

## Speaker age and vowel perception

Katie Drager

University of Hawai‘i at Mānoa

Running head: Speaker age and vowel perception

key words: speech perception, sociophonetics, exemplar theory

Corresponding author:

Katie Drager  
University of Hawai‘i at Mānoa  
Department of Linguistics  
561 Moore Hall  
Honolulu, Hawai‘i 96822  
<katie.drager@hawaii.edu>

## Acknowledgements

This research was conducted as part of a MA thesis at the University of Canterbury under the supervision of Jen Hay. The paper has benefited from the insights of Laurie Bauer, Gerry Docherty, Andy Gibson, Christian Langstrof, Margaret Maclagan, Aaron Nolan, ‘Ōiwi Parker Jones, Brynmor Thomas, Lauren Hall-Lew, Rebecca Greene, and the two reviewers, Benjamin Munson and Kyoko Nagao. I am especially grateful to Jen Hay for her guidance and suggestions throughout the course of the study. Thank you also to the anonymous participants for making this research possible and to Bob McMurray for making Klattworks available to me. Of course, all faults remain my own.

## **ABSTRACT**

Recent research provides evidence that individuals shift in their perception of variants depending on social characteristics attributed to the speaker. This paper reports on a speech perception experiment designed to test the degree to which the age attributed to a speaker influences the perception of vowels undergoing a chain shift. As a result of the shift, speakers from different generations produce different variants from one another. Results from the experiment indicate that a speaker's perceived age can influence vowel categorization in the expected direction. However, only older participants are influenced by perceived speaker age. This suggests that social characteristics attributed to a speaker affect speech perception differently depending on the salience of the relationship between the variant and the characteristic. The results also provide evidence of an unexpected interaction between the sex of the participant and the sex of the stimulus. The interaction is interpreted as an effect of the participants' previous exposure with male and female speakers. The results are analyzed under an exemplar model of speech production and perception where social information is indexed to acoustic information and the weight of the connection varies depending on the perceived salience of sociophonetic trends.

[194 words]

## **Speaker age and vowel perception**

Social information is part of a speaker's communicative competence (Hymes 1972). It is related to which sounds, words, and structures are used during speech production (Wolfram 1969, Labov 1972, Trudgill 1972) and recent research suggests that it can also affect how sounds are perceived (Strand 1999, Niedzielski 1999, Hay, Warren and Drager 2006b). Because social information patterns with linguistic variation in predictable ways, socially-conditioned variation can be accounted for in linguistic models, thereby providing a more complete picture of linguistic variation and a better understanding of the human mind.

One theory that has made inroads in terms of integrating social factors is Exemplar Theory (ET). In ET, utterances are stored in the mind as complete, acoustically-rich exemplars that are indexed to other information (such as information about the speaker) stored at the time of the utterance (Johnson 1997, Pierrehumbert 2001). The amount of attention paid to a particular component of the incoming signal (e.g. the formant values in a vowel) affects perception (Nosofsky 1986, 49). A memory is stronger (it has a greater weight) if more attention is paid when it is stored and, therefore, not all stored exemplars of a vowel, even within the same word, affect speech perception in identically the same way (Johnson 2006, 493).

During speech perception, exemplars are activated based on their context-dependent similarity to the incoming utterance (Nosofsky 1986). Activation of information indexed to the acoustically-rich exemplar can result in greater activation of that exemplar, biasing perception toward the variant it represents. During speech production, speakers activate multiple exemplars that are indexed to relevant social information. All of the activated exemplars contribute to the variant that is ultimately produced. Thus, ET predicts that social information indexed to an

acoustically-rich exemplar will, if activated, influence speech perception during a lexical categorization task.

This paper reports on results from a speech perception experiment investigating the degree to which individuals are affected by their perception of a speaker's age when identifying tokens from along a continuum of vowels involved in an ongoing shift. The results are discussed within the context of an exemplar-based model of speech production and perception, where the role of salience is not limited to the exemplars themselves but is also coded on the index between stored social and acoustic information.

## **SOCIOPHONETIC VARIATION**

### **Speech Perception**

Although the link between social information and linguistic variables is well-established in speech production, only recently have researchers begun to investigate an analogous effect in speech perception; social characteristics attributed to a speaker appear to influence how their speech is perceived (Johnson, Strand, and D'Imperio 1999, Niedzielski 1999, Strand 1999, Hay, Nolan and Drager 2006a, Hay, Warren and Drager 2006b).

In both Detroit and Canada, speakers produce raised variants of /au/, a trend commonly referred to as 'Canadian Raising'. Speakers in Detroit associate the stereotype of raised variants with Canadians and are not aware that Detroiters also produce raised variants. Niedzielski (1999) conducted an experiment in Detroit where she played recorded sentences from a Detroiters who produces raised variants. The sentences contained words with the /au/ diphthong and the target word was underlined on an answer sheet provided to the participants. Participants were asked to

match the vowel in the underlined word to one from a 6-step continuum of synthesized tokens, ranging from un-raised to raised variants. There were two experimental conditions. In one condition ‘Michigan’ appeared at the top of the response sheet and in the second condition ‘Canada’ appeared at the top of the response sheet. Participants were led to believe that the speaker they would hear was from the location at the top of their answer sheet. Niedzielski found that participants who were in the ‘Canada’ condition were more likely to respond with a raised token from the continuum, despite the fact that both Detroiters and Canadians produce raised variants and despite all sentences in both conditions being read by the same Detroiters. Niedzielski argues that participants are influenced by their expectations regarding a speaker’s dialect area combined with the stereotypes associated with that dialect.

Hay, Nolan, and Drager (2006a) conducted an experiment based on the same paradigm. In contrast to Niedzielski’s study, the target vowel, /ɪ/, in Hay et al. (2006a) differed between the varieties of English spoken in the two test regions: New Zealand and Australia. Another difference between their study and Niedzielski’s was that participants were not told any information about the speaker. Speakers from Australia and New Zealand are aware that their realizations of the vowel /ɪ/ differ: it is raised in the speech of Australians and centralized in the speech of New Zealanders. For the experiment, participants from New Zealand were played sentences of natural speech produced by a New Zealander. Words containing the target vowel were underlined on an answer sheet and participants were asked to match the vowel in the underlined word to one from a 6-step continuum of resynthesized speech that ranged from an Australian-like token (raised /ɪ/) to a New Zealand-like token (centralized /ɪ/). Again, there were two conditions, one with ‘Australian’ at the top of the response sheet and one with ‘New Zealander’. After finishing the perception task, participants completed a questionnaire where they

indicated what social characteristics they had attributed to the person who had read the sentences, including the speaker's country of origin. Hay et al. (2006a) found that female participants in the 'Australian' condition were significantly more likely to respond with a raised, more Australian-like token of /ɪ/. The results for the male participants were inconclusive and, if anything, were in the opposite direction. In the post-experimental survey, all but one of the participants in the Australian condition indicated that they knew that the speaker was a New Zealander. Hay et al. (2006a) suggest that participants may not need to believe a speaker is from a dialect region in order for that dialect to influence perception. In a follow-up study, Hay and Drager (to appear) found that stuffed toy kangaroos and koalas (associated with Australia) and stuffed toy kiwis (associated with New Zealand) influenced perception of /ɪ/ in much the same way as in the original regional label experiment. Participants completed the same task as in Hay et al. (2006a) except that no regional label was written on the answer sheet. Instead, when the experimenter retrieved the answer sheets from a cupboard, she also revealed the stuffed toys. To avoid questions, she feigned surprise at seeing the toys and set them on the counter, where they could be seen by the participant throughout the perception task. Hay and Drager (to appear) found that females were more likely to respond with a raised token if shown a kangaroo and males were more likely to respond with a raised token if shown a kiwi. It seems that mere reminders of a dialect area can be enough to affect how vowels are perceived but the directionality of the shift appears to be different for different groups of participants.

