

A VOT analysis of apraxic/aphasic voicing errors

J. L. WAMBAUGH, J. E. WEST and P. J. DOYLE

Highland Drive VA Medical Center, Pittsburgh, PA, USA

Abstract

The primary purpose of this investigation was to perform a VOT analysis of correct and incorrect productions elicited from an apraxic/aphasic speaker (S1), who consistently devoiced stop and affricate targets. A secondary objective was to analyse the VOTs of an apraxic/aphasic speaker (S2) who produced stop and affricate voicing distinctions correctly, as determined by broad phonetic transcription. Seven hundred and ninety monosyllabic words, representing four voiced-voiceless cognate pairs, were analysed for each subject. VOT results indicated that S1's error productions (i.e. [t] for /d/; [p] for /b/; [k] for /g/; and [tʃ] for /dʒ/) differed significantly from his correct, voiceless, homonymous productions (i.e. [t] for /t/; [p] for /p/; [k] for /k/; and [tʃ] for /tʃ/). Therefore, S1's sound errors were not considered to be sound substitutions, as had been indicated perceptually. Overlapping VOT values were found for all of S2's voiced-voiceless cognate pairs, despite perceptually accurate productions, indicating possible speech motor control problems.

Introduction

The analysis of voice-onset time (VOT) has been used in numerous investigations to examine the speech production of apraxic and/or aphasic persons (Baum *et al.* 1990, Blumstein 1977, 1980, Hoit-Dalgaard *et al.* 1983, Itoh *et al.* 1982, Kent and McNeil 1987, Shewan *et al.* 1984). In general, non-fluent and/or apraxic speakers have been found to have more variable VOT values than normal speakers with less clear separation of values for voiced-voiceless cognates. However, exceptions to these findings (Baum *et al.* 1990, Seddoh 1996a, b, Shewan *et al.* 1984) highlight the need to examine acoustic data on an individual rather than group level.

Typically, data from normal speakers have been used as a basis of comparison for apraxic/aphasic VOT findings. This comparative approach allows a determination of normalcy, but offers limited information about an individual's sound system, specifically in terms of determining the nature of sound errors.

The purpose of this investigation was to examine the VOTs of an apraxic/aphasic speaker who consistently produced voiced stop and affricate targets as what were perceived as their voiceless cognates (e.g. 'bet' → 'pet'; 'gill' → 'kill'). Specifically, we compared the error productions (i.e. incorrect productions of voiced targets) to homonymous, correct, voiceless productions (e.g. 'pet' → 'pet';

Address correspondence to: Julie L. Wambaugh, who is now at the Department of Communication Disorders, 1201 Social and Behavioral Science Building, University of Utah, Salt Lake City, Utah 84112, USA.

'kill' → 'kill') in order to determine if the error sounds should be considered substitutions or distortions. For example, we compared the [p] in 'bush' → 'push' to the [p] in 'push' → 'push'. We hypothesized that if the sound errors were substitution errors, then error and correct voiceless productions should appear very similar, acoustically. Conversely, if the errors were related to a disturbance in motor control, we thought it likely that an intact phoneme selection system would attempt to avoid homonymy, and that we would observe a differentiation between error and correct voiceless sounds.

A secondary objective of this investigation was to perform an examination of the VOTs of an apraxic/aphasic speaker who consistently produced correct voicing distinctions in stops and affricates.

Method

Subjects

Subject 1 was the individual who produced voiced stops and affricates incorrectly. He was 58 years old and presented with moderate to severe apraxia of speech and non-fluent aphasia. He was 26 months post-onset of a left-hemisphere cerebral vascular accident (CVA) and exhibited right hemiparesis. Subject 2 was also apraxic and aphasic, but produced stop and affricate voicing distinctions correctly. He was 53 years old and 45 months post-onset of a left-hemisphere CVA and was very similar to Subject 1 in terms of type and extent of aphasic and apraxic impairments. For both subjects, motor speech examination revealed no significant abnormalities in muscle tone or strength or any classifiable dysarthrias (Darley *et al.* 1975). Test results for Subjects 1 and 2 are shown in Table 1.

Both subjects' speech was characterized by trial-and-error groping, difficulty initiating speech, numerous perceived sound substitutions, occasional sound distortions, and variability and inconsistency in multisyllabic word productions over repeated trials (i.e. inconsistency of error location and variability of error type on successive productions of multisyllabic words during administration of the Apraxia Battery for Adults—(Dabul 1979). Sound errors in monosyllabic words tended to be non-variable, in manner, place, and voicing, across repetitions.

Subject 2 was quite accurate in identifying sound errors and frequently attempted to correct errors. Subject 1 appeared to be much less aware of and/or bothered by his sound errors than S2, and rarely re-attempted erroneous productions spontaneously.

