Sociophonetic Variation in Speech Perception
Katie Drager*
Department of Linguistics, University of Hawai‘i at Mānoa

Abstract
In this paper, I review previous research that investigates sociophonetic variation in speech perception. I also argue that more speech perception work is needed to inform work on variation in speech production, particularly in the areas of language change and stereotype formation. Exploring the mental representations and processing of social and linguistic information as well as treating phonetic and social factors as multidimensional and interacting will take future work in sociophonetics in new and exciting directions.

Introduction
The vast amount of work which examines the relationship between social factors and phonetic variation has focused on variation in speech production. However, a growing body of work demonstrates a link between social factors and variation in perception; not only do individuals produce phonetic variables in socially meaningful ways, they perceive speech differently depending on trends in their own production, their previous experience with other dialects, and the social characteristics that they attribute to the speaker. Additionally, listeners make judgements about a speaker based only on hearing their speech, attributing social characteristics and assigning broad social categories in consistent ways.

In this paper, I step through previous socioperceptual work and discuss some of the benefits of examining variation in speech perception. Perception studies provide an opportunity to ask research questions in addition to those explored in work on speech production. Production and perception studies can complement each other to aid our understanding of how language change occurs, how stereotypes are formed, how linguistic variation is stored in the mind, and what kinds of cognitive processes underlie language use. While this paper is primarily a review of previous literature, it also discusses the advantages of conducting socioperceptual research, and it suggests some research questions for future work. Discussions of previous work investigating sociophonetic variation in perception can also be found in Thomas (2002) and Foulkes et al. (2010b).

1. Why Perception?
Phonetic variation in speech production is now known to correlate with a number of factors, including social characteristics of the speaker and the formality of the situation (Labov 1972), token frequency (the number of times a speaker has encountered a word) (Bybee 2002), and how predictable a word is given its position in a sentence (Jurafsky et al. 2002). Furthermore, there is evidence that some of this information is stored and affects speech processing (Jurafsky 2003).

Similarly, phonetic variation in speech is correlated with the speaker’s social characteristics (see e.g., the contributions to Chambers et al. (2002)). Speakers are not constant in
the phonetic realisations they produce but shift depending on the style they construct in a given context. The work presented in the following sections helps to demonstrate how individuals have (not necessarily conscious) knowledge of sociolinguistic variables and that they can use this knowledge during the perception of speech. While the majority of linguistic theories do not attempt to account for socially conditioned variation, socioperceptual studies, together with sociolinguistic work in production, provide evidence that sociolinguistic variation is not systematic ‘noise’ that is filtered out during the processing and storage of speech events; the relationship between social and linguistic information must be stored in the mind in such a way that it can be accessed during speech perception. Thus, if linguistic theories aim to provide an accurate representation of linguistic variation, they must account for the richness of social theory, the gradience of phonetic detail, and the probabilistic distribution of linguistic variables.

Furthermore, results from socioperceptual studies also have implications for our understanding of sound change and of language-based prejudices: if both a listener’s expectations regarding a speaker and the objects in the speaker’s environment can affect how their speech is perceived by others, what does this say about sounds that are stored and associated in memory with the speaker (and other individuals who are perceived to be similar to the speaker)? In combating ‘linguicism’, a form of prejudice based on language, it is crucial that we understand the relationship between how sounds are perceived and what types of social judgements are made based on exposure to those sounds. Likewise, if we hope to understand how sound change occurs, we must explore the extent to which stored memories of sounds are affected by factors in the listener’s environment as well as the cognitive processes underlying the perception of speech.

2. Production, Perception, and Exposure

The phonetic variants observed in the speech of an individual are reflected in the way that the individual perceives sounds. For example, an individual’s regional origin can predict not only how they pronounce sounds but also how they perceive them (Ladefoged and Broadbent 1957; Willis 1972; Fridland and Okamoto 2009).

However, the link between an individual’s production and their perception is not entirely straightforward. Some research suggests that speech perception may not always be affected when there is a change in production, even when the change in production is a potentially long-term shift (Evans and Iverson 2007). Some groups of listeners appear to use phonetic cues during perception that they do not use in their own production but that are used by other groups (Thomas 2000). Thomas (2000) found that listeners of different ethnicities and from different regions used some of the same phonetic cues in the preceding diphthong to identify whether a consonant was voiced or voiceless (as in tide and tight). In regard to the effect of diphthong qualities, the responses during the perception experiment were similar despite only one of the subject groups in the production experiment demonstrating much stronger patterns relating the phonetic cues of the diphthong to the voicing on the consonant. This provides evidence that individuals can have the ability to use phonetic cues during perception even if they do not use those cues in production. In a perception study of sounds undergoing a merger in Swedish, Janson and Schulman (1983) found that listeners who maintained a distinction were not always accurate at identifying tokens, even when they were produced by someone who also maintained a distinction. Similarly, Hay et al. (2006b) conducted an experiment in New Zealand which also investigated the perception of sounds that were involved in an ongoing merger. In contrast to Janson and Schulman’s (1983) results, Hay et al. (2006b) found
that listeners were highly accurate at identifying distinct tokens, even if they did not maintain a distinction in their own speech.

