

# 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

BENJAMIN MUNSON

*University of Minnesota*

## 1. Introduction

One central goal in the laboratory phonology (labphon) research community is to understand variation in the sound structure of language. A focus of labphon that sets it apart from other work on variation is the emphasis on understanding how knowledge of variation might be represented cognitively. A growing consensus among labphon researchers is that at least some aspects of acoustic and articulatory variation are represented cognitively, rather than being derived by universal principals (Keating 1988; Pierrehumbert 2002; Beckman et al. 2007; and Cutler et al. 2006). Variation arises from many sources. Another substantial contribution of the labphon community has been to document the various sources of variation in speech. Variation in parameters like voice onset time and vowel-space dispersion has been shown to relate to prosodic structure (e.g., Beckman and Edwards 1988; Keating et al. 2002; *inter alia*), the structure of the lexicon (Wright 2002; Munson 2007), and morphological structure (Steriade 1995), among other factors.

An emergent theme in the labphon community is the topic of phonetic variation related to social categories, which has come to be referred to as *socio-phonetics*. Laboratory phonologists have used instrumental techniques to examine variation at a more-detailed level than could be achieved with transcription alone. Representative studies here include Yaeger-Dror's (1994) study of the influences of lexical class and prosodic structure on vowel variation in Montréal French; Stuart-Smith's (2007) study of the influence of age and social class on sex differentiation in sibilant fricative production in Glasgow; Docherty et al.'s (2006) study of variation in the phonation type of medial stops in Tyneside English; and Harrington's (2007) study of sound change in Her Majesty Queen Elizabeth II's Christmas radio broadcasts. Other studies have examined how this variation might affect people's speech-sound perception and spoken-word recognition. One representative study of this is Warren et al.'s (2007) study, which showed that New Zealand English listeners' knowledge of a socially stratified ongoing sound change affects their

perception of the vowels in the NEAR and SQUARE lexical sets (using Wells' 1982 notation).

The topic of social variation is notable in that it is not one of the traditional areas of study in formal phonology. Prosodic structure, morphological structure, and the lexicon all are questions that have been studied by phonologists working in traditional generative frameworks. Both generative and psycholinguistic approaches (i.e., Levelt 1989; Levelt et al. 1999) offer a rich framework for understanding the types of variation described in many of the papers cited above. In contrast, the most-established models of social variation are from the field that studies social variation primarily, sociolinguistics. Sociolinguistics is a large, diverse field with many disciplinary goals. The cognitive representation of variation has received relatively less attention in sociolinguistics than have other topics (though see Campbell-Kibler 2007; and Mendoza-Denton et al. 2003). In contrast, arguably the central goal in the labphon community *is* to develop a model of how variation in the sound structure of language is represented cognitively. Given this, it is imperative that laboratory phonologists have a coherent theoretical model of how socioindexical variation might be represented, and how it might interact with other aspects of phonological knowledge.

The purpose of this commentary is fourfold. The first section outlines parallels between sociophonetic variation and other types of phonological variation. The second section examines the five papers in this section in light of the differences and similarities between these types of variation. The third section provides a model for understanding how variation of indexical variation might be represented cognitively. This model is an extension of a model first presented by Beckman and Edwards (2000), and developed in greater detail in Beckman et al. (2007). The fourth section proposes a research agenda for laboratory phonologists working on sociophonetics.

## **2. Regular Meaning and Social Meaning: A Comparison**

Even a cursory glance at the literature on sociophonetic variation and social meaning suggests that developing a comprehensive model of relationships between the two would be a daunting task. One simplifying assumption that laboratory phonologists might take in this endeavor is to hypothesize that social variation and social meaning might pattern more-or-less similarly to regular variation and regular semantic meaning. This section considers this hypothesis. Consider first a basic fact about phonology: the relationship between form (Saussure's *signifiant*) and meaning (Saussure's *signifié*) is arbitrary. No attribute of the object shown in Figure 1 is iconically associated with a dorso-palatal closure concurrent with vocal-fold vibration, an alveolar nasal plosive,



Figure 1. *Felis catus*

or a voiceless palatoalveolar fricative (as in the Italian *gatto* [gat:o], Japanese 猫 [neko], or French *chat* [ʃa]). The nature of form-meaning relationships is in part the purview of formal semantics. As summarized by Smith et al. (this issue), semanticists apply formal tests of entailment to determine a variety of facts about meaning, such as the degree to which a particular form-meaning association is conventionalized, and the strength with which different meanings for a form are invoked.

Listeners need to learn a daunting number of such arbitrary form-meaning pairings to be mature speakers of a language. Moreover, they need to both perceive and produce these labels. These labels have a rich multi-layered internal structure. At the most basic level, representations consist of instance-based encodings of the tactile, kinesthetic, somatosensory, auditory, and visual characteristics of speech that has been heard, seen, and produced. These representations are indexed to many other types of information, such as characteristics of the talkers that produced the utterances that are heard and seen, and sensory characteristics of whatever meaning the utterance references.

The first level of structure that is imposed on these instance-based representations is a consequence of the fact that listeners must be skilled in both production and perception. A young learner's ability to reproduce a signal that they have heard enforces the development of discrete sub-parts of representations. Imagine a Japanese-acquiring child mapping the /o/ from a token of 猫 that was produced by an adult with an adult-sized vocal tract, with adult-scaled acoustics and a visual image of an adult, to an /o/ produced with a child-sized vocal-tract, child-scaled acoustics, and the somatosensory feedback associated with the articulatory maneuvers for /o/. This mapping happens when a child repeats an adult /o/. Learning these stable mappings enforces the development of /o/ as a category. We can impute the existence of stable

mappings in the middle to the end of the first year of life, when infants' vocalizations begin to mirror the timbre of the productions of the adult speakers of the ambient language (de Boysson-Bardies et al. 1984).