Experimental stimuli

One hundred and fifty-eight monosyllabic real words served as stimuli (see Appendix). These words were composed of eight sets: 20 beginning with /b/, 20 beginning with /p/, 20 beginning with /g/, 20 beginning with /k/, 20 beginning with /d/, 20 beginning with /t/, 19 beginning with /dʒ/, and 19 beginning with /tʃ/. For each set of sound cognates (e.g. p-b; t-d; k-g; tʃ-dʒ), the words in the lists were matched such that they formed minimal contrast pairs (e.g. 'buy-pie'). These sounds were selected for study because S1 consistently produced the voiced consonants incorrectly (i.e. as voiceless stops/affricates) during a 200-item consonant production probe.

Table 1. Subjects' test results

Variable	Subject 1	Subject 2
Porch Index of Communicative Abilities (Porch 1981)		
Overall score	10.46	10.49
Western Aphasia Battery (Kertesz 1982)		
Aphasia quotient	34	29.3
Western Aphasia Battery		
Classification	Broca's	Broca's
Test of Adult/Adolescent Word Finding (German 1990)		
Raw score	2	1
Apraxia Battery for Adults (Dabul 1979)		
Apraxia severity	Moderate/severe	Moderate/severe

Speech samples and recording

The 158 monosyllabic words were used to elicit a total of 790 single-word productions from each subject. The 79 words beginning with voiced sounds were grouped together and a list was prepared in which each word occurred five times in pseudo-randomized order (i.e. none of the 79 words appeared twice in a row). The principal investigator read each word individually to each subject and asked him to repeat the word. No feedback concerning the correctness of the subjects' productions was provided. Repetitions of the presented word were provided upon request by the subject. If the investigator was unsure that the subject had attempted production of the correct target word, the investigator repeated the target and requested another attempt by the subject. This repetition procedure was not employed with productions that were ambiguous because of voicing errors, and occurred on only 5% of attempts with S1 (primarily due to occurrences of [b] for /g/ and [t] for /k/). Because S2 sometimes produced /k/ as [t] and /t/ as [k], repetitions were required on approximately 10% of attempts when potentially homonymous productions created uncertainty about the intended target. These data were collected over two sessions, with 5-10-min breaks provided after production of approximately 60 words. This procedure was then repeated with the words beginning with voiceless consonants.

All speech samples were collected in a quiet room with the examiner seated facing the subject. A Sony EMC 155 lapel microphone was used with a Sony TC-D5M cassette recorder.

Perceptual analysis

Productions were transcribed on-line, using broad phonetic transcription (Shriberg and Kent 1982). Audio-recordings were used to confirm transcriptions. A staff member (ASHA-certified speech-language pathologist) independently transcribed every item using the audio-recordings. A third staff member transcribed each production for which there was a disagreement, and the majority transcription was used. Disagreements between the first and second transcriber occurred for only 20 of the 1580 items.

For each of the four sets of sound cognates the subjects' productions were grouped into one of three categories: (a) correct production of the voiceless target,

Table 2. Subject 1: means and standard deviations of VOT values (ms)

Category	p/b	t/d	k/g	tʃ/dʒ
Correct voiceless	M = 74.9 SD = 19.2	M = 81.1 SD = 17.4	M = 80.8 SD = 20.9	M = 92.1 SD = 16.9
Error (voiced → voiceless)	M = 45.2 SD = 11.8	M = 49.8 SD = 10.3	M = 58.2 SD = 13.6	M = 73.6 SD = 13.7
Correct voiced	M = 18.2 SD = 7.1	M = 25.8 SD = 7.9	M = 38.7 SD = 5.6	M = 49.0 SD = 11.7

(b) correct production of the voiced target, and (c) incorrect production of the voiced target (i.e. error production). Subject 1 produced the voiceless targets correctly for 393 of the 395 words. However, as expected, he rarely produced the voiced targets correctly: (a) /d/ → [d] 17 of 100 times, (b) /b/ → [b] 46 of 100 times, (c) /g/ → [g] 8 of 100 times, and (d) /dʒ/ → [dʒ] 20 of 95 times. His incorrect productions of the voiced target were always perceived as productions of the voiceless cognate. Subject 2's productions were always transcribed as belonging to the targeted sound category. That is, there were no perceived voicing errors. As noted previously, both subjects' sound errors were quite consistent at the monosyllabic word level. S1's difficulties with stops and affricates in monosyllables were primarily confined to voicing errors. Whereas S2 occasionally produced [t] for /k/ and [k] for /t/, he readily corrected productions upon request for repetition; therefore only correct productions of targets were analysed for S2.

