

Chapter 18

Tongue trajectories in North American English /æ/ tensing

Christopher Carignan

North Carolina State University

Jeff Mielke

North Carolina State University

Robin Dodsworth

North Carolina State University

Ultrasound imaging is of interest to many dialectologists, due to the relative transportability and low cost associated with this technique for imaging the tongue. The current study introduces a method for examining the temporal dynamics of articulatory correlates of sociolinguistic variables directly from ultrasound video. This technique is demonstrated with data from North American English /æ/.

Short-a tensing is one of the most well-studied regional variables in North American English (NAE): /æ/ is variably realized as [æ], [eə], [eɪ], etc. depending on region and segmental context (Ash (2002); Becker & Wong (2009); Boberg (2008); Boberg & Strassel (2000); Labov, Ash & Boberg (2006); Plichta (2005), *inter alia*). The reason for these differences remain largely unknown (e.g., why /æ/ tenses before /d/ in some Mid-Atlantic varieties but before /g/ in some Northern varieties). The dimension of /æ/ tensing is commonly operationalized as the front diagonal of the acoustic vowel space—i.e., along a line approximating the axis between [a] and [i] quantified as Z2-Z1 (normalized F2 - normalized F1, see Labov, Rosenfelder & Fruehwald 2013). /æ/ has also been observed to involve both falling and rising diphthongal qualities along this axis (Assmann & Katz 2000; Fox & Jacewicz 2009; Labov, Ash & Boberg 2006). The realization of a particular variant of /æ/ along this acoustic diagonal is often interpreted as a correlate of tongue fronting and raising, but the F1 and F2 differences can be due to other factors such as nasalization (De Decker & Nycz 2012) or changes in oral



cavity shape due to velum lowering (Baker, Mielke & Archangeli 2008). Therefore, while Z2-Z1 is clearly an appropriate measure for describing /æ/ tensing, the effort to understand the phonetic motivations for /æ/ tensing is aided from observing lingual articulation directly and the ability to derive time-varying measures of tongue postures.

De Decker & Nycz (2012) showed from single ultrasound frames that speakers vary in the amount of tongue raising associated with their acoustic tensing. The current study extends this approach to allow the study of tongue trajectories. Selecting one time point to represent a token of a speech sound is a common simplifying technique in acoustic and articulatory studies of variation. For ultrasound studies, this simplification is often motivated by the amount of work involved in data processing, which typically involves tracing of the tongue surface in individual video frames (Li, Kambhamettu & Stone 2005). The method described here applies principal component analysis (PCA) and linear regression to derive linguistically meaningful time-varying articulatory signals directly from ultrasound video, in order to characterize the temporal aspects of /æ/ tensing.

Ultrasound and acoustic data was collected from 20 speakers (13 male, age range 20–72) from regions of North America previously known to exhibit distinct regional patterns of /æ/ tensing: seven from the Southern U.S. (tensing before nasals expected), seven from the Northern U.S. and three from Ontario (tensing before nasals and /g/ expected), one from Newfoundland (no tensing expected), and two from Philadelphia (Philadelphia system expected). Participants read a randomized list containing three repetitions of 120 monosyllabic words, while lingual ultrasound images were captured at 60 frames per second using a Terason T3000 ultrasound system with a 8MC3 microconvex array ultrasound probe, Ultraspeech software (Hueber et al. 2007), and an Articulate Instruments ultrasound stabilization headset. Audio was collected using an omnidirectional microphone attached to the headset and recorded through a USB preamplifier in Audacity. A phone-level segmentation of each audio recording was carried out using the Penn Phonetics Lab Forced Aligner (Yuan & Liberman 2008). All images within one second of segmented speech were subject to analysis.

The initial feature extraction from the ultrasound video uses the technique described by Hueber et al. (2007): images for each participant's session are filtered to reduce noise and enhance edges, cropped, and downsampled, and then PCA is applied.¹ Similar techniques have been described by Story (2007) for point-tracking data, and Carignan et al. (2015) for MRI data. The PCA model yields

¹ Matlab scripts and instructions for performing PCA of ultrasound video as described here are available from <http://phon.wordpress.ncsu.edu/lab-manual/>

principal components (PCs) that represent independent axes of variation within a set of speech data, and scores for each PC for each image, which represent how strongly an individual image is correlated with each PC. Thus, the analysis allows an ultrasound video to be represented by a matrix with a column for every PC the researcher wishes to retain, and a row for every video frame. This is comparable to deriving a matrix of linear predictive coding (LPC) or Mel-frequency cepstral (MFC) coefficients from a waveform, both in data compression and in the interpretability of the coefficients.