The large number of words enforces the development of another, more-abstract level of representation. The task of learning so vast a number of words is arguably facilitated if the names that individuals assign to objects re-use sub-parts of existing words. For example, French learners who can interpret the label [ʃiẽ] as containing part of an existing label (the [ʃ] from [ʃɑ]) have an arguable advantage over learners who cannot. The existence of these levels of representation can be inferred from learners' performance on a number of tasks. These labels facilitate young children's ability to judge the sound-similarity of different words (Storkel 2002), and to repeat nonwords containing sequences of sounds that do not exist in real words in the mental lexicon, or which occur in only a few words (Edwards et al. 2004). The reciprocal nature of the relationship between lexical development and the development of these categories can be inferred from cross-sectional investigations of the relationship between measures of vocabulary size and experimental measures of the degree of granularity in children's phonological representations (Edwards et al. 2004; Munson et al. 2005) and measures of children's ability to access different types of representations in word-learning tasks (Storkel and Hoover 2006).

### 2.1 *The Socioidindexical Signifié*

One tactic for understanding how sociophonetic variation and social meaning might be represented cognitively is to compare them to the models of the representation of regular phonological variation and regular semantics that were just reviewed. First, is the *signifiant-signifié* for social meaning arbitrary, as it is for regular meaning? To answer this question we must first understand the nature of social meaning itself. Broadly speaking, two tactics have been used to study social meaning. The first is illustrated by the many studies showing phonetic variation stratified by macrosociological categories like gender, age, social class, regional origin, and sexuality. Many of these studies also show that listeners group talkers into these categories when presented with speech that has particular variants used by people of particular genders, ages, social classes, etc. One such study is Munson et al. (2006), which found differences between self-identified gay and heterosexual men in their production of the vowel /æ/, among other sounds. Listeners' ratings of how gay-sounding a man's voice was were significantly correlated with the acoustic characteristics in /æ/. Munson et al. tentatively concluded that different variants of /æ/ were markers of men's sexuality and that one of these, low-retracted /æ/, bore the meaning 'gay'. In a strict analogy with regular semantic meaning, this association between the *signifiant* 'retracted [æ]' and the *signifié* 'gay' is an item in

a socioindexical lexicon. This finding and others like it lead to a very simple hypothesis: social meanings are essentially the labels of the social categories by which variation is stratified, like age, social class, ethnicity, regional geographic origin, and sexuality.

This conjecture is immediately challenged by another finding in Munson et al. (2006). Here, Munson et al. showed that /æ/ variation also predicts listeners' ratings of speech clarity and perceived height. Mack and Munson (2008) showed that they predict ratings of perceived age. This apparent multiplicity of meanings is not limited to /æ/ variation in Minnesota. Campbell-Kibler (2007) presents similar results for the perception of variants of the present progressive morpheme *-ing*, which conveys a variety of different indexical attributes, including education, formality, and regional dialect. There are two possible interpretations of this apparent multiplicity of meanings. The first, conjectured by Munson et al., is that this is evidence that the construct *gay* has an internal structure, comprised of features like *speaks clearly*, *sounds younger*, and *sounds diminutive*. /æ/ variation predicts gayness because it cues judgments of the attributes that comprise the broader construct of gayness. Consider the following parallel with regular semantic meaning. Imagine listeners being presented with items on a /kæt/-/kæp/ continuum. In one block, the listeners would be asked in the extent to which the item indexed by the token is *feline*, *domesticated*, *mischievous*, and *fuzzy*. In the other block, they would be presented with the picture in Figure 1 and a picture of a cap. Listeners' responses on the two blocks would likely parallel one another closely. In the first block, tokens whose acoustics indexed a final alveolar gesture would elicit judgments of feline, domesticated, mischievous, and fuzzy. In the second block, the same tokens would prompt listeners to point to the cat. This isn't because word-final apico-alveolar gestures are iconically fuzzy or mischievous, but because the constellation of articulatory and auditory events in [kæt] is indexed to the class of objects *felis catus* and their associated attributes. In this view, variants like retracted [æ] appear to cue different judgments because the different judgments that listeners give are actually sub-attributes (diminutive, young, etc.) of a single superordinate attribute (being gay).

The second explanation, argued by Campbell-Kibler (2007) and recently formalized by Eckert (2008), is that the association between different phonetic variants and social meanings is more malleable, and that the particular meaning that a variant activates is dependent on a complex set of factors related to listeners' expectations and the communicative context. In the above analogy, the specific meaning invoked by a [kæp]-like token (which might index a *head covering*, *particularly one with a brim a device that covers or protect something*, or *an upper limit on expenditures*) depends on the communicative context. Depending on the talkers' intention, the conversational topic, and the listeners' expectations, the secondary entailment of the word-form [kæp] might be the *head covering* meaning or the *upper limit* meaning. That

is, variants like retracted [æ] give rise to different judgments because they are homophones.

Regardless of the explanation for the multiplicity of meanings of different variants, we can advance a revised hypothesis that the inventory of social meanings is an inventory of speaker attributes. The term *attributes* is not meant to imply that these only refer to fixed characteristics of speakers (like whether they grew up in Scotland or Buffalo, New York), but that they might also index something about the speaker's internal state or communicative intentions (such as stances that speakers want to convey, or attitudes a speaker holds toward a listener). This broad orientation allows us to discuss a variety of phenomena, including cases of extremely specific associations between specific phonetic variants and a speaker attribute, as well as the constellations of such associations that can be described as a *style*. As argued in detail in Eckert (2000) and in numerous papers in Eckert and Rickford's (2001) volume, the stratification of variants by macrosociological categories occurs not as an inevitable consequence of membership in the categories themselves, but because people in different categories participate in different communities, referred to as *communities of practice*. The variants index the specific stances and attitudes to which members of these communities typically ascribe.

