets closure voicing as a cue for voicedness. In otherwise ambiguous combinations, other secondary cues have an effect as well.

Secondary cues affect the percept when the stimulus lacks primary cues, and also when the stimulus configuration is unnatural or implausible. In a sufficiently implausible stimulus, primary cues have less effect than usual. For example, a long voicing lag is normally a strong primary cue to voicelessness. However, if the transition onset that follows the lag is taken from a different stimulus that had no such lag, the effect of the lag cue is weakened. This result indicates that speech cues are not assessed independent of the context. Rather, they are assessed in a speech mode that takes into account possible productions and is sensitive to discontinuities in the stimulus. (Another possible explanation of this effect is that the perceptual system does not process totally unfamiliar patterns in the same way as more familiar speech stimuli.) Whatever set of weightings is normally brought to bear in assessing the total configuration of cues that co-occur in a stimulus, it is practically abandoned in the face of a truly equivocal stimulus. Any cue can then contribute to the percept.

Polish listeners are quite sensitive to the plausibility of a stimulus presented in a perceptual test. They are more sensitive than English listeners to the range of VOT values presented. If the experimental parameters used are not suitable, Polish listeners do not respond to VOT in a linguistically-relevant way, i.e., in relation to the categories found in their language. Rather, they respond in terms of the psychoacoustic boundary found for non-speech stimuli at +20 msec separation between onsets. The Polish voicing categories are not aligned with any natural psychoacoustic boundary. On the other hand, the English [t]-[d] boundary at about +35 msec VOT corresponds to a second psychophysical boundary, one for synthetic speech stimuli, as demonstrated by chinchillas (Kuhl and Miller, 1978). Thus there is no reason for English listeners to show boundary shifts, for whether they respond on the basis of linguistic or psychoacoustic categories, their categorizations should remain the same.

A second goal of this study was to explore the advantages and disadvantages of each of the available voicing contrasts, prevoiced-short lag and short lag-long lag. In the choice of voicing contrasts, a language seems to face a trade-off between production and perception stability. The English [t]-[d] contrast, with the category boundary at about +35 msec VOT, is more stable perceptually, with a better psychophysical basis, but its production may require a more precise control of timing. In contrast, the Polish [t]-[d] contrast, with the category boundary at about 0 to +5 msec VOT, is fairly straightforward to produce, but it seems to be more difficult to perceive consistently under varied circumstances.

Both the Polish and English voicing contrasts use the short lag category, which is marked by neither closure voicing nor aspiration. The production of short lag stops depends on having the vocal cords in their phonation-neutral position at the time of the release of the stop occlusion. It takes about 20 msec for phonation to get underway once air starts flowing through the glottis, which it will do when the supraglottal air pressure drops at stop release (Lieberman, 1967). This timing relation corresponds exactly to the VOT values measured for Polish [t], and it will be recalled that the mean VOT for [t] is quite consistent across positions and contexts.

Thus languages start with this one highly-valued VOT category and with it build a voicing contrast. In Polish, this is done by exploiting voicing-lead values along the VOT dimension, giving a stable, easy-to-produce contrast. In English, this is done by exploiting the psychoacoustic categorization of the VOT dimension into short-lag and long-lag values.

5.2. Feature Representations of the Voicing Contrast

In Ch. 1 it was proposed that at one level of feature representation, all two-category voicing contrasts should be represented as [±voice], regardless of their phonetic specifications. The intention of this proposal is that phonological rules which refer to voicing be equivalent across languages at a higher level than the phonetic. Thus voicing assimilation and devoicing processes will be described similarly across languages, regardless of the actual phonetic contrasts involved. The feature representation of three- and four-category contrasts would be determined by phonological evidence in each language involved, possibly using a non-binary voicing feature, or possibly two separate features, e.g., voicing and aspiration.

To allow this equivalent representation, the phonological features cannot be provided by the lowest level of phonetic description and then be carried up to the morphophonemic level. The phonological features are in this sense abstracted from the phonetic detail. However, this use of non-physical features is not meant to be arbitrary. The phonological features must still have phonetic content. The mapping of the voicing feature into the phonetic dimension of VOT provides such content, although other dimensions may be needed for certain categories such as voiced aspirates.

