

## Using Social Information in Language Processing

LAURA STAUM CASASANTO  
*Stanford University*

### 0. Introduction

Since its inception, the study of sociolinguistic variation has been primarily the study of sociolinguistically conditioned variable *production* (Labov, 1966, 1972; Rickford, 1987). The variationist project has been documenting and describing the way different groups of people produce language, studying the factors that condition variation at all levels of linguistic description, from phonetics to discourse patterns. More recently, sociolinguists have begun to analyze the social hierarchies and relationships that underpin this variable production (Eckert, 1989). However, very little is known about the *comprehension* of this variable linguistic behavior - what do listeners do with all this structure observed in socially conditioned variable production? If this information is somehow monitored, do listeners use the knowledge they accumulate?

One study of the effects of the sociolinguistic variable ING (e.g. *walkin'* vs. *walking*) on listeners' attitudes suggests that listeners do make use of linguistic variation to make judgments about speakers. In the matched guise paradigm (Lambert, Hodgson, Gardner and Fillenbaum, 1960), only the critical feature is manipulated between speech samples, which are then evaluated by naïve listeners. Manipulating the realization of the final nasals in ING influenced listeners' judgments about the person who used it – the alveolar nasal (n) makes speech sound more casual and less educated/intelligent, while the velar nasal (ng) makes speakers sound more formal and more educated/intelligent (Campbell-Kibler, 2006). Listeners use the realization of ING to make attitude judgments about speakers, and they have explicit and implicit beliefs about who uses this variable and what it means about them. But what about the other way around – do people use information about speakers to make judgments about speech?

Listeners use information about speakers in making metalinguistic judgments. Strand (2000) showed that seeing a picture of a man or a woman affects how people categorize ambiguous stimuli between sibilants *s* and *esh* and between the back vowels in *hood* and *hud*, and more stereotypical men and women elicit stronger effects than less stereotypical men and women. Because these variables

are associated with gender through vocal tract characteristics, it's unclear to what extent the effect relates to sociolinguistic knowledge. The social meaning of fricative frequency and back vowel formant frequencies is not something well-understood. - we don't know to what extent speakers use them as sociolinguistic resources and how variation in these domains is structured.

More importantly, because phone categorization is not a part of normal language comprehension, this study leaves open the question of whether people use information about speakers when they're not just making a metalinguistic judgment, but actually trying to understand a speaker in real time. Is social information one of the clues listeners use when figuring out the puzzle of spontaneous speech?

The ideal sociolinguistic variable for a study of language comprehension is one that has been well-studied from the point of view of production. Consonant cluster reduction (a.k.a. *t/d* deletion) is a phonetic variable in English in which final coronal stops in consonant clusters may be deleted in some environments. This variable makes a good test case because "...over the past thirty-five years, this phenomenon has been studied in more detail than probably any other variable phonological phenomenon" (Coetzee, 2004).

In addition, consonant cluster reduction has a very convenient property, which is that the deletion can sometimes cause ambiguity between two words. For example, the word *mast* produced without its final consonant becomes ambiguous with the word *mass*. This situation provides a good opportunity to see the effects of contextual information on the resolution of this ambiguity.

This reduction is conditioned by several aspects of the linguistic environment, including features of the segment before the stop (as in *last night* vs. *hard night*), features of the segment following the stop (*fast car* vs. *fast action*), and the morphological status of the stop (*past resolution* vs. *passed resolution*) (Labov et al., 1968; Fasold, 1972). The possible realizations of the final consonant vary along a continuum from an aspirated *t* with a strong release burst to a completely deleted *t* which leaves few signs that it ever existed in the acoustic signal. While there is some evidence that many different realizations of this variable may be socially meaningful in various ways (Podesva, 2006), the standard taxonomy of consonant cluster reduction distinguishes only between two variants: the deleted and non-deleted variants. It is these two categories of consonant realization that have been studied extensively from the perspective of production, and so that is the distinction I consider in the experiments presented in this paper.