If social meanings consist of an inventory of speaker attributes, we can then ask whether the link between forms and meanings is arbitrary or iconic. This question is essentially impossible to answer without an understanding of the set of speaker attributes that can potentially be indexed phonetically. Finding such a set is indeed a gnarly endeavor. Consider the case of another variable that can be conveyed and perceived phonetically, emotion. There exists a large literature examining the extent to which emotions are conveyed and perceived through phonetic variation (see Banse and Scherer 1996 for a review), and these are arguably at least partly due to physiological reactions to emotional states. A considerable challenge to reviewing this literature is that the terms used to describe emotion vary from study to study. This is true despite the fact that the most influential and widely held view in personality psychology is that emotion can be described using two factors, arousal (high to low), and affect (positive and negative) (Watson and Tellegen 1985). That is, characterizing the meanings of different emotional states is dependent at least in part on the choice of the inventory of emotions that researchers select when asking people to judge emotional states from speech samples. Listeners who are presented with sets of labels like {*kindly, bored, fearful, calm, delighted, tired, angry, at ease*} might be able to parse phonetic variation as indexing each of these eight categories, but this would miss the generalization this set contains: four pairs of two labels that putatively describe the endpoints of a two-dimensional affect-by-arousal space (Watson and Tellegen 1985: 226).

Without an inventory of possible speaker attributes that could be conveyed phonetically, studies run the risk of positing a larger set of attributes than are

actually represented cognitively. With this criticism in mind, existing data showing apparent cross-linguistic differences in the phonetic parameterization of speaker attributes become less convincing than they might seem initially. Consider Johnson's (2006) review of crosslinguistic variation in the magnitude of sex differences in vowel formants. Johnson shows that male-female differences in the first and second formant frequencies of vowels differ substantially across a wide range of (mostly Germanic) languages. At first glance this suggests that the phonetic instantiation of gender is relatively arbitrary, supporting the notion that socioindexical *signifié-signifiant* associations pattern like regular semantic meanings. Without knowing the meanings of these sex-related differences across languages, we cannot make definitive conclusions about this. These cross-linguistic differences may reflect cross-cultural differences in the attributes that men and women typically project. This is made more plausible by the suggestion that these formant differences actually reflect men and women's different engagement in certain types of ongoing sound changes, a tendency that has been noted previously (e.g., Labov, 2000, *inter alia*). These changes might affect the vowel spaces of the languages differently:<sup>1</sup> people who engage in the front vowel chain shift in New Zealand have lower average first-formant frequencies than people who do not; people who engage in the Northern California chain shift would have higher second-formant frequencies and lower first-formant frequencies than those who do not. The tendency to engage in ongoing sound changes might reflect a social meaning that is more-or-less stable across languages and cultures. If so, there may be more consistency across languages (and cultures) than a simple comparison of men and women's vowel production would suggest. By raising this point I do not mean to advance this hypothesis seriously, but rather to say that in the absence of an agreed-upon set of speaker attributes, the existing data are ambiguous. Clearly, further research is needed to examine these possibilities in detail. Research to date has not established definitively whether social meaning parallels that of regular semantic meaning. To clarify these issues, we need first a set of agreed-upon speaker attributes that can be indexed phonetically, paired with rigorous studies to examine the extent to which the phonetic parameterization of these attributes varies across languages and cultures.

## 2.2 *The Socioindexical Signifiant*

Consider next the indexical *signifiants*. The phonetic parameterization of different speech styles has shown that almost every parameter that is involved in a regular lexical contrast can be used to convey a speaker attribute. Consider Podesva's studies of gay male speech styles (Podesva 2004, 2006, 2007). Podesva shows that these speech styles are instantiated phonetically through



the release of final stop consonants, falsetto voice quality, and high-rising glissandi at the end of utterances. The former two features are regularly used across languages to convey lexical contrast, and the latter feature is used in the discourse-pragmatic system of English and many other languages.

Podesva (2004, 2006) argues that the extent to which a phonetic feature is used relates gradually to the presumed extent to which a speaker presumably wants to convey the attributes that it indexes. That is, a talker who uses longer final-stop releases and more high-rising glissandi more strongly indexes gayness than one who uses them less. This parallels the use of pitch range to convey pragmatic functions (i.e., Liberman and Pierrehumbert, 1984). However, it stands in sharp contrast to “regular” *signifiants*, in which producing a hyper-articulated variant of a word, i.e., producing /kæt/ with a /k/ with an especially long VOT, an /æ/ with more-extreme formant frequencies, and a released final /t/ would not index a more prototypical *cat* than a hypoarticulated form.

What is the structure of socioindexical *signifiants*? As described earlier, the *signifiants* for ‘regular’ lexical items—even those as simple as the constellation of articulatory~acoustic events transcribed as [kæt]—have a rich internal structure. This structure arises (a) as a consequence of the drive to both produce and to perceive items, and (b) as a consequence of the large number of *signifiant-signifié* mappings that individuals must learn to be mature speakers of a language. How might these factors influence the representation of the socioindexical *signifiants*? First, consider the production-perception asymmetry that drives the discretization of the ‘regular’ *signifiants* early in development. It is potentially the case that there is a much greater asymmetry between the number of distinct attributes speakers need to perceive than those they need to produce. For example, speaker-hearers of English are able to identify a set of variants as coding different sexualities (gay, straight), but they can’t necessarily produce all of the variants for a particular style (Gordon 2008). If this were true, then there would be at least one fewer sources of pressure to phonologically analyze socioindexical *signifiants* in as great of detail as there is for regular *signifiants*.

Moreover, the number of different distinct speaker attributes that a person must understand or produce is likely to be much smaller than the number of arbitrary *signifiant-signifié* mappings for regular semantic meaning they must command as speakers of a language. Estimates of the size of the mental lexicon vary, but even the most conservative estimates posit that speakers know many thousands of words. It is difficult to imagine a similar number of speaker attributes. This again reduces the pressure to analyze the *signifiant* into “phonological” units analogous to the phonemes, syllables, stress feet, and the like that arise as a consequence of the expansion of the regular lexicon. However, it is notable again that both of these conjectures about the socioindexical *signifiant* are crucially linked to a particular model of the cognitive representation of social meaning.



### 3. Commentary

Section 2 argued that the structure of social meanings differs substantially from the structure of regular semantic meaning, and that representations of social meaning are likely then to be very different from regular lexical representations. This section considers the five papers in this section in light of the discussion above. Each of these papers makes a substantial, original contribution to our understanding of sociophonetic variation. Moreover, the papers illustrate the diversity of methods that can be used to study sociophonetics, including production and perception studies at multiple levels of analysis. In sum, they are an impressive contribution to the sociophonetics literature and to the laboratory phonology literature more generally.

