9-2020

Cot in the Act: Ethnicity and Age Affects Phonemic Perception of the Low-Back Merger in New York City English

Omar Ortiz
The Graduate Center, City University of New York

Cot in the act: Ethnicity and age affects phonemic perception of the low-back merger in New York City English

by

Omar Ortiz

A master's thesis submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2020

Cot in the act: Ethnicity and age affects phonemic perception of the low-back merger in New York City English

by

Omar Ortiz

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the thesis requirement for the degree of Master of Arts.

_____	_____
Date	William Haddican Thesis Advisor
_____	_____
Date	Gita Martohardjono Executive Officer

The City University of New York

ABSTRACT

Cot in the act: Ethnicity and age affects phonemic perception of the low-back merger in New York City English

by

Omar Ortiz

Advisor: William Haddican

This paper is an experimental study on how perceptions about a speaker's age and ethnicity may influence whether listeners perceive the THOUGHT / LOT distinction. The macro-categories of age and ethnicity have been found to correlate with the lowering of raised THOUGHT (Wong 2012, Becker 2010) and the favoring of the merged vowels in perception (Haddican et al. 2016). This thesis examines whether images of faces associated with different age and ethnicity categories condition perception of auditory stimuli as belonging to either the LOT or THOUGHT class. This thesis builds on previous results suggesting that non-linguistic information influences speech perception (Rubin 1992, Hay, Warren and Drager 2006), and one specific line of work in this area has shown that being shown a photograph of a person of a particular ethnic background affects listeners' comprehension of their speech.

ACKNOWLEDGMENTS

I would like to express my first thanks to Bill Haddican, without him this thesis would not be possible. Thank you Bill for helping me throughout this entire thesis, and always being so encouraging and kind. I also want to thank Charlotte Vaugh and Melissa Baese-Berk who first got me interested in Linguistics in undergrad, and who gave me my first paid job working in a linguistics lab. I would like to thank Michael Stern, LeeAnn Stover, Kyle Gorman and Cass Lowry for all your help with stats, big and small. Thank you to my dear friends Tysean Bucknor and Meredith Hilliard who studied with me almost every night our first year together. A special thank you to my dog Kuro, who left this world too early and was supposed to be with me all through graduate school. Kuro, you got me through many nights of studying, your comfort made it easier to not get too stressed, I love you. Lastly, I would like to thank my mother, sister, brother, and step-father who have supported me in my long academic journey.

Contents

List of Tables

List of Figures

1. Introduction

1.1 Raised THOUGHT and Low-back Merger

A well-known traditional feature in New York City English (NYCE) is having a distinction between THOUGHT ([ɔ]) and LOT ([ɑ]) lexical sets. The low-back merger (THOUGHT and LOT having the same value in production and/or perception) has been documented in many parts of the United States and is reported to be spreading at the expense of distinctions in many areas (Labov, Ash, and Boberg 2006, Wong 2012). The resistance to the low-back merger (keeping the distinction) by native New-Yorkers have also been documented (Newman 2014) but has been reevaluated by recent work (Haddican et al. 2016) that show a diffusion of the low-back merger in apparent time across all five boroughs, with Man
hattan and Queens having the highest merger rates. (See also Johnson 2010.) Merger is the process in which a phonemic contrast (two distinct phonemes) come to be realized and/or perceived as one single phoneme (Hay, Warren, and Drager 2005).

Among conservatie NYCE speakers, THOUGHT is sharply raised. Raised THOUGHT can be calculated by looking at F1 values, with values less than 700hz being considered raised (Becker 2010). The lowering of raised THOUGHT has been documented across apparent time among Chinese speakers (Wong 2012) but also by young, White, and Jewish speakers (Becker 2010). In both studies, the younger speakers have the smallest distance and distinction between the vowels (almost merged). Interestingly, in both studies, the distinction is there for almost all participants (except for two younger speakers in Wong's (2012) results who have fully merged the vowels). However, the lowering of raised THOUGHT through apparent time can suggest a diffusion of a low-back merger to come.

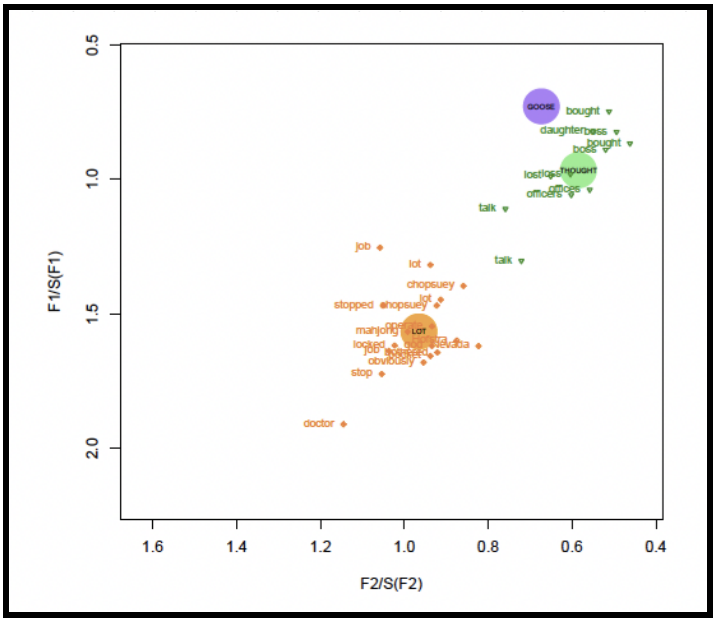

Figure 1: From Wong (2012). Participant born in 1949 with a high THOUGHT (green).

Figure 1 shows an older participant in a study by Wong (2012). The green bubble THOUGHT is much higher than the orange bubble LOT. This participant has a robust distinction between THOUGHT and LOT vowels.

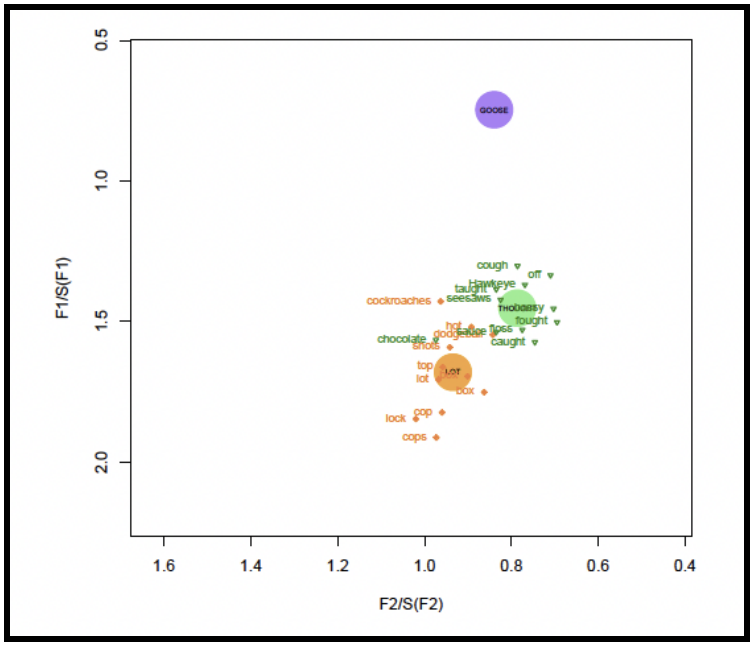

Figure 2: From Wong (2012). Participant born in 1997 with a much lower THOUGHT (green).