Consonant cluster reduction is also conditioned by many stylistic and social factors, and this conditioning has been studied extensively in a variety of social groups. The deleted variant is less formal than the non-deleted variant, younger people use the deleted variant more than older people do, men use the deleted variant more than women do, and African Americans use the deleted variant more than Anglo Americans do (Wolfram, 1969). Because ethnicity is a very robust conditioning factor, and because it is possible to manipulate the purported ethnic-

## *Using Social Information in Language Processing*

ity of a speaker by showing pictures, it is the relationship between ethnicity and consonant cluster reduction that I examine.

Experiment 1 investigates whether listeners have implicit knowledge of the relationship between ethnicity and *t/d* deletion in American English. Experiment 2 determines whether listeners make use of this knowledge to resolve ambiguity.

Although there is copious information available about what kinds of speakers use the deleted and non-deleted variants and in what circumstances, very little is known about how listeners interpret them. Campbell-Kibler's 2005 dissertation includes a pilot study comparing listeners' reactions to release vs. non-release of /t/, which may be related to the way listeners interpret *t/d* deletion; the results of this study suggested that released /t/ influenced how smart, casual, formal, and relaxed listeners believed a speaker to be, but not how educated and wealthy. Because these are both non-deleted *ts*, it is not clear how well these findings correspond to deleted vs. non-deleted *t/d*, and they don't provide much information about the stereotypes listeners have about who uses what variants of *t/d* overall.

While many social factors correlate with *t/d* deletion, the fact that linguists can attend to and measure rates of *t/d* deletion and find correlations with social factors does not mean that listeners regularly do this. Campbell-Kibler (2005) provided some of the first evidence that listeners find changes in single variants meaningful; however, in the case of each variant or variable, the burden of proof is on the linguist to demonstrate that these correlations have some psychological reality for the listener. Thus, before addressing the question of whether listeners use information about social characteristics of the speaker to resolve ambiguities caused by sociolinguistic variation, it is necessary to establish that they have some way of linking speaker social characteristics to linguistic behavior.

In principle, listeners could attend to and store information about all the relationships that exist in the world between social characteristics and linguistic variation. This may in fact be the case; however, it is also possible that listeners are insensitive to much of this variation. Which variables they attend to and store information about could be constrained by a variety of factors. Perhaps listeners attend only to variation that is above a certain threshold of linguistic salience. They might attend only to variation that is relevant to their social lives in certain ways. Listeners might attend only to linguistic variables that they themselves use. The possible constraints on this attention are nearly limitless.

Because it is impossible to consider all sources of socially meaningful variation at once, a fruitful way to pursue the question of how sociolinguistic variation affects the listener is to frame empirical questions in limited domains. The results of experimental investigations of individual variables can be combined to construct a larger understanding of how listeners map the landscape of sociolinguistic variation, and how this information is incorporated into the mechanisms of language understanding.

Two experiments investigate whether listeners have implicit knowledge of the relationship between ethnicity and *t/d* deletion in American English, and if so, how they may use this knowledge in language processing.

## 1. Experiment 1

Experiment 1 was designed to determine whether listeners have explicit knowledge about the relationship between *t/d* deletion and ethnicity. If listeners keep track of the distribution of deleted final consonants in the input in some way, they should associate deleted final stops more with African Americans than with white speakers. This is an important prerequisite for Experiment 2 - if listeners have knowledge about the distribution of the variants with respect to ethnicity, then it will be possible to find out if they use this information in processing language. If listeners do not appear to keep track of which types of speakers produce the different variants, then they have no way to link social characteristics of the current speaker to probabilities of hearing specific variants. If they cannot make this link, then they cannot use this information to help them understand language.

Experiment 1 is designed as a modified Matched Guise study (Lambert et al. 1965, Lambert 1967), in which minimal pairs of utterances are constructed to create “guises” that differ only in the feature relevant to the study. If listeners have unconscious associations between deleted *t/d* tokens and African American speakers, they will more often rate sentences with deleted tokens likely to have come from African Americans than sentences with non-deleted tokens.

### 1.1 Participants

111 Stanford University undergraduates received course credit in an introductory psychology class for their participation in this study. Participants were of both genders and a mixture of ethnicities.