Acoustic analysis

The recorded speech data were sampled at 20 kHz and stored using the Computerized Speech Lab (Kay Elemetrics 1994). Spectrographic and oscillographic displays were produced and linked to determine VOTs. VOTs of the initial stop and affricate consonants were measured as the time from the onset of the noise burst to the onset of periodic energy. Twenty-five per cent of the speech samples were selected through stratified random sampling for remeasurement by a second investigator. The Pearson product moment correlation for these remeasurements was $r = 0.98$. In examining point-to-point agreement we found 94% agreement within ± 5 ms, with the average difference in measurement being 2 ms. The majority of disagreements occurred in cases of double bursts.

Results

The means and standard deviations of VOT values for each pair of cognates for Subject 1 are presented in Table 2. The correct voiceless VOT values, in the top row of Table 2, were in the ranges that would be expected for a normal speaker (Blumstein *et al.* 1980, Shewan *et al.* 1984). Additionally, lag times increased from labials to velars, as would be expected (Blumstein *et al.* 1980).

The mean values for the correct voiced productions, in the bottom row of Table 2, were also similar to values seen for normals, with perhaps slightly longer lag times for [g] and [dʒ] than normally seen.

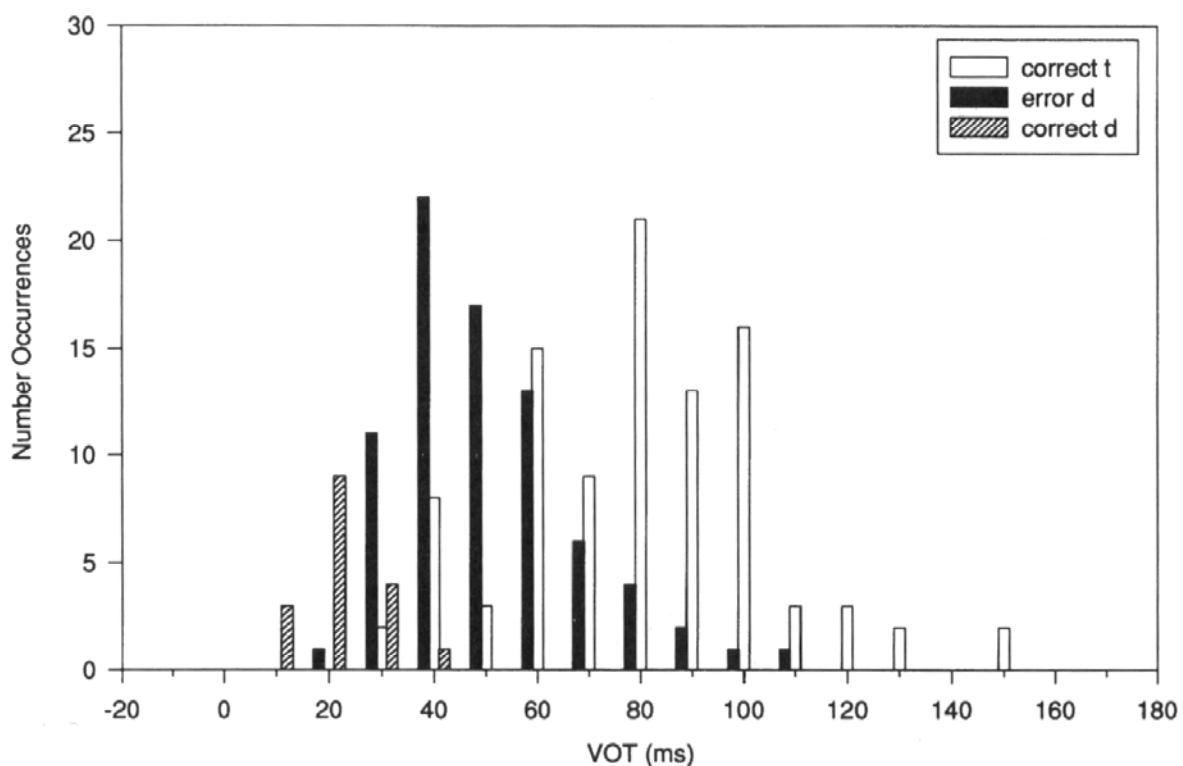


Figure 1. Subject 1: distribution of VOT values for [t] and [d].