Perceived speaker gender has also been found to influence perception. Strand and Johnson (1996) and Strand (1999) played participants tokens from a /s/-/ʃ/ continuum and matched the voices with photographs of different individuals who were judged as being more or less stereotypically female. They found that participants were more likely to indicate that they had

heard /ʃ/ if they were shown a more stereotypically female face. Females have a higher acoustic boundary between the phones than males do and these results suggest that individuals used information about a speaker's gender in order to determine which sound they heard. In other words, perception of the fricatives was affected by the perceived social characteristics of the speakers, removing the need for an independent process of vocal-tract normalization. Similarly, Johnson, Strand, and D'Imperio (1999) found that the apparent gender of a speaker influenced vowel categorization when exposed to a synthesized vowel along a could/cud continuum. Additionally, they found that the imagined gender of a speaker (when participants were not shown a photograph but were instructed to imagine either a male or a female talker) affected vowel categorization and that the effect was strongest at the beginning of the experiment (when participants were probably paying more attention to the imagined gender) and at the end of the experiment, immediately prior to the questionnaire about the imagined speaker (which participants had been informed they would complete). These results provide additional evidence that the gender attributed to the speaker can influence perception and that the amount of attention paid to social characteristics of the speaker can play a role.

Hay, Warren, and Drager (2006b) investigated the effect of a speaker's age and social class on the perception of variants undergoing a merger in progress. In New Zealand English (NZE), the centering diphthongs /ɪə/ and /eə/ are undergoing a merger, so that older speakers of NZE are more likely to maintain a distinction than younger speakers. The merger has been led by females and members of lower socioeconomic groups (Maclagan and Gordon 1996, Warren, Hay and Thomas 2007). The degree of merger is predictable on the basis of social class and age. Hay et al. (2006b) played the participants recordings of words containing /ɪə/ and /eə/ produced by NZE speakers who maintained the distinction. Participants were asked to indicate which word they

heard in a two-alternative forced-choice task. The auditory stimuli were identical across the five conditions; only the photograph with which a voice was matched differed. Across the five conditions, the apparent age and socioeconomic status of the person pictured was varied: there were two conditions in which the age was manipulated (so that if a voice was paired with an older face in condition one, it was paired with a younger face in condition two), two conditions where the apparent socioeconomic status was varied (so that if a voice was paired with a photograph of someone dressed to look like they came from a lower socioeconomic group in condition three, that voice was paired with a photograph of the same person dressed to look like they came from a higher socioeconomic group in condition four), and one control condition where there were no photographs shown. Hay et al. (2006b) found that the social characteristics attributed to the person in the photograph influenced participants' perception of the diphthongs; participants who self-reported that they would maintain a distinction for the word-pair were more accurate when shown a photograph of someone who was judged to be older and, when responding to one of the stimulus voices who maintained the weakest distinction between the diphthongs, unmerged participants were more accurate when shown a photograph of someone who was judged to be from a higher socioeconomic group. These results provide evidence that social information attributed to a speaker can affect the perception of sounds produced by that speaker.

Results from an experiment examining the perception of Texan English provide additional evidence that the perceived age of a speaker can affect lexical access within the context of a vowel merger (Koops, Gentry and Pantos 2008). In Houston, the pre-nasal vowels /ɪ/ and /ɛ/, as in *pin* and *pen*, are merged for middle-aged and elderly speakers (Gentry 2006), but a growing number of young females appear to maintain a distinction (Pantos 2006). Thus, there is a

difference between younger and older female speakers of Houston English, but it is in the opposite direction as the merger investigated by Hay, Warren and Drager (2006b); in Houston, older speakers are less likely to maintain a distinction between pre-nasal /ɪ/ and /ɛ/ than younger speakers. Koops, Gentry and Pantos (2008) played participants distinct auditory tokens that were matched with photographs of women of different ages: young, middle-aged, and elderly. Using an eye-tracking device to examine duration of fixation of near-neighbor words, they found that listeners were more likely to fixate longer on the lexical competitor when shown a photograph of an older “speaker” than when shown a photograph of a middle-aged “speaker”. This suggests that listeners were sensitive to the age of the person in the photograph when perceiving speech. This paper reports on an experiment investigating the extent to which the age of a speaker can influence perception of another change in progress, namely a chain shift in progress.

## **The Short Front Vowels in New Zealand English**

The short front vowels, as in the words *had*, *head*, and *hid*, are involved in an ongoing chain shift in NZE. The lexical set labels TRAP, DRESS, and KIT (Wells 1982) are used to refer to the vowels in each of the respective words. These vowels have been involved in a push chain, such that TRAP has raised into the space of DRESS, DRESS has raised into the space of KIT, and KIT has centralized (Gordon, Campbell, Hay, Maclagan, Sudbury and Trudgill 2004, Langstrof 2006). The change is ongoing; in modern New Zealand English, younger speakers of NZE have higher variants of both DRESS and TRAP than older speakers (Maclagan, Gordon and Lewis 1999, Maclagan and Hay 2007).

The shift has been led by females (Gordon, Campbell, Hay, Maclagan, Sudbury and Trudgill 2004, Maclagan, Gordon and Lewis 1999, Maclagan and Hay 2007). Within each age group

analyzed by Maclagan, Gordon and Lewis (1999), females produced a larger percentage of raised (innovative) tokens of DRESS and TRAP than males did. The only exception was the group of young non-professional males, who produced the largest percentage of innovative tokens among all the sampled categories. In the study by Maclagan and Hay (2007), DRESS was so raised in the speech of young males and females that it was completely overlapping with FLEECE in terms of their F1-F2 space. In fact, DRESS was actually higher than FLEECE for two speakers in their dataset, both of whom were young, non-professional females (Maclagan and Hay 2007, 10).

Drager (2006) investigated the extent to which listeners were sensitive to this sociophonetic variation in NZE by conducting a speech perception experiment using ten-step resynthesized vowel continua between TRAP and DRESS for two male speakers. After completing the experiment, participants estimated each speaker's age based on their most *bad*-like token. The results indicate that listeners identified more tokens as *bad* when they believed the speaker was younger. This finding is consistent with production; young NZE speakers not only produced variants of DRESS and TRAP that were more raised than those produced by older speakers of NZE (Maclagan, Gordon and Lewis 1999), but listeners were more likely to identify an ambiguous token as TRAP if they believed that they were listening to a younger NZE speaker. This suggests that participants were aware, at some level, that younger NZE speakers' productions of TRAP are more raised than older NZE speakers' productions are. Because the voices differed in acoustic characteristics, such as F0, that can affect vowel categorization (Traunmüller 1981), the directionality of the effect is unknown: perceived age may have been extracted based on acoustic cues that differed between the voices and then used to inform vowel categorization or, alternatively, listeners may have extracted age information based on the perceived quality of the vowel; if they perceived a relatively raised token as *bad*, they may have assumed that the speaker must be younger.

In Drager (2006), perceived speaker age was extracted solely from acoustic cues in the signal. This paper reports on a speech perception experiment that was conducted in order to test the effect of perceived age on vowel categorization when the perceived speaker age was manipulated systematically using photographs. Additionally, it was conducted to examine how listeners respond to female voices, as results for female voices in Drager (2006) were inconclusive.

## **METHOD**

### **Stimuli**

#### **Auditory Stimuli**

Using the program Klattworks (McMurray, in prep), resynthesis was conducted on natural tokens of the word *bad* produced by four different speakers: two males (M1 and M2) and two females (F1 and F2). The voices upon which the resynthesis was based were taken from the Canterbury Corpus, part of the Origins of New Zealand English (ONZE) archives held at the University of Canterbury. All four people whose voices were used were speakers of New Zealand English and were under the age of 30 at the time the experiment was conducted.

Tokens of the words *bad* and *had* were extracted from wordlists read by each of the speakers. Resynthesis was conducted on the vowel from *bad* in order to create a 9-step vowel continuum between DRESS and TRAP for each of the voices. The vowel was extracted at the nearest zero crosspoint after voicing had begun following the initial burst and before a sudden drop in amplitude indicating closure for the /d/.

Only the first and second formants (F1 and F2) of the vowels were manipulated; other formant measurements were not modified so that any results could be attributed to differences in F1 and F2. Additionally, using different-sounding voices was desired so that listeners would

believe they were different speakers. The pitch was also maintained in an effort to produce a more natural sounding voice. Duration was constant across all tokens in a single continuum but differed for the different voices. The F0 and F3 values at the vowel midpoint and the vowel durations for tokens from each voice are shown in Table 1.

insert table 1 about here
---------------------------

There is evidence that in the Intermediate Period of NZE (speakers born between the 1890s and 1930s) there was little difference in duration between DRESS and TRAP (Langstrof 2006). If this trend is still present in modern NZE, vowel categorization may not be affected as strongly as it would be for varieties where there is a large durational difference between the vowels. Furthermore, the main question of interest in the experiment was whether the ages of people in photographs could affect vowel categorization, and the effect of a single photograph was calculated across responses to both sex-matched voices. Therefore, it was not a problem for the aims of the experiment that the tokens varied in vowel duration across the different voices and were constant within a single continuum.