### 1.2 Materials

24 sentences were constructed so that each included a word with a consonant cluster that could be subject to *t/d* deletion (e.g. *mast*, *least*, *wind*). These consonants were primarily in phonological environments that promote consonant cluster reduction, such as following a fricative or a nasal and preceding a stop or a glide. Each sentence appeared in two versions, the standard version and the deleted version. In the standard version, this word was presented with its normal orthography (Fig. 1A). In the deleted version, this word appeared with its final stop replaced by an apostrophe (Fig. 1B), indicating a *t/d* deletion.

Participants received one of two questionnaires. Questionnaire A contained twelve target items, all presented in normal orthography (standard version), and twelve similar sentences presented in normal orthography as filler items. Questionnaire B contained twelve target items, all written with an apostrophe (deleted version). Questionnaire B also contained an equal number of fillers (twelve), all of which contained another nonstandardism (such as *coffee* spelled *cawfee*). Underneath each nonstandardism in Questionnaire B, participants saw a ‘translation’ of the nonstandardism to help them understand what it was intended to mean. In the case of *t/d* deleted items, this consisted of the same word written in standard orthography. Because the standard versions of the sentences appeared only in

## *Using Social Information in Language Processing*

Questionnaire A and the deleted versions appeared only in Questionnaire B, the difference between standard and deleted versions was a between-subjects comparison. Items were counterbalanced across versions of each questionnaire so that all twenty-four items were seen in both standard and deleted versions.

Pictures of potential speakers were taken from a database of University of Pennsylvania ID photos (Killgore et al., 2000), and included the shoulders and head of college-aged individuals, on a white or neutral background (Fig. 1). Four black and four white male individuals were used for both the target and the filler items, with a variety of pairings of black and white individuals instantiated.

### **1.3 Procedure**

58 participants filled out Questionnaire A, which had the following instructions:

Below each pair of pictures is a sentence. For each pair, please circle the person you think is more likely to have said the sentence.

The remaining 53 participants filled out Questionnaire B, which contained the nonstandard items, and had the following instructions:

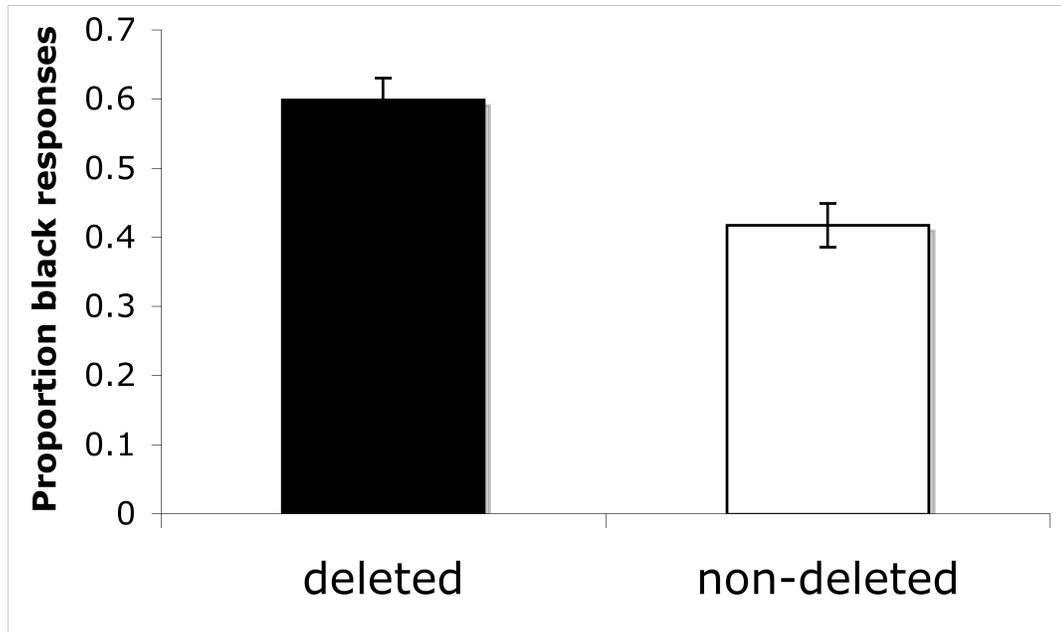
The sentences below are transcribed from natural speech. Some of the words speakers used are non-standard. ‘Translations’ for these words are given in parentheses. Try to imagine how they might have sounded in your mind’s ear. For each sentence, decide which person pictured above it is more likely to have said it.