It is proposed that the nature of the VOT dimension itself--a continuum of timing relations--accounts for the cross-linguistic equivalences. The VOT dimension provides three potentially-contrastive categories, and a given language with two or three voicing categories uses the middle (short lag) one and either or both of the extreme ones for its own contrast, although the exact location of each category may vary across languages. There are then three reference points along the VOT continuum that can be used in describing cross-linguistic voicing processes. They are the middle category, the positive endpoint, and the negative endpoint. That is, the phonological voicing rules can be thought of as shifts towards one of the endpoints, or towards the (phonetically unmarked) short lag category. The $[\pm \text{voice}]$ feature is seen to stand for a <u>relative</u> index of more- or less-voicedness. A given language's <u>starting point</u> along the continuum will not affect the process. The VOT dimension allows a degree of phonetic content but still abstracts away from non-contrastive variation specific to individual languages or speakers.

At some lower level still, that phonetic detail must be provided. The VOT dimension is well-suited to feed into a physical level of description, since it is a cover-term for acoustic consequences of articulatory timing relations. As such, it allows the acoustic correlates of the various voicing contrasts to be described systematically, relating production and perception.

APPENDIX: Polish sentences constructed by MJM and read by five speakers in Łódż, Poland and three in Providence (see Ch. 2)

1. Czyja to gazeta?

2. Czy ktoś dzwonił dzisiaj?

3. Zaraz wychodzimy.

4. Idziemy do kina?

5. Jedziemy szybko do Warszawy.

6. Czemu nie zadzwoniłeś?

7. Zaraz wychodzimy.

8. Listonosz przychodzi raz dziennie.

9. Byłeś wczoraj w Warszawie?

10. Żle się czuję dzisiaj.

11. Na zabawie było fajnie.

12. Mogę wyjść?

13. Czy Andrzej już wrócił?

14. Kto był na wykładzie?

15. Na zabawie było fajnie.

16. To moja gazeta.

17. Jadę zaraz do domu.

18. Jakie masz plany na jutro?

19. Listonosz przychodie raz dziennie.

20. O której wychodzimy?

21. Wracamy do domu?

22. Nie moglem wczoraj zadzwonić.

23. Jak bylo na zabawie?

24. Nie mam planów na jutro.

25. Dostałeś bilety?

26. Czy on ma samochód?

27. Jak się czujesz dzisiaj?

28. Nikogo nie było na wykładzie.

29. Jedziemy szybko do Warszawy.

30. Nie moglem wczoraj zadzwonic.

31. Czy to był twój plan?

32. Jak czesto przychodie listonosz?

33. To moja gazeta.

34. Nikogo nie było na wykładzie.

35. Ktorędy wracamy do Warszawy?

36. Nie mam planów na jutro.

37. Czy dostałes już paczkę?

38. Kiedy wracasz do domu?

39. Żle się czuję dzisiaj.

40. Jadę zaraz do domu.

BIBLIOGRAPHY

Abercrombie, David (1967). Elements of General Phonetics.Chicago: Aldine Atherton, Inc.

Abramson, Arthur S. (1977). Laryngeal Timing in Consonant Distinctions. Phonetica 34: 295-303.

Abramson, Arthur S., and Leigh Lisker (1970). Discriminability along the voicing continuum: cross-language tests. Proc. 6th Intern'l Cong. Phonetic Sci. Prague, 1967. Pp 569-73. Prague:Academia.

Abramson, Arthur S. and Leigh Lisker (1972). Voice-timing perception in Spanish word initial stops. J. Phonetics 1: 1-8.

Bell-Berti, Fredericka (1975) Control of pharyngeal cavity size for English voiced and voiceless stops. J. Acoust. Soc. Am. 57: 456-61.

Blumstein, S.E., W.E. Cooper, E.B. Zurif, and A. Caramazza (1977). The Perception and Production of Voice-Onset Time in Aphasia. Neuropsychologia 15: 371-83.

Caramazza, Alfonso and Grace Yeni-Komshian (1974). Voice onset time in two French dialects. J. Phonetics 2: 239-45.

Chen, Matthew (1970). Vowel length variation as a function of the voicing of the consonant environment. Phonetica 22: 129-59.

Chomsky, Noam, and Morris Halle (1968). The Sound Pattern of English New York: Harper and Row.

Darwin, C.J., and S.A. Brady (1978). A range effect in the perception of voicing. J. Acoust. Soc. Am. 63: 1556 - 58.

Delattre, P. (1962). Follow-up to House, "On Vowel Duration...". J. Acoust. Soc. Am. 34: 1141-43 (L).

Dent, Laurel (1976). Voice Onset Time of Spontaneously Spoken Spanish Voiceless Stops. Presented at the 91st meeting of the Acoustical Society of America, April, 1976.

