Are Tense [æ]s Really Tense? An Ultrasound Study

Paul M. De Decker and Jennifer R. Nycz*

1 Introduction

In certain dialects of American English, the lax low front vowel [æ] (as in *cat* and *have)* coexists with a variant that may impressionistically be described as "raised" or "tense" (Ferguson 1975, Labov 1989). In the mid-Atlantic dialects, one striking aspect of this alternation is the varied subset of environments in which the tense variant [æ̝] may occur. In Philadelphia, [æ̝] occurs before [+anterior] voiceless fricatives (e.g. [pæ̝θ] *path* and [pæ̝s] *pass*), before [+anterior] nasals ([hæ̝m] *ham* and [pæ̝n] *pan*), and in the words *mad, bad,* and *glad* (Ferguson 1975, Labov 1989). In New York City, this set is expanded to include the voiceless fricative [ʃ] ([læ̝ʃ] *lash*) and the rest of the voiced stops (Labov 1994). New Jersey speakers, however, exhibit a variety of tensing patterns (Ash 2002). For instance, some speakers have tense [æ̝] only in syllables which are closed by nasals, while others exhibit the Philadelphia or New York systems.

2 [æ̝] or [æ̃]? [æ̝] *and* [æ̃]?

As the term "tensing" implies, this shift is commonly understood to involve a change in tongue position, such as advancement of the tongue root and/or raising of the tongue body. The result is a vowel similar to a lengthened [ɛ:] or [ɛ:ə]. However, as Labov points out:

> . . . It should be clear that *tense* is a cover term for [a] combination of impressionistic and acoustic features. Though the advanced forms of /aeh/ appear to require more muscular effort, *tense* cannot be used to refer to this property until we have electromyographic data to support this impression (Labov 1994:505).

*We would like to thank Lisa Davidson, Diamandis Gafos, William Labov, members of the Ph-Lab group at NYU and the PLC 29 audience for helpful comments on this work.

In fact, a change in lingual gesture is not the only way to produce what is perceived as a tense [æ̝]. Krakow et al. (1988) have shown that nasalization can mimic the acoustic effects of tongue raising, lowering the F1 and raising the F2 of low vowels. Others studies have indicated that nasalized variants of /æ/ are perceived as being higher (Wright 1975, 1986). More recently, Plichta's (2002) study of nasal airflow in Northern Cities-shifted speech has revealed that [æ] is produced with high nasal airflow in both oral and nasal environments in these dialects; he suggests that these facts may have led to a premature description of the Northern Cities-shifted [æ] as being raised. In general, it seems that phonemically non-nasal low vowels are characterized by some degree of inherent phonetic nasality, as the amount of velopharyngeal port opening is typically greater during production of an oral low vowel as compared to an oral high vowel (e.g. Clumeck 1975).

Given that there are two possible sources of tense [æ̝], it is reasonable to ask whether speakers who produce tense [æ̝] articulate this phone in the same way in all relevant environments. This is a complicated question, because articulation encompasses several parameters (lingual articulation, velum raising/lowering, and possibly others). In this paper, we focus on the following question: To what extent is [æ] "tensing" associated with changes in tongue position? Because multiple sources of [æ]-tensing exist, we might expect to find variation in the role of the tongue. Some tense [æ̝]s may be produced using a lingual tensing gesture, but others may lack such a gesture, resulting instead from nasalization.

3 Method

To address this question, we designed an experiment using ultrasound technology, which allows non-invasive, real-time visualizations of tongue motion during speech (Stone 1999). Ultrasound has been used in speech studies to examine, for example, the production of non-native consonant clusters (Davidson 2004), Hungarian vowel harmony (Benus 2005), and /l/ vocalization (Wrench & Scobbie 2003).

Tokens of the low front vowel were elicited from four speakers, all of whom are from New Jersey, and all of whom have been observed to use tense [æ̝]. Speaker 1 is a 22 year old male from Union County. Speaker 2 is a 20 year old female, also from Union County. Speaker 3 is a 21 year old female from Monmouth County, and Speaker 4 is a 22 year old male from Union County. The stimuli of interest consisted of 4 target words — the minimal quadruple *pat, pad, pass* and *pan* — embedded in the carrier phrase

"Say ___ very loudly" (see Table 1). These words were chosen to yield tokens of [æ] produced in several contexts, three of which have been shown to condition tensing in Mid-Atlantic dialects (*pan, pass, pad*) and one of which typically yields a lax [æ] (*pat*). Eight additional CVC filler words were also presented in this carrier phrase. Stimuli were presented in 12 blocks of 12 sentences each, and the order of sentences within each block was randomized. Tokens from Blocks 2 to 11 are analyzed here.