Amount of exposure to other dialects affects both speech processing and the ability of listeners to identify social characteristics of the speaker (Bowie 2000; Clopper and Pisoni 2004). Clopper and Pisoni (2004a) found that amount of previous exposure to a dialect affected accuracy at identifying the regional origin of the speaker. In phoneme and word identification tasks, even relatively small amounts of exposure to speech can cause listeners to adjust their categorisation (Norris et al. 2003). However, listeners only appear to be affected by such exposure if the variable is interpretable as an idiolectal (speaker-specific) feature rather than a dialectal (generalisable) feature (Kraljic et al. 2008). While differing degrees of exposure to other dialects can affect how sounds are perceived, previous exposure affects perception differently for different types of tasks depending on whether they require short-term or long-term storage (Sumner and Samuel 2009).

3. Attribution of Social Information

Sociolinguistic work on identity construction often focuses on the social meaning of linguistic variables from the point of view of the speaker. But in order for the variable to become imbued with social meaning, there must be a listener present to hear the token and attribute meaning to it. In investigating the role of the perceiver in the construction of socially meaningful phonetic factors, it is necessary to examine what kind of social information is attributed to a speaker based only on exposure to phonetic information and what kind of phonetic cues can become socially meaningful.

There is ample research which provides evidence that social information (both personality traits as well as broad social categories) can be extracted from auditory input in fairly consistent (though not necessarily accurate) ways. Researchers have investigated the degree to which listeners can identify a speaker’s ethnicity (Buck 1968; Tucker and Lambert 1969; Purnell et al. 1999), gender and sexuality (see Munson and Babel (2007)), regional dialect origin (Bush 1967; Preston 1999; Williams et al. 1999), and the consistency with which listeners, based solely on short clips of speech, attribute personality traits such as friendliness to the speaker (Lambert et al. 1969, Addington 1968; Preston 1999; Bayard 2000). In an experiment examining listeners’ perceptions of a speaker’s regional origin, Clopper and Pisoni (2004b) found that although listeners were not especially accurate when identifying the actual regional origin of the speaker, they responded in ways consistent with one another and appeared to be using a small set of phonetic cues when assigning a region. In speech perception experiments where listeners were asked to identify the speaker’s ethnicity, accuracy was inversely proportional to the number of ‘atypical’ variants: variants considered atypical for speakers of that ethnicity (Thomas and Reaser 2004). Foulkes et al. 2010a conducted an experiment in Tyneside where they played tokens of children’s speech to adults. They found that when identifying the sex of a child, listeners appeared to use a number of different acoustic cues, including whether or not the child produced a laryngealised voiceless stop, a variant associated with males in the region. When listeners from other regions were asked to complete the same task, they did not display the same sensitivity to the realisation of the stops. This provides evidence that individuals are sensitive to socially conditioned variation when perceiving speech. The ability to associate some perceived characteristic of the speaker with the phonetic variants that speaker tends to produce appears to be highly productive. There is some evidence that even a small amount of exposure to phonetic patterns distributed over
novel, previously unencountered groups can affect listeners’ categorisation of a speaker into one of the groups (Docherty et al. 2008).

There is some evidence that different perceived speaker characteristics can interact with each other in interesting ways. For example, Williams et al. (1999) found that listeners who were asked to identify the regional origin of a speaker were more likely to claim that the speaker was a member of their in-group if that speaker was judged to be likable. The field would benefit from a greater amount of work which follows this line of enquiry: What kinds of social judgements can affect each other and to what extent is the categorisation of encountered sounds affected by these judgements?

There are also experiments investigating the interaction of multiple phonetic variables. Levon (2007) conducted a speech perception experiment where participants listened to one of four guises that were counterbalanced across variables and included voices of both homosexual and heterosexual males. Participants were asked to rate the voices on a number of scales, including Straight/Gay and Effeminate/Masculine. Levon (2007) found that participants’ ratings were predictable depending on a combination of the phonetic cues manipulated and the actual sexuality of the speaker. This result helps to demonstrate how listeners make judgements about a speaker depending on multiple phonetic cues in the speech signal.