#### 3.1 Drager

Consider first Drager's study of the relationship between vowel quality, final-consonant glottalization, and social-group membership in high school girls. Drager's previous work had shown a relationship between /k/ glottalization in the word *like* and both the function of *like* and the girls' social group, indexed by the spaces in which they socialized (or failed to socialize, as one of her groups was characterized by a much looser social structure than the other). In this paper, she showed that listeners were very poor at inferring the grammatical function of *like* and the social-group membership of the producer, performing at only slightly better-than-chance levels when asked to do so. However, the factors that biased their responses suggested that they were aware of the patterns of variation in the population. Drager's study illustrates how variation in perception and production can be studied in tandem without relying on macrosociological categories. Notably, this finding required intensive ethnographic analysis to determine the relevant social groups in this community, as well as detailed instrumental analyses to uncover the distinctive phonetic variants associated with these categories.

Two critiques of Drager's study come to mind. The first concerns her grammatical-category experiment. This experiment presented single tokens of *like* and asked listeners to judge which type of sentence they might occur in. This presumes that the distinctive phonetic variants are not due to the prosodic structure of the different grammatical types, a critique that is only partly addressed by Drager's reporting that the grammatical types did not differ in duration. Perhaps listeners' accuracy would have been higher overall if they were given a spoken context that was neutral with respect to grammatical category, but which allowed listeners to assess the extent to which the form was concordant with the entire prosodic structure.

Second, Drager's experiment examining the extent to which girls in the high school could identify social groups based on the production of *like* is

somewhat at odds with the claim that such groups use distinctive variants to convey different stances and attitudes. Perhaps an experiment examining listeners' associations between these phonetic variants and the actual stances and attitudes that are characteristic of different social groups would have elicited greater-than-chance performance. These criticisms aside, Drager's study is a landmark in sociophonetics, in that it shows the value of intensive microanalytic studies of phonetic variation secondary to social groups other than those that derive from macrosociological categories.

### 3.2 *Smith, Hall and Munson*

Smith et al. present a review of tools used by formal semanticists to determine whether a word's meaning is a primary entailment (i.e., whether a particular interpretation addresses the question under discussion) or a secondary entailment, and whether it is conventional or conversational (i.e., whether the meaning is consistent across grammatical structures or if it is only true in a subset of contexts). They further apply these tests to social meanings. The motivation for this investigation is the suggestion from Potts (2003) that honorifics (as in Japanese) might pattern as *conventional implicatures*, a type of meaning that is both conventionalized and a secondary entailment. Because honorifics are typically thought to be social meanings, it is natural to wonder whether other social meanings might pattern similarly. The determination of whether something is a conventional implicature is based on a variety of tests typically administered with written stimuli, using native-speaker judgments. One such test, the *hey wait a minute* test (Shannon 1976), is employed by Smith et al. to show that the association between monophthongization and the perception of being from the Southern United States (shown previously by Plichta and Preston 2005) patterns as a secondary entailment. A further test of whether something is a secondary entailment is to examine whether a given meaning survives in different syntactic structures. An experiment described briefly in Smith et al. (described in greater detail in Smith et al. 2008) examined whether the association between /æ/ variants and men's sexuality are robust in different syntactic contexts.

I have two critiques of this endeavor. The first echoes the critique of Drager's paper: tests of the nature of social meaning should test the meanings that are represented cognitively. Following Munson et al., Smith et al. asked about a demographic variable, sexuality, when the actual meaning that listeners invoke when they are asked to label a talker's sexual orientation might be very different. Interestingly, the key finding of Smith et al. (2008) is a failure to replicate Munson et al.'s earlier association between /æ/ variation and men's sexuality, but a strong finding across syntactic contexts that /æ/ variants are related to other speaker attributes. The second is that Smith et al.'s application of conventional semantic tests to examine social meaning was two degrees

removed from the tests' original uses. First, they examined social meanings rather than regular semantic meaning, and second, they used spoken stimuli rather than written ones. A natural follow-up to this study is to examine whether tests of regular semantic meaning are robust when spoken stimuli are used.

### 3.3 *Mack*

Mack presents a study of the relationship between acoustic variation and the perception of a variety of demographic characteristics of speakers of Puerto Rican Spanish. She showed that listeners associate unique sets of acoustic parameters with judgments of talkers' height, social class, sexual orientation, and age. Mack further showed that listeners fell into distinct groups based on the acoustic parameters that were associated with their judgments of perceived sexual orientation. Mack's study was very much in the mold of other studies on the perception of talker characteristics, such as Munson et al. (2006) and Smyth et al. (2003). Critiques of that endeavor were made earlier in this paper and will not be replicated here.

One additional critique comes to mind. Though Mack's study is not based on production data, it is natural to assume that listeners' responses in this experiment reflect their experiences hearing speakers of Puerto Rican Spanish convey these attributes. Clearly this argument would be strengthened if Mack found such differences in production. In the absence of this, though, Mack's argument would be strengthened if she could show that the association between acoustic features and speaker attributes was specific to that dialect. A similar tactic was used by Foulkes et al. (in press) to examine whether listeners from Newcastle were aware of gender differences in the use of preaspiration in medial stops. They found that listeners in Newcastle labeled the gender of children whose speech was otherwise acoustically gender-ambiguous differently depending on whether the token had been produced with preaspiration. Listeners from dialect regions in England in which preaspiration is not linked to gender did not associate that variable with gender, nor did listeners in the US.

### 3.4 *Clopper, Pierrehumbert, and Tamati*

Clopper, Pierrehumbert, and Tamati showed that North American English listeners systematically misperceive words produced by talkers of a different dialect, reflecting in part mismatches between the vowel systems of the talkers' and listeners' dialects. They also show that listeners' tendency to perceive a shifted vowel in a word is considerably reduced if the sequence with the shifted vowel is a nonword.