Tokens for the female and male voices were resynthesized so that they had the formant values shown in Tables 2 and 3, respectively. Initially, the F1 and F2 values at the two ends of the continua were based on the formant values of DRESS and TRAP presented by Maclagan (1982). However, playing the tokens to NZE speakers prior to running the experiment revealed that only rarely were any of the tokens identified as DRESS. Thus, the end-points of the continua were changed so that one end of the continuum was unambiguously DRESS and the other was

unambiguously TRAP for the NZE speakers consulted prior to running the experiment. Individuals who gave feedback on the continua did not take part in the actual experiment.

Following resynthesis, the initial consonants /b/ and /h/ and the final consonant /d/ were attached to the vowels. These were extracted from the original (natural) tokens of *bad* and *had* produced by the same speakers who produced the vowels. The /b/ was extracted immediately after the burst (the final tokens had a VOT of zero) and the boundary following the /h/ was made before periodic energy from the vowel was evident. The /d/ boundaries were made at the beginning of the closure period and following the burst. Formant transitions were included in the synthesized vowels in order to improve clarity and naturalness. In order to clarify the perception of the following consonant, the last 10 ms of all vowels had been resynthesized to have an F2 that lowered by 100 Hz. No further manipulation took place for /h/-initial stimuli. For /b/-initial stimuli, F2 in the vowel was manipulated in the 15 ms closest to the /b/ so that F2 rose by 200 Hz as it approached the vowel's target.

insert table 2 about here
---------------------------

insert table 3 about here
---------------------------

### **Visual Stimuli**

Photographs of people of different ages were used in order to manipulate the apparent age of the speaker. A photograph of an older male (OM), an older female (OF), a younger male (YM), and a younger female (YF) were taken for the purposes of this experiment. Dress and background were controlled as much as possible. The photographs are shown in Figure 1.

insert figure 1 about here

During the perception task, the voices were paired with one of the sex-appropriate photographs; a photograph of a male voice was matched with a male voice and a photograph of a female was matched with a female voice. The photograph with which a voice was paired varied depending on the condition in which a subject was participating. For example, if a voice was paired with OF in condition one, it was paired with YF in condition two. This was done in order to investigate the effect of the age of the person in the photograph independent of the voice with which it was matched.

## **Procedure**

The experimenter met with participants individually in a quiet room at the University of Canterbury. Auditory stimuli were presented one at a time over headphones and were played at a comfortable hearing level. Responses were collected automatically by the presentation program, MediaLab (Jarvis 2002).

Participants listened to tokens from the DRESS-TRAP continua and indicated what word they heard. The two response options, *bed* and *bad* or *head* and *had*, appeared on the screen and participants were instructed to use the mouse to click on the word that they believed the auditory token sounded most like. The tokens were pseudo-randomized and alternated between male and female voices. No fillers were used. The order of the words on the screen was pseudo-randomized: the TRAP word, *bad* or *had*, was listed above the DRESS word half of the time. The

order of the auditory stimuli was identical across the two conditions and different listeners took part in each of the conditions.

Each token was paired with a photograph of the “speaker”; voices were matched with photographs so that, within a condition, a single voice was always matched with the same photograph. For example, in condition 1, all tokens produced by M1 were played when the photograph of the older male (OM) was shown on the screen, and in condition 2, all of the tokens produced by M1 were accompanied by the photograph of the younger male (YM). This way, participants were provided with consistent visual evidence of the age of each voice, but the conditional manipulation allowed for the effect of the photograph to be investigated independently from the voice with which it was matched. The sound file was played approximately 14ms after the photograph appeared on the screen and the photograph remained on the screen until the participant responded.

During the first half of the experiment, participants responded to all 72 tokens (9 continuum steps x 2 preceding contexts x 4 voices). They were then invited to break from the task for 1-2 minutes. Upon resuming the task, all 72 tokens were again presented in the same order as in the first half of the experiment. The only difference between the two halves was the order in which the words appeared on the screen, so that if the DRESS word was located above the TRAP word in the first half, it was located below the TRAP word in the second half. This was done in order to balance a potential response bias resulting from a tendency toward responding with the word listed on top. In the interest of time, participants responded to each of the /b/ or /h/ initial tokens only twice. While a larger number of responses to each token would be desirable, this is less of a concern than it might be for other experiment designs because participants responded to each of the tokens in each /b/-initial and /h/-initial continuum twice. The effect of apparent age can be tested across all tokens of a continuum and across both environments for preceding context.

Before beginning the perception task, participants completed a short training exercise. The training session had the same design as the perception experiment, but different voices and faces were used.

## **Participants**

All 24 participants were native speakers of NZE and reported that they did not recognize the people in the photographs.

In order to test whether participants in the two conditions were matched for socioeconomic background, a participant's social class was calculated by combining the New Zealand Socioeconomic Indices assigned to each of the participant's parents' occupations (Davis, McLeod, Ransom, Ongley, Pearce and Howden-Chapman 1997, Davis, Jenkin and Coope 2003). The combined score, referred to here as a participant's socioeconomic index (SEI), could potentially range from 40 (for subjects from lower socioeconomic backgrounds) to 180 (for participants from higher socioeconomic backgrounds). A summary of the distribution of participant characteristics is shown in Table 4. Across the two conditions, participants were well-matched for sex, age, and SEI. In both conditions, the mean age of participants was low (mid-twenties); only seven participants were 30 or older. The uneven distribution of participant ages within each condition was due to the method of recruitment: posters on the University of Canterbury campus and emails to university departments other than the Department of Linguistics.

insert table 4 about here
---------------------------

Through the paper, the terms *older* and *younger* are used to refer to the participants as well as the people in the photographs. These are not used to denote specific age groups but to describe a participant's age relative to the other participants; age is treated as a continuous factor. As a result, *older* refers to participants (and stimuli) who were middle-aged rather than elderly; they were older than other participants in the study.

## **Calculating Photo-Age**

After finishing the perception task, participants completed a short questionnaire where they were asked to indicate the gender, age, education, and occupation of the people in all four photographs. In estimating the age, participants were asked to select the category they felt best approximated the age of the person in the photograph:

(a.) 18-25 (b.) 26-35 (c.) 36-45 (d.) 46-55 (e.) 56-65 (f.) 66+

The value in each age group that was divisible by ten was treated as the assigned age for a photograph. To calculate perceived PHOTO-AGE in a manner that was independent of the voice with which the photograph was matched, the average age assigned to a photograph in condition one was averaged with the average age assigned to the photograph in condition two. The ages assigned to the photographs are shown in Table 5 and are referred to as each photograph's PHOTO-AGE. PHOTO-AGE is not intended to represent the actual age attributed to the person in the photograph but instead should be viewed as a ranking of the different photographs based on age.

insert table 5 about here
---------------------------

Using a numeric value for PHOTO-AGE rather than a binary measure (older and younger) allowed a comparison of the four different photographs' perceived ages rather than comparing OM and OF with YM and YF. However, the binary measure was also tested as a factor in order to ensure that it behaved as would be expected. Because the binary measure and the numeric value behave similarly, only the model for the numeric value is presented in this paper.

Participants identified OF and YF as female and OM and YM as male 100% of the time. The assigned occupation and education of the people in the photographs was highly correlated with age; higher levels of education and jobs with higher incomes were assigned to OM and OF. Because of this correlation, these were not tested as predicting factors.

## **RESULTS**

First, the results are presented by plots of vowel categorization across the different token numbers of each continuum. Second, statistical significance and the degree of any effects are tested through a logistic regression model with mixed effects. The estimated 50% crossover points for each participant, calculated based on coefficients from the logistic regression model, are also provided.