Figure 2 shows a younger participant in a study by Wong (2012). The green bubble THOUGHT is not much higher than the orange bubble LOT. This participant does not have a robust distinction between THOUGHT and LOT vowels. Although this participant still has a distinction in production, it is much smaller, showing the lowering of raised THOUGHT.

The macro-categories of year of birth and ethnicity have been found to correlate with the lowering of raised THOUGHT (Wong 2012, Becker 2010) and the favoring of the merged vowels (Haddican et al 2016). Younger Chinese and White speakers tend to favor lowering in studies by Wong (2012) and Becker (2010), and younger Latinos and East Asians favoring the merger in studies by Haddican et al (2016). Although White speakers had the lowest THOUGHT (Becker 2010), they were followed by Puerto Ricans (Latinx) and Chinese (East Asian), which is in line with those that favor the merger, being Latinx and East Asians (Haddican et al 2016). Wong (2012) found that the older participants had much higher THOUGHT vowels (Figure 1) compared to the younger participants (Figure 2). As previously mentioned, although only two participants were fully merged (where THOUGHT and LOT overlapped on a graph), the lowering of raised THOUGHT by younger speakers could imply a merger-in-progress of THOUGHT and LOT vowels.

Social correlates of linguistic change, or resistance to change, have been described in many studies, most influentially by Labov's 1963 Martha's Vineyard study, where positive attitudes towards the island had correlations with keeping the local areas distinct language features in people's speech. An explanation for this change is proposed by looking into third wave

sociolinguistics. “The third wave focuses on the indexical nature of sociolinguistic variation and on the stylistic practice in which variables gain their meaning” (Eckert 2016). A matched-guise design was used by Becker (2010) to find that raised THOUGHT indexed a *Classic New Yorker* who is older, white, mean and aloof. An explanation for the distancing from raised THOUGHT and potentially a merger is to create a new identity and distance oneself from the overarching negative traits associated with these once prominent NYCE features.

1.2 Social information and speech perception

Many previous studies have shown that non-linguistic information influences speech perception (Rubin 1992, Hay, Warren and Drager 2005). When told a speaker is from Canada, participants have reported hearing speech sounds associated with Canadian-English (raised diphthongs) compared to if told the person was from Detroit, although both speech samples were from the same Detroit born speaker (Niedzielski 1999). One specific line of work in this area has shown that being shown a photograph of a person of a particular ethnic background affects listeners’ comprehension of their speech. Babel and Russell (2015) report that listeners shown a Chinese-face reported less intelligibility over a White-face, and the Chinese-face was reported to be more accented than the White-face, although both speech samples were from the same Canadian born speaker .

Other vowel features have been examined like TRAP backing, which looks at the distinction between words like *sack* and *sock* (D’Onofrio 2015). Findings show that not only macro-categories like place of birth (California) are more associated with TRAP vowels compared to someone not given any speaker information, but that *persona-based* information

is highly salient in the connection to the TRAP vowel and while being told the person was a *valley-girl* had the strongest affect associated with the TRAP vowel.

From this, we can see that many pieces of listeners' prior knowledge come into play when listening to speech. The use of photographs has the assumption that participants' stereotypes about each ethnicity are triggered with the presentation of the photographs. However, looking at one macro-category like ethnicity is not enough, as we know that individuals are more than just their ethnicity.

1.3 Goals of this study

The goal of this study is to examine whether native New Yorkers give merged or distinct judgements on THOUGHT and LOT vowels while looking at specific photographs. Specifically, this study examines whether photographs linked to non-linguistic features like specific ethnicity, age, and place of birth have an effect on judgements of phonemic categories. This will also allow us to examine not only if NYCE speakers are aware of the potential merger, but if they are conscious or subconsciously aware of who is leading the change. In particular the expectation raised by previous production studies (Wong 2010, Becker 2010), matched guise (Becker 2014), and homophony judgements studies (Johnson 2010, Haddican et al. 2016) is that the younger looking photographs, and the East Asian and Latinx photographs will be rated as merged more often than older, and White faces.

The first hypothesis tested in this study, then, is that ethnicity is correlated with the perception of the low-back merger, specifically, that Latinx and East Asian faces will have more merged ratings than White faces. The second hypothesis is that age is correlated with the

perception of the low-back merger, specifically, that younger looking photographs will have more merged ratings than older looking photographs.

2. Methods

2.1 Design

Using Qualtrics, an online survey generator, participants (Explained in section 2.2) took part in a binary forced perception task, modeled after Hay, Warren, and Drager's (2015) study of NEAR and SQUARE vowels (Explained in section 2.5.1). All guises with photographs (see section 2.4) were noted being native New Yorkers. All six voices have a distinction between the low-back vowels THOUGHT and LOT (see section 2.3). Lastly, participants selected whether pairs of LOT/THOUGHT words sound the "same" or "different" to them. This is done to see if any correlations exist between individual participants who perceive the merger in homophony judgments and those who have the distinction on the perception task.

2.2 Participants

Participants were recruited through the Human Subjects Pool at Queens College, where a total of 140 students took the survey, in the spring of 2020. Removing subjects who didn't complete the entire survey, completed the survey too quickly to be valid, or answered our filler question incorrectly, left analyzable data from 84 subjects for Task 1, and 74 subjects for task 2 (10 people who completed Task 1 did not complete Task 2).

The only strict requirements to take the survey was that they had to be a native New Yorker and between the ages of 18 and 65. This also allowed us to have a wide range of ages

and ethnic backgrounds. Participants were awarded a required course credit for their participation in research at Queens College.

2.3 Audio stimuli

Voices from six female volunteers [1] that have the low-back distinction were recorded. They were all native New Yorkers from across all the boroughs, in their late twenties to early forties. They were recorded in a quiet environment using a Zoom H4n Pro Handy Recorder. They read a word list consisting of six THOUGHT and six LOT vowels, randomized with twelve other words (that do not have the vowels in question) totaling twenty-four words (Table 1). They also read twenty-four sentences, twelve having words with THOUGHT and LOT vowels, and twelve that do not (Table 1).

Wordlist			Sentences	
Auto	Goat		I said the word auto twice	I said the word goat twice
Ban	Goose		I said the word ban twice	I said the word goose twice
Bot	Otto		I said the word bot twice	I said the word otto twice
Bought	Peace		I said the word bought twice	I said the word peace twice
Bug	Plate		I said the word bug twice	I said the word plate twice
Cat	Rag		I said the word cat twice	I said the word rag twice
Caught	Sam		I said the word caught twice	I said the word sam twice
Cot	Strut		I said the word cot twice	I said the word strut twice
Dawn	Talk		I said the word dawn twice	I said the word talk twice
Don	Taught		I said the word don twice	I said the word taught twice
Face	Tock		I said the word face twice	I said the word tock twice
Fleece	Tot		I said the word fleece twice	I said the word tot twice

Table 1: Wordlist created from most common THOUGHT and LOT vowels used in previous research, with twelve additional arbitrary words. Sentence structure "I said the word ___ twice" has been used by Haddican (2016).