Another study along these lines was conducted by Campbell-Kibler (2007), who played groups of participants clips of speech and asked them to comment on the speakers (e.g., What can you tell me about Jason?). Across the two experimental conditions, the clips of speech used were identical except that word-final nasals in gerunds, present participles, and the words nothing and something were manipulated: in one condition the alveolar nasal [n] occurred in a word (e.g., fishin’ in the -in guise) and in the other condition the velar nasal [ŋ] occurred in that word (e.g., fishing in the -ing guise). Although all other aspects of the utterances were identical, speakers were more likely to be rated as educated and articulate when in the -ing guise than when in the -in guise. But the variable did not affect the perception of social characteristics equally for all voices; participants were more likely to identify one speaker in particular as gay, especially when in the -ing guise. These results provide evidence that social characteristics attributed to a speaker can shift depending on what other socially meaningful phonetic cues are present.

More experiments along these lines are necessary; particularly needed are ones which control the patterning of phonetic variables. While the design of such an experiment becomes increasingly complex as the number of variables increases, such investigations are promising ways to shed light on how a listener’s perception of another’s persona is formed.

4. Social Information Affecting Speech Perception

The relationship between phonetic and social information during speech perception is not unidirectional; the phonetic variants perceived can affect what characteristics are attributed to a speaker, and the characteristics attributed to the speaker can influence how sounds are perceived. For example, within the speech of a single individual, the focus of the aperiodic energy of the alveolar fricative /s/ is higher than for the palatal fricative /ʃ/. The acoustic boundary between /s/ and /ʃ/ depends in part on vocal tract size and therefore tends to be higher for females than males. This means that it is possible for a token of /s/ produced by a male to have its turbulence focused in a similar frequency range as a female’s token of /ʃ/. In an experiment where video clips of men and women were matched with gender-ambiguous tokens from an auditory /s/- /ʃ/ continuum, Strand (1999, 2000) found that participants were more likely to perceive a token as /ʃ/ if shown...
a video of a female. In other words, the same fricative was perceived differently depending on the face with which it was paired. These results provide evidence that perceivers attribute social characteristics to a speaker and then use this information to help identify sounds produced by that speaker. Results from a related experiment suggest that a greater amount of attention paid to the visual cues may increase the effect (Johnson et al. 1999).

There is evidence that the perception of phonetic variables can also be affected by other social characteristics attributed to the speaker, including dialect area (Niedzielski 1999; Hay et al. 2006a), socio-economic status (Hay et al. 2006b), age (Hay et al. 2006b; Koops et al. 2008, Drager 2011), and ethnicity (Staum Casasanto 2008, in press). The centring diphthongs /əʊ/ and /eə/, as in the words near and square, are undergoing a merger in New Zealand English. This change has been led by members of lower socio-economic groups; while some New Zealanders maintain the distinction, the diphthongs are merged in the speech of many New Zealanders who are young and/or members of lower socio-economic groups. Using photographs to manipulate the perceived socio-economic status and age of speakers in a perception experiment, Hay et al. (2006b) found that participants’ accuracy at identifying distinct tokens of the diphthongs depended on the social characteristics of the person in the photograph. Likewise, Drager (2011) found that the age of the person in a photograph could influence perception of variants undergoing a chain shift in progress. Results from both of these studies provide further evidence that stored social information attributed to a speaker can affect the perception of sounds they produce.

In both Detroit and Canada, speakers produce variants of the diphthong /aʊ/, as in the word mouth, with a raised nucleus. Speakers from Detroit associate this variant with Canadians and are not aware that they also produce raised variants. Niedzielski (1999) conducted an experiment where participants were asked to match a vowel from natural speech to one from a synthesised vowel continuum ranging from raised variants to standard American English variants. She found that participants were more likely to respond with a raised token from the continuum if they were in the condition where Canada appeared at the top of the response sheet than if they were in the condition where Michigan was at the top of the response sheet. Niedzielski argued that participants’ responses shifted due to their expectations regarding the speaker’s dialect area. Recent work also provides evidence that such effects are not limited to phoneme categorisation tasks and phonetic-similarity matching tasks. Using an eye-tracking device to investigate whether perceived speaker age affects processing of vowels involved in a change in progress, Koops et al. (2008) found that participants fixated longer on a lexical competitor when played a less conservative token and shown an older face. In other words, processing of the word appeared to take longer when the sound encountered was inconsistent with the most probable variant to be produced by the person in the photograph.

In addition to effects from social information attributed to the speaker, there is evidence that the dialect of an experimenter can affect performance on perception tasks (Hay et al. 2006b; Brunelle and Jannedy 2007). In a study on the perception of tones in Vietnamese, Brunelle and Jannedy (2007) found an interaction between the dialect of the experimenter, the dialect of the participant, and the tone. Similarly, Hay et al. (2006b) report an interaction between the dialect of the experimenter and the participant’s degree of merger.