The articulatory PCs themselves are hard to interpret in any meaningful way. To generate an articulatory signal representing the lingual contribution to tensing (tongue body raising/fronting), the PC score vectors are transformed to correlate with the front diagonal of the acoustic vowel space (Z2-Z1). To find the range of lingual configurations correlated with this acoustic front diagonal, F1 and F2 were measured every 6 ms in all vocalic intervals. The articulatory PCs were interpolated temporally in order to match the acoustic time points, and a linear regression was performed with dependent variable Z2-Z1 and independent variables PCs 1-20. Data was every frame during a vowel lying along the front diagonal [a æ ε e ɪ i]. The coefficients from the linear regression model are used to transform the articulatory PC score matrix to match the articulatory diagonal, resulting in a tongue height signal composed of a single score for each ultrasound frame. For any given ultrasound frame, the higher the score is the more raised and fronted the tongue body is. Since it is derived from ultrasound images instead of acoustic data, the articulatory signal is continuous throughout the recording, even during consonant and silence intervals.

We present a sampling of the results in selected contexts for particular speakers, to demonstrate how ultrasound-derived articulatory signals can be applied to questions about the phonetic origins of the observed dialect patterns. Figures 1–3 display examples of this lingual tensing signal for /æ/ before different coda consonants, using Smoothing-Spline ANOVA (Gu 2002) to compare articulatory trajectories across contexts. In these figures, the tensing signal is on the *y*-axis and normalized time is on the *x*-axis, with the interval [0,1] representing the vowel. Unlike formants, ultrasound-derived articulatory signals are observable during all kinds of consonants.

For most of the speakers, pre-/m n/ tensing is characterized by a peak aligned approximately with the vowel midpoint, while there is no such peak before /ŋ/. These observations are consistent with previously observed formant trajectories for /æ/, and they are illustrated for three speakers in Figure 1. The coda consonant closure intervals are to the right of time=1.0 in the figures. Because it involves

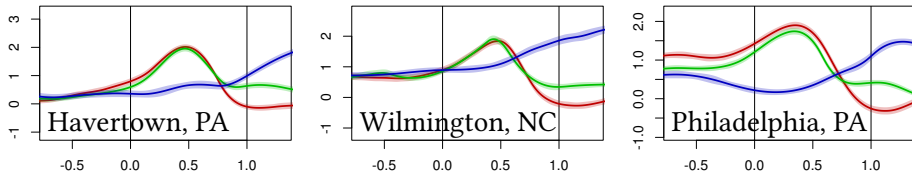


Figure 1: Tongue height for /æ/ before /m/ (red), /n/ (green), and /ŋ/ (blue).

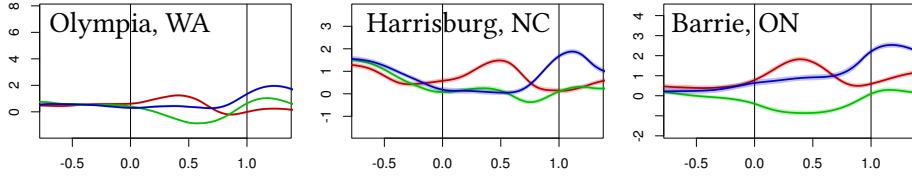


Figure 2: Tongue height for /æ/ before /n/ (red), /d/ (green), and /g/ (blue).

tongue raising, /ŋ/ is high on the tongue raising scale. The raising observed at the end of /æ/ before /ŋ/ can be interpreted as coarticulatory, but the tongue raising in /æ/ before the other nasals is clearly due to a distinct phonetic target. While the *consonant* /n/ trivially involves higher tongue position than /m/, this is clearly not the basis of /æ/ tensing before /n/, which is the same as before /m/. Pre-/ŋ/ tongue raising is less extreme than pre-/n/ tongue raising, except for the speaker from Fargo, ND, whose /æ/ has higher tongue position before /ŋ/ than before /n/, and the speaker from Barrie, Ontario, who raises to the same degree in both contexts.

Figure 2 shows /æ/ before /n d g/ for three speakers. This is an important comparison because pre-/n/ tensing is widespread, pre-/d/ tensing is observed primarily in the Mid-Atlantic region, and pre-/g/ tensing is observed primarily in parts of the North and Canada. All speakers in the sample manifest higher tongue position before /g/ than before /d/, at least by the end of the vowel. This is consistent with /g/ being a velar consonant. For some speakers, including the Philadelphia speaker who tenses more before /d/ in select lexical items, this is true *only* at the very end of the vowel, and it appears to be a purely coarticulatory raising of the tongue body in anticipation for [g]. For the others, the tongue body raising starts earlier in the vowel, which we interpret as a pattern that is not solely due to anticipatory velar coarticulation. This greater tongue height before /g/ compared to /d/ begins in the second half of the vowel for the mid-Atlantic and Buffalo speakers and all but one of the Southern speakers, and from