The percentage of tokens identified as TRAP and DRESS is shown in Figures 2-5 for each step along the continua and for each photograph: the older female (OF), the younger female (YF), the older male (OM), and the younger male (YM). These values were averaged over all participants and were based on raw values. If the perceived age of a speaker based on the photographs affected perception, we would expect that the crossover point between DRESS and TRAP would occur at lower token numbers when the voices are paired with the younger photographs compared to when they are paired with the older photographs. Because the assigned formant values differed between the male and female voices and because photographs were only

matched with same-sex voices (i.e., female photographs were matched with female voices and not male voices), cross-gender comparisons of PHOTO-AGE effects could be misleading and are not discussed here.

In comparing responses to stimuli from the female voices, the crossover points for both voices are between tokens 4 and 5. However, the crossover point is slightly closer to token 4 for the younger photo, whereas it is slightly closer to token 5 for the older photo. This trend is in the expected direction. The same is true when comparing responses to the male voices; the crossover point is between tokens 4 and 5 (and closer to 5) for the older photograph and at token 4 for the younger photograph. This suggests that the perceived age of the person in the photographs affected responses in a way that was consistent with production; younger speakers are, in general, more likely to produce raised variants of DRESS and TRAP and participants' perceptual boundaries between the vowels appear to be slightly more raised when they are shown a photograph of a younger "speaker" than when they are shown a photograph of an older "speaker".

insert figure 2 about here

insert figure 3 about here

insert figure 4 about here

insert figure 5 about here

To determine the robustness and significance levels of PHOTO-AGE within the context of other factors affecting perception, the data were analyzed under a binary mixed-effects logistic regression model, which was fit to the data by hand using the statistical tool R (R Development Core Team, 2007). Logistic regression produces predicted probabilities that are very similar to those produced by probit analysis, a technique often used in the perception literature. In a logistic regression, a positive coefficient indicates a positive correlation between the dependent variable and the effect and, conversely, a negative coefficient indicates a negative correlation. An analysis using boundary values (as is standard) and a logistic regression analysis (as presented here) would both lead to the same conclusions and the coefficients presented for a logistic regression can be used to calculate the estimated perceptual boundary (when log odds equal zero) (Benkí 2001, 10).

In a perception experiment, individual participants may respond differently from one another. For example, in the experiment presented in this paper, some participants may have been biased toward responding *bed* while others may have been biased toward responding *bad*. But the question of interest is whether the trends observed are generalizable to listeners from New Zealand beyond only those who took part in the experiment. Simple linear and logistic regression models treat differences between responses from a single participant (or two realizations of vowel produced by a single individual) no differently than differences between participants. Random effect models can help to remedy this without resorting to averaging over responses made by a participant, as this could potentially mask certain trends present in the data. When the subject is included as a random effect, each individual subject is assigned their own coefficient

and that coefficient is associated with each response made by that subject. Thus, including the subject as a random effect reduces the risk that a single subject can carry a trend (Baayen 2008).

From a statistical standpoint, one challenge of working with human subjects is missing data: participants may fail to respond to a question on an experiment or, for a study examining phonetic variation of vowels from spontaneous speech, a speaker may not produce tokens of the vowel in a certain phonological environment. Whereas some statistical methods (e.g., MANOVA) are susceptible to missing data, mixed effects models appear to perform well. Baayen, Davidson, and Bates (2008) found that when datapoints were randomly removed and the two methods of statistical analyses were performed, the mixed effects analysis more closely approximated the probabilities that were calculated based on the entire dataset (Baayen et al. 2008, 402).

For the model presented here, the subject and question number were included as random effects; each subject and each question number was assigned their own coefficient. The dependent variable was whether a token was identified as TRAP. A number of effects were tested as potential predictors, including social characteristics of the participant (e.g. age, socioeconomic index, and sex) and stimulus characteristics (e.g. the token number and the PHOTO-AGE). Only fixed effects that reached a significance level of  $p < 0.05$  or smaller were included in the model. All figures presented are within the context of the model; they represent the model's predictions, in log odds, of a factor's effect while holding the other fixed effects in the model constant. The model's fixed effects are shown in Table 6, alongside the factors' estimated coefficients and their predictive significance.

The model presented in Table 6 predicted the likelihood that a particular token would be perceived as TRAP (*bad* or *had*) rather than DRESS (*bed* or *head*). Included in the model was whether a token was h-initial or b-initial (INITIAL C) and the token number (TOKEN), which

served as an indication of where in the synthesized continuum a particular token was ranked. While it was not surprising that token number influenced responses, this was included in the model as a way to control for token number when investigating the effects of the other factors. Also included in the model were two interactions, one between the sex of the stimuli (SPEAKER-SEX) and the self-reported sex of the subject (SUBJ-SEX) and one between the averaged age of the people in the photographs (PHOTO-AGE) and the self-reported age of the subject (SUBJ-AGE).

Positive coefficients, listed in the ‘Estimate’ column, indicate that a token was more likely to be perceived as TRAP with the factor listed (for categorical factor groups) or with a higher value (for continuous factor groups). The default factor (when ‘Estimate’ equals zero) for discrete factor groups is in alphabetical order. Therefore, the default values in this model are /b/-initial contexts as opposed to /h/-initial contexts and ‘female’ for both participant sex and the sex attributed to the photograph. The default value for continuous factors (e.g. PHOTO-AGE) is zero.

The coefficients can be viewed as an indication of effect size. For continuous factors such as SUBJ-AGE, the product of the coefficient and a participant’s age is added to the estimated coefficient in order to determine the log odds of a TRAP response for participants of that age. For example, the log odds of a 56 year old female participant identifying token 4 as TRAP when shown the photograph of the older female is -3.15. (She is not very likely to identify the token as TRAP.) The log odds of an 18 year old female participant identifying that same token as TRAP are -1.35. (She is unlikely to identify the token as TRAP but is more likely than the older participant.) The positive difference between these log odds (1.8) is an indication of the size of the effect when comparing an 18 year old participant and a 56 year old participant while keeping the other factors constant. For comparison, the difference in log odds between two consecutive

tokens from the continuum is 0.96 (the factor's estimated coefficient) and the difference in log odds between tokens from each end of the 9-step continuum is 7.68.

insert table 6 about here
---------------------------

The estimated 50% crossover points for each participant's response to the different photographs are shown in Table 7. These values are based on the coefficients in Table 6 and the random effect coefficients of each participant.

insert table 7 about here
---------------------------

The continuum was designed so that smaller token numbers were more DRESS-like and larger token numbers were more TRAP-like. The model predicts that the larger token numbers will be perceived as TRAP ( $p < 0.0001$ ). This prediction is an indication that the design of the continuum was successful. Although this is not a surprising result, the factor is included in the model as a control.

Another unsurprising result is that the linguistic environment appears to play a role in perception. Participants were significantly more likely to perceive TRAP when the token they were responding to was in the /hVd/ frame rather than in the /bVd/ frame ( $p < 0.0001$ ). In other words, if the token was /h/-initial, participants were more likely to indicate that they had heard *had*, and if the token was /b/-initial, participants were more likely to indicate that they had heard

*bed*. /b/-initial vowels were resynthesised to have transitions, but the targets of both /b/- and /h/-initial words were identical. In rapid natural speech, vowels that follow labials have a lower F2 than when the same vowel follows /h/ (Stevens and House 1963) and during perception, listeners normalize accordingly (Lindblom and Studdert-Kennedy 1967). Thus, a vowel with set values for F1 and F2 could be perceived as TRAP when it is /h/-initial and as DRESS when it is /b/-initial. Including the initial consonant as a predicting factor in the model also serves as a control; the model holds the initial consonant constant when testing the other factors included in the model. As an additional test, separate models were fit to the subsets of data with the different initial consonants. The trends in these two models were the same as those described in the remainder of this section but, due to fewer datapoints, were less robust.

The age of a participant affected how that participant categorized the vowels. Older participants were less likely to respond with TRAP than younger participants ( $p < 0.0001$ ). This is in the expected direction given trends in production. The age attributed to a photograph also influenced perception; participants were more likely to categorize a vowel as TRAP if they were shown a photograph of a younger individual ( $p < 0.0001$ ). However, as shown in Figure 6, this tendency was carried entirely by the older participants. This demonstrates the importance of examining factors within the context of the interactions in which they appear.