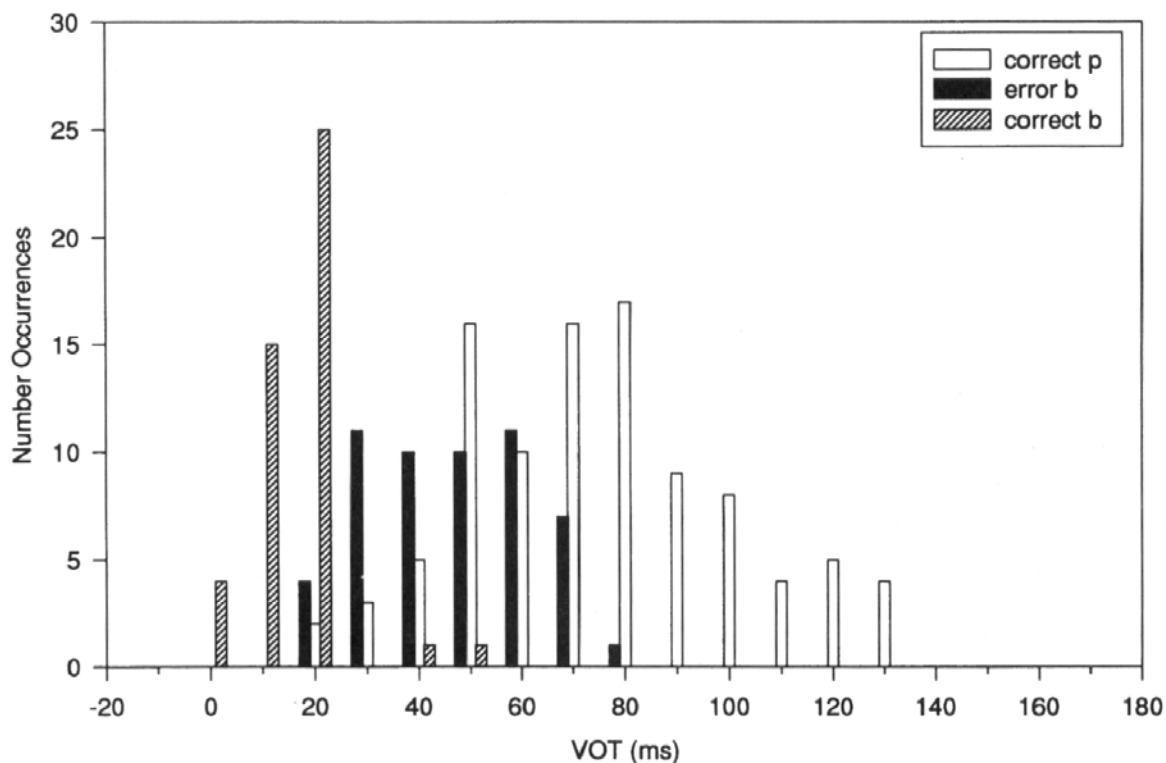


Figure 2. Subject 1: distribution of VOT values for [p] and [b].

The mean VOT values for the error sounds (voiced → voiceless) are shown in the middle row of Table 2. These values fell in the ranges expected for voiceless stops, which corresponded to our perceptions of these sounds being voiceless substitutions.

The distributions of VOT values for Subject 1's productions are represented graphically in Figures 1 through 4, with one cognate group displayed per graph.

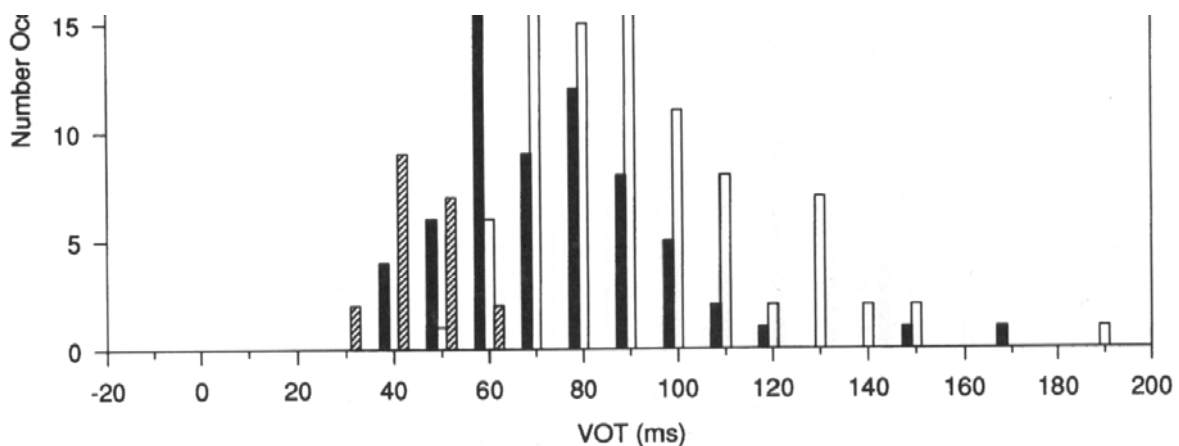


Figure 4. Subject 1: distribution of VOT values for [tʃ] and [dʒ].

For each cognate group the three categories of sounds (i.e. correct voiceless, error, correct voiced) overlapped. In all cases the distribution of error productions appeared to be shifted towards shorter lag times than the correct voiceless productions, which corresponded to the reported mean values.

