NATURALLY OCCURRING VOT CONTINUA IN APHASIC SPEECH: PERCEPTUAL CORRELATES

WOLFRAM ZIEGLER PHILIP HOOLE

Neuropsychological Department
Max-Planck-Institute for Psychiatry
D-8000 München 40

ABSTRACT

Using plosive-vowel stimuli spliced out of aphasic speech VOT-continua were prepared and presented to experienced listeners for identification of the voicing of the initial consonant. Possible range effects on listeners' judgments were studied. Stimuli with unexpected response rates are discussed.

INTRODUCTION

Errors of consonant voicing are considered to play an important role in the categorization of aphasic production deficits. They have often been reported in transcription studies (e.g. /1/) and have, at least with respect to certain aphasic syndromes, been interpreted as reflecting a deficit in the temporal coordination of articulatory and laryngeal gestures.

A major shortcoming of the perceptual assessment of consonant voicing is that it leads to a description based on discrete entities where continuous measures would probably be more adequate. Therefore, disturbances of plosive voicing have been adressed in voice-onset time studies of various aphasic syndromes, revealing that aphasic speakers can vary considerably in their VOT /2/.

Yet, to our knowledge there has not yet been any attempt to systematically examine a particular corpus of aphasic speech material both from the point of view of judgments of discrete categories and measurements of a continuous variable such as VOT. Such an inquiry, however, would be of great interest since it cross-relates the results obtained on different methodological bases. One would predict that the perception of stop consonants produced by aphasic patients is largely conditioned by the measured VOT and that only productions with voice onset times near the category boundary are ambivalent in perception. Yet, other cues to stop voicing may be present in aphasic speech which override the perceptual significance of voice onset time.

The fact that VOT in some aphasic patients covers a broad range of considerable density was used in this study to construct sets of plosive-vowel stimuli along VOT continua. On the basis of such continua, category boundaries were estimated for the productions of three aphasic patients in identification experiments, asking whether these boundaries are invariant over the different subjects. The possible influence of different VOT ranges covered by a stimulus set /3/ was also investigated. Stimuli which did not attract the responses expected from their VOT are discussed with regard to possible acoustic features that override the VOT cue.

EXPERIMENT 1

Method, subjects and material

The material for the perception experiments was taken from speech utterances of 3 female patients (aged 46, 55 and 64) suffering from apraxia of speech following occlusion of the left middle cerebral artery. In a production experiment reported in detail elsewhere the patients had been required to produce 20 repetitions of the nonsense words /dabo/, /dapo/, /tabo/ and /tapo/. For these words the length of the segment from release burst of the initial consonant to onset of periodicity (VOT) was determined by obtaining a consensus of opinion from three examiners (/d/ generally has no voicing lead in South German). The initial consonant and 20 periods of the vowel /a/ were then digitally spliced out, after eliminating those words showing gross deviations in place of articulation for the consonant, or in loudness or fundamental frequency for the vowel. The vowel length of 20 periods was chosen so that the vowel would be as long as possible, but without including the transition to the following plosive. These spliced segments represent the actual stimuli used in the experiments; a total of 68 stimuli spanning the VOT continuum were obtained for patient A1, 56 for A2 and 70 for A3. In contrast to classical synthetic continua the stimuli were not absolutely equidistant along the VOT axis.

A panel of listeners with considerable experience in perception experiments took part in the listening tests; six listeners for Patient A1 and five listeners for A2 and A3. Each listener was tested separately. In each session the randomized stimuli of one patient were presented to the listeners, with 4 repetitions of each stimulus. Each patient was presented in 3 sessions with a different randomization each time; thus each listener heard each stimulus of each patient 12 times. The listeners were required to identify the stimuli as /da/ or /ta/ (forced-choice). Presentation of stimuli was performed by a laboratory computer using a 20 kHz sample rate for D/A conversion with a 9kHz low-pass filter prior to output over high-quality loudspeakers.