Using the coefficient values in the ‘Estimate’ column in Table 6, Figure 6 plots the log odds of perceiving TRAP for participants of different ages depending on the photograph shown. Other factors are held constant (e.g., SUBJ-SEX = F).

insert figure 6 about here

PHOTO-AGE played little role in the perception of the vowels for the younger participants. In contrast, older participants were much more likely to perceive TRAP when shown a photograph of a younger individual than an older individual. It seems that the older participants were influenced by the perceived age of the speaker but younger participants were not. The interaction between the age of the participant and the age attributed to the person in the photograph is highly significant ( $p < 0.0001$ ). People of varying ages seem to perceive the vowels differently depending on whether they are shown a photograph of an older individual or a photograph of a younger individual. This is discussed in the following section.

There was also an effect of participant sex and stimulus sex. Male participants perceived fewer tokens as TRAP than did female participants ( $p < 0.05$ ). This is consistent with trends in production: females have led the chain shift and tend to produce variants of DRESS and TRAP that are more raised in respect to the rest of their vowel space than variants produced by male participants (Maclagan and Gordon 1996). In general, male speakers are less likely to *produce* raised variants of TRAP and results from this experiment suggest that males are less likely to *perceive* an ambiguous (raised) vowel as TRAP.

The results also indicate that participants perceive more TRAP when listening to a female voice than when listening to a male voice ( $p < 0.05$ ). Given work by Strand (1999) and Johnson, Strand, and D'Imperio (1999), it would be predicted that listeners would identify a greater number of tokens as TRAP if produced by female voices than by male voices. The results presented here are consistent with this prediction. However, while it is possible that listeners used perceived speaker sex to determine the likelihood of the speaker producing a raised variant, this observation could also be a result of how the continua were designed: different formant values

were used for the male and female voices. If directly comparing the formant values from the male and female continua, the most TRAP-like token for the male voices falls between tokens 5 and 6 in terms of F1. That a larger number of tokens were identified as TRAP for the female voices may be due to the discrepancy in formant values.

There is a significant interaction between whether the stimulus voice was male or female and whether the participant was male or female ( $p < 0.001$ ); male and female participants seem to have responded differently to male and female voices. A graph of this interaction is shown in Figure 7. Female participants perceive more tokens as TRAP than males do when responding to a female voice and males perceive more tokens as TRAP when responding to a male voice. While the interaction is statistically significant, the effect size is small; in terms of the log odds of responding TRAP, the positive difference between male and females responding to a female stimulus is only 0.37 and for a male stimulus 0.41. Although the effect of the interaction is relatively small and should be tested and replicated in future work, possible interpretations of this result are discussed in the following section.

insert figure 7 about here

## **DISCUSSION**

The results provide evidence that (1) males and females perceive tokens produced by males and females differently and that (2) participants of different ages vary in their categorization of vowels that are undergoing change, depending on the perceived age of the speaker.

## Interpreting the Sex-based Interaction

Female participants identified more tokens as TRAP than males did when listening to a female voice and male participants identified more tokens as TRAP than females did when listening to a male voice. The effect of the stimulus sex cannot be attributed to normalization due to vocal-tract normalization; the effect is in the opposite direction for two different groups of participants and the effect for male participants is in the opposite direction than what would be expected with vocal-tract normalization. In general, females produce vowels with higher formant frequencies than males because they tend to have shorter vocal tracts. Thus, an F1 value that is observed for DRESS produced by a female might be identical to an F1 value for TRAP produced by a male. This would mean that, with only vocal-tract normalization, a token with the same F1 value might be perceived as DRESS when the speaker is female and TRAP when the speaker is male. Female listeners, however, responded in the opposite direction; they identified a greater number of tokens as TRAP when listening to the female voices. This result cannot be due to vocal tract normalization.

Why did the male and female participants respond differently to the male and female stimuli? Remember, these results should be viewed with some caution. While statistically significant, the size of the effect is relatively small and replication of the results is needed. Despite this, some possible explanations of the interaction is discussed here.

One possible explanation is that male and female participants attended to different acoustic cues in the voices. Without vocal-tract normalization, F1 values for the female voices were more TRAP-like than those for the male voices, while F2 values for the female voices were more DRESS-like than those for the male voices. It is possible that female participants based their vowel categorization more on vowel height whereas male participants based theirs more heavily

on vowel frontness. Further work is required to determine whether male and female listeners from New Zealand attend to different acoustic cues.

Another possible explanation is that participants' perception of the vowels was influenced by their exposure to different distributions of male and female talkers in everyday life. Many of the participants were in their first year of university. In New Zealand, individuals who attend single-sex schools are more likely to pursue a university degree than those who attend co-educational schools (Gibb, Fergusson and Horwood 2008). A study based in Christchurch (where the perception study was conducted) found that a larger number of high school students attended co-educational schools but roughly equivalent numbers of students from single-sex and co-educational schools attended university (Gibb, Fergusson and Horwood 2008). The reasons behind this discrepancy are complex but it means that a sizeable proportion of the participants in the current experiment probably came from single-sex schools. As a result, young females who attended single-sex schools would have had exposure to a vast number of other young females and young males who attended single-sex schools would have had ample exposure to a number of other young males. Of course, each would also have exposure to young members of the opposite sex but the distribution of ages across the opposite-sex groups would vary; older family members, teachers, and community members would make up a greater proportion of the interactions with members of the opposite-sex. Thus, stored speech from older speakers would play a greater role in the perception of speech produced by members of the opposite sex. To test this as a possible explanation, the model was fit to responses from participants who were 25 or younger and to responses from participants who were 30 or over. If exposure during childhood and adolescence affects perception in this way, we would expect to observe the sex-based interaction only for the younger participants. In fact, this is what was observed; the interaction was significant among the younger participants ( $p < 0.01$ ) but not among the older participants

( $p > 0.1$ ). The sex-based interaction observed in perception appears to be strongest among the younger participants, supporting the claim that this effect was a result of exposure to different distributions of ages (and therefore stages of the vowel shift) for members of the opposite sex than for members of the same sex. Unfortunately, information regarding the type of school attended by participants was not collected.

## **Interpreting the Age-based Interaction**

Also observed in the data was an interaction between SUBJ-AGE and PHOTO-AGE, where older participants were more likely to perceive TRAP when shown a younger face than when shown an older face but younger participants were not. While the size of the effect is greater than for the sex-based interaction, it is still a subtle effect and appears even weaker when viewed in the context of the raw data when other factors (like SUBJ-SEX) are not taken into account. While further experimentation is required, this result supports the hypothesis that the age attributed to a speaker can affect the perception of vowels undergoing change. That this effect was only found among older participants may be explained in terms of how salient, for a given participant, the relationship was between variant produced and the age of a speaker. Presumably, older participants had not only been exposed to a range of speakers from different generations, but had experienced the progression of the change in a way that is very different from how it was experienced by the younger participants; older participants would have had more exposure to the vowel shift as it has progressed. Therefore, the relationship between variant produced and speaker age may have been more salient for the older participants. Here, the conception of salience does not entail overt recognition or knowledge of the relationship but is a gradient measure of importance or prominence in relation to its neighbors.

In general, listeners are good at judging the age of a speaker (Huntley, Hollien and Shipp 1987, Neiman and Applegate 1990) and previous work provides evidence that the age attributed to a voice (based only on acoustic cues and no visual information) is linked to categorization of vowels that are undergoing change (Drager 2006). Although the experiment conducted by Drager (2006) investigated perception of the same vowel shift examined in the experiment presented here, she did not observe the age-based interaction; vowel category assignment shifted in the expected direction regardless of the listener's age. The directionality of the effect based only on acoustic information is unclear; do vowel category boundaries shift as a result of the age attributed to the speaker, or does the apparent age of the speaker shift depending on whether (based on other factors such as F0) the vowel is perceived as TRAP or DRESS? That the age-based interaction was only observed when speaker age was manipulated using visual information suggests that the results presented in Drager (2006) may have been due to the latter: the age attributed to the speaker shifted as a result of how the vowel was perceived. Future work would benefit from collecting information about the perceived age of the voices in addition to using visual stimuli.