Subject 1's data were subjected to one-way analyses of variance with repeated measures to determine the differences in mean VOT values across each cognate

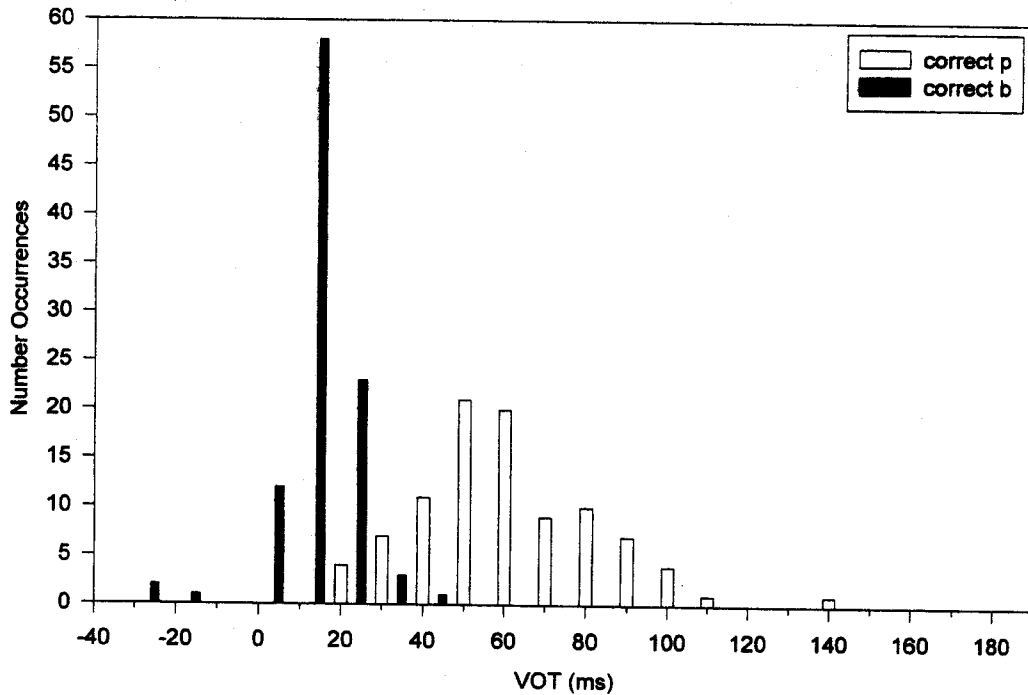


Figure 6. Subject 2: distribution of VOT values for [p] and [b].

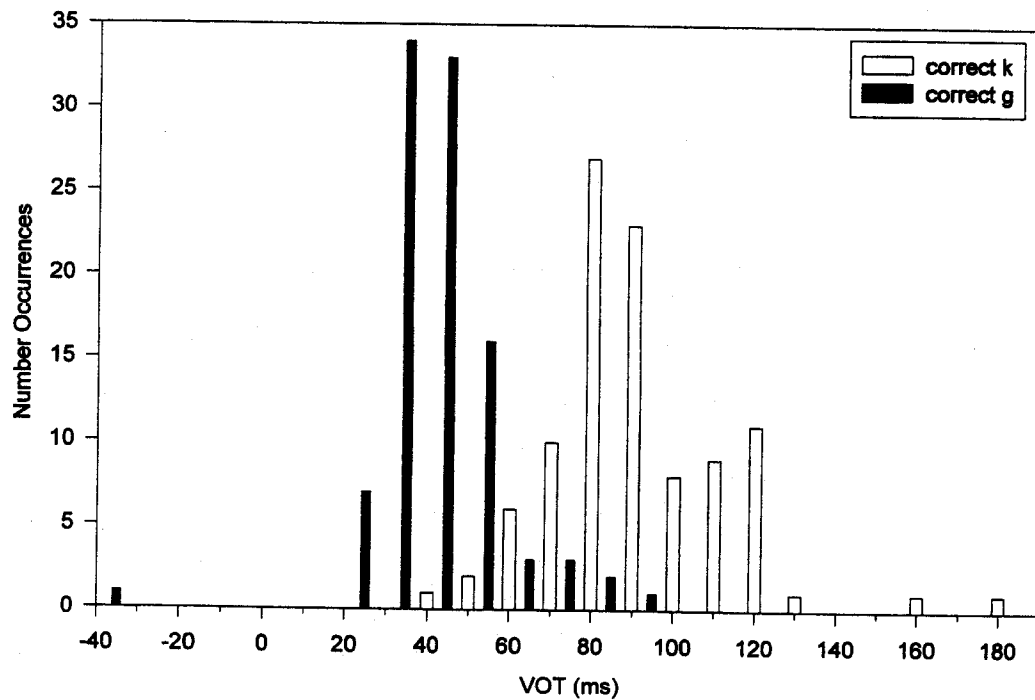


Figure 7. Subject 2: distribution of VOT values for [k] and [g].

