University of Pennsylvania Working Papers in Linguistics

Volume 18
Issue 2 *Selected Papers from NWAV 40*

Article 18

9-1-2012

The Lowering of Raised-THOUGHT and the Low-Back Distinction in New York City: Evidence from Chinese Americans

Amy Wing-mei Wong
New York University, amywong@nyu.edu

The Lowering of Raised-THOUGHT and the Low-Back Distinction in New York City: Evidence from Chinese Americans

Abstract
This paper examines the production of the *thought* and *lot* vowel classes by New Yorkers of Chinese heritage. Sixteen New York-born Chinese American males between the ages of 11 and 61 were sampled. About 600 *thought* and *lot* tokens were instrumentally measured and normalized for statistical analyses and plotting. A linear regression analysis and a correlation test find evidence of the reversal of *thought*-raising. The height (normalized F1) of *thought* lowers as speaker's year of birth increases. In other words, older Chinese New Yorkers are more likely to produce *thought*-raising than the younger ones. The finding corroborates Becker's (2010) results from European New Yorkers.

To determine how the lowering of *thought* may have affected the low back distinction in New York City English, this study utilized the Pillai-Bartlette trace and the Euclidean distance between *lot* and *thought* as measurements of the magnitude of the low back distinction, along with visual examination of individual vowel plots. Despite the lowering of *thought* across apparent-time, most, if not all, speakers continue to maintain the low back distinction. However, the *lot* and *thought* classes for a few younger speakers are very close in the vowel space with some overlapping tokens. Their low back vowels configuration resembles the patterns exhibited by the "transitional speakers" in the Midland area in Labov et al. (2006), whose *thought* and *lot* classes are neither completely merged nor completely distinct. These results call for further work on the low back vowels of speakers of other social and ethnic groups in order to investigate the future trajectory of the *thought* vowel vis-à-vis the robustness of the low back distinction in the English of New York City.

The Lowering of Raised-THOUGHT and the Low-Back Distinction in New York City: Evidence from Chinese Americans

Amy Wing-mei Wong*

1 Introduction

English in New York City (NYCE hereafter) has traditionally been characterized as one of the dialects of North American English that has resisted the low-back merger of the THOUGHT vowel (e.g., *caught* and *talk*) and the LOT vowel (e.g., *top* and *dock*) (Labov, Ash, and Boberg 2006). Labov et al. (2006) argue that NYCE's resistance to the merger-in-progress is closely linked to another defining feature of the dialect—the raising of the THOUGHT vowel, which was found in Labov (1966) to be a change from below with the expectation that THOUGHT-raising would persist in the region.

Recent work, however, has suggested that the THOUGHT-raising documented earlier is no longer a stable feature of NYCE, nor is it a uniform feature across New Yorkers of different ethnicities. Becker (2010), for instance, finds apparent-time evidence of the reversal of THOUGHT-raising in the Lower East Side, with a general trend for younger Lower East Siders to produce lower THOUGHT than older speakers. Additionally, Becker finds a statistical interaction between age and ethnicity: African American and Latino speakers of all ages continued to produce raised-THOUGHT while young European speakers produced THOUGHT with little or no raising. Although the correlation between age and the height of THOUGHT for Asian American speakers in Becker (2010)—most of whom are Chinese Americans—did not reach statistical significance, the correlation showed a similar trend as that found for white speakers. Given the argument that NYCE's resistance to the low-back merger is tied to the raising of THOUGHT-raising, the recent findings on the lowering of raised-THOUGHT, thus, brings up the question concerning how this reversal of change—which potentially removes the "block" to the merger—impacts the low-back distinction in NYCE.

This paper investigates the production of the THOUGHT and LOT classes by a group of second generation Chinese Americans in New York City, focusing on the height of the THOUGHT vowel and the status of the maintenance between THOUGHT and LOT. Two main research questions will be addressed: (1) Is there apparent-time evidence of the reversal of THOUGHT-raising among New Yorkers of Chinese heritage, the one group that did not reach statistical significance in Becker (2010)? (2) For those speakers who produce the lowest THOUGHT vowel, is their THOUGHT class distinct from their LOT class?

2 The Sample

The analysis is based on a speech sample gathered from sixteen male speakers of Chinese descent, born and raised in four of the five boroughs of New York City (except the Bronx). All of them reported speaking English natively, with many being bilingual in Cantonese as well. The year of birth of the sixteen speakers ranges from 1949 to 1998. At the time of the fieldwork, the oldest speaker was 61 and the youngest 11.