[1] volunteers recruited through word of mouth, with no compensation given.

For each speaker, one token each was selected for six THOUGHT and six LOT words, from either the word lists or sentences. The final selections were made by choosing the most clear sounding, and matching the degree of loudness and similarity to one another. Praat was used to import the audio files, create textgrids, and splice the tokens (bolded words from Table 1) to extract twelve total words from each volunteer. Filler words (and sentences) were used to try and elicit natural sounding speech. No filler words were used in the binary forced choice task.

F1 and F2 formants were extracted from all six voices using DARLA, which is a web-based application for extracting vowel formants from audio samples (Rosenfelder et al. 2014). The acoustic analysis revealed that indeed all six of our speakers had the distinction, and were not merged. Although all six speakers are not merged, the plots show us that some are more distinct than others, figure 3 shows two of our speakers who we could consider to show signs of lowering of raised THOUGHT, while still resisting the low-back merger completely. Figures 4 and 5 show that for the other four speakers, they are quite distinct and have resisted the low-back merger.

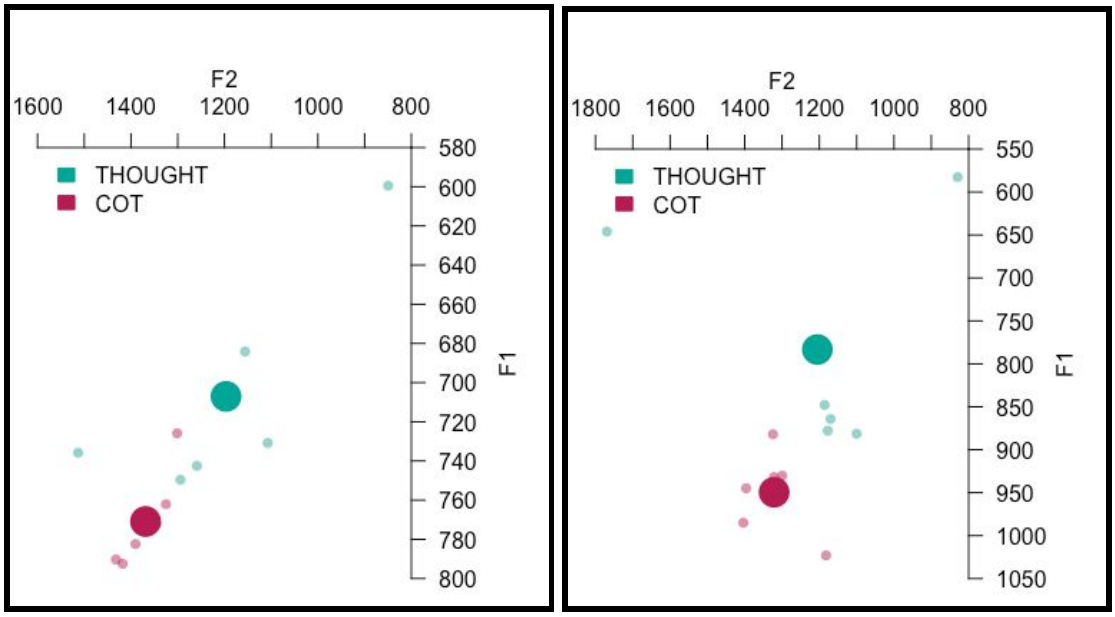

Figure 3: Speaker 1(left) and Speaker 2(right) who are distinct, although raised THOUGHT does appear to be lowering.

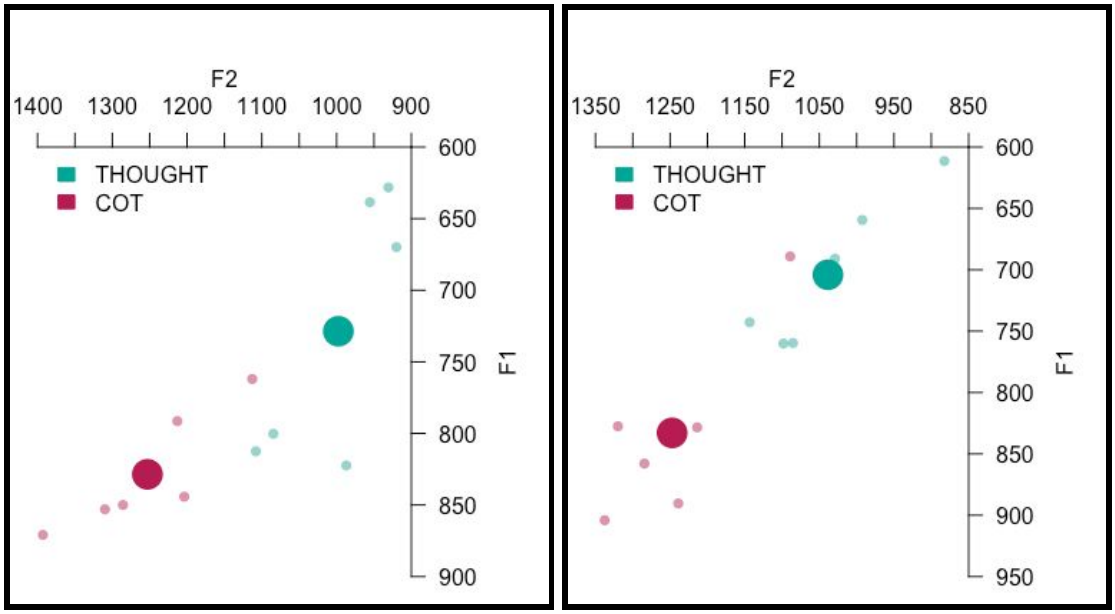

Figure 4 : Speaker 3 and Speaker 4's F1 and F2 plots, showing they produce the THOUGHT/COT distinction.

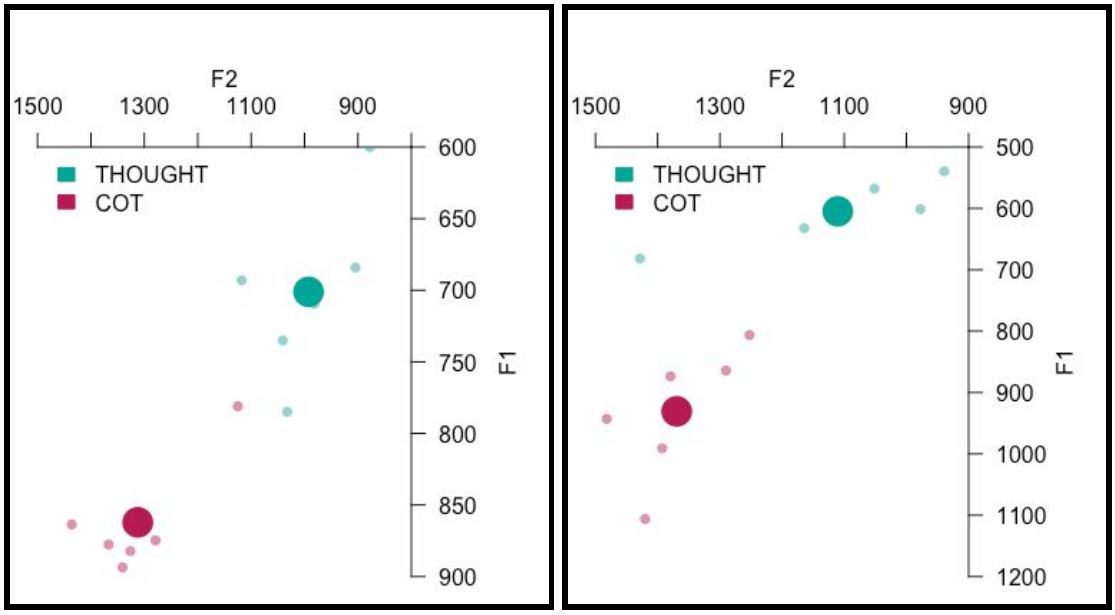

Figure 5 : Speaker 5 and Speaker 6's F1 and F2 plots, showing they produce the THOUGHT/COT distinction.