For Subject 2, overlapping VOT values were found for all voiced/voiceless cognate pairs, with mean VOT values being significantly different between those pairs.

The standard deviations of the VOT values for both subjects fell within the ranges of reported values for normals (Shewan *et al.* 1984).

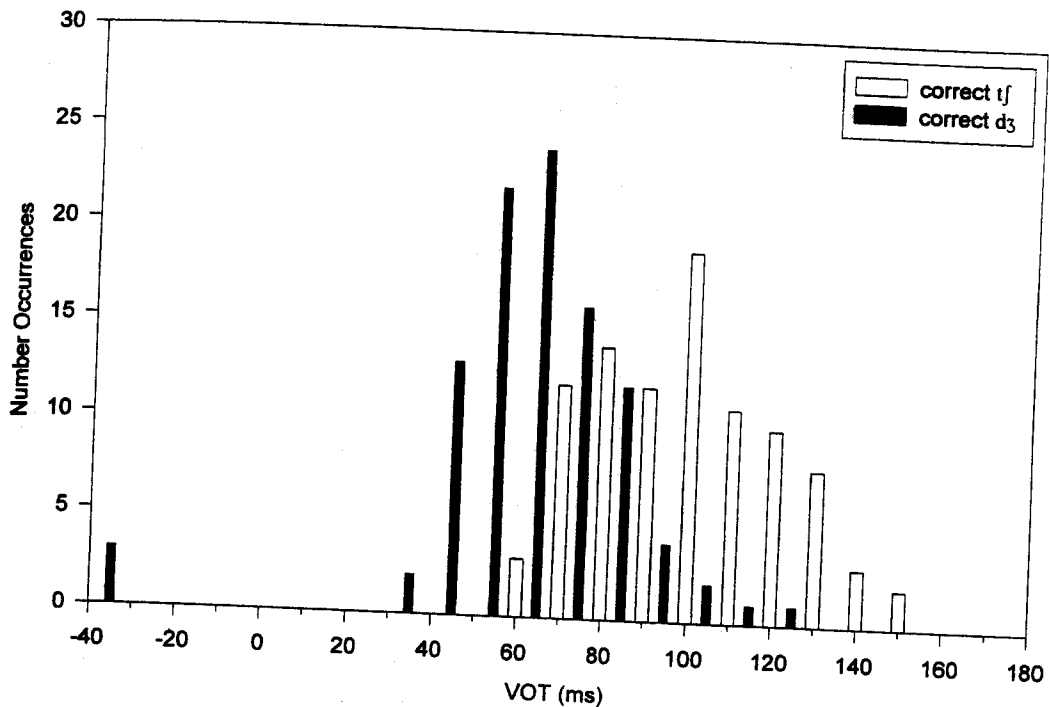


Figure 8. Subject 2: distribution of VOT values for [tʃ] and [dʒ].

Discussion

The findings from this investigation suggest that S1's sound errors (those we examined) did not stem from problems with sound selection. On the basis of this VOT analysis we can see that his incorrect productions of voiced targets were significantly different from his correct voiceless sounds. Whereas we perceived this subject to be substituting voiceless sounds for voiced, the acoustic analysis did not support our perceptions. It appeared that his erroneous productions were differentiated from their correctly produced homonyms, but at a subphonemic level.

This investigation highlights the need to analyse an apraxic/aphasic speaker's incorrect sound productions in relation to his/her correct productions. If we had examined S1's VOT data without considering correctness (i.e. voiced attempts versus voiceless attempts), we would have observed a lack of separation of discrete VOT distributions. This finding would not have been surprising and would have been predicted by previous research (Blumstein *et al.* 1980, Itoh *et al.* 1982). By further comparing S1's VOT values to those of normal speakers, as has frequently been done, we would have seen that his incorrect attempts at voiced stops usually fell within the normal range for voiceless stops, and we may have been inclined to call his errors phonemic (after Blumstein *et al.* 1980). However, by examining S1's errors in relation to his homonymous correct productions, we could see clearly that these sound groups were produced differently. It was apparent that S1 had attempted to produce his voiced and voiceless stops/affricates differentially, and had succeeded to the extent that his erroneous voiced attempts were statistically significantly different from his voiceless productions. Unfortunately these differences were not great enough to be perceived as clinically significant, phonemic differences.

Subject 2's data were collected primarily to serve as a reference for comparison

and his VOT analysis was essentially an indirect replication of previous studies. We had chosen to include S2 in this study as a direct contrast to S1, in order to illustrate the across-subject variability seen in apraxia of speech in terms of subject behaviours (i.e. S1 had obvious voicing problems, whereas S2 had no apparent difficulties with voicing). We were surprised to see the amount of overlapping of VOT values for voiced and voiceless cognates with this subject, whose productions were perceived by the authors as being accurate. The findings from this subject emphasize the importance of providing perceptual data along with acoustic findings. That is, one cannot assume that abnormal VOT distributions are indicative of perceptually inaccurate productions.