Speech data analyzed for this paper were elicited in two stylistic contexts. Targeted vowels from the eight younger speakers (born between 1995–1998) were taken from the reading context. One adult (George, born in 1949) only provided tokens under the interview context. Vowel tokens of the remaining seven adults (born between 1952 and 1990) were taken from both interview and reading contexts. T-tests on each of these seven speakers found that the formant values of a given

*The material reported here is based upon work supported by the National Science Foundation under Grant No. 0951841. Special thanks to Kara Becker, Lauren Hall-Lew, Dan Johnson, Jennifer Nycz, Marcos Rohena-Madrazo, John Singler for useful discussion. Thanks also go to Libby Coggshall, Emily Nguyen, Luiza Newlin Lukowicz, Cara Shousterman and the audience of NWAV40 for valuable comments on an earlier draft. All errors remain my own.

word class were not statistically distinct between the two stylistic contexts.

Tokens of THOUGHT and LOT were the analytical focus of the paper and were measured following the protocol set up in Thomas (2011) and Labov et al. (2006). GOOSE, FLEECE, and TRAP tokens were also measured for normalization purposes. Tokens that were not stressed, or tokens followed by a nasal, liquid, or glide were omitted from analysis due to known coarticulation and reduction effects that would affect the formant values. Given NYCE's split short-a system (Cohen 1970, Labov et al. 2006), TRAP tokens in potentially tensing environments (i.e., preceding a voiced stop or a voiceless fricative) were also excluded.

Following the practice of Labov et al. (2006), token measurements were taken at a point that best represents the central tendency of the nucleus of each vowel. For tokens of LOT, GOOSE, FLEECE, and TRAP, point measurements were taken at the F1 maximum. For THOUGHT tokens, measurements were taken at the point with minimum F2. As a general principle, no fewer than 9 tokens in each of the LOT and THOUGHT classes were analyzed for a given speaker. In total, 608 tokens of THOUGHT and LOT and another 616 tokens of GOOSE, FLEECE, and TRAP were measured in Praat (Boersma and Weenink 2010). Formant values were normalized using the modified Watt and Fabricius method (Fabricius, Watt, and Johnson 2009, Watt and Fabricius 2002) available through the NORM suite (Thomas and Kendall, 2007).

Regression analysis of the normalized F1 values of THOUGHT was used to establish the extent of change in vowel height over time. The Pillai-Bartlett trace (more to follow), the Euclidean distance between LOT and THOUGHT and visual inspection of individual vowel plots are used to evaluate the status of the maintenance of the low-back distinction for each speaker. Statistical results were evaluated and interpreted in conjunction to visual inspections of vowel plots.

3 Height Analysis

Table 1 below is a summary of the mean normalized F1 (height) of THOUGHT averaged across different phonological contexts for each speaker, arranged in ascending order. Speakers producing the highest THOUGHT in the vowel space (i.e., lower normalized F1) appear at the top of the table.

Speaker	Year of Birth	Mean F1/S(F1) (Height)	Standard Deviation	N
George	1949	0.966	0.159	11
Paul	1951	0.997	0.122	28
Norman	1969	1.103	0.136	26
Ernie	1980	1.153	0.131	24
John	1957	1.166	0.134	31
Joseph	1974	1.214	0.213	21
Avery	1995	1.246	0.086	11
Chris	1966	1.258	0.087	30
Tommy	1998	1.278	0.151	11
Tim	1990	1.319	0.053	23
Michael	1997	1.321	0.132	12
Nick	1995	1.338	0.054	9
Kelvin	1998	1.360	0.127	13
Simon	1997	1.392	0.091	11
Jeff	1997	1.417	0.096	11
James	1997	1.452	0.096	11

Table 1: Mean height of THOUGHT arranged by normalized F1 in ascending order.