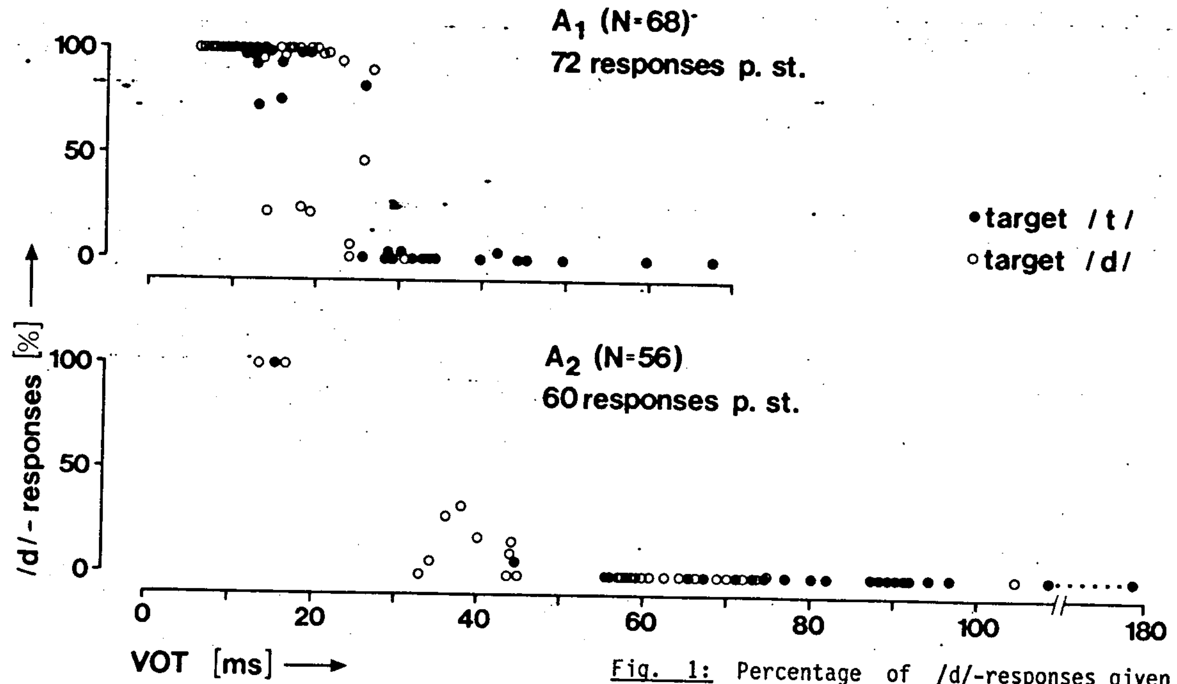

Fig. 1: Percentage of /d/-responses given by experienced listeners to /da/, /ta/ stimuli spliced out from utterances of patients A1 and A2.

Results

In Fig. 1 the percentage of /d/-responses is displayed versus measured VOT for patients A1 and A2. The figure also indicates whether the production target had been /t/ or /d/. For patient A1 most VOT values lie between 5 ms and 50ms. Intended /t/-productions mainly occupy the range above 25 ms. Responses are unanimously /d/ below about 10 ms and /t/ above 30 ms. The category boundary would seem to lie between 20 and 25 ms, but with some ambivalent stimuli in the region below this diverging widely in the number of /d/-responses. The responses were obviously not conditioned by the actual production target, since intended /t/ productions, in particular, were often unambiguously perceived as /d/. There are also several intended /d/ productions with low VOT that are nonetheless almost unanimously identified as /t/. Productions of Patient A2 covered a much wider range of VOT values, with a general shift to higher values. There are unfortunately some gaps in the continuum; only three productions received 100% /d/-responses. It is also noticeable that several stimuli with a VOT of about 40 ms still received ambivalent responses while for Patient A1 /t/-responses in this range had already risen to 100%. Thus the category boundary seems to be further to the right for A2. The question of whether this might be due to the different VOT ranges was addressed in the second experiment outlined below.

The results of the first twenty stimuli of the continuum from Patient A3 are given in Tab.I with the corresponding VOT value. All further stimuli i.e. nos. 21-70, with VOTs up to almost 200 ms gave 100% /t/-percepts. Perhaps the most surprising aspect of the response pattern is that even at very low VOT values very high numbers of /t/-responses were obtained for some stimuli. The possible acoustical reasons for this will be examined in the discussion.