Though Clopper and Pierrehumbert's results are modest in size, they illustrate the importance of recognizing that studies of the capacity of humans to

deal with variability in speech have the most ecological validity if the kind of variability that they deal with is like that present in the signals that they encounter. Consider this recommendation in light of Kraljic et al.'s (2008) study of listeners' perceptual adaptation to different types of variation in anterior sibilant fricatives. Kraljic et al. studied two types of variation, (faux) dialectal variation and (faux) idiosyncratic speaker-related variation. Interesting asymmetries in listeners' perceptual adaptation were found between the two types of variation. However, the type of variation that Kraljic et al. examined was not based on *attested* dialectal and socioindexical variation in anterior sibilant fricatives. One potential follow-up would be for Kraljic et al. to examine adaptations to the kinds of dialectal and seemingly speaker-specific variation that is attested in the natural speech, like the corpora from which Clopper et al. drew their stimuli.

### 3.5 *Foulkes*

Foulkes's paper is a spectacularly detailed and comprehensive review of work on sociophonetic variation. It will undoubtedly serve as a definitive reference for researchers interested in this area. Two issues that Foulkes discusses warrant particular mention. The first is the discussion of the contrast between overt evaluation and tacit stereotypes. Foulkes advances the exciting hypothesis, based in part on Labov (2006) and Silverstein (2003), that the extent to which variation invokes a conscious evaluation is an indication of the extent to which it is represented in long-term memory. This is a very concrete and very testable prediction, and methods for testing it are well within the realm of expertise of laboratory phonologists.

Second, Foulkes advances the equally exciting hypothesis that stylistic variation may have a facilitative influence in language acquisition. His argument surrounds the well-attested positive effect that child-directed speech (CDS) has on some aspects of language learning. While I concede that studies of CDS have shown it to affect language acquisition positively, I would emphasize the similarly widely cited finding that CDS is not universal across cultures, and that children in cultures that do not use CDS acquire language typically nonetheless (Schieffelin and Ochs 1986). I would advance the alternative hypothesis that stylistic variation affects language acquisition because it helps children to interpret linguistic variation. That is, learners who can interpret variable forms as indexing a speaker attribute might be able to more easily parse that form than those who treat variation as random. In this way, speaker-attribute perception functions much in the same way sublexical phonological representations do: listeners who can interpret a novel label for a novel object as being comprised of known categories is likely to be a more-successful learner than one who treats every new string as a more-or-less unanalyzed whole.

#### 4. A Proposal

The studies and viewpoints reviewed and advanced in Section 2, combined with the papers in Section 3, set the stage for a new research program for laboratory phonologists concerned with sociophonetics. The take-home message in this paper thus far has been the complexity of this topic. Clearly, the existing models of phonology and semantics cannot be applied to socioindexical variation and social meaning without some substantial expansion. In hopes of advancing work on this topic, this section proposes a model of how socioindexical knowledge might be represented linguistically by extending a model that has been developed to account for regular phonology and regular semantics.

The starting point is Beckman et al.'s (2007) model of speech-sound knowledge, shown in Figure 2. This model was intended to account for a variety of findings in phonological development and adult phonological competence. This model shares many features in common with an independently developed proposal regarding phonological development by Pierrehumbert (2003), and is qualitatively very similar to a proposal by Mestala and Walley (1998) to account for the development of metaphonological abilities in young school-aged children. The essence of this model is that it proposes that phonological knowledge is multi-level, and that there is a developmental progression from

#### Levels of knowledge about speech sounds

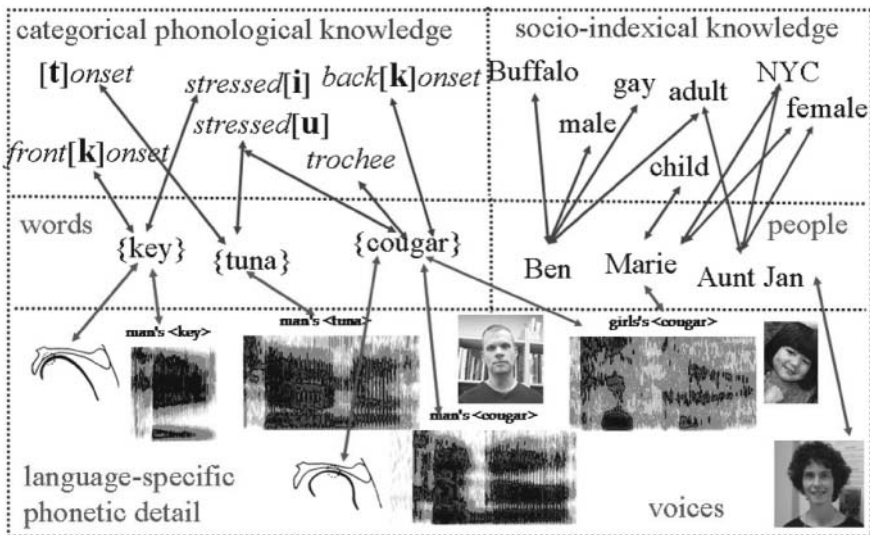


Figure 2. A proposed model of levels of phonological knowledge.

more specific and concrete knowledge to more abstract knowledge of sound structure. At the lowest level there is a rich set of encodings of the auditory-acoustic, visual, and articulatory-kinematic characteristics of utterances that the person has produced and heard. The nature of these encodings is currently a topic of serious and spirited debate, with some researchers favoring models in which these are highly veridical records of experienced instances, and others, including the architects of this model, favoring the view that these are representations of probability distributions in an  $n$ -dimensional, multisensory space.

Parametric representations are indexed to the basic inventory of concepts that they convey. Again, the nature of the inventory of form-concept mappings is a matter of debate. We assume that it is likely that this is similar to the dictionary-like mental lexicon that is assumed in standard linguistic theory. This lexicon is then the basis over which higher-order phonological generalizations are made. These generalizations include both the basic generalization that word-forms can be analyzed into units (i.e., [k<sup>h</sup>ɪrɪ] can be analyzed into a trochaic stress foot, each syllable of which is comprised of two phones) as well as generalizations about the constraints on those structures (i.e., that the medial flap can only occur in the onset of a stressless syllable).