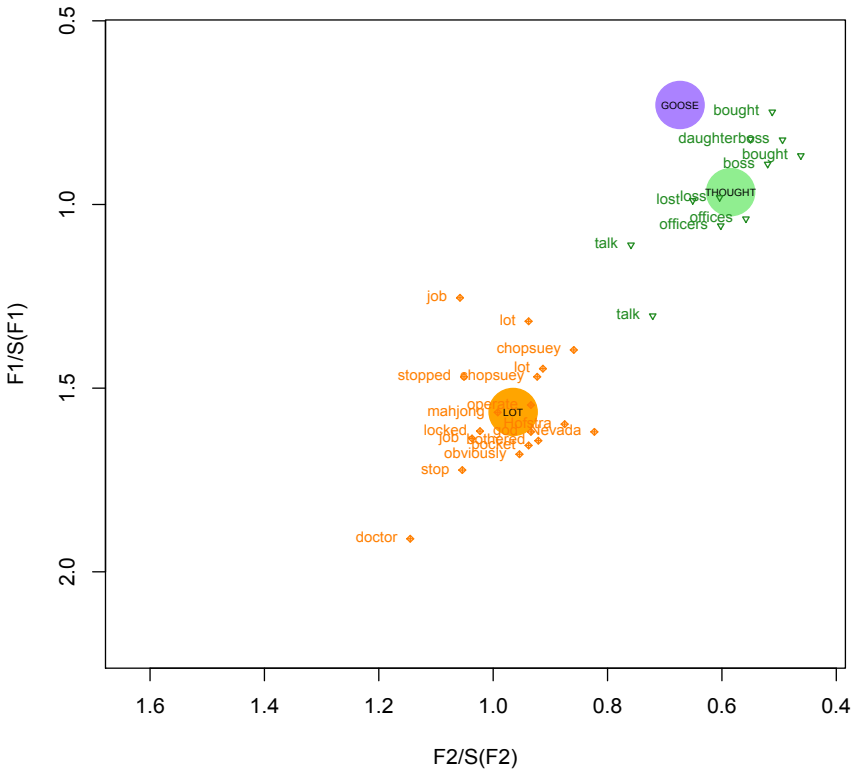

Figure 1: LOT, THOUGHT tokens of George (born 1949) against normalized F1/F2 means of GOOSE.

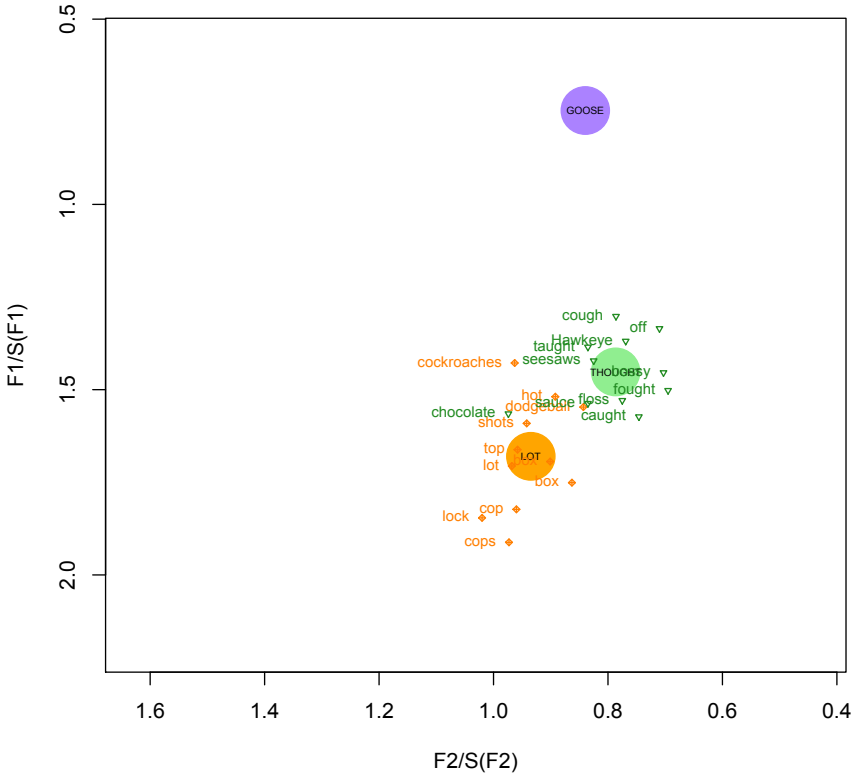

Figure 2: LOT, THOUGHT tokens of James (born 1997) against normalized F1/F2 means of GOOSE.

The results show that the oldest speakers, George and Paul, produced the highest THOUGHT whereas the younger speakers such as Jeff and James produced the lowest THOUGHT. Figure 1 is a plot of George's THOUGHT tokens against his LOT and GOOSE vowels.

It is obvious from the plot that most of George's THOUGHT tokens were produced with raising. Some of his THOUGHT tokens (e.g., *bought*, *boss,* and *daughter*) are almost as high as the mean height of his GOOSE vowel. This is in sharp contrast to the vowel plot of one of the younger speakers, James, who produced THOUGHT with no raising at all (Figure 2). Figure 3 plots the height of all THOUGHT tokens analyzed (on the y-axis) against speaker's year of birth (on the x-axis), fitted with a regression line. There is a significant and positive correlation ($r = .621$, $p < .0001$) between speakers' year of birth and the F1 of THOUGHT. As speaker's year of birth increases, the lower the THOUGHT class is in the vowel space (i.e., the higher the normalized values on the y-axis).