Carrier Phrase	**Environment**
Say pat very loudly.	pre-voiceless stop
Say pad very loudly.	pre-voiced stop
Say pass very loudly.	pre-voiceless fricative
Say pan very loudly.	prenasal

Table 1. Elicited sentences, and the phonological environments which they represent.

Each speaker was recorded while reading the stimuli in a soundproof booth, using an Audio Technica condenser microphone powered by an Aphex thermionic 2-channel microphone pre-amplifier. The audio signal was digitized to a Dell Optiplex GX270 Pentium 4 hard drive running Adobe Premier 6.0 in Windows using a Canopus ADVC 1394 audio/video capture card. In addition to the audio recording, a midsagittal image (see Fig. 1 for an example)[1] of the speaker's tongue was recorded using a Sonosite Titan ultrasound machine at a rate of 30 frames/second.

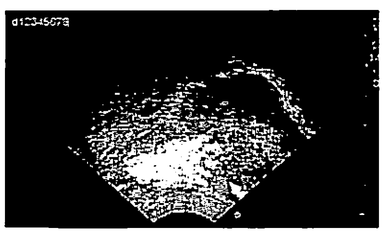

Figure 1. An ultrasound still image. The tongue dorsum is located at the left side of the image.

[1]The ultrasound transducer is placed under the chin, where it sends ultra high frequency sound waves up through the soft tissue. When the waves reach an interface between surfaces of different densities (in this case, the interface of the tongue surface and the air pocket above the tongue), they are reflected back to the transducer and are used to create an image.

Real-time images of the tongue were digitized using the same audio/video capture card used to save the audio signal. Synchronized image and audio were saved as .avi files and converted to .mov format using Quicktime, which was then used to extract the audio files and still images used in the acoustic and articulatory analyses.

4 Data Analysis

4.1 Acoustics

Acoustic analysis of each speaker's [æ] tokens was carried out to establish which kind of tensing pattern each speaker exhibits. The freeware phonetics program Praat (version 4.2.21, Boersma & Weenik 2005) was used to measure the first four formants of each [æ] token at three points: the onset of voicing and the regular periodic wave, the point at which the periodic wave ends or where a qualitatively different wave begins (in the case of a following /s/ or /n/), and the temporal midpoint between these two landmarks. Words were tested for statistically significant differences in F1 and F2 at the temporal midpoint, using a one-way ANOVA and Tukey post-hoc tests.

4.2 Articulation

To address the question introduced in Section 2 — to what extent is [æ] tensing associated with a change in tongue position? — we must be able to compare tongue shapes across the different word classes. To do this, a series of JPEG image sequences of each token were extracted from the ultrasound movie files, and tongue contours for each image were traced using Edgetrak (Li, Akgul, and Kambhamettu 2004). From each of these image sequences, the tongue contour showing the most retraction (the target gesture for [æ], following Wood 1979) was used for analysis: for each of the four environments, the target tongue contours of [æ] were combined in Excel to yield an average tongue shape for that environment. Thus, for each speaker we obtained four average tongue shapes for [æ], which could then be compared to determine whether there were lingual differences in the production of [æ] across different contexts.

5 Results

5.1 Acoustics

In this section, we present the results of the acoustic analysis of each speaker's [æ] tokens. Tables 2 through 9 summarize the results of the formant analysis taken at the temporal midpoint of each vowel. The numbers in the last four columns represent the differences in mean F1 and F2 values, comparing the words in the left-most column with the words across the last four cells of the top row. ANOVA and post-hoc tests reveal significant differences in the mean formant values between at least two vowel subsets for each speaker; significant differences ($p < .05$) are printed in bold.