## **Salience in an Exemplar Model**

The results provide evidence that a combination of social characteristics of the listener and the stimulus can influence vowel categorization. How might these results be accounted for in a model of speech perception? One model of speech perception which predicts that social information attributed to a speaker will affect vowel categorization is an exemplar model of speech production and perception. In an exemplar model, utterances are stored in the mind as separate exemplars, complete with phonetic detail (Pierrehumbert, 2001). The exemplars are

indexed to social characteristics of the speaker and other information stored at the time of the utterance (Johnson, 1997; Foulkes & Docherty, 2006; Hay, Warren & Drager, 2006).

All of this information is stored initially but decays with time (Lacerda, 1995, in press; Pierrehumbert, 2001, 2002; Hawkins, 2003). Decay of exemplars is slowed by frequent and/or recent activation. During production, a speaker samples over the frequently and recently activated exemplars. The actual variant produced is the result of averaging over activated exemplars. Activation is gradient; exemplars can be activated to varying degrees, contributing to the ultimate variant produced or perceived to a degree that is proportional to the amount of activation of the exemplar. Exemplars reach full activation fastest if they are indexed with social information related to how the speaker wants to portray themselves, how they see themselves, and even characteristics of someone who they are speaking to or about (Foulkes and Docherty 2006).

During perception, exemplars are activated depending on their similarity to the incoming utterance. This pertains to both their acoustic similarity as well as the similarity of other information (such as social characteristics of the speaker) to which the exemplar is indexed. As a result, perception is biased toward variants produced by the listener and by individuals who share similar social characteristics with the speaker. This predicts that the age of a speaker will affect perception; a perceiver's exemplars indexed with a similar age to the speaker would be activated. The incoming acoustic information would then activate exemplars with which it is most closely matched and the exemplars that are already partially activated through the perception of social information reach full activation fastest. This biases perception toward variants produced by individuals with an age similar to the person in the photograph; individuals should perceive more tokens as TRAP when shown a photograph of a younger "speaker". However, this effect was only observed for the older speakers in the experiment. Why were the older participants

influenced by the perceived age of the people shown in the photographs while the younger participants were not? And how might exemplar models be developed to account for this result?

Different individuals may consider different sociophonetic relationships to be more salient: they are more explicitly aware that individuals with certain social characteristics tend to speak a certain way. It would not be entirely surprising if variants undergoing a chain shift were particularly salient for those people who had actually observed the progress of the change. The older NZE speakers would have had the opportunity to observe the chain shift, while the younger NZE speakers would not have had as much of an opportunity and may not be as aware of how the variants pattern with age. Salience of the relationship between a social characteristics and phonetic variables has not yet been implemented in an exemplar model.

One way that exemplar models could address this would be to have a greater amount of weight on the relationship between the social information and the phonetic information when the relationship is more salient for that individual speaker-hearer. The differing weights would result in varying amounts of activation when the social information is activated; salient social information would influence perception more strongly than a relationship with less salience. If the relationship between age and the realization of DRESS and TRAP is more salient for older participants than younger participants, a speaker's age would play a larger role for older participants than younger ones.

This interpretation assumes abstraction of social categories (e.g. age) from the rich social information expressed by an individual in the construction of their identity. Another possible way to account for the results would be for these abstract representations of social categories to be formed only upon a gradient measure of salience (based on rich speaker-specific social information) reaching a threshold, perhaps upon the individual speaker-hearer's awareness of a relationship. Thus, stereotypes (based on highly salient categories) of how certain groups of

people speak would affect perception to a greater degree than when a perceiver is not aware of a relationship between social and phonetic information. While this may suit some stereotyped information, it seems unlikely that the age of an individual would not form an abstract category for everyone.

## **CONCLUSION**

The results presented in this paper provide evidence that social characteristics of both the speaker and perceiver can influence vowel perception. Depending on the age of the participant, the age attributed to a speaker influences the categorization of vowels which are undergoing change and, depending on the sex of the participant, the sex of a speaker affects vowel categorization differently. These interactions are interpreted as a result of exposure to a particular distribution of speakers for the sex-based interaction) and how this could lead to certain relationships between social and phonetic information being more salient than others. Future work would benefit from collecting responses from both a larger number of older participants and information about the high school that the participants attended.

Regardless of the particular speech perception model preferred, results from this experiment add to previous work demonstrating how stored social and phonetic information must have a direct link in the mind (Strand 1999, Johnson, Strand, and D’Imperio 1999, Hay, Warren and Drager 2006b). Variability that does not pattern according to linguistic factors is often ignored in formal linguistic theories, being treated as “noise” that does not need to be accounted for. This ignores the fact that some of this “noise” in production and perception becomes predictable when social factors are included in the analysis. The work presented in this paper provides evidence that social information is stored alongside fine-grained phonetic detail and discusses how this stored social information is accessed during speech perception.

## References

- Baayen, R., Davidson, D. and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items, *Journal of Memory and Language*, 59, 390–412.
- Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics*. Cambridge: Cambridge University Press.
- Benkí, J. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, 29, 1–22.
- Davis, P., Jenkin, G., & Coope, P. (2003). *NZSEI-96: an update and revision of the New Zealand Socioeconomic Index of Occupational Status*. Wellington: Statistics New Zealand.
- Davis, P., McLeod, K., Ransom, M., Ongley, P., Pearce, N., & Howden-Chapman, P. (1997). *The New Zealand Socioeconomic Index of Occupational Status (NZSEI): Research report #2*. Wellington: Statistics New Zealand.
- Drager, K. (2006). From bad to bed: the relationship between perceived age and vowel perception in New Zealand English. *Te Reo*, 48, 55-68.
- Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34, 409–438.
- Gentry, E. (2006). Hovering between South and West: Houston's merged dialect. *Paper presented at NWA V 35*. OH: The Ohio State University.
- Gibb, S. J., Fergusson, D. M., & Horwood, L. J. (2008). Effects of single-sex and co-educational schooling on the gender gap in educational achievement. *Australian Journal of Education*, 52, 301–317.

- Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.
- Hay, J., & Drager, K. (to appear). Stuffed toys and speech perception. To appear in *Journal of Linguistics*.
- Hay, J., Nolan, A., & Drager, K. (2006a). From fush to feesh: exemplar priming in speech perception. *The Linguistic Review*, 23, 351–79.
- Hay, J., Warren, P., & Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34, 458–484.
- Huntley, R., Hollien, H., & Shipp, T. (1987). Influences of listener characteristics on perceived age estimations. *Journal of Voice*, 1, 49–52.
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: selected readings* (pp. 269–293). Harmondsworth: Penguin.
- Jarvis, B. G. (2002). Medialab (version 2002) [computer software].
- Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & J. Mullennix (Eds.), *Talker variability in speech processing* (pp. 146–165). San Diego: Academic Press.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *Journal of Phonetics*, 34, 485–499.
- Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27, 359–384.
- Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas, *University of Pennsylvania Working Papers in Linguistics: Selected papers from NWA 36*, 14, 91–101.
- Labov, W. (1972). *Sociolinguistic patterns*. PA: University of Pennsylvania Press.

- Langstrof, C. (2006). *Vowel change in New Zealand*, PhD thesis, University of Canterbury. Unpublished PhD thesis.
- Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, *42*, 830–843.
- Maclagan, M. (1982). An acoustic study of New Zealand English vowels. *New Zealand Speech Therapists Journal* *37*, 20–26.
- Maclagan, M. A., & Gordon, E. (1996). Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change*, *8*, 125–147.
- Maclagan, M. A., Gordon, E., & Lewis, G. (1999). Women and sound change: conservative and innovative behaviour by the same speakers. *Language Variation and Change*, *11*, 19–41.
- Maclagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English “short” front vowel shift. *Language Variation and Change*, *19*, 1–25.
- Neiman, G. S., & Applegate, J. A. (1990). Accuracy of listener judgments of perceived age relative to chronological age in adults. *Folia Phoniatrica*, *42*, 327–330.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*, 62–85.
- Nosofsky, R. M. (1986). Attention, similarity, and identification-categorization relationship. *Journal of Experimental Psychology*, *115*, 39–57.
- Pantos, A. J. (2006). Redefining the South: Teenage Houstonians and the Southern Shift. *Paper presented at NWA 35*, OH: The Ohio State University.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. J. Hopper (Eds.), *Frequency effects and emergent grammar* (pp. 137–158). Amsterdam: John Benjamins.

- Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research*, 6, 111–128.
- Strand, E. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, 18, 86–99.
- Strand, E., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural Language Processing and Speech Technology* (pp. 14–26). Berlin: Mouton de Gruyter.
- Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America*, 69, 1465–1475.
- Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, 1, 179–195.
- Warren, P., Hay, J., & Thomas, B. (2007). The loci of sound change effects in recognition and perception. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* pp. 87–112. Berlin: Mouton de Gruyter.
- Wolfram, W. A. (1969). *A sociolinguistic description of Detroit Negro speech*. VA : Center for Applied Linguistics.

## Tables

Table 1: Vowel durations, F0, and F3 for all tokens of each voice.

voice	F0 (Hz)	F3 (Hz)	duration (ms)
F1	167	2949	13.8
F2	168	3020	20.6
M1	103	3264	23.4
M2	125	2728	17.4

Table 2: F1 and F2 for female voices. Token 1 is the most DRESS-like token and token 9 is the most TRAP-like token.

token	F1	F2
1 (most DRESS-like)	472.99	2559.6
2	504.2	2548.23
3	525.67	2527.26
4	554.22	2510.66
5	591.44	2495.48
6	631.45	2480.53
7	665.91	2465.84
8	685.58	2460.69
9 (most TRAP-like)	714.21	2450.06

Table 3: F1 and F2 for male voices. Token 1 is the most DRESS-like token and token 9 is the most TRAP-like token.

token	F1	F2
1 (most DRESS-like)	422.89	2141.65
2	440.63	2118.25
3	486.09	2095.7
4	513.79	2070.96
5	521.58	2043.99
6	527.59	2019.74
7	546.01	1995.74
8	590.51	1971.54
9 (most TRAP-like)	616.73	1948.29

Table 4: Distribution of participant characteristics across conditions.

participant characteristic	cond 1	cond 2	total
total participants	14	13	27
number of females	9	8	17
min age	18	18	18
mean age	27.3	26.4	26.9
max age	62	56	62
min SEI	57	62	57
mean SEI	104.7	110.5	107.5
max SEI	143	154	154

Table 5: Summary of perceived PHOTO-AGE.

photo	average
OM	48.3
OF	45.8
YM	20.8
YF	25.4

Table 6: Coefficients of fixed effects; higher coefficients indicate a greater likelihood of a TRAP response.

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-7.329	0.458	-16.006	<0.0001
TOKEN	0.961	0.031	30.592	<0.0001
INITIAL C = /h/	0.659	0.116	5.669	<0.0001
PHOTO-AGE	0.054	0.010	5.262	<0.0001
SUBJ-AGE	0.078	0.013	5.867	<0.0001
PHOTO- AGE:SUBJ- AGE	-0.003	0.000	-7.131	<0.0001
SUBJ-SEX = M	-0.370	0.153	-2.413	0.0158
SPEAKER- SEX = M	-0.328	0.137	-2.393	0.0167
SUBJ-SEX = M:SPEAKER -SEX = M	0.747	0.202	3.708	0.0002

Table 7: Estimated 50% crossover point in order of participant age, for questions matched with the photograph shown.

AGE	SEX	COND	OM	YM	OF	YF
18	F	1	5.351	5.528	5.026	5.158
18	F	2	6.204	5.449	5.795	5.234
18	F	2	5.476	5.724	5.158	5.342
18	M	2	5.319	4.994	5.680	5.439
19	F	1	6.095	4.838	5.640	4.707
19	F	1	6.729	4.253	6.895	5.057
19	F	2	6.164	6.269	5.833	5.911
19	F	2	5.270	5.088	5.644	5.510
19	M	1	5.757	5.934	5.432	5.564
19	F	2	5.469	5.646	5.144	5.276
20	F	1	5.181	5.430	4.863	5.048
20	F	2	6.322	5.854	5.939	5.591
21	M	1	5.950	6.055	5.618	5.697
22	M	1	4.844	5.092	4.526	4.710
23	F	1	5.191	5.153	5.578	5.550
24	M	2	5.318	5.495	4.994	5.125
24	M	1	4.486	4.663	4.893	5.024
25	M	1	5.209	5.028	5.584	5.449
26	M	2	6.343	6.090	6.711	6.523
28	F	1	5.073	5.250	4.748	4.880
30	M	2	5.587	5.620	5.981	6.006
32	F	2	5.074	5.696	5.455	5.373
39	F	1	4.853	5.101	5.266	5.451
43	F	1	5.938	5.326	6.273	5.819
44	F	2	7.185	4.278	6.581	4.422
56	M	2	6.476	4.86	5.989	4.789
62	F	1	6.846	5.301	6.365	5.218

## Figure Titles

Figure 1: The photographs used as visual stimuli. Clockwise from top left: older male (OM), older female (OF), younger female (YF), and younger male (YM).

Figure 2: Percent of DRESS (grey) and TRAP (black) responses when a voice was matched with the photograph of the older female (OF).

Figure 3: Percent of DRESS (grey) and TRAP (black) responses when a voice was matched with the photograph of the younger female (YF).

Figure 4: Percent of DRESS (grey) and TRAP (black) responses when a voice was matched with the photograph of the older male (OM).

Figure 5: Percent of DRESS (grey) and TRAP (black) responses when a voice was matched with the photograph of the younger male (YM).

Figure 6: The likelihood of perceiving TRAP depending on participant age and PHOTO-AGE. The lines represent participants' responses to the different average ages assigned to the photographs.

Figure 7: The likelihood of perceiving TRAP for male and female participants when responding to a female photograph (grey) and a male photograph (black).

## Figures



Figure 1: The photographs used as visual stimuli. Clockwise from top left: older male (OM), older female (OF), younger female (YF), and younger male (YM).

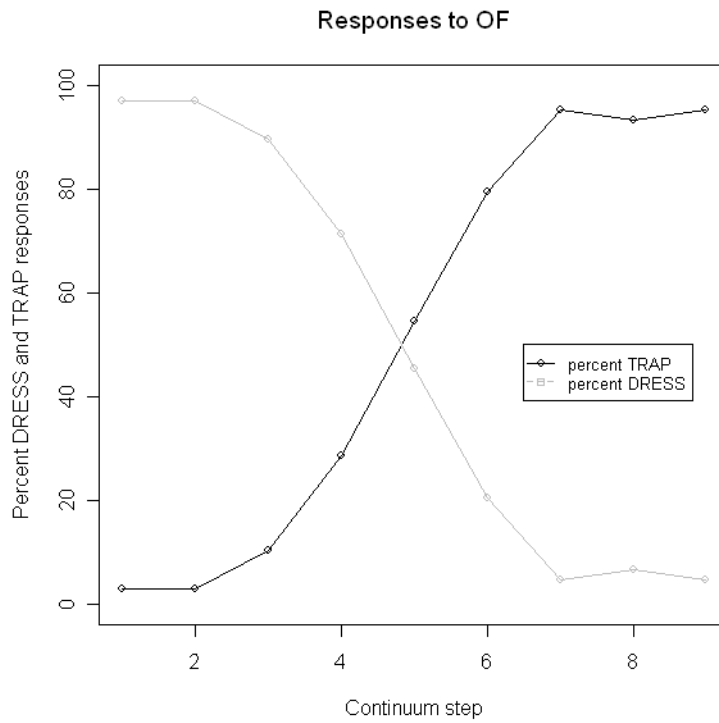


Figure 2: Percent of DRESS (grey) and TRAP (black) responses when a voice was matched with the photograph of the older female (OF).