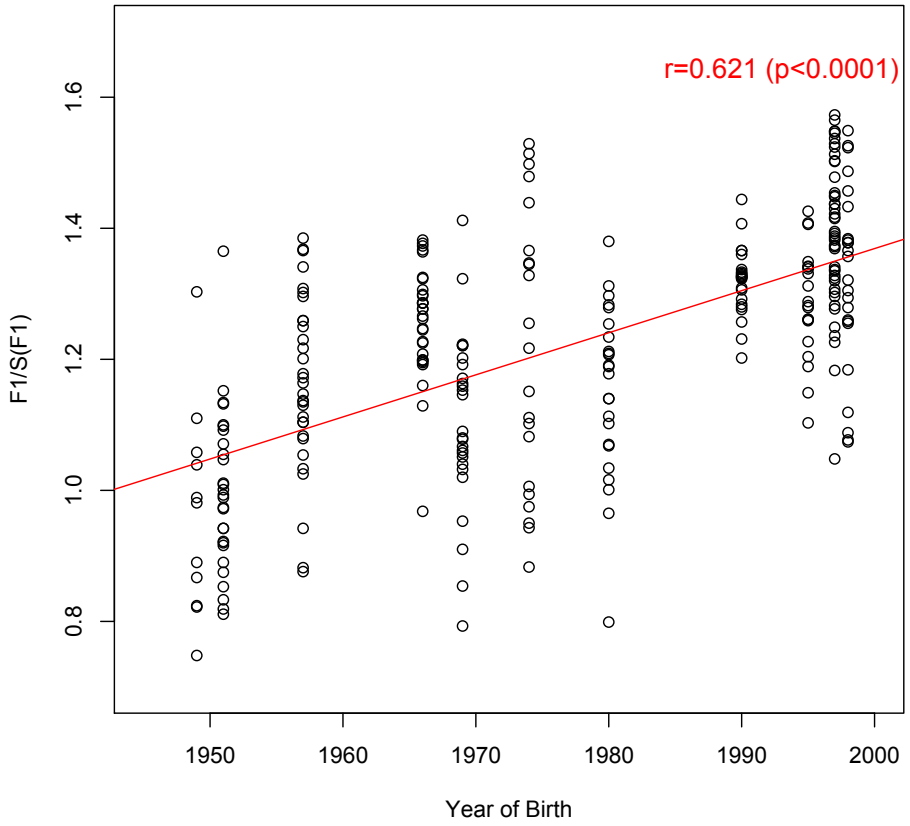

Figure 3: Height of all THOUGHT tokens by speakers' year of birth.

A mixed model regression analysis was performed in Rbrul (Johnson 2009) to more accurately model the effects of speakers' year of birth on the height of THOUGHT while taking into consideration possible phonological conditioning and individual variation due to sampling effects. The response variable was the continuous normalized F1 values. Four predictor variables were considered: (1) following phonological environment, (2) year of birth, (3) stylistic context, and (4) an interaction between following phonological environment and year of birth. Speaker and lexical item were entered as random effects.

The model that best accounts for the variation in the height of THOUGHT contains three fixed predictors: year of birth, following phonological environment, and the interaction between year of birth and following phonological environment. Stylistic context was not a significant predictor, further justifying grouping these tokens in the analysis. The model begins with an intercept (0.982) which is set at 1940 as the year of birth. The coefficients within a given significant predic-

tor are then added to this intercept. Since the analysis finds a significant interaction effect between speaker's year of birth and following phonological environment, considering year of birth and following phonological environment as separate main effects would provide an incomplete picture. It is therefore necessary to consider these two main effects in conjunction. One way to do this is to revise the interaction coefficients by adding the year of birth coefficient to each of the interaction coefficients. The revised interaction coefficients are given in Table 2b. They tell us how each increase in speaker's year of birth affects the phonological environments differently.

The fact that all the revised interaction coefficients are positive provides apparent time evidence of change towards lowered THOUGHT regardless of phonological environments. However, the extent of lowering over time differs between phonological environments. THOUGHT followed by /k/ shows less drastic lowering over apparent time as indicated by the smallest positive coefficient while THOUGHT followed by /t/ or in the word final position shows the most lowering over time.