High frequency words are recognized faster than low frequency words (Hulme et al, 1997). For this reason, it should be noted that all of our THOUGHT words (besides "dawn") have a higher word frequency than our LOT words. Although high frequency words are recognized faster than low frequency words, this is just one of many reasons as to why it might be possible that the THOUGHT words will be perceived "correctly" (i.e. matching the speaker's target) than the LOT words. The mergers in progress recognize raised THOUGHT, so LOT words are the ones being merged, therefore they can be perceived as LOT or THOUGHT. This naturally will affect LOT accuracy, as we know those words can be considered merged, and our recorded audio stimuli volunteers all produce raised THOUGHT and aren't expected to be heard as LOT.

Word Frequencies			Log Outputs		
LOT	THOUGHT		LOT		THOUGHT
bot <- 72974	bought <- 1273993		log(bot)		log(bought)
			11.19786		14.05767
cot <- 23869	caught <- 682381		log(cot)		log(caught)
			10.08034		13.43334
don <- 285048	dawn <- 199695		log(don)		log(dawn)
			12.56041		12.20455
tot <- 17788	taught <- 629471		log(tot)		log(taught)
			9.786279		13.35264
tock <- 2810	talk <- 2087909		log(tock)		log(talk)
			7.94094		14.55167
otto <- 36575	auto <- 572336		log(otto)		log(auto)
			10.50712		13.25748

Table 2: Word frequencies and (log) outputs from COCA (Corpus of Contemporary American English).

2.4 Visual stimuli

The photographs were collected from various websites that allow them to be used freely, modified, and shared commercially. The only modifications to the photographs were that they were cropped from the waist up, and all set to a blank white backdrop. All photographs are similarly dressed, and have similar poses. A pre-study design was created and shared among volunteer participants that rated the photographs on perceived ethnicity, age, and believability to be a native New Yorker. The pre-study was implemented twice, with a final total of fifty participants. The final photographs were selected based on the pre-screen, where participants answered questions while looking at a total of twenty photos.

Question 1: What ethnicity do you believe this person to be? (Must only give one answer)

Question 2: How old do you believe this person to be? (Must type one number, no ranges)

Question 3: Do you believe this person could be a native New-Yorker? (Only answer Yes or No)

The final photographs were selected by taking the photographs most likely to be a native New- Yorker, having an ethnicity match to either White, (where answers like "Caucasian" and "European" were considered acceptable) East Asian, (where answers like "Japanese" and "Korean" were considered acceptable) or Latina (where answers like "Mexican" and "Hispanic" were considered acceptable). The original goal for the younger photographs was 18 years old and the older photographs 50 years old. The average age for the three younger photographs was (23.4) and the average age for the three older photographs was (46.5).

For the Younger and Older White faces, 100% of participants selected "White" for ethnicity.[2] For the Younger and Older Latinas, the final two faces selected were the two considered to be the most Latina, with over 80% of participants selecting "Latina" as ethnicity. [3] For the Younger and Older East Asian faces, the final two faces selected were the two considered to be the most East Asian, with over 85% of participants selecting "East Asian" as ethnicity. [4]

[2] Answers like "Caucasian" and "European" were accepted as correct answers for "White".
[3] Answers like "Mexican" and "Hispanic" were accepted as correct answers for "Latina".
[4] Answers like "Japanese" and "Korean" were accepted as correct answers for "East Asian".

Figure 6: Younger (left) and Older (right) White faces.

Figure 7: Younger (left) and Older (right) Latina faces.

Figure 8: Younger (left) and Older (right) East Asian faces.

2.5 Tasks

2.5.1 Task 1 (Instructions)

Participants took the binary forced choice task designed through Qualtrics remotely on their own personal device. Each participant judged all six guises. However, Unlike Hay, Warren, and Drager's (2015) study where participants only viewed one guise, each participant in this current study listened and rate all guises, meaning, each participant viewed an older and younger East Asian face, Latina face, and White face. The six photographs were blocked and pseudorandomized, having a subset of six unique groups that match one photograph paired with one unique voice, meaning, an even number of participants saw Guise 1 paired with Voice 1, and others seeing Guise 1 paired with Voice 2 and so on and so forth. The goal was that each photograph was heard with all six voices equally.

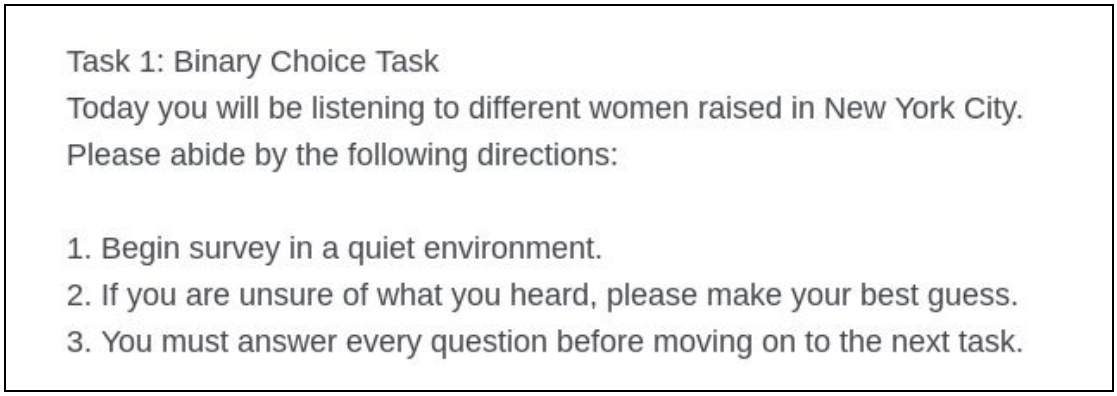

Figure 9: Task 1 instructions

They were told that they were listening to native women from New York, and to select what word they heard out of two choices. Accompanying the photographs were one of the two most common names in the White, Latinx, and East Asian community.

The use of stereotypically common names was used by (Babel and Russell 2015) in their study on Expectations and Speech Intelligibility . They were told to not overthink their judgements, and if not sure to make their best guesses. The photograph remained on screen during the

duration of the task, with their name printed underneath the photograph. They heard a total of twelve different words (six THOUGHT and six LOT), six times, totally seventy-two words heard. In short, they saw the face (with bio), hear the voice, and then choose which word they heard (e.g. LOT or THOUGHT).