In both cases, participants’ task was to circle the picture in the pair above the sentence of the person they think is more likely to have said the sentence. Each sentence had one picture of an African American and one picture of an Anglo American above it. Using written stimuli allows this experiment to address the influence of *t/d* deletion without the influence of auditory cues to ethnicity. Because participants’ knowledge of the relationship between *t/d* deletion and ethnicity might not be conscious, participants were not made aware that consonant cluster reduction or ethnicity were of interest. Because other non-standardisms were present in the fillers, *t/d* deletion was not particularly salient; however, because all pairs of faces contained one white and one black face, participants may have realized that ethnicity was of interest in the experiment.

### **1.4 Results and Discussion**

Participants attributed 60% of the ‘deleted’ sentences, represented with apostrophes, to the African American person pictured, and the remaining 40% to the Anglo American person pictured. By contrast, they attributed only 42% of the non-deleted sentences, with normal orthography, to the African American, attributing the remaining 58% to the Anglo American. This pairwise difference between proportions was significant ( $t(1,109)=4.86$ ,  $p<.001$ ) (Fig. 1, below), indicating that participants do associate *t/d* deletion more with African American speakers than with white speakers.

While the absolute frequency with which participants associated *t/d* deletion with black and white speakers may be specific to the types of sentences the deletions occurred in, the fact that participants showed the same association between *t/d* deletion and ethnicity that occurs in production strongly suggests that they track probabilities related to social factors during language comprehension.



**Figure 1.** Participants selected the African American from the pair of pictures 60% of the time when they saw the deleted versions of the sentences, but selected the African American only 42% of the time when they saw the non-deleted (standard) versions of the sentences. Error bars represent s.e.m.

### 1.5 Summary

Experiment 1 demonstrated that listeners associate *t/d* deletion more strongly with African Americans than they associate this variant with whites, which is consistent with the facts about production of this variable. While this experiment confirms that listeners have access to statistical relationships between *t/d* deletion and ethnicity, they leave an important question unanswered: do listeners use this knowledge when understanding language? Experiment 2 investigated whether listeners use the linguistic stereotypes identified in Experiment 1 automatically in language processing tasks.

## 2. Experiment 2

## *Using Social Information in Language Processing*

Experiment 2 investigated this question using pairs of sentences that can be temporarily ambiguous when spoken aloud. The source of the ambiguity between the two sentences in each pair is the deletion of the final *t* or *d* segment of a word (*t/d* deletion), such that, for example, the word *mast* becomes confusable with the word *mass*. Each pair contains one word (the *t*-word) that has a final coronal stop (*t* or *d*) and one word (the non-*t*-word) that contains the same string of phones as the first word, but without the final coronal stop. When the *t*-word (*mast*, below) is pronounced, the final coronal stop may be deleted, yielding a phone string that is identical to that of the standard pronunciation of the non-*t*-word (*mass*, below). The sentences will remain ambiguous between the meanings of these two words for several words, and then be disambiguated by later words in the sentence:

- 1a) [The mast probably lasted] through the storm.
- 1b) [The mass probably lasted] an hour on Sunday.

In the above example, a listener can not tell whether the word they heard was *mass* or *mast* until they get the information *through the storm* or *an hour on Sunday*, at which point they can tell from this context which lexical item was intended by the speaker.

If listeners use their knowledge of sociolinguistic variation when they understand sentences, then they will be more likely to predict a consonant cluster reduction when they believe the speaker is black than when he is white. Thus, they should reach the *mast* interpretation faster when they believe the speaker is black than when they believe he is white, because reaching this interpretation involves inferring a deleted stop. By contrast, they will reach the *mass* interpretation faster when they believe the speaker is white than when they believe he is black, because reaching this interpretation involves rejecting the alternative with a deleted stop.

### **2.1 Participants**

40 American native English speakers from the Stanford University community participated in this study in exchange for payment.

### **2.2 Materials**

24 pairs of sentences were constructed which were identical for the first few words (the section underlined in 2a and 2b below) except for a critical word (italicized). The critical words in each pair of sentences were identical except for the presence or absence of a stop at the end of a final consonant cluster:

- 2a. The *mast* probably lasted through the storm.
- 2b. The *mass* probably lasted an hour on Sunday.