Similarly, one cannot assume that relatively normal-appearing VOT data reflect normal physiology. Like other investigators (Seddoh *et al.* 1996, Shewan *et al.* 1984), we found our subjects' VOT standard deviations to be within normal limits. We cannot infer that normal movement patterns were responsible for achieving these relatively consistent VOTs. Additionally, by grouping S1's voiced attempts into correct and incorrect categories we may have artificially reduced his VOT variability. Specifically, if S1's voiced attempts had been analysed as single groups of sounds, it is likely that his standard deviations would have been larger.

Whereas VOT is obviously an important cue to the perception of initial stop voicing, it may not always be highly predictive of such perception (Forrest and Rockman 1988). Additional factors such as fundamental frequency, F1 cutback, burst amplitude, and aspiration amplitude may also contribute significantly to perceived voicing (Forrest and Rockman 1988, Kent and Read 1992, Ohde 1985). These factors may account for our perceptions of accurate productions by S2 in the face of abnormal VOT data. They may also account for some additional seeming disparities between the subjects' VOT data and our perceptual categorizations of their sound productions. For example, the mean of S1's /d/ → [t] productions was 49.8 ms and these productions were perceived as voiceless (this perception would be predicted on the basis of normative VOT data). In contrast, the mean for S2's /d/ → [d] productions was 47.7 ms and these productions were perceived as voiced. It appears likely that factors other than, or in addition to, VOT were responsible for our perceptions of S2's [d] productions being correct. Obviously, additional acoustic analyses of both subjects' productions are warranted.

We can see from the preceding discussion that acoustic analyses, such as VOT, are currently limited in their clinical utility with respect to determining sound 'correctness'. Ultimately, perceptual judgements must serve as the basis for making clinical diagnoses of speech abnormality. However, as pointed out by Weismer (1984), 'an important part of phonological description is the contrastive function of linguistic units' (p. 50) and consistent acoustic contrasts may be evident despite a lack of perceptual contrasts. Whereas our perceptual analysis may have had more ecological validity in terms of measuring our speakers' productions in comparison to the linguistic community, our VOT analysis (and not the perceptual analysis) allowed us to see that S1 was attempting to contrast his stop/affricate productions in the direction expected by that community.

Our VOT analysis also allowed us to see that S2 probably did not achieve his voicing distinctions in the same manner as normal speakers, and because of his abnormal VOT distributions we may suspect a problem with speech motor control (especially in light of his other apraxic behaviours). However, either the problem is relatively minor in terms of voicing control, or sufficient compensations are

being made by the subject to allow perceptually accurate productions. We do not consider S2 disordered in voicing control, and would certainly not advocate the diagnosis of a motor speech disorder on the basis of these data. Further, we could not attribute these VOT deviancies to S2's apraxia of speech. Further acoustic analyses of S2's data may reveal the reasons why he sounded so normal in his stop/affricate voicing, and may aid in providing direction for treatment with S1. We would like to quote an anonymous reviewer, who stated that the findings from these subjects are interesting examples of 'struggling to be normal when one is aphasic (or apraxic)'.

In terms of treatment this analysis allows us to see that a treatment focused upon promoting correct sound selection (e.g. homonymy avoidance, perceptual training) would probably not be a suitable choice for S1. Since he was already attempting to differentiate voiced stop/affricate targets from potential homonyms, treatment designed to facilitate or maximize sound differentiation would seem to be most appropriate (e.g. reducing fundamental frequency of vowels following voiced stops, promoting prevoicing). Additionally, the use of real-time spectrography for biofeedback would intuitively appear to have utility with this type of speaker. Regardless of treatment approach, follow-up acoustic and perceptual analyses of correct productions should provide insights for treatment of apraxic/aphasic voicing errors.

Acknowledgements

This research was supported by Rehabilitation Research and Development, Department of Veterans Affairs. The authors thank Michelene Kalinyak for her assistance with data entry.