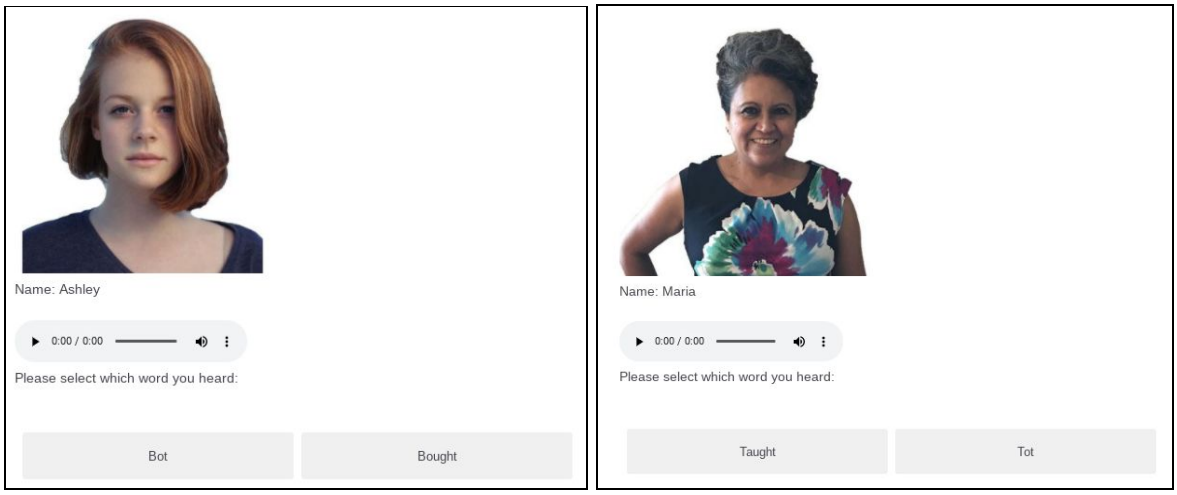

FIgure 10: Examples of the seventy-two binary forced choice decision tasks.

2.5.2 Task 2 (Instructions)

After completing Task 1, participants completed a self-assessment on their own productions of LOT/THOUGHT tokens.

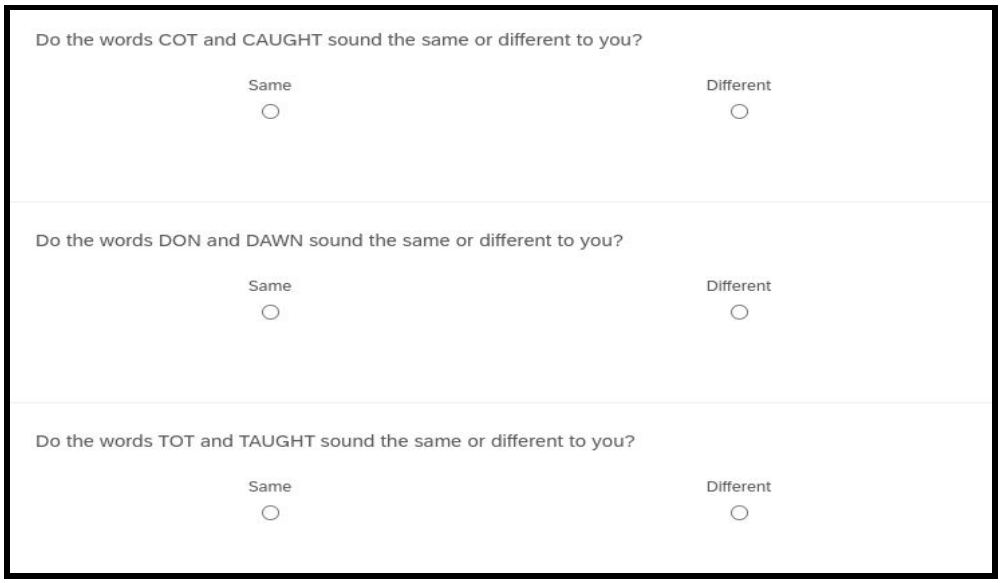

Figure 11: Task 2 example

Participants who answered "Same" to at least 2 of the 3 questions were considered Merged. And those who answered "Different" to at least 2 of the 3 questions were considered Not Merged (Distinct). We used this information to run the statistical models to see if individual merged or not merged had an effect on judgements and to see if they performed differently from one another.

2.5.3 Task 3 (Questionnaire)

The final part of this survey was to collect some information on the participants. We asked them what their ethnicity was, age, in which borough they were born, and their gender. This information was collected, but not used to influence the results.

3. Results

3.1 Binary Forced Choice Perception Task

Data were analyzed by fitting a generalized linear model to the data with Judgment (Correct vs. incorrect) as the dependent variable, and Face ethnicity, Face age, vowel stimulus (LOT vs. THOUGHT) and Participant merger status (merged vs. unmerged) as independent variables.

We found that in Task 1 subjects viewing East Asian and Latina faces are less accurate than those viewing White faces.[5] East Asian faces (p = 0.028) Latina faces (p = 0.0119). Inaccurate judgments are effectively judgments of the stimuli as instantiating a merged system. This is in line with the first hypothesis, that speaker ethnicity is correlated with the perception of the low-back merger, where East Asians and Latinas are perceived as more merged than Whites.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.6975	0.2449	2.848	0.0044
face_ethnAsian	-0.323	0.1478	-2.186	0.0288
face_ethnLatina	-0.3712	0.1476	-2.515	0.0119
face_ageYounger	-0.2078	0.1425	-1.458	0.1449
vowel_typethought	1.2909	0.3293	3.92	8.87E-05
face_ethnAsian:face_ageYounger	0.1558	0.1837	0.848	0.3964
face_ethnLatina:face_ageYounger	0.3831	0.1838	2.085	0.0371
face_ethnAsian:vowel_typethought	0.3034	0.1914	1.586	0.1128
face_ethnLatina:vowel_typethought	0.2067	0.191	1.082	0.2792
face_ageYounger:vowel_typethought	0.0766	0.1562	0.49	0.6238

Table 3: Results using Treatment Coding

[5] max.model = glmer(dep ~ face_ethn * face_age + vowel_type + vowel_type:face_ethn + vowel_type:face_age + (1 | subj_id) + (1 | answer), data = dat.1, family="binomial") summary(max.model)

Table 2 shows all the p-values, Std. Errors, and z values for all the variables. We use α = 0.05 to determine which p-values show significance. Using "White" as a baseline, we see that "face_ethnAsian" gives us a p-value which is less than α = 0.03. For Latina faces "face_ethnLatina" our p-value is even smaller at α = 0.02. The next significance comes from "vowel_typethought" which was so robust that ($p < 0.001$). Somewhat surprising in light of previous research on this merger is that subjects viewing the Younger Latina face (using "older" as another baseline) were more accurate than those viewing older faces. "face_ethnLatina:face_ageYounger" was significant at α = 0.04.