These nearly identical sections (underlined above) would be ambiguous when spoken aloud if a speaker used the deleted variant of a word like *mast*. Because

the deleted variant is a possibility, a listener would not be able to tell whether the word *mass* or *mast* was intended at this point in the sentence. The pairs of sentences, however, are all disambiguated by the endings of the sentences, which are only consistent with one of the interpretations of the beginning. For example, *through the storm* is only consistent with the *mast* interpretation of the beginning, and *an hour on Sunday* is only consistent with the *mass* interpretation of the beginning.

24 filler pairs were created that also contained an ambiguity that was resolved later in the sentences:

- 2a) While Bill hunted the deer ran into the woods.
- 2b) While Bill hunted the deer we made the fire.

None of these ambiguities were related to *t/d* deletion. While some of these temporary ambiguities could be disambiguated by prosody, care was taken to select recordings in which the prosody would be appropriate for both readings of the ambiguous portion. These ambiguous fillers masked the experimental sentences; when a subject encountered a temporarily ambiguous sentence, it was only an experimental sentence half the time. This prevented the experimental sentences from standing out from the fillers, as well as preventing them from exclusively constituting the most difficult sentences to understand and respond to.

In addition, 48 unambiguous filler sentences were constructed of similar length and complexity. Thus the full set of stimuli includes 24 pairs of experimental stimuli, 24 pairs of ambiguous fillers, and 48 unambiguous fillers which were unpaired. The total number of sentences used in the experiment was 144; however, because each subject only heard one sentence from each pair, this resulted in 96 total sentences presented to each subject, half of which were ambiguous. The ratio of fillers to target stimuli was 3 to 1.

I recorded approximately 25 Stanford undergraduates, both male and female, of various ethnicities, reading all 144 of these sentences aloud. Each sentence was preceded by a context sentence to make the participants' reading of the experimental and filler sentences more natural:

- 4a) I went to a new church last week that has very short services. (context)  
The mass probably lasted an hour on Sunday. (target)

- 4b) I hope my old boat wasn't damaged by the wind last night. (context)  
The mast probably lasted through the storm. (target)

Most of these speakers were recruited from the Stanford University Linguistics Department subject pool; because the subject pool did not contain many African American participants, additional speakers were recruited via email and paid ten dollars for their time. Recordings were digital, made in a sound-proof booth. Sen-

## *Using Social Information in Language Processing*

tences were displayed to the subjects using the experimental software PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Macintosh computer.

Recordings of white and African American males were used for the target stimuli. The targets were restricted to males for two reasons. First, only one gender could be used because mixing genders would likely create implausible pairings between pictures and voices. Second, males were chosen over females because they tend to have higher rates of *t/d* deletion overall than females. Using males for the stimuli decreased the likelihood of encountering a floor effect in listener responses, which would result if listeners did not predict the deleted word in any of the cases, because they found it too unlikely.

Each target item was recorded once by a black speaker, and once by a white speaker. The actual ethnicity of the speaker who recorded the sentence was counterbalanced across items, so that the pairing of voice and face was equally felicitous across conditions, on average.

The speaker pictures from Experiment 1 (4 black males and 4 white males) were used for the critical trials and one third of the fillers (24 targets and 24 fillers), while 8 females of various ethnicities were displayed with the other two thirds of the fillers (48). This resulted in each subject seeing a female face in half the trials and a male face in the other half of the trials. Within each subject, each face was paired with only one voice, to maximize the plausibility of the premise that the pictures represented the speakers. The face/voice pairings were counterbalanced across subjects.

Because all subjects saw a mix of genders and ethnicities in the experiment, they were unlikely to be able to deduce that the experiment concerned only white and African American males. Thus, unlike in Experiments 1 and 2, neither *t/d* deletion nor black and white ethnicities were salient in this experiment.

### **2.3 Procedure**

Participants were instructed to listen to a short sound clip while looking at a picture of a face, which they were told represented the speaker of the clip. They heard the ambiguous portion of one of the sentence pairs, which contained no final stops. Although both sentences in each pair were recorded by the readers, the sound files used in this experiment were only excerpted from recordings of the version of each sentence that never contained a final stop (e.g. the *mass* version). Participants never heard any version of the experimental sentences that contained an underlying *t/d*, so that there are no cues in the speech stream indicating the presence of a deleted stop.

