Acoustic Structure of Stops Produced by Tracheoesophageal Speakers

Philip C. Doyle, Ph.D., School of Communication Sciences and Disorders, Voice Production Laboratory, The University of Western Ontario, London, Ontario.

Overview

Use of the tracheoesophageal (TE) voice prosthesis (Singer & Blom, 1980) for postlaryngectomy voice restoration has become standard practice across North America. TE voice production involves use of pharyngoesophageal muscular tissue as a vibratory source following removal of the larynx. The prosthesis permits pulmonary air to act as an aerodynamic driving source to this vicarious voicing mechanism. Tracheoesophageal (TE) speech has provided an added rehabilitation option to individuals who undergo laryngectomy . TE speech is supplied by pulmonary air, thus, distinguishing it aerodynamically from esophageal speech. Use of pulmonary air has been shown to favorably affect acoustic aspects of TE voice (Robbins, 1984). Yet concerns about the relative impact of a pulmonary air source on temporal features of TE speech have been raised (Doyle, Danhauer, & Reed, 1988; Weinberg, Horii, Blom, & Singer, 1982). It has been suggested that increased access to pulmonary air may allow the PE segment to initiate and terminate vibration more rapidly (Doyle et al., 1990; Robbins, Christensen, & Kempster, 1986), and hence, may potentially result in unique perceptual confusions (Doyle et al., 1988).

Perceptually, data have shown that TE speakers exhibit some unusual voicing patterns for cognate phonemes (Doyle et al., 1988; Doyle et al., 1990, Doyle & Haaf, 1989; Gomyo & Doyle, 1989) with a tendency toward voiceless-for-voiced cognate errors. Further, simple patterns of voice onset time do not appear to correlate well with the perceived phoneme in symmetrical CVC constructions. Thus, the purpose of this investigation was to identify and describe the temporal acoustic structure of stop production within an intervocalic stimulus context. These acoustic measures were obtained from a small group of excellent TE speakers.

Method

Three adult male TE speakers who were selected from a larger pool of 31 speakers judged to be "excellent" speakers. The larger group of speakers were initially identified and referred by experienced SLPs as being among the best TE speakers they had encountered. These 3 speakers were consistently identified by at least 2 independent judges as "superior" speakers within the group.

Speech Stimuli

The stimuli under investigation in the present work were comprised of the six English stops (/p,t,k,b,d,g/). Each stop was produced within a nonsense CVCVC construction. The stimuli were all initiated with the nasal /m/, followed by one of the vowels; the mid-consonant (one of the target stops) was then followed by the same vowel and then terminated with the nasal (Doyle et al., 1988). The nasal /m/ was chosen because of its demonstrated ease for alaryngeal speakers, and the two vowels were used because they have been shown to exhibit the greatest amplitude for esophageal vowels. Six samples of each stop with each vowel were obtained from each speaker. Stimuli were produced in the phrase " _____ is a word".

Procedure and Data Analysis

Speakers were recorded in a sound suite on research quality equipment with the microphone at a fixed distance. All speaker samples were perceptual evaluated using an open-set response paradigm. Stimuli also were acoustically analyzed using the Canadian Speech Research Environment (CSRE) software (Jamieson & Nearey, 1988). CSRE provided broadband spectrograms with concommitant amplitude displays for analysis. The temporal acoustic measure of voice onset time (VOT) was measured according to the procedures outlined by Lisker and Abramson (1967) and used by Robbins et al. (1986).

Perceptual Assessment

The entire pool of speaker stimuli were randomized and submitted to perceptual evaluations. Listeners were 10 naive listeners who had no prior exposure or experience with alaryngeal speech. Listeners were requested to transcribe their identification of the middle consonant in the stimuli. These data were then collated and placed in a confusion matrix for further evaluation of intelligibility.

Results

Perceptual Evaluation

Based on the confusion matrices generated, it was determined that 47% of the voiceless stops were correctly identified by listeners. In contrast, 79% of voice stops were correctly identified. Thus, a clear advantage in production and perception of voiced targets when compared to their voiceless cognates was observed.. This finding is inconsistent with earlier data (Doyle et al., 1988; Doyle & Haaf, 1989; Gomyo & Doyle, 1989).