5.1.1 Speaker 1

As the vowel plot in Figure 2 shows, Speaker 1 shows a gradual increase in acoustic tenseness from *pat* to *pan*. Table 2 shows that the differences between these word classes are significant for F1 ($F(39)=49.181$, $p<.001$) and F2 ($F(39)=16.050$, $p<.001$). Tukey post-hoc tests reveal that while *pan* and *pat* are each significantly different from all other words with respect to F1, there is no F1 difference between *pass* and *pad*. With respect to F2 (Table 3),

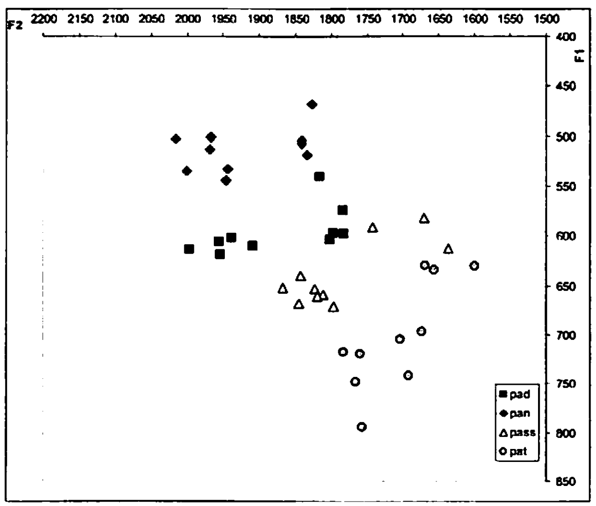

Figure 2. [æ] produced by Speaker 1

no word class is significantly different from the word classes that are immediately adjacent to it in the vowel space (e.g., *pan* is not distinct from *pad*, though *pan* is significantly different from each of the other two word classes; *pad* is not distinct from *pan* or *pass,* though it is significantly different from *pat*, etc.)

According to these results, Speaker 1 makes a three-way [æ] distinction along the dimension of height in the vowel space: prenasal [æ] is most tense, the [æ]s in *pass* and *pad* are somewhat less tense, and pre-voiceless stop [æ] is relatively lax. Along the backness dimension, the distinctions are less clear, though the general pattern remains: *pan* is most advanced, followed by *pad, pass,* and *pat*, however, individual steps in the continuum do not constitute significant differences.

Word	Mean F1	StdDev		*pan*	*Pad*	*pass*	*pat*
pan	513Hz	22Hz	*pan*	*			
pad	596Hz	23Hz	*pad*	**83Hz**	*		
pass	639Hz	32Hz	*pass*	**126Hz**	43Hz	*	
pat	702Hz	56Hz	*pat*	**189Hz**	**106Hz**	**63Hz**	*

Table 2. First Formant ANOVA F(39) = 49.181, p < .001

Word	Mean F2	StdDev		*pan*	*pad*	*pass*	*pat*
pan	1919Hz	74Hz	*pan*	*			
pad	1874Hz	84Hz	*pad*	-45Hz	*		
pass	1786Hz	77Hz	*pass*	**-133Hz**	-88Hz	*	
pat	1707Hz	60Hz	*pat*	**-212Hz**	**-168Hz**	-79Hz	*

Table 3. Second Formant ANOVA F(39) = 16.050, p < .001

5.1.2 Speaker 2

Speaker 2 exhibits the nasal system described by Ash (2002) for some speakers of New Jersey. The vowel plot in Figure 3 reveals a clear distinction between prenasal [æ] and this vowel in other environments. ANOVA results indicate distinctions between the vowel sets for both F1 (F(39)=77.136, p<.001) and F2 (F(39)=79.741, p<.001). Post-hoc tests reveal two significantly different sets of comparisons for this speaker, as shown in tables 4 and 5. The first, like speaker 1's findings, shows strong distinctions between the average F1 and F2 values for *pan* compared to all other words. The only other significant difference is between *pass* and *pat* with respect to F1.