The work outlined above helps exemplify just how malleable speech perception can be. Depending on what information is present at the time of perception, listeners can shift in how they perceive sounds. While any model of speech perception is bound to be incomplete, that social information can affect how sounds are perceived suggests that it should ideally be included.
5. **Agency and Automaticity**

The effect of social information on the perception of sounds has been interpreted as a result of the listener’s expectations regarding the speaker (e.g., Niedzielski 1999: 69): if the listener believes the speaker has certain characteristics, they might expect the speaker to sound a certain way and therefore be more likely to hear the speaker that way. However, some research suggests that objects in the listener’s environment can affect how sounds are perceived (Hay and Drager in press), a phenomenon similar to other well-known priming effects. In an experiment that was based on Niedzielski’s paradigm and that manipulated whether ‘New Zealand’ or ‘Australia’ was written at the top of the answer sheet, Hay et al. (2006a) found results similar to those of Niedzielski (1999). In contrast to the variable in Niedzielski’s study, the target vowel / was one with different realisations in the two dialects. While participants in the Australian condition were more likely to respond with an Australian token from the continuum, all but one of the participants indicated that they knew that the voice was a New Zealander. Hay et al. (2006b) argued that instead of expectations regarding a speaker’s dialect area affecting performance on the task, the mere mention of another dialect area was enough to orient perception towards that dialect. To test this explicitly, Hay and Drager (in press) conducted a follow-up study where subjects were surreptitiously exposed to stuffed toys of either kangaroos and koalas (from Australia) or kiwis (from New Zealand) prior to completing the same task reported in Hay et al. (2006a). Despite the fact that participants were unaware that the toys had anything to do with the experiment, their responses patterned in much the same way as in the original experiment. This finding supports the claim that shifts in speech perception can be observed even without overt expectations by the listener.

6. **Style and Speech Perception**

Taken together, results from sociophonetic perception experiments provide evidence that characteristics attributed to a speaker are accessed during perception and can affect how sounds are perceived. However, for simplicity of both experimental design and statistical analysis, much of the work in perceptual sociophonetics treats social categories as discrete and absolute (e.g., male versus female, Australia versus New Zealand). Consistent with arguments made by Munson (2010), a next step is to find ways where more subtle social information relating to a speaker’s construction of their identity can be incorporated into work on speech perception, investigating questions such as: how is a listener’s perception of a speaker’s style created? How do the different styles and stances of the listener affect speech processing? What is the relationship between the variant perceived, the interactional context, and the perceived persona of the speaker? And what is the relationship between the perceived variant and the styles and stances attributed to the speaker? Through asking questions such as these, we can ultimately address the nature of how rich social information is stored in the mind, how speaker characteristics are generalised to other individuals, and how both speaker-specific and stereotyped social information are used during speech processing.

7. **Conclusion**

In sum, speech perception experiments provide a means of investigating research questions concerned with language change, linguistic variation, and the storage of linguistic variables in memory, questions that could not be addressed as completely by focusing
solely on speech production. The better we understand how sounds are stored in memory and the more we uncover about the nature of the indexation between the mental representations of linguistic variables and social information, the better we will be able to interpret sociolinguistic patterns in speech production and the more complete our understanding of sound change and stereotype formation will be.

Short Biography

Katie Drager recently completed her PhD at the University of Canterbury in New Zealand and is now Assistant Professor in Sociolinguistics at the University of Hawai‘i at Mānoa. Her research interests include the use of stylistic variation in the construction of social personae and the effects of extralinguistic information on speech processing. She is particularly interested in how social information and linguistic variants are stored in the mind and accessed during the production and perception of speech. Recent and forthcoming articles can be found in Linguistics, Laboratory Phonology, and Language and Speech.

Note

* Correspondence address: Katie Drager, University of Hawai‘i at Mānoa, 561 Moore Hall, Honolulu, HI 96822, USA. E-mail: katie.drager@hawaii.edu

Works Cited


Campbell-Kibler, Kathryn. 2007. Accent, (ING), and the social logic of listener perceptions. American Speech 82. 32–64.


Drager, Katie. Forthcoming. Speaker age and vowel perception, Language and Speech 54.


Munson, Benjamin. to appear in 2010. Levels of phonological abstraction and knowledge of socially motivated speech-sound variation: A review, a proposal, and a commentary on the papers by Clopper, Pierrehumbert, and Tamati; Drager; Foulkes; Mack; and Smith, Hall, and Munson. To appear in Laboratory Phonology 1.

——, and Molly Babel. 2007. Loose lips and silver tongues, or, projecting sexual orientation through speech. Language and Linguistics Compass 1. 416–49.


Norris, Dennis, James M. McQueen, and Anne Cutler. 2003. Perceptual learning in speech. Cognitive Psychology 2. 204–38.