Deviance	**(df)**	**Intercept**		**Grand Mean**	
-414.83	**13**	**0.982**		**1.219**	
Predictors	**Significance**	**Linear Coefficients**		**N**	**Raw Mean**
Year of Birth	$p < .0001$	+1	0.007		
Following Environment	$p = .0003$	__t (*taught*, *fought*)	-0.134	86	1.164
		__# (*saw*, *law*)	-0.114	8	1.159
		__s (*cause*, *boss*)	-0.021	92	1.207
		__f (*cough*, *off*)	0.058	48	1.233
		__k (*talk*, *chalk*)	0.212	49	1.331
Year of Birth * Following Environment	$p = .0002$	+1 * __k (*talk*, *chalk*)	-0.003		
		+1 * __f (*cough*, *off*)	-0.001		
		+1 * __s (*cause*, *boss*)	0		
		+1 * __t (*taught*, *fought*)	0.002		
		+1 * __# (*saw*, *law*)	0.003		

Table 2a: Results from the best run in a mixed model regression analysis on THOUGHT height.

Predictor	**Significance**	**Revised Linear Coefficients**	
Year of Birth * Following Environment	p = 0.0002	+1 * __k (*talk*, *chalk*)	0.004
		+1 * __f (*cough*, *off*)	0.006
		+1 * __s (*cause*, *boss*)	0.007
		+1 * __t (*taught*, *fought*)	0.009
		+1 * __# (*saw*, *law*)	0.010

Table 2b: Revised interaction coefficients from the mixed model regression analysis in 2a.

This section on the height of THOUGHT shows that older New Yorkers of Chinese background produced raised THOUGHT, similar to older New Yorkers of other ethnic backgrounds sampled in earlier studies (Becker 2010, Labov 1966, Labov et al. 2006, *inter alia*). In addition, there is a statistically significant lowering of raised-THOUGHT across time among these speakers of Chinese descent, corroborating the trend observed in Becker (2010) for Asian American speakers. The lowering of raised-THOUGHT by younger speakers raises the question of whether these speakers continue to maintain the low-back distinction. The following section examines the position of THOUGHT in relation to LOT for these sixteen speakers.

4 Distinction Analysis

To quantify the magnitude of distinction (or overlap) between a speaker's LOT and THOUGHT tokens that are unevenly distributed across phonological contexts, this study adopts the Pillai-

Bartlett trace measurement. The use of Pillai-Bartlett trace as a measurement of vocalic distinction was introduced by Hay, Warren, and Drager (2006) in their study of the NEAR-SQUARE merger in New Zealand English and later adopted by Hall-Lew (2009, 2010) to examine the low-back merger and the fronting of high and mid-back vowels in a San Francisco neighborhood. The MANOVA analysis takes into consideration both between-group and within-group variability along the height and anteriority dimensions simultaneously. To examine the magnitude of LOT/THOUGHT distinction for each speaker, normalized F1 and F2 were entered as the response variables to be considered simultaneously in the MANOVA test. The two predictor variables were following phonological environment and word class. Crucially, following phonological environment was entered into the model before word class so that variation due to known phonological conditioning would be accounted for before determining if there is a significant contrast between the two word classes.

Two relevant outputs from the analysis include the *F*-value and the significance level. The *F*-value (henceforth, Pillai score, following Hay et al. (2006) and Hall-Lew (2009)) can be interpreted as a summary of the extent to which the two word classes are statistically distinct. The Pillai score ranges from 0 to 1; a speaker who receives a Pillai score that is closer to 1 suggests that he maintains a relatively robust distinction between LOT and THOUGHT. The distinction is robust either because there is a greater distance between the two word classes, or because there is little variability within each word class, or because of both. Conversely, a speaker with a Pillai score closer to 0 suggests that he has a weaker LOT/THOUGHT distinction, either due to smaller differences between the two classes, or because of high degree of variability within one or both classes. Another relevant output from the MANOVA is the *p*-value. The *p*-value estimates whether the difference between tokens can be significantly predicted by word class membership. A significant *p*-value identifies those speakers whose LOT and THOUGHT are statistically distinct (see Hall-Lew 2010 for more details).

Table 3 is a summary of the MANOVA results. The MANOVA results are arranged by speaker's year of birth, with the oldest speaker, George, at the top and the youngest two speakers, Tommy and Kelvin, at the bottom. The asterisks (*) after a Pillai score indicates that it is statistically significant. Non-significant Pillai scores are in gray. In addition to the Pillai score and the *p*-value, the Euclidean distance (ED) between LOT and THOUGHT was also included as a reference. The ED in this study was calculated using each speaker's normalized F1 and F2 means of LOT and THOUGHT, averaged across different phonological environments.