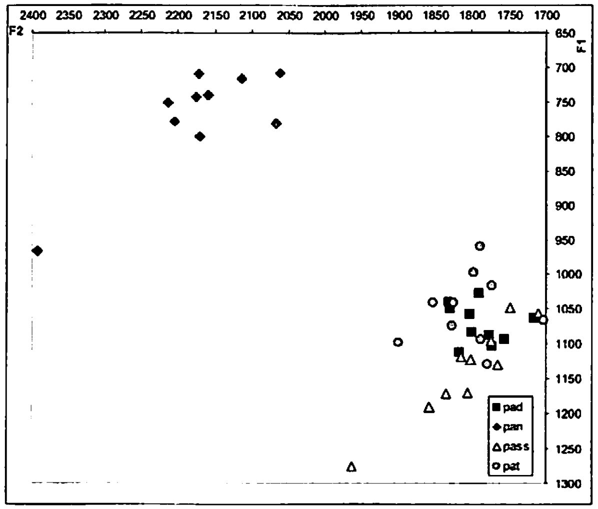

Figure 3. [æ] produced by Speaker 2

Word	Mean F1	StdDev		*pan*	*pad*	*pass*	*pat*
pan	770Hz	76	*pan*	*			
pad	1072Hz	28	*pad*	**302Hz**	*		
pass	1138Hz	68	*pass*	**368Hz**	66Hz	*	
pat	1052Hz	51	*pat*	**282Hz**	-20Hz	**-86Hz**	*

Table 4. First Formant ANOVA F(39) = 77.136, p < .001

Word	Mean F2	StdDev		*Pan*	*pad*	*pass*	*pat*
pan	2174Hz	93	*pan*	*			
pad	1789Hz	35	*pad*	**-385Hz**	*		
pass	1807Hz	70	*pass*	**-367Hz**	18Hz	*	
pat	1803Hz	52	*pat*	**-371Hz**	14Hz	-4Hz	*

Table 5. Second Formant ANOVA F(39) = 79.741, p < .001

5.1.3 Speaker 3

Like Speaker 2, Speaker 3 exhibits a nasal system. In this case, there is an even clearer two-way distinction: the prenasal [æ] in *pan* is different from each of the other word classes with respect to both F1 and F2, and there are no significant differences between any of the pre-oral environments. Tables

6 and 7 show significant differences for mean values of F1 (F(39)=68.481, p<.001) and F2 (F(39)=23.276, p<.001).

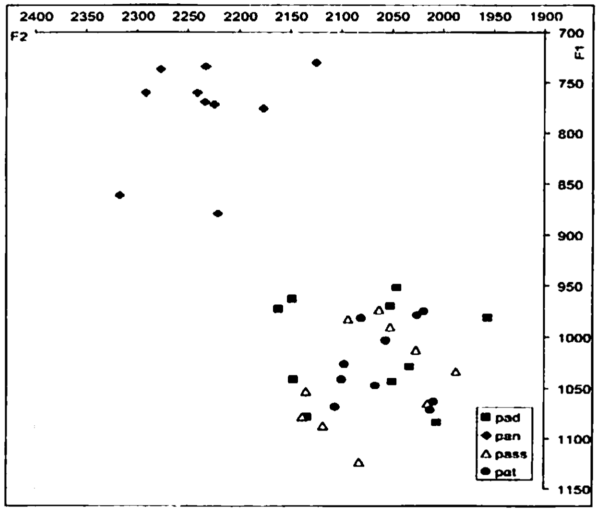

Figure 4. [æ] produced by Speaker 3

Word	Mean F1	StdDev		*pan*	*pad*	*pass*	*pat*
pan	778Hz	51	*pan*	*			
pad	1012Hz	50	*pad*	**234Hz**	*		
pass	1040Hz	50	*pass*	**262Hz**	28Hz	*	
pat	1026Hz	39	*pat*	**248Hz**	14Hz	-14Hz	*

Table 6. First Formant ANOVA, F(39) = 68.481, p < .001

Word	Mean F2	StdDev		*pan*	*pad*	*pass*	*pat*
pan	2234Hz	55	*pan*	*			
pad	2073Hz	70	*pad*	**-161Hz**	*		
pass	2071Hz	52	*pass*	**-163Hz**	-2Hz	*	
pat	2057Hz	38	*pat*	**-177Hz**	-16Hz	-14Hz	*

Table 7. Second Formant ANOVA, F(39) = 23.276, p < .001

5.1.4 Speaker 4

Speaker 4 shows an acoustic tensing cline similar to that of Speaker 1: There is a three-way distinction in tensing, with *pat* being lax/least tense, *pan* being most tense, and *pass/pad* occupying an intermediate spot in the vowel space. (Figure. 5). Tables 8 and 9 show significant differences in F1 ($F(39)=49.181$, $p<.001$) and F2 ($F(39)=16.050$, $p<.001$).