There are two aspects of this model that we have tested previously. The first is the prediction—argued strongly by Pierrehumbert (2003)—that development should involve the building of gradually more abstract structures. This assumption appears to be amply justified. As shown by empirical works reviewed by Beckman et al. (2007), Pierrehumbert (2003), and Metsala and Walley (1998), there is indeed evidence that phonological development involves the building of gradually more abstract knowledge about the sound structure of language, and that detailed encodings in the parametric phonetic space ground this learning. There is considerable evidence that these encodings continue to play a role in adults' ongoing perception of language. That is, adults use far more fine phonetic detail in ongoing speech perception than can be captured in even detailed categorical phonetic transcriptions (e.g., Goldinger and Azuma 2004; McLennen and Luce 2005). The second is that phonological generalizations should track the statistics of the lexicon. Again, this facet of the model appears to be amply justified, as shown by numerous investigations, such as Edwards et al. (2004), and Frisch et al. (forthcoming).

Previous presentations of this model speculated that a similarly hierarchical structure might underlie individuals' knowledge of socioindexical variation. This facet of the model posits that there is a parallel set of encodings of events in the real world, a universe of items to which these are indexed, and a set of generalizations over this universe of items. The standing hypothesis in Beckman et al. (2007) and related papers is that the same detailed encodings are linked to a "lexicon" of talkers. The generalizations projected from this lexicon would be the inventory of speaker attributes hypothesized in Section

2. Though Figure 2 shows these as macrosociological categories, these can be represented equally well as having an internal structure of the type proposed by Eckert (2008), consisting of attributes organized into fields that cohere into a particular style. They would, in effect, be the “primitives” of the socioindexical *signifié* discussed earlier. The generalization from a lexicon of talkers to an inventory of speaker attributes we presume would be a self-organizing process, much like the self-organizing processes that have been argued to give rise to categories like /s/ and /ʃ/.

This remains very much a hypothesis to be tested actively. One attractive facet of this model is that it allows us to test the extent to which socioindexical learning parallels regular phonological learning. For example, it predicts an interaction between the level of granularity in the socioindexical *signifié* and the size of the lexicon of talkers that a person holds in memory. Indeed it is possible that there are substantial individual differences in the size of this lexicon. The differences in the size of the lexicon are reflected in studies reviewed by Foulkes showing substantial individual differences in the ability to recognize and remember voices. It also predicts that the individual differences in the type of generalizations that listeners make (i.e., the inventory of attributes that they generalize) should be based on the composition of the lexicon.

## **5. Conclusion: A Research Program in Sociophonetics for Laboratory Phonologists**

### *5.1 Measuring Meaning*

This last section discusses three areas of research in sociophonetics that I view as crucial for laboratory phonologists who are interested in this area to pursue. The first of these, emphasized throughout this paper, is to work with our colleagues in related fields to determine the types of speaker attributes that can be conveyed phonetically. Clearly, we need to go beyond the tactic of simply constructing experiments eliciting judgments on a pre-selected set of attributes. Ethnographic studies, including those by Drager, Eckert, Podesva, and others, reviewed earlier, are crucial here. These intensive observations of the use and perception of variables in their social context can add critical insights on social meaning. A critical charge to these ethnographers is to continue to meta-analyze the results of different ethnographic studies so that we can build a richer typology of social meanings across languages and cultures.

Equally crucial, though, is to collaborate with scholars in the social sciences to get a better sense of how general social-cognitive skills might influence the set of attributes that a person projects from a lexicon of talkers. An entire branch of psychology, impression formation, studies this topic. Even a cursory exploration of this literature suggests that while the specific constructs that it



examines (i.e., personality) are largely foreign to laboratory phonologists, the general principles underlying this work (i.e., categorization and cue-weighting) are reassuringly familiar. One specific example from this literature warrants mention in this commentary. A classic finding in this literature is that individuals' overall impression of an individual is generally more negative than the sum of their impression of individual attributes of that person, and that extreme values disproportionately influence the final categorization (e.g., Anderson 1974, *inter alia*). Consider how this finding might influence two listeners' perception of the attributes of a particular set of talkers. If both of the listeners identify the talkers' sexual orientation, but one person perceives gay people more negatively than the other, then the ultimate extraction of the attributes of those talkers might be more weighted toward those attributes than the listener whose view of sexual orientation is less negative and less extreme. This finding has further implications for how speaker attributes might evolve in individuals across development as their social-cognitive knowledge develops and changes.

## 5.2 *Understanding the Signal*

The second area I regard as critical is to utilize our expertise as laboratory phonologists to best understand how attributes are conveyed phonetically. While this area undoubtedly capitalizes on our expertise more than studying social meanings does, a closer look at this problem suggests it is surprisingly more daunting than we might think. Consider again Docherty et al.'s (2006) and Foulkes et al.'s (in press) findings regarding preaspiration in Tyneside English. It is noteworthy that this feature is very clearly socially stratified and very salient to listeners—both to adult listeners in Foulkes et al., and to children listeners in Docherty et al., as evidenced by their early acquisition of it. It is also noteworthy that this feature was only discovered by Docherty, Foulkes, and colleagues through detailed spectrographic analysis of speakers' productions. It is, in a sense, the proverbial needle in the haystack. Studying sociophonetic variation, then, is a little like doing fieldwork on a very under-documented language. In the absence of an agreed-upon inventory of speaker attributes that can be conveyed phonetically, it is like fieldwork on a language whose lexicon is unknown, or at least whose lexicon has not been lemmatized.

A starting point in systematizing the *modus operandi* for this search is to consider what a viable cue for a speaker attribute might be. Again, laboratory phonologists can benefit from our link to colleagues doing phonetic fieldwork. Presumably, if a phonetic parameter is robust enough to code a lexical contrast in a language then it is within our species' capability to use it to convey a speaker attribute. Another area of inquiry that laboratory phonologists could examine systematically is whether languages privilege as cues to speaker identity phonetic parameters that are already used to convey lexical contrast, or

whether they privilege those that are not used for that purpose. One could make a principled argument that these phonetic parameters should tend to be the same as those used to convey lexical meaning, as this would capitalize on the learner's attention to them. That is, children learning Mandarin would be more likely to manipulate sibilant fricatives' centroid frequencies to convey speaker attributes (as they were shown to do by Li et al. 2008), as this parameter is already used to differentiate among the three anterior sibilant fricatives in that language. One could also make a principled argument that languages should favor using parameters that are not used in phonemic contrasts, as orthogonal cues would be easier to disentangle in perception than convolved ones. Future research should address whether there are predictable cross-linguistic tendencies in the use of different phonetic parameters to convey speaker attributes.