Donald, Susan Lea (1976). Selective adaptation of labial stops for Thai and English speakers. Presented at the annual meeting of the Linguistic Society of America, Dec. 1976.

Donald, Susan Lea (1978). Discrimination of Subphonemic Phonetic Distinctions . Haskins Labs. Status Report on Speech Research SR - 54.

Eimas, P.D., E.R. Siqueland, P. Jusczyk, and J. Vigorito (1971). Speech perception in infants. Science, 171: 303-306.

Fudge, E.C. (1967). The Nature of Phonological Primes. Journal of Linguistics 3:1 - 36.

Finney, D.J. (1952, 1971). Probit Analysis. Cambridge: Cambridge University Press.

Fujimura, Osamu (1971). Remarks on consonants - synthesis experiments and acoustic cues. IN L.L. Hammerisch, R. Jakobson, and E. Zwirner (eds.) Form and Substance - Phonetic and Linguistic Papers Presented to Eli Fischer-Jorgensen. Copenhagen: Akademisk Forlag, pp. 221 - 232.

Ganong, W.F. (1978). A word advantage in phoneme boundary experiments. J. Acoust. Soc. Am. 63, Suppl. 1, S 20 (A).

Haggard, Mark, S. Ambler and M. Callow (1970). Pitch as a voicing cue. J. Acoust. Soc. Am. 47: 613 - 17.

Halle, Morris and K.N. Stevens (1972). A Note on Laryngeal Features. Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T., 101: 198 - 213.

Heffner, R-M.S. (1950, 1969). General Phonetics. Madison: The U. of Wisconsin Press.

Jakobson, R.;G. Fant & M. Halle (1952). Preliminaries to Speech Analysis: The Distinctive Features and their Correlates. Cambridge: MIT Press.

Klatt, D.H. (1973).Interaction between two factors that influence vowel duration. J. Acoust. Soc. Am. 54: 1102-04.

Klatt, Dennis (1975).Voice onset time, frication, and aspiration in word-initial consonant clusters. J. of Speech and Hearing Research 18: 686 - 706.

Kuhl, Patricia and J.D. Miller (1975). Speech perception by the chinchilla: Voiced - voiceless distinction in alveolar plosive consonants. Science 190: 69 - 72.

Kuhl, Patricia and J.D. Miller (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. J. Acoust. Soc. Am. 63: 905 - 917.

Ladefoged, P. (1971). Preliminaries to Linguistic Phonetics. Chicago: Univ. of Chicago Press.

Lasky, R.E., A. Syrdal-Lasky, and R.E. Klein. (1975). VOT discrimination by four and six and a half month old infants from Spanish environments. J. of Experimental Child Psychology,20: 215 - 225.

Lieberman, P. (1967). Intonation, perception, and language. Cambridge: MIT Press.

Lieberman, Philip. (1970). Towards a Unified Phonetic Theory. Linguistic Inquiry 1, 307 - 22.

Lieberman, P. (1977). Speech Physiology and Acoustic Phonetics. New York: Macmillan Publishing Co.

Lisker (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. Language 33: 42 - 49.

Lisker, Leigh (1975). Is it VOT for a first - formant transition detector? J. Acoust. Soc. Am. 57: 1547 - 51.

Lisker, Leigh (1977). Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. J. Acoust. Soc. Am. 62 (Suppl. 1): 577 (A).

Lisker, Leigh & Arthur S. Abramson (1964). A cross-language study of voicing in initial stops: acoustical measurements. Word 20: 384 - 422.

Lisker, Leigh & Arthur S. Abramson (1967). Some effects of context on voice onset time in English stops. Language and Speech 10: 1 - 28.

Lisker, Leigh and Arthur S. Abramson (1970). The voicing dimension: some experiments in comparative phonetics. Proc. 6th Int.Cong. Phonetic Sci., Prague 1967. Prague: Academia.

Lisker, Leigh, and Arthur S. Abramson (1971). Distinctive Features and Laryngeal Control. Language 47 : 767 - 85.

Mikoś, Michael J. (1977). Problems in Polish Phonology. Ph.D. dissertation, Brown University.

Mikoś, Michael, Patricia Keating, and Barbara Moslin (1978). The perception of voice onset time in Polish. J. Acoust. Soc. Am. 63: S 19 (A).

Miller, J.D., C.C. Wier, R. Pastore, W.J.Kelly, and R.J. Dooling (1976). Discrimination and labeling of noise - buzz sequences with varying noise - lead times: An example of categorical perception. J.Acoust. Soc. Am. 60: 410 - 17.