References

- BAUM, S. R., BLUMSTEIN, S. E., NAESER, M. S. and PALUMBO, C. L. (1990) Temporal dimensions of consonant and vowel production: an acoustic and CT scan analysis of aphasic speech. *Brain and Language*, **39**, 33-56.
- BLUMSTEIN, S. E., COOPER, W. E., GOODGLASS, H., STATLENDER, S. and GOTTLIEB, J. (1980) Production deficits in aphasia: a voice-onset time analysis. *Brain and Language*, **9**, 153-170.
- BLUMSTEIN, S. E., COOPER, W. E., ZURIF, E. B. and CARAMAZZA, A. (1977) The perception and production of voice onset time in aphasia. *Neuropsychologia*, **15**, 371-383.
- COMPUTERIZED SPEECH LAB 4300B [Computer hardware and software] (1994) Lincoln Park, New Jersey: Kay Elemetrics.
- DABUL, B. L. (1979) *Apraxia Battery for Adults* (Pro-Ed., Austin, TX).
- DARLEY, F. L., ARONSON, A. E. and BROWN, J. R. (1975) *Motor Speech Disorders* (W. B. Saunders, Philadelphia, PA).
- FORREST, K. and ROCKMAN, B. K. (1988) Acoustic and perceptual analysis of word-initial stop consonants in phonologically disordered children. *Journal of Speech and Hearing Research*, **31**, 449-459.
- GERMAN, D. J. (1990) *Test of Adolescent/Adult Word Finding* (DLM, Allen, TX).
- HOIT-DALGAARD, J., MORRAY, T. and KOPP, H. G. (1983) Voice onset time production and perception in apraxic subjects. *Brain and Language*, **20**, 329-339.
- ITOH, M., SASANUMA, S., TATSUMI, I. F., MURAKAMI, S., FUKUSAKO, Y. and SUZUKI, T. (1982) Voice onset time characteristics in apraxia of speech. *Brain and Language*, **17**, 193-210.
- KENT, R. D. and MCNEIL, M. R. (1987) Relative timing of sentence repetition in apraxia of speech and conduction aphasia. In J. H. Ryalls (Ed.) *Phonetic Approaches to Speech Production in Aphasia and Related Disorders* (College-Hill Press, San Diego, CA), pp. 221-242.
- KENT, R. D. and READ, C. (1992) *The Acoustic Analysis of Speech* (Singular, San Diego, CA).

- KERTESZ, A. (1982) *Western Aphasia Battery* (Grune & Stratton, New York).
- OHDE, R. N. (1985) Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children. *Journal of the Acoustical Society of America*, **78**, 1554-1561.
- PORCH, B. E. (1981) *The Porch Index of Communicative Ability* (Consulting Psychologists Press, Palo Alto, CA).
- SEDDOH, S. A. K., ROBIN, D. A., SIM, H.-S., HAGEMAN, C., MOON, J. B. and FOLKINS, J. W. (1996a) Speech timing in apraxia of speech versus conduction aphasia. *Journal of Speech and Hearing Research*, **39**, 590-603.
- SEDDOH, S. A. K., ROBIN, D. A., HAGEMAN, D., SIM, H.-S., MOON, J. B. and FOLKINS, J. W. (1996b) Temporal control in apraxia of speech: an acoustic investigation of token-to-token variability. In M. L. Lemme and D. A. Robin (Eds), *Clinical Aphasiology* (Pro-Ed., Austin, TX), pp. 65-81.
- SHEWAN, C. M., LEEPER, H. A. JR and BOOTH, J. C. (1984) An analysis of voice onset time (VOT) in aphasic and normal subjects. In J. C. Rosenbek, M. R. McNeil and A. Aronson (Eds) *Apraxia of Speech: Physiology, acoustics, linguistics, management* (College-Hill Press, San Diego, CA), pp. 197-220.
- SHRIBERG, L. D. and KENT, R. D. (1982) *Clinical Phonetics* (Macmillan, New York).
- WEISMER, G. (1984) Acoustic analysis strategies for the refinement of phonological analysis. *ASHA Monographs*, **22**, 30-52.

Appendix: Stimuli

k	g	tʃ	dʒ	p	b	t	d
cab	gab	chain	Jane	pat	bat	tab	dab
cage	gauge	char	jar	patch	batch	tad	dad
came	game	chaw	jaw	pay	bay	tale	dale
cap	gap	chaz	jazz	pear	bear	tame	dame
cash	gash	cheap	jeep	peg	beg	tear /ter/	dare
cave	gave	cheer	jeer	pen	ben	teal	deal
coal	goal	chess	jess	pest	best	tear /tir/	dear
coast	ghost	chest	just	pet	bet	ted	dead
coat	goat	chet	jet	pie	buy	teen	dean
cold	gold	chew	jew	pig	big	ten	den
come	gum	chill	jill	pill	bill	tie	die
cord	gourd	chin	gin	pole	bowl	till	dill
core	gore	chinks	jinx	poor	bore	time	dime
cot	got	chock	jock	puff	buff	tip	dip
could	good	choice	joyce	pun	bun	tock	dock
curl	girl	choke	joke	push	bush	too	due
cut	gut	chug	jug	putt	but	tore	door
kate	gate	chump	jump	pail	bail	tot	dot
kay	gay	chunk	junk	part	bart	touch	dutch
kill	gill			pan	ban	tuck	duck