Speaker 1

Speaker 1 showed consistent increases in VOT (Table 1) as loci for stops moved from front-to-back for both V+ and V- stops. This was most apparent with the vowel /i/. Relating to the voiced-voiceless distinction, Speaker 1 tended to exhibit relatively rapid VOTs (<25 msec) for the more anterior voiceless targets (/p/ and /t/).

Table 1

Means, standard deviations and ranges of VOT (in msec): Speaker 1

Stop	Vowel	M VOT	SD	Range
/p/	/i/	10.37	5.31	5.2-15.8
	/u/	10.40	2.61	8.7-13.4
/t/	/i/	25.30	6.54	17.8-29.8
	/u/	22.63	4.54	17.8-26.8
/k/	/i/	46.67	6.61	40.3-53.5
	/u/	18.43	9.26	8.9-27.4
/b/	/i/	8.03	6.78	3.3-15.8
	/u/	14.00	4.76	9.1-18.2
/d/	/i/	31.80	3.55	27.7-33.9
	/u/	17.73	6.73	13.7-25.5
/g/	/i/	46.43	9.26	37.0-55.5
	/u/	47.67	6.37	41.9-54.5

Speaker 2

Speaker 2 exhibited similar mean VOT (Table 2) patterns across stimuli, with values ranging from +30 to +60 msecs. Data also indicate that Speaker 2 increased VOT based on articulatory loci for V+ targets (front to back), but was inconsistent for V- stops. Overall, Speaker 2 exhibited VOTs >25 msec for all stops.

Table 2

Means, standard deviations and ranges of VOT (in msec): Speaker 2

Stop	Vowel	M VOT	SD	Range
/p/	/i/ /u/	52.13 39.73	14.46 5.33	43.0-68.8 35.8-45.8
/t/	/i/ /u/	33.43 50.60	3.76 6.66	30.6-37.7 45.8-58.2
/k/	/i/ /u/	60.70 51.80	8.82 13.09	54.5-70.8 42.7-66.8
/b/	/i/ /u/	31.10 31.83	5.95 5.09	25.1-37.0 28.6-37.7
/d/	/i/ /u/	35.13 33.70	9.75 3.38	26.9-45.9 29.8-35.7
/g/	/i/ /u/	51.80 41.17	10.28 1.69	40.2-59.8 39.3-42.6

Speaker 3

Speaker 3 exhibited consistent decreases in mean VOT (Table 3) for V+ stops, regardless of vowel. While fairly consistent patterns of VOT were noted for bilabial and velar cognate pairs regardless of vowel, relatively greater VOT differences were noted for alveolars..

Table 3

Means, standard deviations and ranges of VOT (in msec): Speaker 3

Stop	Vowel	M VOT	SD	Range
/p/	/i/ /u/	17.23 19.83	4.16 6.36	12.6-22.8 12.5-23.8
/t/	/i/ /u/	28.80 46.17	5.98 7.40	25.1-35.7 38.7-53.5
/k/	/i/ /u/	29.33 25.40	6.15 10.50	23.3-35.6 13.7-34.0
/b/	/i/ /u/	10.33 14.57	2.66 7.91	8.7-13.4 9.9-23.7
/d/	/i/ /u/	23.97 34.23	2.87 5.66	22.3-27.3 29.1-40.3
/g/	/i/ /u/	24.67 15.50	8.04 5.98	17.4-33.3 9.9-21.8

VOT by Temporal Cluster

In order to provide a comparative index of each speaker's productive performance, VOT data were clustered into arbitrarily selected time categories for further evaluation (0-25 msec, 26-50 msec, and 51-75 msec). The data reveal that all three speakers exhibited +VOT values (i.e., post-stop release). These values were highly individual to speakers. Speaker 1 produced a majority of stops with VOTs between 0-50 msec, Speaker 2 within the range of 26-75 msec, and Speaker 3 within the range of 0-50 msec.