Moslin, Barbara J. (1978). The role of phonetic input in the child's acquisition of the voiced-voiceless contrast in English stops: a voice onset time analysis. Ph.D. dissertation, Brown University.

Myerow, S. and Millwood, R. (1978). SPLIT -- A Sound Editor for a PDP-8 Computer. Behavioral Instrumentation and Research News 10: 281 - 84.

Pisoni, David B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. J. Acoust. Soc. Am.61: 1352 - 61.

Port, Diana Kewley and Malcolm S. Preston (1972). Early apical stop production: a voice-onset time analysis. Haskins Lab. Status Report on Speech Research SR 29/30.

Port, Robert F. (1977). The Influence of Speaking Tempo on the Duration of Stressed Vowel and Medial Stop in English Trochee Words. Ph.D. Dissertion, U. Conn., 1976, distributed by Indiana U. Linguistics Club.

Postal, Paul (1968). Aspects of phonological theory. New York: Harper and Row.

Repp, Bruno H. (1979). Perceptual trading relation between aspiration amplitude and VOT. J. Acoust. Soc. Am. 65, Suppl. 1: S8 (A).

Sawusch, J.R., and D.B. Pisoni (1974). Category boundaries for speech and non-speech sounds. J. Acoust. Soc. Am. 55, Suppl. 1: S436 (A).

Sharf, D. (1962) Duration of post-stress inter-vocalic stops and preceding vowels. Language and Speech 5: 26 - 30.

Simon, Claude and Adrian Fourcin (1978). Cross-language study of speech-pattern learning. J. Acoust. Soc. Am. 63: 925 - 35.

Simon, Helen J., and Michael Studdert-Kennedy (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. J. Acoust. Soc. Am. 64: 1338 - 57.

Sommerstein, Alan H. (1977). Modern Phonology. Baltimore: University Park Press.

Stevens, Kenneth S. (1972). Quantal Nature of Speech. IN E.E.David, Jr. and P.B. Denes (eds.) Human Communication: A unified view. New York: McGraw-Hill.

Stevens, K.N. and S.E. Blumstein (1978). Invariant cues for place of articulation in stop consonants. J. Acoust. Soc. Am.64: 1358 - 68.

Stevens, Kenneth N. & Dennis Klatt (1974). Role of formant transitions in the voiced - voiceless distinction for stops. J. Acoust. Soc. Am. 55, 653 - 59.

Streeter, Lynn A. (1976) Kikuyu labial and apical stop discrimination. J. Phonetics 4, 43 - 49.

Streeter, Lynn A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. Nature 259: 39 - 41.

Streeter, Lynn A. and T.K. Landauer (1976). Effects of learning English as a second language on the acquisition of a new phonemic contrast. J. Acoust. Soc. Am. 59: 448 - 61.

Summerfield, A.Q., and M.P. Haggard (1974). Perceptual processing of multiple cues and contexts: effects of following vowel upon stop consonant voicing. J. Phonetics 2: 279 - 95.

Summerfield, Quentin & Mark Haggard(1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. J. Acoust. Soc. Am. 66: 435 - 48.

Trubetzkey, N.S. (1969). Principles of Phonology. Trans. Christiane A.M. Baltaxe. Berkeley: Univ. of California Press.

Yeni-Komshian, Grace, Alfonso Caramazza, and Malcolm S. Preston (1977). A study of voicing in Lebanese Arabic. J. Phonetics, 5: 35 - 48.

Waters, R.S. and W.A. Wilson, Jr. (1976). Speech perception by rhesus monkeys: the voicing distinction in synthesized labial and velar stop consonants. Perception & Psychophysics 19: 285 - 289.

Westbury, J.R., and S. Niimi (1979). An effect of phonetic environment on voicing control mechanisms during stop consonants. J. Acoust. Soc. Am. 65, Suppl. 1: S 23 (A).

Williams, Lee (1974). Speech Perception & Production as a Function of Exposure to a Second Language. Ph.D. dissertation, Harvard Univ.

Williams, Lee (1977). The voicing contrast in Spanish. J. Phonetics 5, 169 - 184 (1979).

Winitz, Harris, Conrad La Riviere and Eve Harriman (1975). Variations in VOT for English initial stops. J. Phonetics 3, 41 - 52.

Zlatin, Marsha A. (1974). Voicing contrast: perceptual and productive voice onset time characteristics of adults. J. Acoust. Soc. Am. 56: 981 - 94.

Zue, Victor W. (1976). Acoustic Characteristics of Stop Consonants: A Controlled Study. Sc.D. dissertation, M.I.T.