Participants then saw one of the sentence endings appear below the picture of the speaker. For example, in one trial, participants heard:

4. The [mas] probably lasted

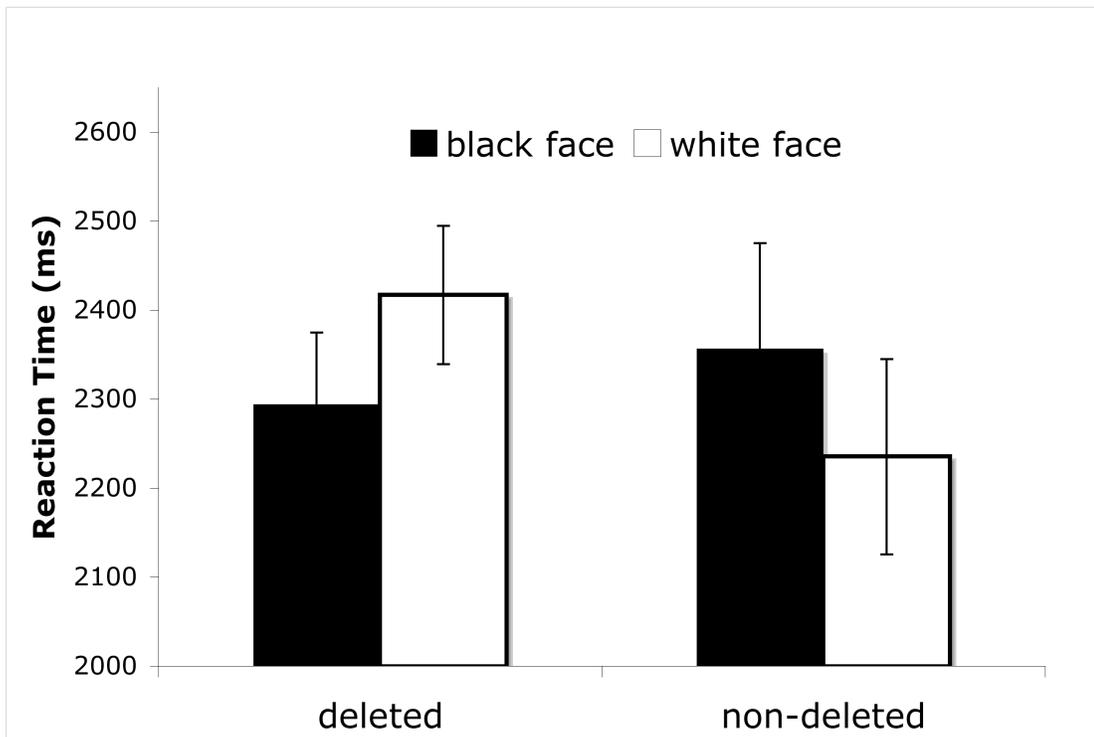
After this clip, one of the following endings appeared on the screen:

- 5a. ...through the storm.
- 5b. ...an hour on Sunday

In half the cases participants saw a continuation that made sense if the ambiguous word had no final stop (e.g. *an hour on Sunday*, which makes most sense if the word was *mass*), and in the other half of cases they saw the other continuation, which made sense if the ambiguous word did have a final stop that had been deleted (e.g. *through the storm*, which makes most sense if the word was *mast*). Participants' job was to assess whether the ending created a sensible sentence in combination with the beginning they had heard, and response times were measured from the time the continuation appeared on the screen.

## 2.4 Results and Discussion

As predicted, participants responded marginally faster to the continuation that was compatible with the word whose underlying phonemic form has a *t* (the *mast* interpretation) when they saw a black face ( $t(1,39)=1.21$ ,  $p=.11$ ,  $t(1,23)=1.8$ ,  $p=.04$ ). However, also as predicted, they responded faster to the continuation that was compatible with the word whose underlying phonemic form does not have a *t* (the *mass* interpretation) when they saw a white face ( $t(1,39)=1.75$ ,  $p=.04$ ,  $t(1,23)=1.81$ ,  $p=.04$ ). This difference of differences yields the predicted ethnicity by *t/d* deletion interaction ( $F(1,39)=5.64$ ,  $p=.02$ ,  $F(1,23)=9.23$ ,  $p=.006$ ) (Fig. 2, below).