Summary and Conclusions

Normal English speakers have been shown to systematically vary VOT by increasing durations as the stop moves from labial to velar loci (Lisker & Abramson, 1967). This variation has been noted with excellent esophageal speakers. Speaker 1 did exhibit this pattern for all V+ stops in both intervowel contexts. However, for V- stops this pattern was only noted with the /u/ vowel. Speaker 3 exhibited this pattern for both V+ and V- in the /i/ vowel context, but was inconsistent with /u/.

When the temporal cluster data were inspected, these speakers exhibited idiosyncratic patterns which are likely related to unique postsurgical productive systems. It should be noted that all speakers produced V+ stops with positive VOTs. This is likely due to the effects of a powerful aerodynamic system on the PE segment. Further, these TE speakers produced unique temporal clusters in relation to VOT. Again, this may be due to aerodynamic influence, postsurgical anatomy, or both. Although limited, the present data suggest that multiple acoustic cues likely signal stop perception for TE speakers. Additional research in our laboratory suggests that proficient TE speakers are able to effectively signal voicelessness, even when acoustic data reveal that no break in voicing exists. This raises questions regarding the use of air in the vocal tract and upper airway turbulence as a compensatory mechanism in these speakers. This finding and the present data suggest that the proficiency of TE speech may be related how the speaker is able to utilize the air source once it trangresses the pharyngoesophageal segment (the muscular tissue which serves as an alaryngeal voice source). Time varied features of stops are worthy of further inquiry (Kewley-Port, 1993).

Previous research with normal speakers has shown that they vary VOT to distinguish prevocalic V- stops from V+ cognates. Speaker 1 did not effect this distinction. In most cases, his VOTs for stops within cognate pairs were similar. This suggests a restricted VOT range. Although Speaker 2's VOTs were produced within a rather narrow range and were always >25 msec, all V- stops except one were produced with VOTs greater than those noted for V+ cognates. It appears that Speaker 2 attempted to make this distinction, but as a result of physiologic constraints, could only do so within a restricted VOT range. Subject 3 produced V- stops /t/ and /k/, but not /p/ with overall VOTs >25msec. Thus, inconsistency in relation to articulatory loci was again noted. This finding further confirms the idiosyncratic patterns that may characterize many TE speakers.

References

Doyle, P.C., Danhauer, J.L., & Lucks, L.E. (1990). SINDSCAL analysis of perceptual characteristics of consonants produced by esophageal tracheoesophageal talkers. Journal of Speech and Hearing Disorders, 55, 756-760.

Doyle, P.C., Danhauer, J.L., & Reed, C.G. (1988). Listeners' perceptions of consonants produced by esophageal and tracheoesophageal talkers. Journal of Speech and Hearing Disorders, 53, 400-407.

Doyle, P.C., & Haaf, R.G. (1989). Pre-and post-vocalic consonant intelligibility in tracheoesophageal talkers. Journal of Otolaryngology, 18, 350-353.

Doyle, P.C., Swift, E.R., & Haaf, R.G. (1989). Effects of listener sophistication on judgments of tracheoesophageal talker intelligibility. Journal of Communication Disorders, 22, 105-113.

Gomyo, Y., & Doyle, P.C. (1989). Perception of stop consonants produced by esophageal and tracheoesophageal speakers. Journal of Otolaryngology, 18, 184-188.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. Journal of the Acoustical Society of America, 72, 322-335.

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

Robbins, J. (1984). Acoustic differentiation of laryngeal, esophageal, and tracheoesophageal speech. Journal of Speech and Hearing Disorders, 27, 577-585.

Robbins, J., Christensen, J., & Kempster, G. (1986). Characteristics of speech production after tracheoesophageal puncture: Voice onset time and vowel duration. Journal of Speech and Hearing Research, 29, 499-504.

Singer, M.I., & Blom, E.D. (1980). An endoscopic technique for restoration of voice after laryngectomy. Annals of Otology, Rhinology, and Laryngology, 89, 529-533.