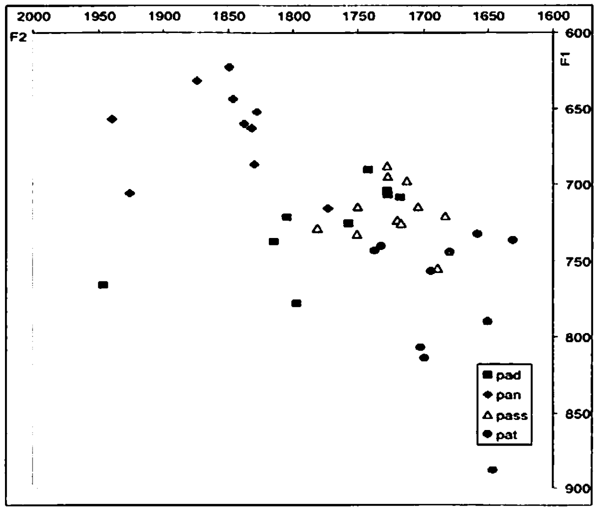

Figure 5. [æ] produced by Speaker 4

Word	Mean F1	StdDev		*pan*	*pad*	*pass*	*Pat*
pan	664Hz	30	*pan*	*			
pad	724Hz	29	*pad*	**60Hz**	*		
pass	720Hz	19	*pass*	**56Hz**	-4Hz	*	
pat	776Hz	50	*pat*	**112Hz**	**52Hz**	**56Hz**	*

Table 8. First Formant ANOVA, $F(39) = 18.182$, $p < .001$

Word	Mean F2	StdDev		*pan*	*pad*	*pass*	*pat*
pan	1854Hz	49	*pan*	*			
pad	1776Hz	70	*pad*	**-78Hz**	*		
pass	1724Hz	30	*pass*	**-130Hz**	-52Hz	*	
pat	1683Hz	36	*pat*	**-171Hz**	**-93Hz**	-41Hz	*

Table 9. Second Formant ANOVA, $F(39) = 22.936$, $p < .001$

In summary, the results from acoustic and statistical analyses of F1 and F2 at the temporal midpoint reveal a significant distinction between tense and non-tense variants of [æ] for all speakers. However, we observe two different acoustic systems among our speakers: Speakers 2 and 3 each have a nasal system contrasting the prenasal [æ] with [æ] in pre-oral contexts, while Speakers 1 and 4 show a three-way distinction between the tense prenasal [æ] in *pan*, the somewhat less tense [æ] of *pass/pad*, and the relatively lax [æ] in the pre-voiceless stop context of *pat*.

We now turn to the results of the articulatory analysis of the tongue shapes associated with these systems. Is it the case that tense [æ]s which are acoustically distinct are also lingually distinct? In other words, are distinct tongue shapes associated with the observed acoustic distinctions?

5.2 Articulation

To compare the lingual articulation of [æ] in different contexts, we plotted the four average tongue contours for each speaker on a single graph (Figures 6 through 9). The rightmost third of each contour represents the position of the tongue body, while the leftmost third represents the position of the tongue dorsum[2].

First, we consider the articulatory patterns of Speakers 1 and 4, who produced a three-way distinction between tense [æ] in *pan*, lax [æ] in *pat*, and the intermediate [æ]s of *pass/pad*.

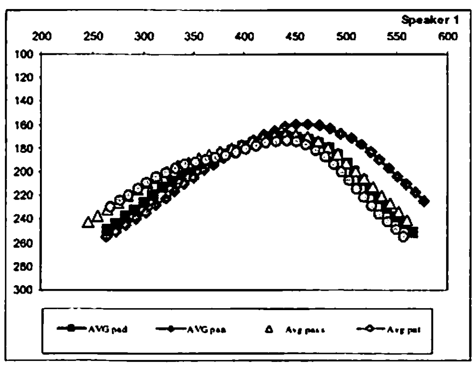

Figure 6: Speaker 1, Average Tongue Contours for All Environments

As Figure 6 shows, Speaker 1 does have a change in lingual gesture for each of these environments. The [æ] in *pan* is the most raised, and the [æ] of *pat*

[2]The tongue tip, which does not image well, is not represented on these figures.

has the lowest tongue body; *pass* and *pad* are both somewhat more raised than *pat*. The articulatory distinctions are less clear with respect to tongue advancement, although this is consistent with Speaker 1's acoustic data, which shows that the four contexts are less distinct from one another in the F2 dimension.