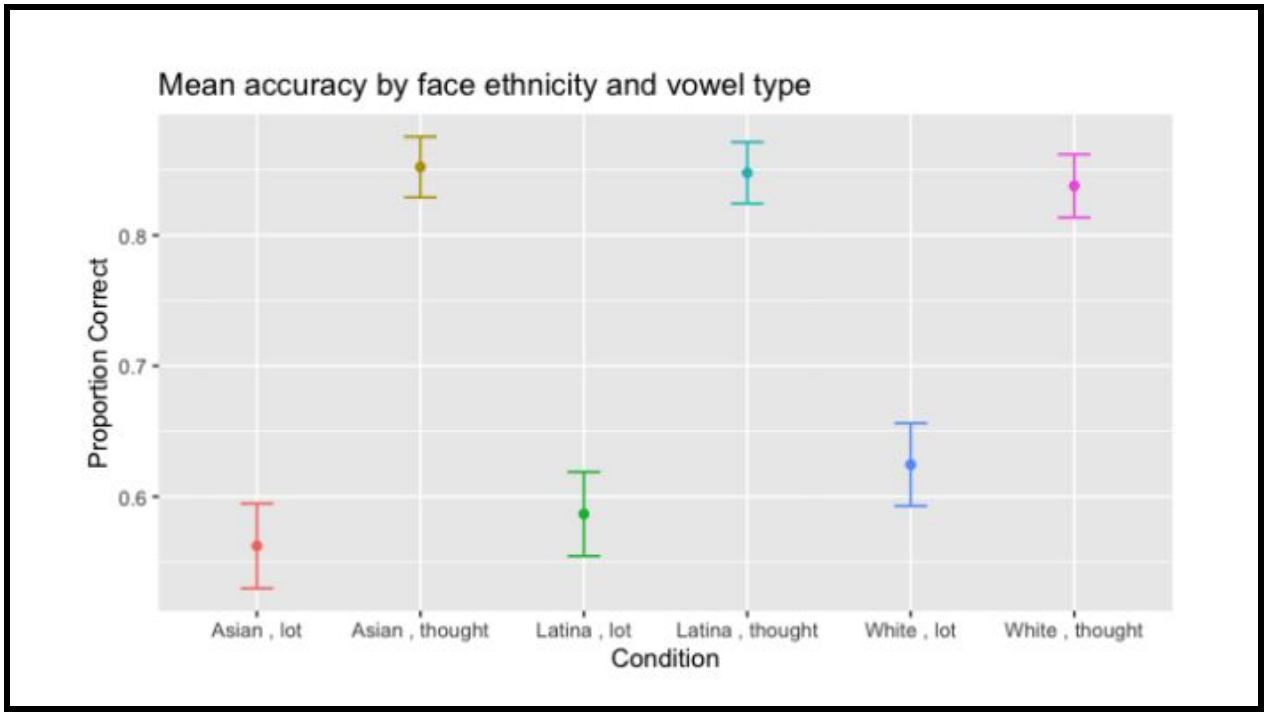

Figure 12: Accuracy based on Ethnicity and vowel type

Figure 9 shows that for the LOT tokens, the blue bar (White) is significantly higher than the green (Latina) and red (East Asian) bars.

We also see that THOUGHT stimuli are judged more accurately than LOT tokens. Vowel type thought ($p = < 0.001$). The finding that THOUGHT tokens are significantly more accurate

than LOT tokens was expected, as LOT tokens are the ones that can potentially be heard as LOT or THOUGHT. Although THOUGHT lowering is an explanation for the merger-in-progress, our audio stimuli volunteers all produced the distinction themselves with raised THOUGHT, so their THOUGHT tokens would not be easily mistaken as LOT tokens.