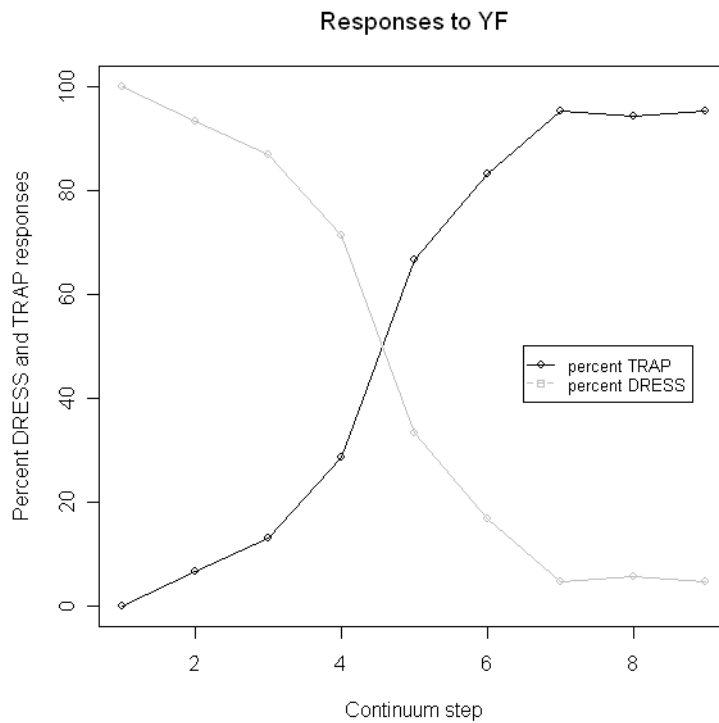


Figure 3: Percent of DRESS (grey) and TRAP (black) responses when a voice was matched with the photograph of the younger female (YF).

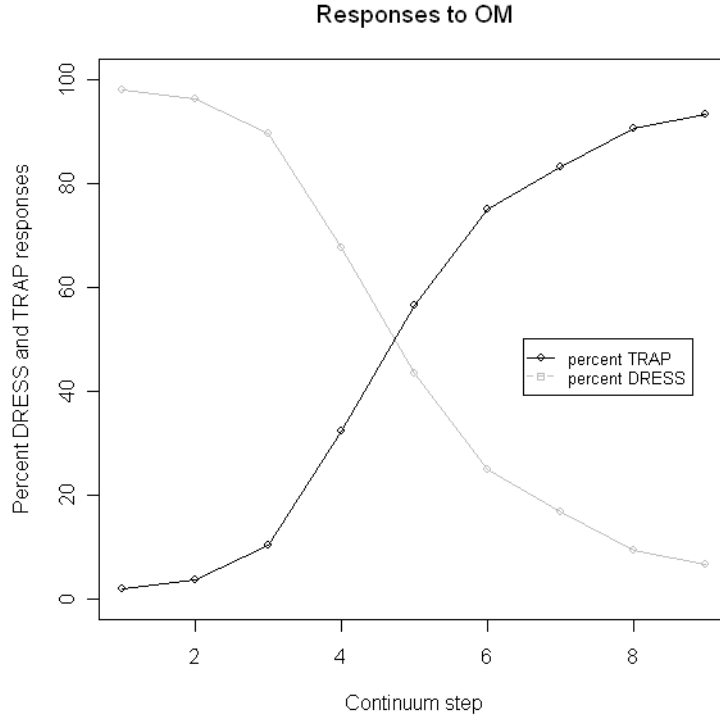


Figure 4: Percent of DRESS (grey) and TRAP (black) responses when a voice was matched with the photograph of the older male (OM).

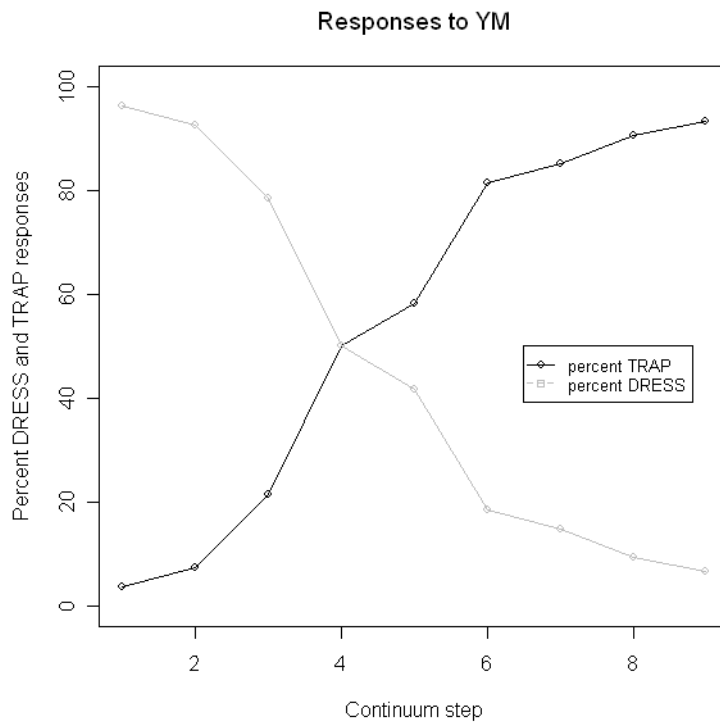


Figure 5: Percent of DRESS (grey) and TRAP (black) responses when a voice was matched with the photograph of the younger male (YM).

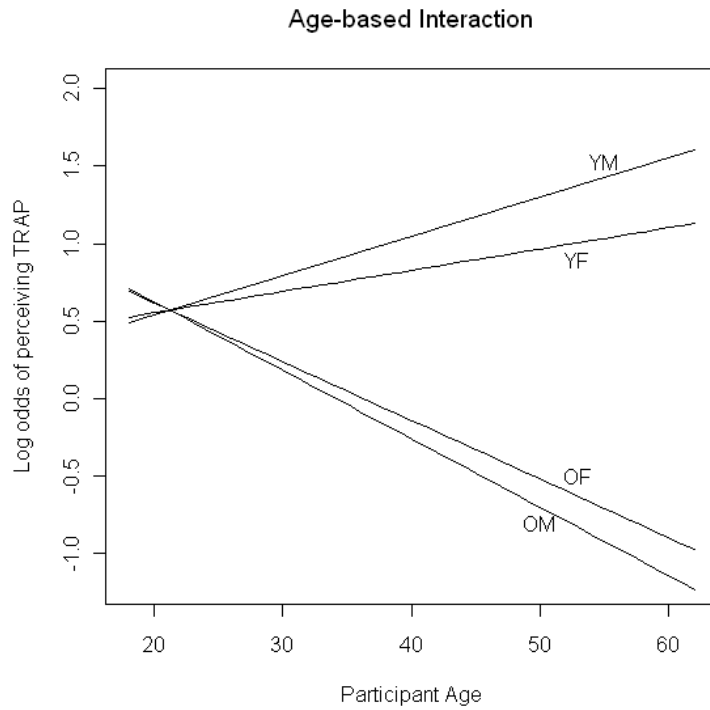


Figure 6: The likelihood of perceiving TRAP depending on participant age and PHOTO-AGE. The lines represent participants' responses to the different average ages assigned to the photographs.

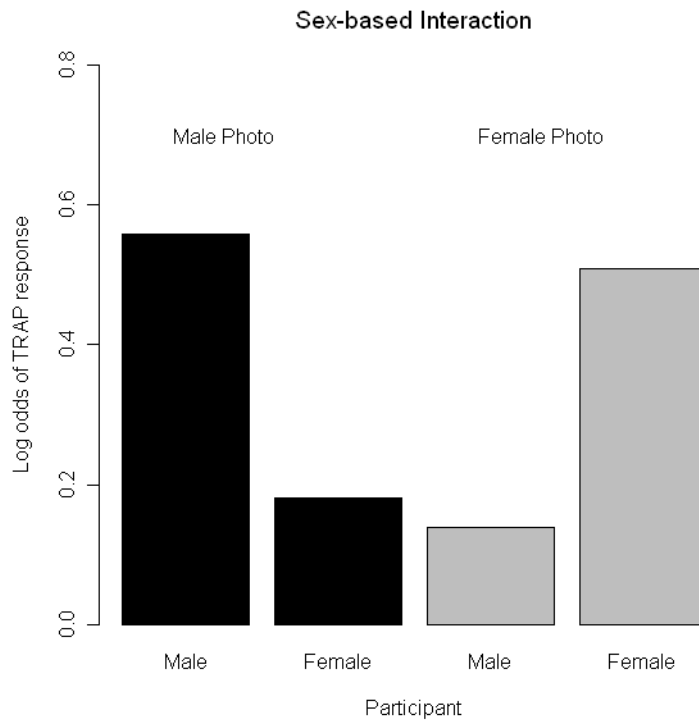


Figure 7: The likelihood of perceiving TRAP for male and female participants when responding to a female photograph (grey) and a male photograph (black).