A Role for the Larynx in Contextual VOT Variation in American English?

Julie Larson
University of California, Berkeley, USA
jlarson@socrates.berkeley.edu

ABSTRACT

Different glottal gestures regulate contrasts between voiced, unaspirated, and aspirated stops in many languages. But what about contextually-predictable variations in VOT, such as longer VOT before close vowels and sonorant consonants than before open vowels: are aerodynamic factors responsible, as has been hypothesized, or could different glottal gestures be the cause? Glottal aperture was monitored in two adult male American-English speakers using photoglottography. The results give no indication that glottal gestures vary systematically with these contextual VOT variations, thus indirectly supporting the aerodynamic hypothesis.

1. INTRODUCTION

Different laryngeal gestures often underlie contrasting stop series in a language. The presence, timing, and relative aperture of a glottal opening and closing gesture, controlled in large part by activity of the posterior cricoarytenoid muscle, is a crucial component in the production of distinct aspirated and unaspirated stops in numerous languages [1, 2]. Löfqvist [2], for example, found that for Swedish the location of peak glottal opening (PGO) with reference to the closure or release of a stop (/t/ in his study), differed appreciably for aspirated versus unaspirated stops. For the unaspirated stops, PGO occurred nearer and was time-locked to the *closure* of the consonant, whereas for the aspirated stops, PGO occurred nearer and was time-locked to the *release* of the consonant. Based on these measurements (closure to PGO or PGO to release), the two classes of sounds fell clearly into separate categories. Additionally, Löfqvist collected electro-myographic data which showed that increases in glottal opening are accompanied by increases in posterior cricoarytenoid activity. Thus, it seems clear that variation in VOT as a mechanism for distinguishing contrastive stop series is to a large extent controlled by the speaker.

The question remains, however, of whether the contextual variation in VOT *within* the class of (intentionally) aspirated or unaspirated stops in a language is also controlled by the speaker. It is generally accepted, for example, that within the class of voiceless aspirated stops, VOT varies depending on both the identity of the stop itself and the identity of the sonorant or vowel that follows it. Specifically, VOT tends to be significantly longer following velar stops than following alveolars or bilabials [3] and VOT tends to be significantly longer before sonorants and close vowels than before open vowels [4].

One hypothesis for the *contextual* variation in VOT is that it could, indeed, be purposeful on the part of the speaker in order to help the listener perceive the correct sound. Another hypothesis for the contextual variation is that it is largely mechanical [5]. Because a certain pressure drop across the glottis is required for voicing, it is reasoned that VOT must be directly affected by the amount of time it takes for air that accumulated in the oral cavity during the stop closure to exit so that a pressure differential across the glottis can be restored. Since higher, or closer, vowels leave a smaller opening through which air can escape, it takes longer for the oral pressure to be released and, thus, longer for voicing to commence for those sounds.

Attempts to find experimental evidence in support of the mechanical account of vowel-dependent VOT have tested how VOT varies with oral pressure decay [6], as well as how it is affected by nasalization of following vowels [7], with mixed results. An important assumption that underlies these and other studies of vowel-dependent VOT variation is that the timing of the laryngeal gesture responsible for aspiration is held constant. In other words, in order to interpret results obtained in the acoustic and/or aerodynamic domain as evidence for a purely mechanical account of VOT variation, one must assume that adjustments are not also made in the articulatory domain.

This study seeks to investigate whether or not speakers of American English adjust the timing of the glottal opening gesture required for aspirated stops depending on the following vowel or sonorant.

2. EXPERIMENTAL DESIGN

2.1. METHOD

To study variations in glottal aperture photoglottography (PGG) was used simultaneously with a standard audio recording. PGG places a light source and light sensor on opposite sides of the glottis such that the detected variations in light flux are related to variations in glottal area. Fig. 1 gives a sample record accompanied by the usual acoustic displays. While the technique cannot be calibrated for glottal area it provides an excellent record of the timing of variations in glottal area [8].