Speaker	Year of Birth	Pillai Score	*p*-value	ED
George	1949	0.887***	< 0.0001	0.705
Paul	1951	0.815***	< 0.0001	0.626
John	1957	0.782***	< 0.0001	0.598
Chris	1966	0.600***	< 0.0001	0.281
Norman	1969	0.804***	< 0.0001	0.618
Joseph	1974	0.704***	< 0.0001	0.493
Ernie	1980	0.795***	< 0.0001	0.455
Tim	1990	0.595***	< 0.0001	0.23
Avery	1995	0.678***	< 0.0001	0.287
Nick	1995	0.685***	< 0.0001	0.274
James	1997	0.470*	< 0.05	0.272
Michael	1997	0.567*	< 0.05	0.244
Simon	1997	0.208	> 0.05	0.146
Jeff	1997	0.245	> 0.05	0.143
Tommy	1998	0.524**	< 0.001	0.391
Kelvin	1998	0.740***	< 0.0001	0.286

Table 3: Pillai score, *p*-value, and Euclidean Distance between THOUGHT and LOT by speaker.

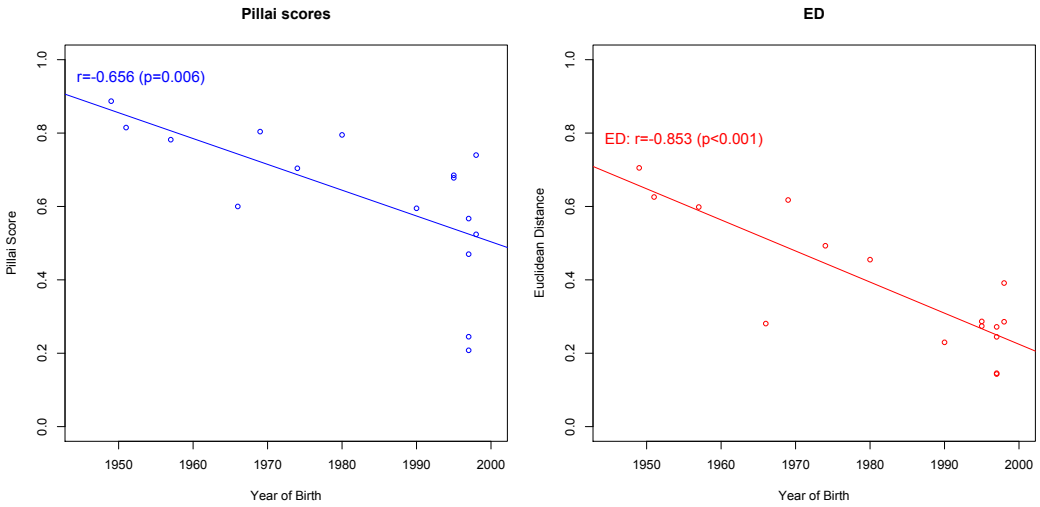

Figure 4: Pillai scores and ED by year of birth.

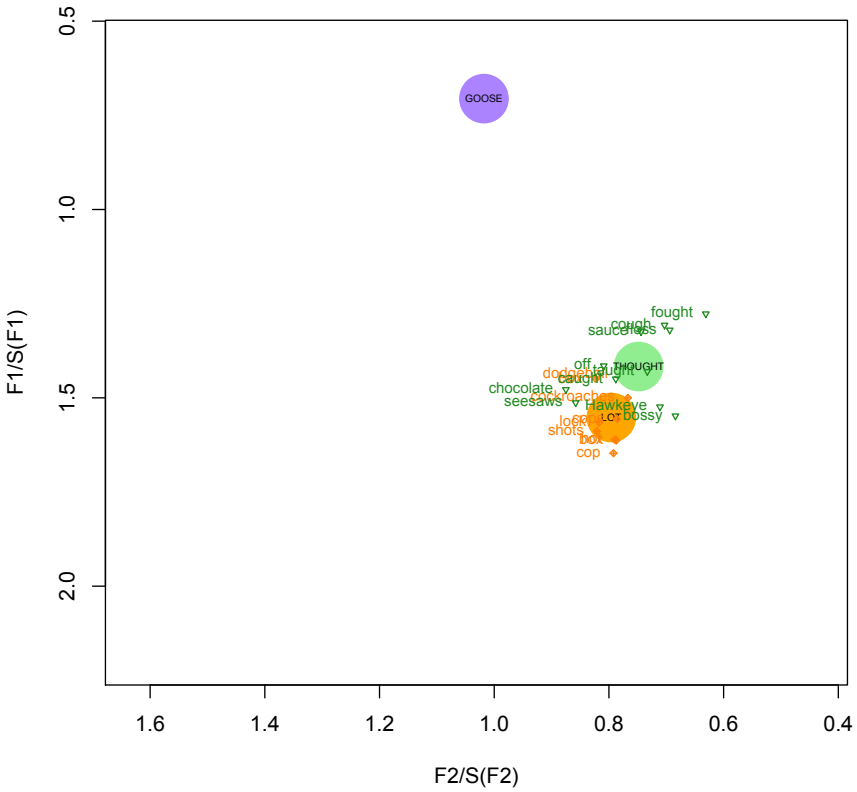

Figure 5: LOT, THOUGHT tokens of Jeff (born 1997) against normalized F1/F2 means of GOOSE.

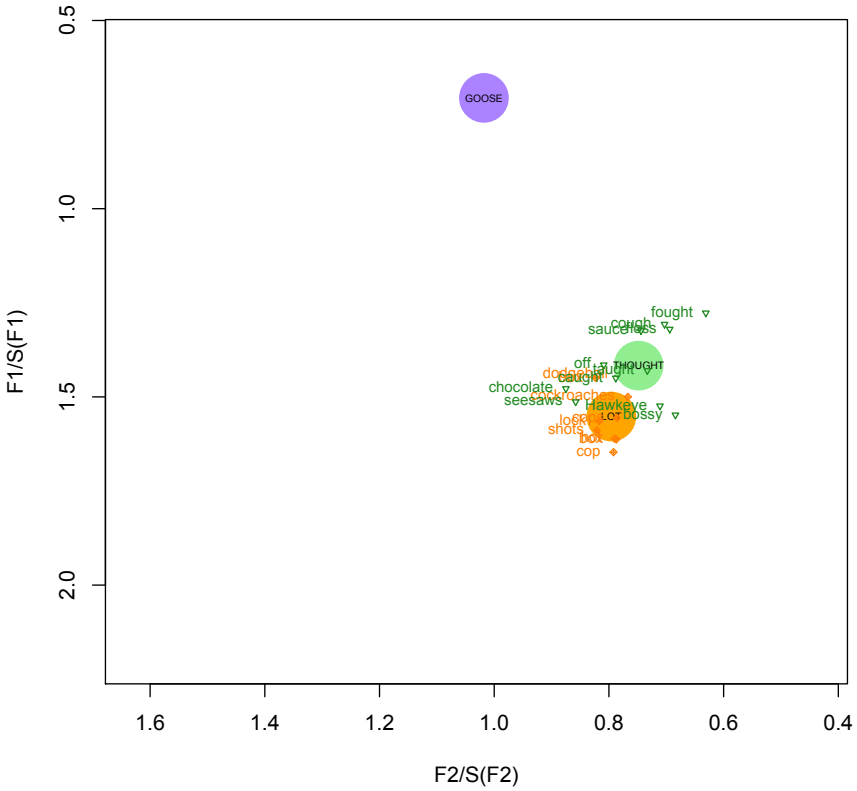

Figure 6: LOT, THOUGHT tokens of Simon (born 1997) against normalized F1/F2 means of GOOSE.

Speaker's year of birth is negatively correlated both with the Pillai scores (r = .656, p = .006) and the ED (r = .853, p < .001), as shown in Figure 4. The younger the speaker is, the smaller the distance and degree of distinction between his LOT and THOUGHT classes. There is also a strong and positive correlation between the Pillai scores and the ED (r = .804, p = .0002). The two measurements, however, are not in a perfectly linear relationship. While a smaller ED generally corresponds to a smaller Pillai score, there are cases in which a small ED is associated with a high Pillai score, as in the cases of Chris and Kelvin. In other words, these speakers appear to be maintaining a rather robust low-back distinction despite having relatively close LOT and THOUGHT classes in the vowel space. When evaluating the degree of distinction or overlap between the vowel classes, the ED and Pillai scores can be used as complementing measurements.