The average tongue contours for Speaker 4 follow a similar pattern. As shown in Figure 7, the [æ] of *pan* is both higher and more advanced. The [æ]s of *pass* and *pat* are slightly higher than those of *pat*, though there seems to be no clear differentiation among these 3 words at the tongue dorsum. Like Speaker 1, however, Speaker 4's distinction between *pat* and *pass/pad* is less robust on the F2 dimension (only *pad* is distinct from *pat*), so this articulatory pattern is not surprising.

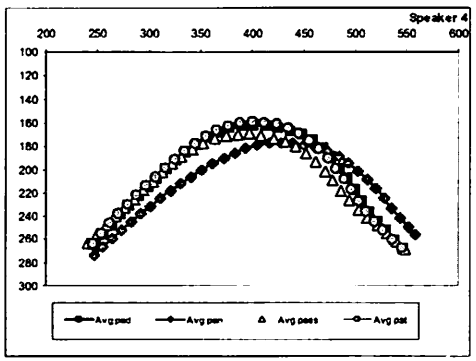

Figure 7: Speaker 4, Average Tongue Contours for All Environments

Let us now look at Speakers 2 and 3, who each produced a two-way (nasal/nonnasal) distinction between tense and lax [æ]. Speaker 2 has clear lingual raising and fronting gestures for [æ] in *pan* (Figure 8), while the pre-oral contours are not distinct; these results are entirely consistent with her acoustic data.

Speaker 3, however, shows a different pattern. For this speaker, the [æ] in *pass* has the highest tongue contour; in addition, while the [æ] in *pan* is somewhat more fronted than the other contours, there does not seem to be a very strong difference between *pan* and the other contours with respect to the position of either the tongue dorsum or the tongue body. Therefore, it is not clear that there is a distinct lingual gesture associated with the tense/lax acoustic distinction present in Speaker 3's speech.

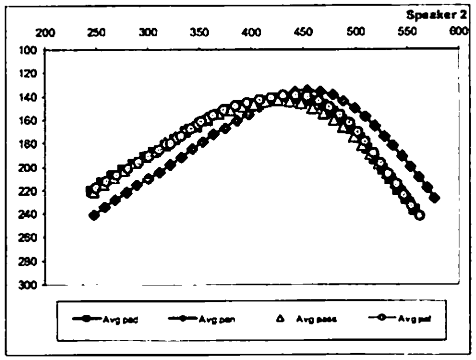

Figure 8: Speaker 2, Average Tongue Contours for All Environments

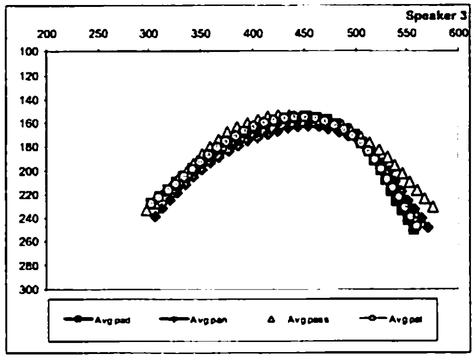

Figure 9: Speaker 3, Average Tongue Contours for All Environments.

6 Conclusion

As we can see from the small sample discussed here, there is variation in how speakers distinguish between tense and nontense [æ]s. This is clearest in the case of Speakers 2 and 3: While both have a clear acoustic difference between [æ] in the prenasal environment and [æ] in oral environments, this difference seems to be the result of different articulations. While Speaker 2's prenasal [æ] is associated with a lingual fronting and raising gesture, it is not clear that Speaker 3's prenasal [æ] is the result of a similar gesture. If Speaker 3's acoustic distinction is not the result of a change in lingual gesture, it must be due to some other articulatory change; we hypothesize that this speaker may be using nasalization to create the acoustic difference. To test such hypotheses, future work will need to incorporate both ultrasound (or some similar) technology and a means of measuring nasal airflow. Such work will also allow us to learn more about the specific relationship between nasalization and lingual raising/fronting with respect to [æ] tensing: Is there is a tradeoff relationship between these two mechanisms, or an enhancement relationship, such that we would ultimately expect coupling of these articulations?