ISBN 1-876346-48-5

2.2. MATERIALS

A photoelectric glottograph was made in-house. A miniature light sensor was inserted in a 46 cm long transparent plastic medical tube, 5 mm outer diameter, leaving approximately 15 cm of tubing beyond the sensor. The output of the light sensor was connected to a PCQuirer® DC input channel feeding a computer running the PCQuirer® data acquisition program. Subjects inserted the PGG themselves via the nose until the light sensor was positioned optimally in the pharynx and the end of the 15 cm portion was swallowed into the esophagus for stabilization. A flashlight held against the neck just below the larynx served as the light source. A noise canceling headset microphone was used for the acoustic recording.

A word list was designed to place sonorants and vowels of varying closeness after a voiceless aspirated stop. The following sextuplet was used: /tid/, /tud/, /tæd/, /tɑd/, /twɑd/, /tɹɑd/. The stimuli were arranged in 10 independently randomized orders then printed in orthography on a single page with the carrier phrase "Say ___ again." above each set.

2.3. SUBJECTS

Subjects were native-English-speaking adults from the Linguistics Department at UC Berkeley. Of several volunteer subjects only two males succeeded in swallowing the end of the PGG tube into the esophagus. Because stabilizing the position of the light sensor was considered essential for obtaining consistent readings, no data were collected from the other volunteers. The two male subjects were tested individually. After achieving the proper positioning of the PGG, each subject was presented with the page containing the randomized lists and asked to read the stimuli (in the carrier phrase) aloud.

2.4. DATA COLLECTION & MEASUREMENT

The audio recording was sampled at 11,000 Hz; the PGG signal was sampled at 1375 Hz. The figure below shows a typical synchronized waveform, photoglottogram, and spectrogram from Speaker 1.

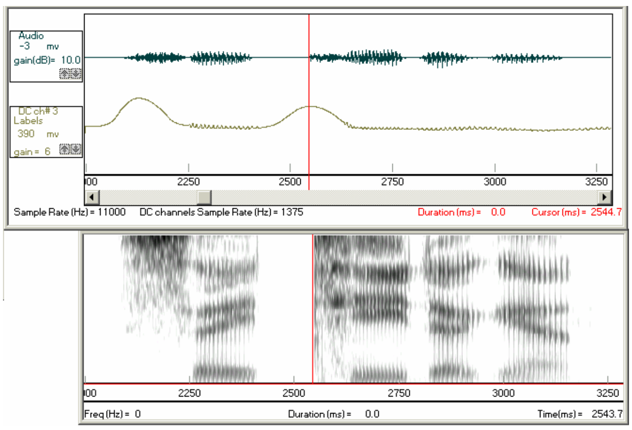

Figure 1: Synchronized waveform, glottogram, and spectrogram from Speaker 1 saying "Say teed again."

Note that glottal opening gestures are clear for both [s] and [t^h] (the opening gesture for breathing, although not shown here, was also easily identifiable where present). Note also that, in addition to the gross opening and closing gestures, vocal cord vibration can be seen in the PGG signal; this was seldom the case for Speaker 2. The cursor is placed at the /t/ release to show its temporal alignment with the PGO.

The following measurements were logged for each token: 1) onset of /t/ closure, 2) /t/ release, 3) onset of voicing as determined from the waveform and spectrogram, 4) peak glottal opening, and 5) degree of glottal opening as reflected by the maximum PGG amplitude in mv. From measurements 1–4, others, like VOT and the temporal distance from PGO to consonant release, could be calculated. Measurements of the relative area of glottal opening were considered; however, due to the difficulty of maintaining a consistent distance between the light source and light sensor, area measurements may not be reliable.

3. RESULTS & DISCUSSION

For a small number of tokens, the PGG signal was saturated. Those tokens are not included in the analysis, resulting in less than 10 tokens for some stimuli. Speaker 2 re-read the first randomized set at the end of his recording session, resulting in 11 tokens for some stimuli. Means and standard deviations for all measurements by subject are available upon request.

3.1. CONTEXT-DEPENDENT VOT

As a preliminary, the data were analyzed to determine whether or not the expected pattern of higher VOT before close vowels and sonorants than before open vowels was exhibited. The results, illustrated in Figure 2, show some deviation from expectations, particularly for the open vowels. Both subjects had slightly longer average VOT before the open vowel /æ/ than before the close vowels /i/ and /u/, and Speaker 2 had longer average VOT before /ɑ/ than any other vowel.

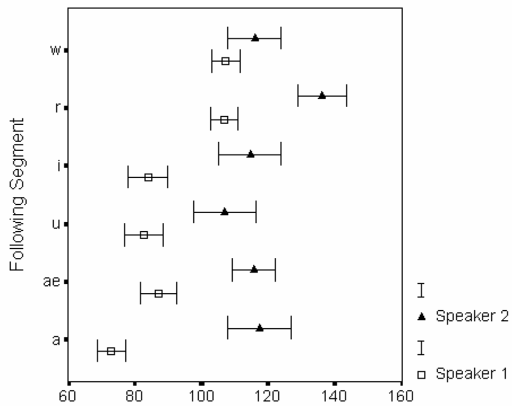

Figure 2: Mean VOT (ms) by following segment +/- 2 SE

At the same time, the error bars in Figure 2 show that few of the differences in VOT by vowel are statistically significant. It may be that to observe a significant effect for this low level variation would require collection of many more than 10 samples per item. Pooling the data for the two subjects to result in 20 tokens does not help in this case, since their individual VOT ranges were so different. On the other hand, Klatt found a significant 15% difference ($p<.01$) between VOT preceding close vowels /i, u/ verses open vowels /ɛ, ai/ based on averages of 30 tokens which had been pooled from three subjects and three consonant places [4]! Actually, for Speaker 1 in this study, one might argue that the sonorants, high vowels, and /ɑ/ fall into three distinct categories as expected. The lax vowel /æ/ calls for further investigation.

3.2. TIMING OF PEAK GLOTTAL OPENING

Because this study concerned aspirated stops, it was expected that a glottal opening and closing gesture would be observed, and that its "peak" (the point at which the gesture changes from abduction to adduction) would be timed to occur closer to the release of the oral constriction than to its closure. This was, indeed, the case, as shown schematically in Figures 3 and 4. In fact, in most cases, the peak glottal opening (PGO) occurred *after* the oral release.

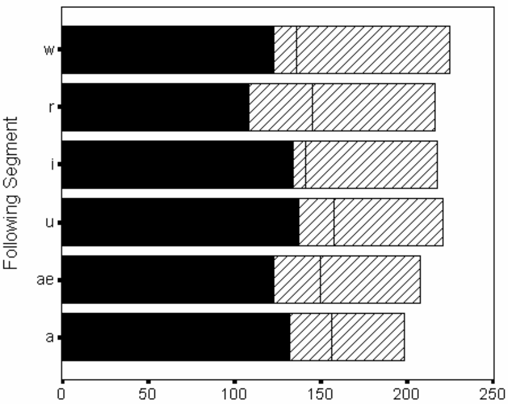

Figure 3: Average timing (ms) of closure (solid), PGO (vertical line), and aspiration (hatched) sequence; Spkr 1

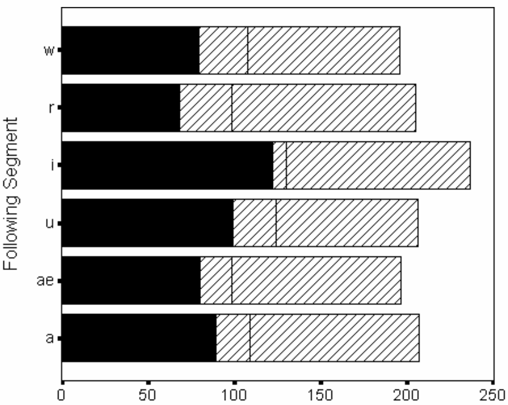

Figure 4: Average timing (ms) of closure (solid), PGO (vertical line), and aspiration (hatched) sequence; Spkr 2

Turning to the primary question being investigated here—whether or not speakers systematically vary the timing of the glottal opening gesture—the PGO lag time was analyzed according to following segment. Figure 5 shows the average PGO lag time by following segment and speaker. The zero line corresponds to the point of consonant release.

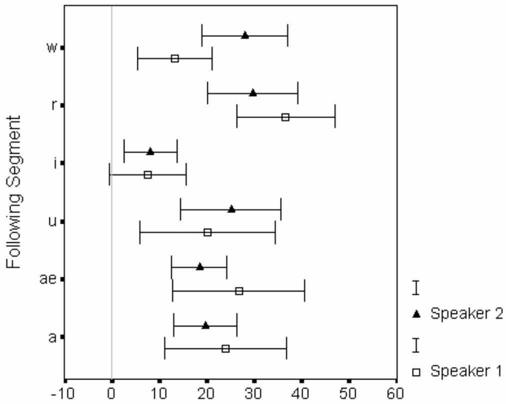

Figure 5: Mean PGO lag time (ms) by following segment +/- 2 SE

The means in Figure 5 provide no evidence that the timing of the PGO is systematically varied according to the following segment. For Speaker 2, the only context for which the timing is significantly different is before /i/; however, this difference is in the opposite direction to that which would be expected if the gesture were timed to enhance the characteristically long VOT for that high close vowel. For Speaker 1, the PGO lag time is substantially longer for /ɹ/, than for /w/ and /i/; however, if one recalls that the VOT for /w/ and /r/ were virtually the same for that subject, it is clear that alignment of PGO with respect to the consonant release is not responsible for the VOT variation.

3.3. PGO TIMING AND VOT

Although the VOT was not always found to vary with vowel closeness as expected, it was nonetheless deemed worthwhile to see how the VOTs that were exhibited by the two speakers covaried with PGO delay. Only very low and non-significant correlations were found between VOT and PGO delay for each speaker (Pearson's $r(55)=.176$ $p=.19$ and $r(59)=.136$ $p=.297$ for Speakers 1 & 2, respectively). Figure 6 shows the scatterplot with regression lines.

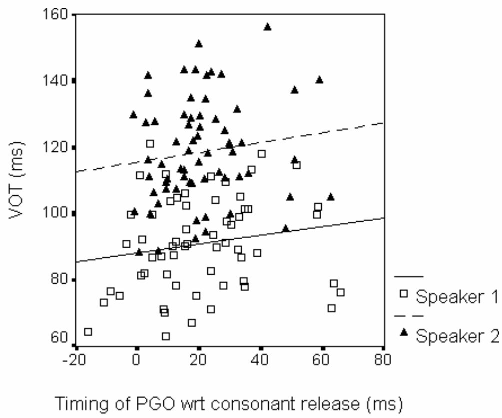

Figure 6: VOT by PGO delay; Speaker 1

ISBN 1-876346-48-5 © 2003 UAB

3.4. DEGREE OF GLOTTAL OPENING

Degree of glottal opening was considered as another factor that speakers might manipulate in order to vary VOT. However, as was the case for the *timing* of the glottal opening gesture, analysis of the data showed no evidence that the *degree* of opening was varied systematically according to following segment by either speaker.

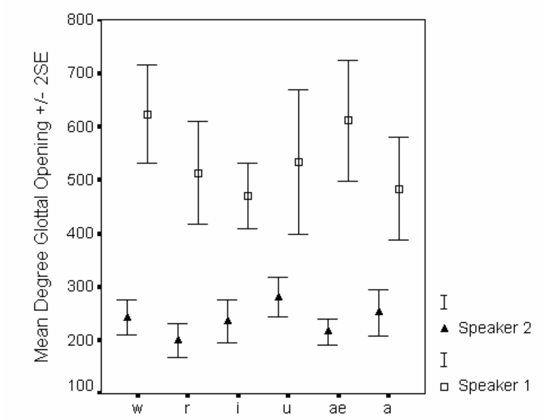

Figure 7: Degree of glottal opening (mv) by following segment +/- 2SE

Again, analysis of the data showed no significant correlation between the glottal gesture (its degree of opening in this case) and VOT (Pearsons r(55)=.167 p=.216 and r(59)=-.005 p=.972 for Speakers 1 & 2, respectively). Figure 8 shows the scatterplot with regression lines.

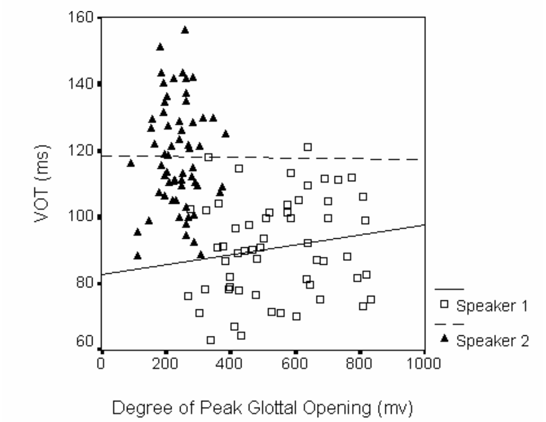

Figure 8: VOT by Degree of Glottal Opening; Speaker 1

It should be noted that, while the *timing* data collected by PGG is considered highly reliable, the "degree of glottal opening" data in Figures 7 and 8 must be interpreted with caution. First, one cannot be certain whether the noticeably higher overall magnitude of glottal opening exhibited by Speaker 1 is due to actual wider opening, to closer relative placement of the light source and sensor, or to a combination of these factors. Second, the variability in degree of glottal opening exhibited by both speakers may represent actual variability in glottal aperture as well as minor flashlight and/or light sensor movements during data collection.

4. SUMMARY

No evidence was found to indicate that speakers make consistent, fine-tuned adjustments to the timing of the peak glottal opening for aspirated stops according to the following segment. The timing differences that were found did not correspond to differences in VOT duration. At this low level, there was no correlation between the timing of the PGO and the duration of VOT. Nor was there a correlation between the degree of glottal opening and VOT, although that finding may be due to the data collection method. VOT before the vowel /æ/ was unexpectedly long for both speakers, warranting further investigation

ACKNOWLEDGEMENTS

I am grateful to John Ohala and Ian Maddieson for advising this project and providing extensive comments on this paper; and to Ronald Sprouse for building the PGG device.

REFERENCES

[1] H. Hirose, Y. Hirohide, and S. Niimi, "A cross language study of laryngeal adjustment in consonant production," *Annual Bulletin of the Research Institute in Logopedics and Phoniatrics*, vol. 12, pp. 61–72, 1978.

[2] A. Löfqvist, "Interarticulator programming in stop production," *Journal of Phonetics*, vol. 8, pp. 475–490, 1980.

[3] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," *Word*, vol. 20 (3), pp. 384–422, 1964.

[4] D. H. Klatt, "Voice onset time, frication, and aspiration in word-initial consonant clusters," *Journal of Speech and Hearing Research*, vol. 18, pp. 686–706, 1975.

[5] J. J. Ohala, "A model of speech aerodynamics," *Report of the Phonology Laboratory* (Berkeley), vol. 1, pp. 93–107, 1976.

[6] S. S. Chang, "Vowel dependent VOT variation," *Proceedings of the XIVth International Congress of Phonetic Sciences* (Volume 2), pp. 1021–1024, San Francisco, 1999.

[7] J. Larson, "The Effect of Nasalization on Voice Onset Time: Data from English, French, and Hindi," Unpublished Linguistics 110 term paper, University of California Berkeley, Fall 2000.

[8] A. Löfqvist, "Laryngeal activity in Swedish obstruent clusters," *Journal of the Acoustical Society of America*, vol. 68(3), pp. 792–801, 1980.