### *5.3 Studying Multiple Levels of Abstraction*

The third area is to determine the extent to which generalizations about socio-indexical variation are independent from generalizations about the 'regular' phonological structure of the language. Clearly, generalizations about parametric phonetic variation are affected by social stratification. Warren et al. (2007), Drager (2006), Niedzielski (1999), and others all show that phonetic perception is biased by presumptions regarding speaker attributes. That is, the generalizations that people make about the range of parametric variation associated with phonological categories is mediated by knowledge of how different groups produce sounds differently. Moreover, one could argue that information about speaker attributes is retained in phonological categories if it were demonstrated that listeners are able to judge attributes in sound combinations that they have never heard, including nonwords. No studies have examined this question systematically with speaker attributes, but one recent study by Simner and Haywood (2009) showed that speakers generalize the association between word-forms and gustatory sensation, such that different synaesthetic tastes are evoked from nonwords. It remains to be seen whether listeners' higher-level phonological generalizations are similarly sensitive to speaker attributes. Such a finding would suggest that knowledge of speaker variation is retained even at relatively abstract levels of representation.

### *5.4. ¡Viva La Variación!*

In sum, the five papers in this session make a substantial contribution to the field of sociophonetics. Each one addresses a topic of critical importance. Equally importantly, the breadth of the topics they cover invite a broader discussion of the direction in which this emerging area of inquiry might move.

## Acknowledgments

I think Mary Beckman and Jan Edwards, as this commentary reflects numerous ideas that have grown out of our collaborations together. I also have had valuable discussions about this topic with Molly Babel, Peter Graff, Jen Hay, Eden Kaiser, and Laura Staum Casasanto.

Correspondence e-mail address: munso005@umn.edu

## Note

1. I thank Jen Hay for pointing this out.

## References

- Anderson, Norman. 1974. Cognitive algebra. In Leonard Berkowicz (ed.), *Advances in Experimental Social Psychology* 7. 1–101. New York: Academic Press.
- Banse, Rainer & Klaus Scherer. 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70. 614–636.
- Beckman, Mary E. & Jan Edwards. 2000. The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development* 71. 240–249.
- Beckman, Mary E., Benjamin Munson & Jan Edwards. 2007. Vocabulary growth and developmental expansion of types of phonological knowledge. In Jennifer Cole & José I. Hualde (eds.), *Laboratory Phonology* 9, 241–264. New York: Mouton de Gruyter.
- Campbell-Kibler, Kathryn. 2007. Accent, -ING and the social logic of listener perceptions. *American Speech* 82. 32–64.
- Cutler, Anne, Frank Eisner, James M. McQueen & Dennis Norris. 2006. Coping with speaker-related variation via abstract phonemic categories. Oral presentation at the tenth conference on laboratory phonology, Paris, France, 1 July.
- Docherty, Gerard J., Paul Foulkes, Jennifer Tillotson & Dominic Watt. 2006. On the scope of phonological learning: issues arising from socially structured variation. In Louis Goldstein, Doug Whalen & Cathy Best (eds.), *Laboratory Phonology* 8, 393–421. New York: Mouton de Gruyter.
- Drager, Katie. 2006. From *Bad* to *Bed*: the relationship between perceived age and vowel perception in New Zealand English. *Te Reo* 48. 55–68.
- Edwards, Jan, Mary E. Beckman & Benjamin Munson. 2004. The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in non-word repetition. *Journal of Speech, Language, and Hearing Research* 47, 421–436.
- Eckert, Penelope. 2000. *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Cambridge, MA: Blackwell.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12. 453–476.
- Eckert, Penelope and John Rickford (eds.). 2001. *Style and Sociolinguistic Variation*. Cambridge, MA: Cambridge University Press.
- Foulkes, Paul, Gerard Docherty, Ghada Khattab & Malcah Yaeger-Dror. Forthcoming. Sound judgments: perception of indexical features in children's speech. To appear in Dennis Preston & Nancy Niedzielski (eds.), *A Reader in Sociophonetics*. Berlin: Mouton de Gruyter.