These results also raise several interesting questions pertaining to inter- and intra-speaker phonetic variation. For instance, do individual speakers have a "choice" with respect to which tensing strategy they employ (as might be the case, for example, with variations of the articulation of American English /r/ (Hagiwara 1994)) or are the two articulatory strategies geographically distributed? Of course, it is possible that individual speakers may employ more than one tensing strategy to achieve the same acoustic end: A given speaker may produce tense [æ]s in context A using a lingual raising gesture, and tense [æ]s in context B using nasalization. The answers to such questions will ultimately bear on issues relevant to the phonetics/phonology interface, such as whether phonetic goals are represented in acoustic or articulatory terms.

References

Ash, Sharon. 2002. The distribution of a phonemic split in the Mid-Atlantic region: Yet more on short a. *U. Penn Working Paper in Linguistics, Volume 8.3*:1–15.

Benus, Stefan. 2005. *Dynamics and transparency in vowel harmony*. Doctoral dissertation, New York University.

Boersma, Paul and David Weenick. 2005. Praat. http://www.fon.hum.uva.nl/praat/

Clumeck, Harold. 1975. A cross-linguistic investigation of vowel nasalization: an instrumental study. In *Nasalfest: Papers from a symposium on nasals and nasalization*, eds. C.A. Ferguson, L.M. Hyman, and J.J. Ohala, 133–152. Language Universals Project, Stanford University.

Davidson, Lisa. 2004. The influence of articulation, perception and coordination on non-native phonotactics and repairs. Talk presented at the "Redefining Elicitation: Novel Data in Phonological Theory" Workshop, NYU, April 9-11, 2004.

Ferguson, Charles. 1975. "Short a" in Philadelphia English. In *Studies in Linguistics in Honor of George L. Trager*, ed. E. Smith. The Hague: Mouton.

Hagiwara, Robert. 1994. Three types of American /r/. UCLA Working Papers in Phonetics 88.

House, A. 1957. Analog studies of nasal consonants. *Journal of speech and hearing disorders* 22(2):190–204.

Krakow, Rena A., Patrice S. Beddor, Louis Goldstein, and Carol Fowler. 1998. Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America*. 83(3):1146–58.

Labov, William. 1989. Exact description of the speech community: Short A in Philadelphia. *Language Change and Variation*, ed. by R. Fasold and D. Schiffrin, 1–57. Amsterdam: John Benjamins.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

Li, Min, Yusuf Akgul and Chandra Kambhamettu. 2004. Edge Extraction and Tracking program. Department of Electrical Engineering and Computer Information Systems, University of Delaware.

Ohala, John. 1975. Phonetic explanations for nasal sound patterns In *Nasalfest. Papers from a symposium on nasals and nasalization*, eds. C.A. Ferguson, L.M. Hyman, and J.J. Ohala, 289–316. Language Universals Project, Stanford University.

Plichta, Bartek. 2002. Coarticulatory Nasalization and the Northern Cities Vowel Shift: Is /æ/ Really Raising? Paper presented at NWAV 31, Stanford University, California.

Stone, Maureen. 1999. Laboratory Techniques for Investigating Speech Articulation. In *The Handbook of Phonetic Sciences*, eds. Hardcastle, WJ. and Laver, J. Oxford: Blackwell Publishers.

Wood, Sydney. 1979. A radiographic analysis of constriction locations for vowels. *Journal of Phonetics* 7:25–43

Wrench, Alan and James Scobbie. 2003. Categorising vocalisation of English /l/ using EPG, EMA and Ultrasound. In *Proceedings of the 6th International Seminar on Speech Production* (ISSP Sydney), eds. S. Palethorpe and M. Tabain, 314–319.

Wright, James. 1975. Effects of vowel nasalization on the perception of vowel height. In *Nasalfest. Papers from a symposium on nasals and nasalization*, eds. C.A. Ferguson, L.M. Hyman, and J.J. Ohala, 373–388. Language Universals Project, Stanford University.

Wright, James. 1986. The Behavior of Nasalized Vowels in the Perceptual Vowel Space. In *Experimental Phonology*, eds. J. Ohala and J. Jaeger. Academic Press.

Department of Linguistics
New York University
719 Broadway, 4th Floor
New York, NY 10003
dedecker@nyu.edu
jennifer.nycz@nyu.edu