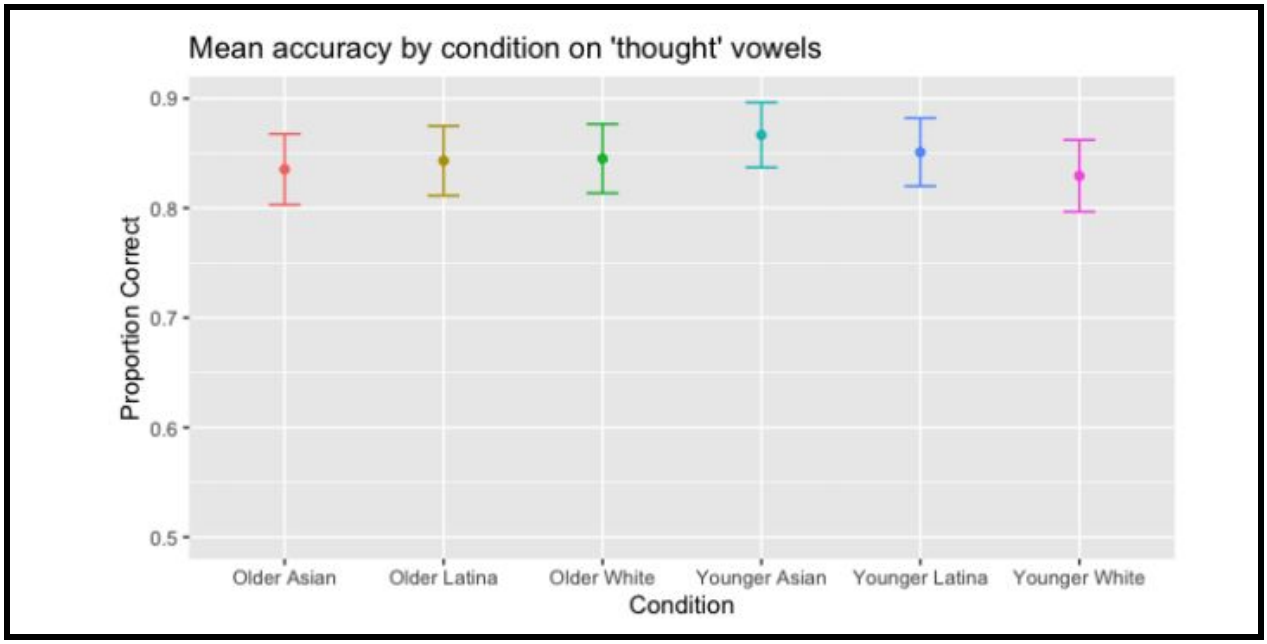

Figure 13: High accuracy across all condition for THOUGHT vowels

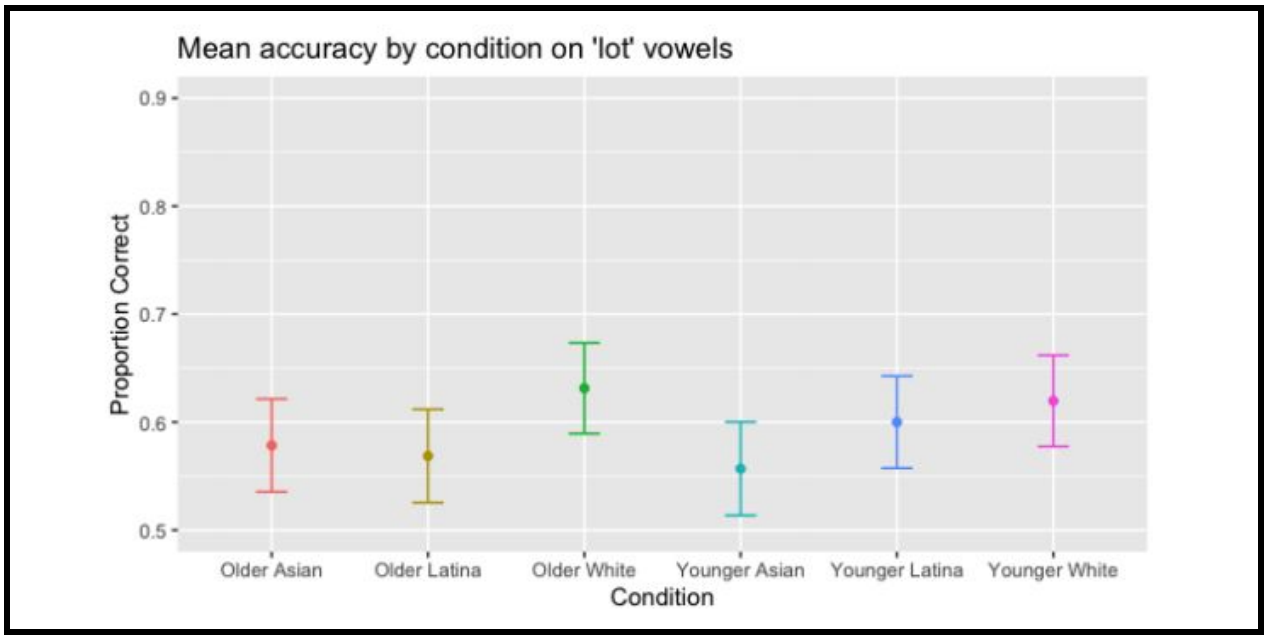

Figure 14: Lower accuracy across all condition for LOT vowels

It was expected for THOUGHT tokens to be much more accurate than LOT tokens, as the LOT words are the ones that are possibly already merged, therefore people can perceive LOT words as THOUGHT words.

We found that among Latina faces, that subjects exposed to younger faces are significantly more accurate than those viewing older faces (p = 0.0371), judgements towards Younger faces being more accurate than older faces means that older faces are perceived as more merged than younger faces. This is not in line with our second hypothesis, that age is correlated with the perception of the low-back merger, where younger faces are perceived as more merged than older faces.

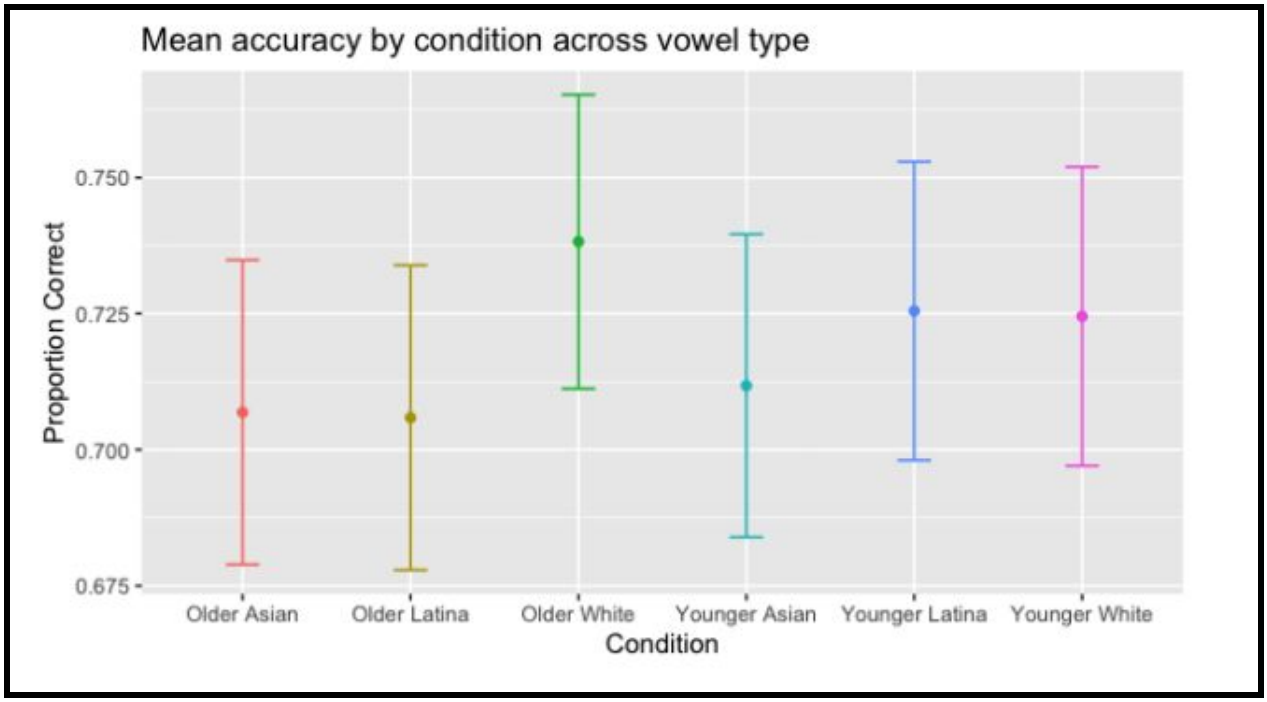

Figure 15: Accuracy based on age.

3.2 Effects of participant's judgements

Using the same Treatment Coding method with the same baseline levels, we found the same effects as those above with slightly different p-values. Judgements towards East Asian and Latina faces are statistically significantly less accurate than White faces. East Asian faces ($p = 0.01297$) Latina faces ($p = 0.02285$).

Judgements towards THOUGHT tokens are significantly more accurate than LOT tokens. Vowel type thought ($p < 0.01$). Among Latina faces, younger faces are significantly more accurate than older faces. Latina face ($p = 0.03674$).

Task 2 was used to find out participants' individual speech, whether they themselves are merged or distinct in regards to the THOUGHT / LOT distinction. Running those models we revealed that those who are not merged (distinct) East Asian and Latina faces less accurate than White faces: East AsianMerged ($p = 0.16607$) LatinaMerged ($p= 0.61819$). Those who were merged, did not show any statistical significance towards either hypothesis.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.77321	0.24967	3.097	0.00195
face_ethnAsian	-0.38266	0.15401	-2.485	0.01297
face_ethnLatina	-0.35036	0.15394	-2.276	0.02285
face_ageYounger	-0.19002	0.14722	-1.291	0.19682
vowel_typethought	1.29187	0.32962	3.919	8.88E-05
merged1	-0.41234	0.27298	-1.511	0.13091
face_ethnAsian:face_ageYounger	0.15467	0.18367	0.842	0.39972
face_ethnLatina:face_ageYounger	0.38452	0.1841	2.089	0.03674
face_ethnAsian:vowel_typethought	0.296	0.1914	1.547	0.12198
face_ethnLatina:vowel_typethought	0.21176	0.19156	1.105	0.26897
face_ageYounger:vowel_typethought	0.07948	0.15639	0.508	0.61128
face_ethnAsian:merged1	0.31985	0.23094	1.385	0.16607
face_ethnLatina:merged1	-0.11345	0.22763	-0.498	0.61819
face_ageYounger:merged1	-0.09299	0.18736	-0.496	0.61968

Table 4: Results using Treatment Coding - including Task 2

Table 3 shows all the p-values, Std. Errors, and z values for all the variables. We use α = 0.05 to determine which p-values show significance. Using "White" as a baseline, we see that "face_ethnAsian" gives us a p-value which is less than α = 0.02 for Distinct participants. For Latina faces "face_ethnLatina" our p-value is also significant at α = 0.03. The next significance comes from "vowel_typethought" which was so robust that ($p < 0.001$). What we didn't expect was for the Younger Latina face (using "older" as another baseline) to be more accurate than older faces. "face_ethnLatina:face_ageYounger" was significant at α = 0.04.