The p-values listed in Table 3 further confirm that most speakers in the sample continue to maintain the low-back distinction, including those who produced THOUGHT with little or no raising. For 14 out of 16 speakers, the two vowel classes are statistically distinct. The main effect of word class fails to reach statistical significance for only two of the younger speakers, Jeff and Simon. The non-significant p-values, however, should not be taken immediately as evidence that their low-back vowels are merged. As pointed out in Hall-Lew (2010), a non statistically significant p-value is not sufficient in teasing apart speakers whose low-back vowels are nearly merged from those who are completely merged. Visual inspection of the vowel plots of these two speakers, in fact, reveal that their low-back vowels, although very close, are not completely overlapping, unlike the low-back vowels configuration in cases of merger (Labov et al. 2006). Jeff's vowel plot (Figure 5) shows a rather dispersed THOUGHT vowel. While some of Jeff's THOUGHT tokens like *fought*, *cough,* and *floss* are higher and more backed than his LOT class, there are also tokens such as *bossy* and *Hawkeye* that are overlapping with his LOT class. Simon's vowel plot (Figure 6) shows an even more dispersed THOUGHT class. While some of his THOUGHT tokens like *bossy, sauce*, *floss* remain higher and more backed than his LOT tokens, there are quite a few THOUGHT tokens that overlap with LOT, such as *Hawkeye*, *chocolate*, and *cough*.

Visual inspection of these vowel plots reveals that the non-significant *p*-values obtained for Simon and Jeff are likely due to the high variability within their THOUGHT class. In other words, even though the main effect of word class does not reach statistical significance for these two speakers, it does not constitute sufficient evidence at this point to claim that these speakers are leading a merger in New York City. The low-back vowels configuration found in Jeff and Simon resembles instead the patterns exhibited by the "transitional speakers" in the Midland area in Labov et al. (2006). The extent to which the high variability in the THOUGHT class is a result of reclassifications of lexical items that have traditionally been in the THOUGHT class (such as *Hawkeye* and *chocolate*) to the LOT class remains an empirical question.

5 Conclusion and Further Questions

The analysis on the height of THOUGHT in this paper provides evidence of the lowering of raised-THOUGHT across apparent time among speakers of Chinese descent. This result corroborates the trend observed in the European American and Asian American speakers in Becker (2010). Results from the distinction analysis find that despite the lowering of THOUGHT, most, if not all, speakers continue to maintain the low-back distinction. Some speakers, especially the ones born after 1990, have LOT and THOUGHT classes that are very close in the vowel space with some overlapping tokens (although the mean normalized F1 and F2 of the two vowel classes are not completely overlapping). The trend identified in this sample of Chinese New Yorkers is therefore better described as a reversal of THOUGHT-raising at this point. One immediate research question that follows is the extent to which the lowering of THOUGHT is driven by the salient indexical meanings associated with raised-THOUGHT (Becker 2010, 2011, Wong and Hall-Lew 2012).

Furthermore, describing the pattern identified in this paper as the reversal of raised-THOUGHT also leaves open the question on the future trajectory of the THOUGHT vowel vis-à-vis the robustness of the low-back distinction in New York. Further work involving New York speakers of other social and ethnic backgrounds is needed to determine if the lowering of THOUGHT among younger speakers is in the direction of close but distinct low-back vowels, similar to the pattern found in the South and the Midlands, or if the lowering will continue in the direction towards merger (Labov et al. 2006).

References

Becker, Kara. 2010. Regional Dialect Features on the Lower East Side of New York City. Doctoral dissertation, New York University.

Becker, Kara. 2011. The social meaning(s) of raised BOUGHT in New York City: A perceptual approach. Paper presented at NWAV40, Georgetown University.

Cohen, Paul. 1970. The tensing and raising of short [a] in the metropolitan area of New York City. M.A. thesis, Columbia University.

Fabricius, Anne, Dominic Watt, and Daniel Ezra Johnson. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change* 21:413–35.

Hall-Lew, Lauren. 2009. Ethnicity and Phonetic Variation in San Francisco English. Doctoral dissertation, Stanford University.

Hall-Lew, Lauren. 2010. Improved representation of variance in measures of vowel merger. *Proceedings of Meetings on Acoustics*.

Hay, Jennifer, Paul Warren, and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34:458–484.

Labov, William. 1966. *The Social Stratification of English in New York City.* Cambridge: Cambridge University Press.

Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English.* Berlin: Mouton de Gruyter.

Thomas, Erik, and Tyler Kendall. 2007. *NORM: The vowel normalization and plotting suite.* URL http://ncslaap.lib.ncsu.edu/tools/norm/.

Thomas, Erik. 2011. *Sociophonetics: An Introduction*. New York: Palgrave Macmillan.

Watt, Dominic, and Anne Fabricius, 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1 ~ F2 plane. In *Leeds Working Papers in Linguistics and Phonetics* 9, ed. D. Nelson, 159–73.

Wong, Amy W.-m., and Lauren Hall-Lew. 2012. Regional variability and ethnic identity: Chinese Americans in San Francisco and New York City. Paper presented at the LSA 86th Annual Meeting, Portland, OR.

Department of Linguistics
New York University
10 Washington Place
New York, NY10003
amywong@nyu.edu