- Frisch, Stefan A., Maria R. Bréa-Spahn & Carla I. Orellana. Forthcoming. Metalinguistic judgments of phonotactics by bilinguals. Oral presentation at the eleventh conference on laboratory phonology. Victoria University, Wellington, New Zealand, 30 June–2 July 2008.
- Goldinger, Stephen & Tamiko Azuma. 2004. Episodic memory reflected in printed word naming. *Psychometric Bulletin and Review* 11. 716–722.
- Gordon, Bryan James. 2008. Gay sounds: A non-discrete model of gay speech. Presentation at the Lavender Languages Conference, American University.
- Harrington, Jonathan. 2007. Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In Jennifer Cole & Jose I. Hualde (eds.), *Laboratory Phonology* 9. 113–124. New York: Mouton de Gruyter.
- Huffman, Marie. 2007. Laboratory phonology and sociophonetics. Partners in a conversational whose time has come: Comments on Docherty, Moreton and Thomas, Stuart Smith, and Warren, Hay, and Thomas. In Jennifer Cole & Jose I. Hualde (eds.), *Laboratory Phonology* 9, 113–124. New York: Mouton de Gruyter.
- Johnson, Keith. 2006. Resonance in an exemplar model of phonology: the emergence of social identity and phonology. *Journal of Phonetics* 43. 485–499.
- Keating, Patricia A. 1990. The window model of coarticulation: articulatory evidence. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology*, 451–470. Cambridge: Cambridge University Press.
- Keating, Patricia A., Taehong Cho, Cecile Fougeron & Chai-Shun Hsu. 2004. Domain-initial articulatory strengthening in four languages. In John Local, Richard Ogden, Roslyn Temple (eds.), *Phonetic Interpretation: Papers in Laboratory Phonology* 6, 143–161. Cambridge: Cambridge University Press.
- Kraljic, Tanya, Susan E. Brennan & Arthur G. Samuel. 2008. Accommodating variation: dialects, idiolects, and speech processing. *Cognition* 107. 54–81.
- Labov, William. 2000. *Principles of Linguistic change. Volume II: Social Factors*. Cambridge, MA: Blackwell.
- Labov, William. 2006. A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics* 34. 500–515.
- Ladefoged, Peter & Donald E. Broadbent. 1957. Information conveyed by vowels. *Journal of the Acoustical Society of America* 29. 98–104.
- Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–75.
- Li, Fangfang, Eunjong Kong, Mary E. Beckman & Jan Edwards, J. 2008. Adult acoustics and developmental patterns for gender-marked phonetic variants in Mandarin fricatives and Japanese stops. Poster presented at the eleventh conference on laboratory phonology, Victoria University, Wellington, New Zealand, 30 June–2 July 2008.
- Lieberman, Mark & Janet B. Pierrehumbert. 1984. Intonational Invariance under Changes in Pitch Range and Length. In M. Aronoff & R. Oehle (eds.), *Language Sound Structure*, 157–233. Cambridge, MA: MIT Press
- Mack, Sara & Benjamin Munson. 2008. Implicit Processing, Social Stereotypes, and the 'Gay Lisp'. Oral presentation given at the annual meeting of the Linguistic Society of America, Chicago, IL.
- Maye, Jessica, Janet Werker & LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82. B101–B111.
- Metsala, Jamie & Amanda Walley. 1998. Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In Jamie L. Metsala & Linnea C. Ehri (eds.), *Word recognition in beginning reading*, 89–120. Mahwah, NJ: Lawrence Erlbaum Associates.

- McLennan, Conor & Paul Luce. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31. 306–321.
- Mendoza-Denton, Norma, Jennifer Hay & Stefanie Jannedy. 2003. Probabilistic sociolinguistics: beyond variable rules. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabilistic Linguistic*, 71–138. Cambridge, MA: MIT.
- Munson, Benjamin. 2007. Lexical access, lexical representation, and vowel production. In Jennifer Cole & Jose I. Hualde (eds.), *Laboratory Phonology* 9. 201–228. New York: Mouton de Gruyter.
- Munson, Benjamin & Molly E. Babel. 2007. Loose lips and silver tongues: projecting sexual orientation through speech. *Linguistics and Language Compass* 1. 416–449.
- Munson, Benjamin, Beth Kurtz & Jennifer Windsor. 2005. The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without language impairments. *Journal of Speech, Language, and Hearing Research* 48. 1033–1047.
- Munson, Benjamin, Elizabeth C. McDonald, Nancy DeBoe & Aubrey White. 2006. Acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics* 34. 202–240.
- Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18. 62–85.
- Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in Laboratory Phonology* VII. 101–139. New York: Mouton de Gruyter.
- Pierrehumbert, Janet B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46. 115–156.
- Plichta, Bartłomiej & Dennis Preston. 2005. The /ay/s have it: The perception of /ay/ as a north-south stereotype in United States English. In Tore Kristiansen, Peter Garrett & Nikolas Coup-land (eds.), *Acta Linguistica Hafniensia* 37, *Subjective Processes in Language Variation and Change*, 107–30. Copenhagen: C. A. Reitzel.
- Podesva, Robert J. 2004. The Significance of Phonetic Detail in the Construction of Social Meaning. Presentation given at the Annual Meeting of the Linguistic Society of America, 8–11 January, Boston, MA.
- Podesva, Robert J. 2006. Intonational variation and social meaning: Categorical and phonetic aspects. *University of Pennsylvania Working Papers in Linguistics Selected Papers from New Ways of Analyzing Variation* 34(12). 189–202.
- Podesva, Robert J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11. 478–504.
- Potts, Christopher. 2003. Expressive content as conventional implicature. In Makoto Kadowaki & Shigeto Kawahara (eds.), *Proceedings of the North East Linguistic Society* 33, 303–322. Amherst, MA: Graduate Linguistics Students' Association.
- Schieffelin, Bambi B. & Eleanor Ochs. 1986. Language socialization. *Annual Review of Anthropology* 15. 163–191.
- Shannon, Benny. 1976. On the two kinds of presuppositions in natural language. *Foundations of Language* 14. 247–249.
- Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23. 193–229.
- Simner, Julia & Sarah L. Haywood. 2009. Tasty non-words and neighbors: the cognitive roots of lexical-gustatory synaesthesia. *Cognition* 110. 171–181.
- Smith, E. Allyn, Benjamin Munson & Kathleen Currie Hall. 2008. Rethinking the meaning of Minnesotan [æ]: sexual orientation or personal well being? Oral presentation at the conference on New Ways of Analyzing Variation (NWAV), Houston, TX.
- Smyth, Ron, Greg Jacobs & Henry Rogers. 2003. Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society* 32. 329–350.

- Staum Casasanto, Laura. 2008. *Experimental Investigations of Sociolinguistic Knowledge*. Dissertation, Stanford University Department of Linguistics.
- Storkel, Holly. 2002. Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language* 29. 251–274.
- Storkel, Holly & Jill Hoover. 2006. Whole-word versus part-word phonotactic probability/neighborhood density in word learning by children. In David Bamman, Tatiana Magnitskaia & Colleen Zaller (eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* 2, 584–594. Somerville, MA: Cascadilla Press.
- Steriade, Donca. 1995. Paradigm uniformity and the phonetics-phonology boundary. In Michael Broe & Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 314–333. Cambridge: Cambridge University Press.
- Stuart-Smith, Jane. 2007. Empirical evidence for gendered speech production: /s/ in Glaswegian. In Jennifer Cole & Jose I. Hualde (eds.), *Laboratory Phonology* 9, 65–86. New York: Mouton de Gruyter.
- