

Modified locus equations categorize stop place in a perceptually realistic time frame

Ariane E. Rhone^{a)} and Allard Jongman

*Department of Linguistics, University of Kansas, 1541 Lilac Lane, Lawrence, Kansas 66044
ariane-rhone@uiowa.edu, jongman@ku.edu*

Abstract: Slope and y -intercepts of locus equations have previously been shown to successfully classify place of articulation for English voiced stop consonants when derived from measurements at vowel onset and vowel midpoint. However, listeners are capable of identifying English voiced stops when less than 30 ms of vowel is presented. The present results show that modified locus equation measurements made within the first several pitch periods of a vowel following an English voiced stop were also successful at classifying place of articulation, consistent with the amount of vocalic information necessary for perceptual identification of English voiced stops /b d g/.

© 2012 Acoustical Society of America

PACS numbers: 43.70.Fq, 43.70.Mn [AL]

Date Received: March 19, 2012 Date Accepted: April 30, 2012

1. Introduction

One of the major goals of phonetic research is to define the relation between the speech signal and phonetic categories. For more than 50 years now, researchers have been debating how the rapidly changing speech stream is mapped onto discrete categories or features. The correspondence between a physical property in the acoustic signal and a phonetic feature is complicated due to individual variation caused by differences in size, age, and gender of a speaker, to name but a few sources of potential variability. The greatest challenge in this domain has been to find stable acoustic properties that correspond to place of articulation in stop consonants, which are transient in nature and heavily influenced by their surrounding vowel contexts.

Early research on the acoustic characteristics of stop consonants suggested that differences in speaker, phonetic context, and speaking rate caused so much variability that consistent acoustic cues corresponding to place of articulation could not be found.^{1,2} Subsequent research suggested that instead of focusing on individual acoustic properties of stop consonants such as burst frequency and formant transitions from a stop into a following vowel, invariant acoustic cues reside in the gross shape of the onset spectrum of a stop-vowel syllable.³ This static measure was later abandoned for a more dynamic approach whereby the way in which the distribution of energy changed from the stop consonant to the following vowel was thought to provide invariant cues to place of articulation.⁴

An alternative approach to the “invariance problem” is the locus equation, which may serve as a source of relational invariance. This metric was originally described by Lindblom⁵ and has been more recently utilized by Sussman and colleagues⁶⁻⁸ to successfully classify word-initial stop consonants across languages and speaking styles, and the basic findings have been replicated across laboratories. While interpretation of the locus equation results remains open for debate (for example, see commentary to Sussman;⁷ also, see Iskarous⁹ for the view that locus equations index degree of and resistance to coarticulation), locus equations clearly distinguish place of

^{a)}Present address: Department of Neurosurgery and Department of Psychology, University of Iowa, E11 Seashore Hall, Iowa City, Iowa 52242.

articulation in stop consonants in English, as well as a variety of other languages.¹⁰ Importantly, locus equations take advantage of the variation in acoustic information present in stop consonant transitions across vowel contexts, rather than being hindered by this variability. Locus equations are able to utilize this variability because they are derived from measurements of the second formant (F2) taken at two points in the vowel following a stop consonant—at the onset of the vowel, and at the “midvowel nucleus” (MVN) which is obtained from the midpoint of the vowel (60–110 ms post-onset). This type of dynamic metric may provide a more complete picture of the acoustic realization of stop consonant-vowel sequences.

For each place of articulation, F2vowel-by-F2onset points are plotted from many vowel contexts, thus capturing the lawful variation in stop consonant acoustics across highly variable environments, arising from both talker-specific factors (e.g., gender, speaking rate) and coarticulation. A linear regression is performed to obtain slope and *y*-intercept values of the line best fit to the data points for each place of articulation, and coefficient values for each place are then compared to determine whether they differ significantly. Slope and *y*-intercept coefficients are then submitted to a discriminant analysis to determine how well the locus equation values correctly classify each place of articulation. Combined slope and *y*-intercept values of this traditional locus equation approach can successfully classify place of articulation at high rates (up to 100% correct classification for labial/alveolar/velar distinctions⁶), suggesting that the locus equation is a potentially viable metric for stop place categorization.

In addition to finding a robust metric, a perceptually relevant account of stop place categorization should incorporate and match what is known about human stop perception. Perceptual studies of stop consonants suggest that a measurement spanning 60–110 ms after the release burst may be an inefficient description of the perception process.^{11,12} A variety of studies have shown that listeners perceive place of articulation on the basis of a much shorter stretch of the speech signal. Accurate perception of place of articulation for English voiced stops seems to require as little as the initial 10–20 ms of stop-vowel syllables.¹³ Resolving the discrepancy between the lengthy temporal interval needed to derive traditional locus equations and the shorter interval found to be sufficient by perceptual studies might lead to further understanding of the cues to stop place categorization as they more realistically correspond to perception and might make locus equations a more realistic candidate to capture these cues. The present research aims to investigate the effectiveness of locus equations derived from shorter temporal intervals between the F2onset and F2vowel measurement points in the classification of English voiced stops /b d g/. The question addressed is whether locus equations derived over a time span that matches that used by human listeners can uniquely distinguish the three places of articulation.

2. Methods

Twenty adult speakers (10 female) were recruited from the University of Kansas community. All speakers were native speakers of American English, reported no speech or hearing impairments, and were compensated or earned course credit for their participation. Participants were recorded using an ElectroVoice RE 20 microphone and Fostex DAT recorder in an anechoic chamber (IAC). The DAT recording was digitized in PRAAT¹⁴ into .WAV files at a sampling rate of 11 025 Hz.

Subjects were recorded speaking CVC syllables at a comfortable rate in the carrier phrase “Say CVC again.” CVC syllables consisted of one initial voiced stop consonant /b d g/, one of ten medial vowels /i ɪ eɪ ε æ ʌ ɔ ou u/, and word-final voiceless alveolar stop /t/. Tokens were randomized and repeated five times, for a total of 150 utterances per subject.

All formant measurements were taken from a wideband spectrogram display generated with PRAAT, with a 6 dB pre-emphasis to enhance higher formant intensities.⁶ As in traditional locus equation measurements, formant values for F2 onset (the Y coordinate of the locus equation scatterplot) for each token were obtained by

recording the F2 value at the first discernible pitch pulse following the release burst. In order to systematically shorten the time frame between onset and vowel measurements, multiple F2vowel (X coordinate) measurement points were used. These measurement points were taken at each pitch period up to and including the MVN. Scatterplots were then generated separately for each pitch period. Because pitch periods correspond to the fundamental frequency of an individual speaker's voice (and are therefore variable across speakers), these measurements reflect a relative progression through the vowel transition. Pitch period durations were estimated by measuring each token's average pitch period up to the time of the MVN for general comparison with previous studies of stop consonant perception.

3. Results

Plots from all speakers showed data points tightly clustered around the regression line, as reported in previous locus equation studies.⁶ Linear regression functions were fit to obtain locus equation slope and y -intercept coefficients for each place of articulation for each measurement point. Average slope values were lower than those reported by Sussman.⁶ However, the overall pattern is consistent: /b/ has the highest slope and lowest y -intercept, /d/ has the lowest slope and highest y -intercept, and /g/ has an intermediate slope and y -intercept.

One-way analyses of variance were performed on slope and y -intercept values obtained from the linear regressions. A significant main effect of place of articulation was found for slope $F(2,57)=42.474$ ($p < 0.001$) and y -intercept $F(2,57)=75.799$ ($p < 0.001$). Scheffé *post hoc* tests showed significant differences ($p < 0.05$) for slope comparisons of /b/ vs /d/ and /g/ vs /d/; comparison of /b/ vs /g/ showed a trend ($p = 0.104$). All y -intercept comparisons were significant ($p < 0.05$). See Fig. 1 for scatterplots for all speakers for each place of articulation.

Discriminant analyses were performed to assess the locus equation coefficients as predictors of stop place category. The leave-one-out method of cross validation was used; each slope/ y -intercept pair was tested against all cases other than that case. Correct classification was 88.3%, which while lower than that reported by Sussman⁶ is very high and well above chance probability (33.3%). In comparison, discriminant analysis performed with F2onset and F2vowel (at MVN) as predictors resulted in 71.6% overall correct classification, supporting the result reported by Sussman⁶ that the higher-order locus equation coefficients categorize place of articulation more successfully than token-level formant values.

To increase statistical power in analyzing the modified locus equation coefficients, the pitch periods corresponding to roughly the same temporal interval from vowel onset for males and females were combined. The average pitch period for female speakers was 4.79 ms and that for male speakers was 9.03 ms. In our analysis, "time point 1" combines male pitch period 1 with female pitch period 2; "time point 2"

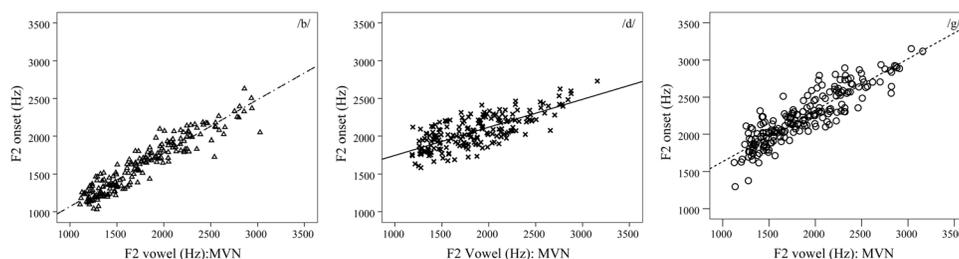


Fig. 1. Locus equation scatterplots obtained from traditional midvowel nucleus (MVN) for all speakers, for bilabial (left panel), alveolar (middle panel), and velar (right panel) stop consonants. One point represents one speaker's average F2 value for all five repetitions of one CV type (e.g., mean F2 values for five repetitions of /bu/ for speaker 1).

refers to male pitch period 2 combined with female pitch period 4, etc. Results from the first three time points (corresponding to approximately 25–30 ms of F2 transition) are discussed.

Coefficient patterns were consistent with the results of traditional locus equation measurements: /b/ was characterized by the highest slope and lowest intercept, /d/ had the lowest slope and highest intercept, and /g/ had slope values close to those for /b/ but with a higher intercept. As the distance between F2 onset and F2 vowel measurement points increased, slope coefficients decreased and y -intercept values increased for all three places of articulation.

Significant main effects were found for slope $F(2,57)=26.340$ ($p < 0.001$) and y -intercept $F(2,57)=37.693$ ($p < 0.001$) coefficients derived as early as the first time point (corresponding to approximately 10 ms from vowel onset). Similar results were found at the second time point [slope: $F(2,57)=28.061$, $p < 0.001$; y -intercept: $F(2,57)=44.557$, $p < 0.001$] and third time point [slope: $F(2,57)=28.964$, $p < 0.001$; y -intercept: $F(2,57)=50.543$, $p < 0.001$]. Scheffé *post hoc* tests showed significant slope differences for pairs /b/ vs /d/ and /d/ vs /g/ ($p < 0.05$); all y -intercept pairs were significant ($p < 0.05$). This pattern is consistent with the results of the traditional locus equation measurements described above, and was maintained for the second and third time points (see Fig. 2).

Discriminant analyses were performed as described above for each time point analyzed, and resulted in correct classification rates of 81.7%, 83.3%, and 83.3% for time points 1, 2, and 3, respectively. As with traditional locus equation analysis, these values were consistently higher than the classification results using F2 onset and F2 vowel as predictors (correct classification results of 76.2%, 76.4%, and 75.2% for the first three time points).

Locus equation coefficients obtained from the traditional MVN measurement point were compared with the corresponding coefficients obtained from each earlier F2vowel time point (e.g., each speaker's /b/ slope at MVN vs their /b/ slope at time point 2). As expected, significant differences were shown for all time points relative to MVN (Wilcoxon signed-rank test, $p < 0.001$ for all comparisons), suggesting that the relative difference between coefficients for each place of articulation may be more informative than absolute slope or y -intercept values.

4. Discussion

The present results show that locus equations derived from measurement points spanning perceptually realistic durations of second formant transition can be used to successfully categorize place of articulation in English voiced stop consonants. The initial tens of milliseconds of vowel transition provide sufficient information for listeners to identify voiced stop sounds, and our modified locus equation method results in successful categorization over comparable durations.

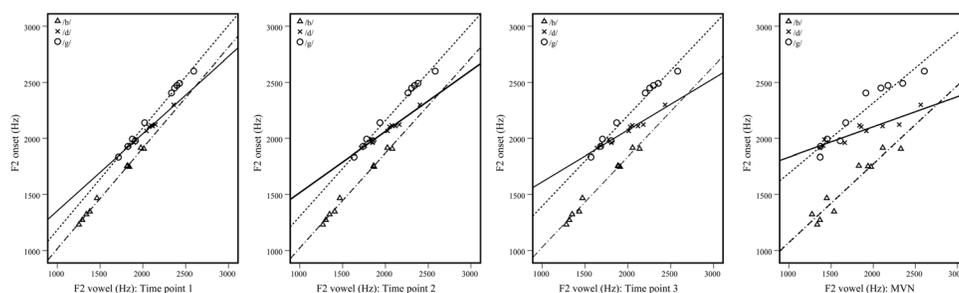


Fig. 2. Group mean locus equation scatterplots obtained from the first three time points (corresponding to approximately 9, 18, and 27 ms into the vowel, respectively) and MVN. Each point represents the average value for one CV type (e.g., mean F2 values for all /bu/ tokens) across all speakers (male and female).

Significant differences in slope and y -intercept—combined with high classification rates from discriminant analyses using these coefficients as predictors—suggest that the locus equation can be computed on a time scale that is commensurate with human perception of place of articulation.

Acknowledgments

Partial support for this work was provided by a University of Kansas Honors Program Undergraduate Research Award (to A.E.R.).

References and links

- ¹A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the speech code," *Psychol. Rev.* **74**, 431–461 (1967).
- ²A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition* **21**, 1–36 (1985).
- ³K. N. Stevens and S. E. Blumstein, "Invariant cues for place of articulation in stop consonants," *J. Acoust. Soc. Am.* **64**, 1358–1368 (1978).
- ⁴A. Lahiri, L. Gewirth, and S. E. Blumstein, "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study," *J. Acoust. Soc. Am.* **76**, 391–404 (1984).
- ⁵B. Lindblom, "On vowel reduction," Report No. 29, The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden (1963).
- ⁶H. M. Sussman, H. A. McCaffrey, and S. A. Matthews, "An investigation of locus equations as a source of relational invariance for stop place categorization," *J. Acoust. Soc. Am.* **90**, 1309–1325 (1991).
- ⁷H. M. Sussman, D. Fruchter, J. Hilbert, and J. Sirosh, "Linear correlates in the speech signal: The orderly output constraint," *Behav. Brain Sci.* **21**, 241–259 (1998).
- ⁸B. Lindblom and H. M. Sussman, "Dissecting coarticulation: How locus equations happen," *J. Phonetics* **40**, 1–19 (2012).
- ⁹K. Iskarous, C.A. Fowler, and D.H. Whalen, "Locus equations are an acoustic expression of articulator synergy," *J. Acoust. Soc. Am.* **128**, 2021–2032 (2010).
- ¹⁰H. M. Sussman, K. A. Hoemeke, and F. Ahmed, "A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation," *J. Acoust. Soc. Am.* **94**, 1256–1268 (1993).
- ¹¹S. E. Blumstein, "The mapping from acoustic structure to the phonetic categories of speech: The invariance problem," *Behav. Brain Sci.* **21**, 260 (1998).
- ¹²A. Jongman, "Are locus equations sufficient or necessary for obstruent perception?" *Behav. Brain Sci.* **21**, 271–272 (1998).
- ¹³M. E. Tekieli and W. L. Cullinan, "The perception of temporally segmented vowels and consonant-vowel syllables," *J. Speech Hear. Res.* **22**, 103–121 (1979).
- ¹⁴P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 4.3.01) [computer program]," <http://www.praat.org> (Last viewed February 2005).