Table I: VOT and percent /d/ responses for the first 20 stimuli from the continuum of Patient A3.

stimulus-no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
VOT (ms)	14	14	17	17	19	20	22	22	23	23	23	24	24	24	25	25	27	30	33	36
/d/-responses (%)	100	13	23	100	100	28	32	2	12	47	22	32	5	3	100	2	10	2	0	0

3 0323

EXPERIMENT 2

In experiment 2 we wanted to investigate whether response behavior to these kinds of continua is susceptible to range effects, thus possibly explaining the apparent difference in category boundary between A1 and A2.
Listeners and methods remained the same, but the range of stimuli presented from each patient was changed in the following way:
For Patient A1 all stimuli with a VOT longer than 28 ms were discarded, i.e. the range in which only /t/-responses occurred in the first experiment. This left a total of 48 stimuli. In Patient A2, a similar criterion led us to discard all stimuli with VOT above 55 ms, which left a total of 16 stimuli.
These manipulations of the stimulus continuum were expected to lead to more /t/-responses, particularly on the ambivalent stimuli.
The stimuli were randomized with 10 repetitions each and presented to the listeners in 2 sessions for A1 and in a single session for A2. These sessions took place several weeks after those of Experiment 1.

Results

Fig. 2 compares the percentage of /d/-responses for the full and reduced continua, with the responses grouped into 5 ms bins for Patient A1 and 10 ms bins for A2. In both cases there is a slight increase in /t/ responses in the reduced condition. But the effect, while going in the expected direction, is clearly rather weak. It appears somewhat stronger if only the ambivalent stimuli are regarded (i.e. those neither 100% /t/ or /d/ in the first experiment). This is displayed in Table II averaged over all listeners and in Fig. 3 for a single listener.

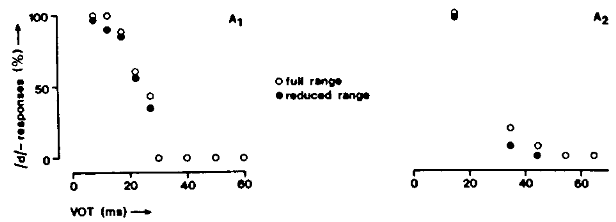

Fig. 2: Change in /d/-response rate from Experiment 1 to Experiment 2 (grouped stimuli, all listeners). Left: Patient A1. Right: Patient A2.

Table II: Percentage of /d/-responses given to ambivalent stimuli of patient 1 (N=19) and patient 2 (N=8) under two different range conditions.

		condition	
		full range	reduced range
patient	1	75	45
	2	14	4

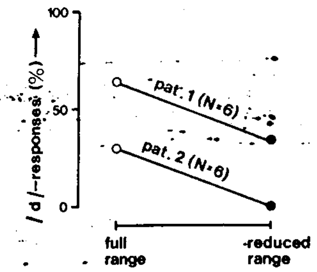

Fig. 3: Change in /d/-response rate from Experiment 1 to Experiment 2 for one listener, restricted to stimuli that were ambivalent in Experiment 1.

DISCUSSION

From the identification results in Fig. 1 and Table I a rough estimate of the category boundary can be made for each patient. However, particularly for Patients A1 and A3 the numerous stimuli not falling on an ideal response curve make it clear that VOT cannot be the only factor conditioning listeners' responses. This is hardly unexpected when one considers that the continua consist of a large number of naturally-occurring stimuli rather than a very restricted number of synthetic or electronically manipulated stimuli. Nevertheless, it is interesting to look closer at those stimuli that fall out of line when ordered according to VOT. Fig. 4 gives one example for each of Patients A1 and A3 in which similar measured VOT values attracted widely differing /d/ response-rates. At the bottom of the figure in the example for A3 both utterances are rather weak in harmonic energy in the high frequencies. The obvious difference, however, is that in the left-hand example the onset of harmonic energy occurs more or less simultaneously at least up to the frequency range of the first formant while in the right-hand example for about 70 ms after voicing onset there is only harmonic energy at the fundamental frequency, presumably because of incomplete glottal adduction. It could be argued that it would be more appropriate to adopt a different criterion for measuring VOT, e.g. the interval from plosive release to onset of harmonic energy in the region of the second or higher formants (cf. Klatt /4/). However, in cases such as the present one even this criterion could be difficult to apply consistently since in the sonagram it is very difficult to discern harmonic energy in the higher formants. This methodological problem would probably be even more acute when measuring in the time-domain. Beyond such technical considerations, however, it seems more important to stress that not all aspects of disturbed glottal adduction can be captured in a single parameter such as VOT, however it is defined.