## *Using Social Information in Language Processing*

**Figure 2.** Participants responded faster to the continuation that was consistent with the deleted interpretation (e.g. *mast*) when they saw a black face representing the speaker, but they responded faster to the continuation that was consistent with the non-deleted interpretation (e.g. *mass*) when they saw a white face representing the speaker. Error bars indicate s.e.m.

### **2.5 Summary**

Experiments 1 and 2 provide evidence that listeners combine their knowledge of sociolinguistic variation, social information from the scene, and auditory information from the speech stream to construct an interpretation of the speech they have heard. The results of Experiment 1 addressed the issue of knowledge – listeners have acquired, though we do not know through what means, implicit knowledge of the relationship between ethnicity and likelihood of *t/d* deletion. This, however, is not sufficient to determine that the social characteristics of speakers will influence language comprehension. Listeners would also need to have access to and interpret social information about the speaker in real time, while they are using all kinds of other information, in order for this knowledge to be of any use to them in comprehension. The results of Experiment 2 indicate that they in fact do this. When social information about the speaker is available to listeners, they integrate this information with information from the speech stream and use their sociolinguistic knowledge about probabilistic relationships between social information and language to understand language. This suggests that information that is not represented linguistically can be integrated into language comprehension, as long as listeners have probabilistic knowledge that coordinates this type of information with linguistic representations.

### **References**

- Campbell-Kibler, K. 2005. *Listener perceptions of sociolinguistic variables: The case of (ING)*. Doctoral dissertation, Department of Linguistics, Stanford University.
- Coetsee, A. 2004. *What it Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. Doctoral dissertation, Department of Linguistics, University of Massachusetts, Amherst.
- Eckert, P. 1989. *Jocks and burnouts: Social categories and identity in the high school*. New York: Teachers College Press.
- Fasold, R. 1972. *Tense marking in Black English: A linguistic and social analysis*. Arlington, Va: Center for Applied Linguistics.

Laura Staum Casasanto

- Labov, W., P. Cohen, C. Robins, and J. Lewis. 1968. *A study of the non-standard English fo Negro and Puerto Rican speakers in New York City*. Final report, Cooperative Research Project 3288. Vols. I and II.
- Labov, W. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Labov, W. 1972. *Language in the Inner City*. Philadelphia: U. of Pennsylvania Press.
- Lambert, W.E., R. C. Hodgson, R. C. Gardner, and S. Fillenbaum. 1960. Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology*, 60, 44-51.
- McGurk, H. and J. McDonald. 1976. Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Nygaard, L and E. Lunders. 2002. Resolution of lexical ambiguity by emotional tone of voice. *Memory and Cognition*, 30, 583-593.
- Podesva, R. 2006. *Phonetic Detail in Sociolinguistic Variation: Its Linguistic Significance and Role in the Construction of Social Meaning*. Doctoral Dissertation, Department of Linguistics, Stanford University.
- Rickford, J. R. 1987. *Dimensions of a Creole Continuum*. Stanford: Stanford University Press.
- Strand, E. 1999. Uncovering the roles of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18. Special Issue: Attitudes, Perception, and Linguistic Features. Lesley Milroy and Dennis R. Preston, Guest Eds.
- Tanenhaus, M., M. Spivey-Knowlton, K. Eberhard, and J. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.
- Trueswell, J. C., I. Sekerina, N. M. Hill, and L. M. Logrip. 1999. The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition*, 73, 89-134.
- Van Berkum, J.J.A., D. Van den Brink, C. Tesink, Kos, M. and P. Hagoort. 2008. The neural integration of speaker and message. *Journal of Cognitive Neuroscience*.
- Wolfram, W. 1969. *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.

Laura Staum Casasanto  
Stanford University  
Department of Linguistics  
Margaret Jacks Hall  
Stanford, CA 94305

lstaum@stanford.edu