Figure 3: Left: Tongue height for /æ/ before /n/ (red), /d/ (green), /g/ (blue) in Newfoundland; Right: Tongue height for /æ/ before /f/ (red), /θ/ (green), /s/ (cyan), and /ʃ/ (violet) in Philadelphia.

the first half of the vowel for all of the other Northern speakers and one Southern speaker. /æ/ before /g/ has higher tongue body throughout the entire vowel for all of the Ontario speakers. The similarity in the rising trajectories before /ŋ/ and /g/ (seen in Figures 1 and 2), indicates that /ŋ/ patterns with the other velar /g/ rather than the other nasals /m n/. Figure 3 shows two more expected patterns. The speaker from Newfoundland shows very little tongue raising even before a nasal (Boberg 2008), and a speaker from Philadelphia shows tongue raising before anterior fricatives, with a trajectory similar to what is observed in other speakers only before /m n/.

The method we have described generates time-varying articulatory signals from ultrasound video at a high frame rate (60 fps for the system used in this study). This allows the observation and quantification of temporal changes and co-articulatory effects, from an articulatory imaging technique that is fairly accessible to linguists, with minimal manual data processing. Unlike formant measurements, these articulatory signals are present in consonants and even silent intervals, and they do not require any kind of tracking. Using this method, we have analyzed temporal characteristics of NAE /æ/ tensing in different segmental contexts for speakers from different regions known to have different /æ/ tensing patterns. At an articulatory level, most speakers' patterns of tongue height before /m n/ suggest a completely different vowel target in these contexts, much like the tense /æ/ observed in a superset of these contexts in the Philadelphia split short-a system. In speakers who have tongue raising before the velars /g N/, a completely different trajectory is observed, in which tenseness increases throughout the vowel. The observed regional variation in the timing of tongue raising before velars suggests that pre-velar /æ/ tensing can be studied as coarticulation that has been phonologized to different degrees in different dialects.

References

- Ash, Sharon. 2002. The distribution of a phonemic split in the Mid-Atlantic region: Yet more on short a. *Penn Working Papers in Linguistics* 8(3). 1–15.
- Assmann, P. F. & W. F. Katz. 2000. Time-varying spectral change in the vowels of children and adults. *The Journal of the Acoustical Society of America* 108. 1856.
- Baker, Adam, Jeff Mielke & Diana Archangeli. 2008. More velar than /g/: Consonant coarticulation as a cause of diphthongization. In Charles B. Chang & Hannah J. Haynie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project.
- Becker, K. & A. W. Wong. 2009. The short-a system of NYC English: An update. *Penn Working Papers in Linguistics* 14. 11–20.
- Boberg, Charles. 2008. Regional phonetic differentiation in Standard Canadian English. *Journal of English Linguistics* 36(2). 129–154.
- Boberg, Charles & S. Strassel. 2000. Short-a in Cincinnati: A change in progress. *Journal of English Linguistics* 28. 108–126.
- Carignan, C., R. Shosted, M. Fu, Z.-P. Liang & B. B. Sutton. 2015. A real-time MRI investigation of the role of lingual and pharyngeal articulation in the production of the nasal vowel system of French. *Journal of Phonetics* 50. 34–51.
- De Decker, P. M. & J. R. Nycz. 2012. Are tense [æ]s really tense? The mapping between articulation and acoustics. *Lingua* 122(7). 810–821.
- Fox, R. A. & E. Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America* 126(5). 2603.
- Gu, Chong. 2002. *Smoothing spline ANOVA models*. Springer Series in Statistics. New York: Springer-Verlag.
- Hueber, T., G. Aversano, G. Chollet, B. Denby, G. Dreyfus, Y. Oussar, P. Roussel & Maureen Stone. 2007. Eigentongue feature extraction for an ultrasound-based silent speech interface. In *Proceedings of the 2007 International Conference on Acoustics, Speech, and Signal Processing*, I-1245–I-1248.
- Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin, New York: Mouton de Gruyter.
- Labov, William, Ingrid Rosenfelder & Josef Fruehwald. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89(1). 30–65.
- Li, M., C. Kambhamettu & Maureen Stone. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* 19(6–7). 545–554.

- Plichta, Bartek. 2005. *Interdisciplinary perspectives on the Northern Cities Chain Shift*. Michigan State University Unpublished Ph.D. thesis.
- Story, Brad H. 2007. Time dependence of vocal tract modes during production of vowels and vowel sequences. *Journal of the Acoustical Society of America* 121(6). 3770–3789.
- Yuan, J. & M. Liberman. 2008. Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics '08*.