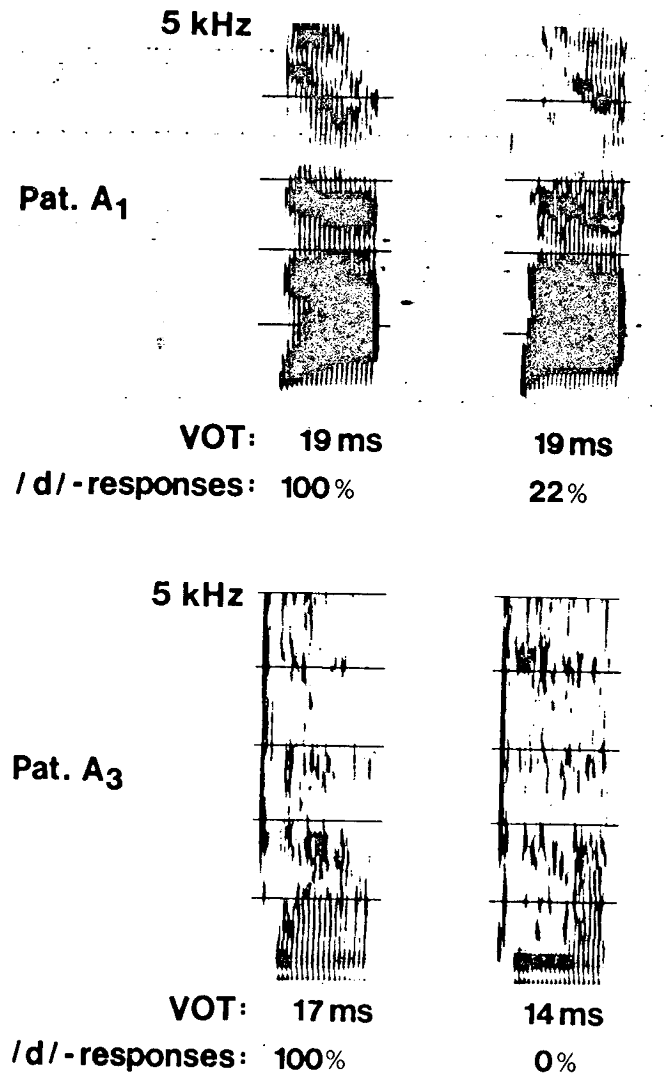

Fig. 4: Two examples of CV-stimulus pairs with similar VOT which attracted different voicing judgments.

The example given in the top half of the figure illustrates a rather different phenomenon. Here it is fairly safe to say that any criterion for VOT measurement would lead to essentially the same value. However, the number of /d/-responses again diverges widely. The most salient difference between the two sonagrams would seem to be the much more visible F1 transition in the production that attracted 100% /d/-responses. This probably reflects differences in the extent of lingual anticipation of the vowel target at the moment of plosive release. One might speculate that a patient like A1 makes active use of such a movement pattern in order to partially compensate for her obvious deficits in the control of VOT. However, the lack of a correlation between intended target and percept in this subject speaks against this view.

Turning to the results of Experiments 1 and 2 with regard to range effects it should first be recalled that for Patient A2 several stimuli still attracted ambivalent reponses in Experiment 1 at a point on the VOT continuum where responses for A1 and A3 were already unanimous (i.e. around 40 ms). Under the assumption that these ambivalent responses to some extent represent the listeners' reaction to the overwhelming number of clear /t/ stimuli it could be expected in Experiment 2 that any manipulation of the VOT range covered in such continua should be reflected in the response-rates, particularly near the category boundaries (cf. /3/). The results are not as clear as one might wish, but there was a trend in the expected direction in both A1 and A2. It is noticeable that the slightly unexpected ambivalent stimuli for Patient A2 almost disappeared in the second experiment.

REFERENCES

/1/ J.E.Trost, G.J.Canter: Apraxia of speech in patients with Broca's aphasia: A study of phoneme production accuracy and error patterns.
/2/ S.E.Blumstein, W.E.Cooper, H.Goodglass, S.Statlender, J.Gottlieb: Production deficits in aphasia: A voice-onset time analysis. Brain & Lang. 9: 153-170 (1980)
/3/ S.A.Brady, C.J.Darwin: Range effect in the perception of voicing. J. Acoust. Soc. Am. 63: 1556-1558 (1978)
/4/ D.H.Klatt: Voice onset time, frication and aspiration in word-initial consonant clusters. J. Speech & Hear. Res. 18: 686-706 (1975)