Adding the variable of mergedness, those who were merged did not show any significance for either hypothesis. "face_ethnAsian:merged1" ($p > 0.05$) and "Latina:merged1" ($p>0.05$).

My First hypothesis is that ethnicity is correlated with the perception of the low-back merger, specifically, that Latinx and East Asian faces will have more merged ratings than White faces My second hypothesis is that age is correlated with the perception of the low-back merger, and specifically, that younger looking photographs will have more merged ratings than older looking photographs.

4. Discussion

4.1 Binary Forced Choice Perception Task

East Asian and Latina faces being less accurate than White faces is in line with our first hypothesis that ethnicity has a correlation with the perception of the low-back merger. We found that the East Asian and Latina faces were significantly less accurate than White faces. For those two ethnicities, LOT words were commonly mistaken for THOUGHT words.

When looking at age alone, we did not find any statistical significance with correctness of the LOT/THOUGHT words. We also didn't find any statistical significance when looking at ethnicity alone. When we looked at how age and ethnicity interacted, we did find that the younger Latina face was more accurate than the older Latina face. Our second hypothesis did not come true that age alone would be correlated with the low-back merger, and it was a surprise to find that the younger face was perceived as more distinct than the older face.

Previous work has found that older Latina faces are heard as more accented and transcribed less accurately than younger Latina faces (Ortiz 2018). This could explain why we got an opposite effect on age with the Latina faces, as older Latina faces are often described as

being more accented, presumed to not be born in The United States, and rated to be harder to understand.

The finding that THOUGHT tokens are extremely more accurate than LOT tokens was highly robust. The inaccuracies came from LOT tokens, where East Asian and Latina faces caused participants to often select THOUGHT. This highlights the idea that THOUGHT is lowering rather than LOT raising/backing. As THOUGHT lowers, these words begin to sound more like LOT words, therefore people perceive / hear LOT words as either LOT or THOUGHT. If it were LOT raising, we would expect to find THOUGHT tokens to be perceived / heard as THOUGHT or LOT words.

Specifically, for our THOUGHT tokens, it was not common for people to mistake them as LOT words because the audio stimuli for each face had the distinction, and the production of all the THOUGHT tokens were highly raised.

4.2 Effects of participant's judgements

We decided to ask participants whether they thought three sets of words were pronounced the same or different to them [cot/caught, don/dawn, tot/taught]. Those who said two or all three were pronounced the same were marked as "merged" participants. Those who said two or all three were different, were marked as "distinct" participants. When we re-ran the models with mergedness as a factor, we found that only distinct people found East Asian and Latina faces less accurate (merged).

It is surprising that only those who are not merged are in line with the previous findings that East Asian and Latina faces are more merged than White faces. One would expect that

those participants who are merged themselves, would select THOUGHT words for LOT words more often, as in their own speech they do that themselves.

5. Conclusion

Our first hypothesis that ethnicity would be correlated with the low-back merger came true, that East Asian and Latina faces would be perceived as more merged than White participants. Our second hypothesis that age would be correlated with the low-back merger did not come true. We expected that younger participants would be perceived as more merged than older participants, but when looking at age with ethnicity, the younger Latina face was perceived as more distinct than the older Latina face. This was predicted to be true because most research has shown that younger people are leading the change to the low-back merger (Labov, Ash, and Boberg 2006, Wong 2012) and lowering of raised THOUGHT (Becker 2010).

Many pieces of listeners' prior linguistic knowledge come into play when listening to speech (Rubin 1992, Hay, Warren and Drager 2006), and the perception of that speech can change depending on who they think they are listening to at the time. We have shown that New Yorkers are aware of the current merger in progress, and are semi-accurate in knowing who is leading the change. Our participants were accurate in stating that East Asian and Latina faces are more merged, and that LOT tokens are the ones to be heard as THOUGHT. Although previous work has shown age being the strongest factor in predicting mergeness (Haddican et al 2016), in this perception task, our participants did not show they were aware of this.

REFERENCES

Babel, Molly & Jamie Russell. 2015. Expectations and speech intelligibility. The Journal of the Acoustical Society of America 137(5). 2823-2833. doi:10.1121/1.4919317

Becker, Kara. 2010. Regional dialect features of the Lower East Side of New York City: Sociophonetics, ethnicity, and identity. New York: New York University dissertation.

Davies, Mark. The Corpus of Contemporary American English. 2008, www.english-corpora.org/coca/.

D'Onofrio, Annette. 2015. Perceiving Personae: Effects of Social Information on perceptions of TRAP-backing. University of Pennsylvania Working Papers in Linguistics. Volume 21 Issue 2 Selected Papers from NWAV 43.

Eckert, Penelope. 2016. Third Wave Variationism. Oxford Handbooks Online. doi:10.1093/oxfordhb/9780199935345.013.27.

Ezra, Daniel. 2007. Stability and change along a dialect boundary: the low vowels of Southeastern New England. University of Pennsylvania dissertation.

Haddican, Bill, Daniel Ezra, Michael Newman & Faith Kim. The diffusion of the low back merger in New York City. Poster presented at: NWAV 45; 2016; NYC

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech Perception in the context of a merger-in-progress. Journal of Phonetics34(4) 458-484. doi:10.1016/j.wocn.2005.10.001.

Hulme, C., Roodenerys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word frequency effects on short-term memory tasks: Evidence for a redintegration process in Immediate serial recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(5), 1217-1232. doi.org/10.1037/0278-7393.23.5.1217.

Labov, William. 1963. The social motivation of a sound change. Word 19:273-300.

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

Newman, Michael. 2014. New York City English. doi:10.1515/978161412127.

Niedzielski, Nancy A. 1999. The effect of social information on the perception of sociolinguistic variables. Journal of Social Psychology 18:62-85.

Ortiz, Omar. 2018. Beyond macro-level ethnicity: multiple social factors that affect perception of Latina women's speech. Undergraduate Honors Thesis, University of Oregon.

Rosenfelder, Ingrid; Fruehwald, Josef; Evanini, Keelan; Seyfarth, Scott; Gorman, Kyle; Prichard, Hilary; Yuan, Jiahong; 2014. FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2 10.5281/zenodo.22281.

Rubin, D.L. (1992). Nonlanguage factors affecting undergraduates' judgements on nonnative English-speaking teaching assistants. Research in Higher education, 33(4), 511-531.

Wing-mei Wong, Amy. 2012. The Lowering of Raised-THOUGHT and the Low-Back Distinction in New York City: Evidence from Chinese Americans. University of Pennsylvania Working Papers in Linguistics. Volume 18 issue 2 Selected Papers from NWAV 40.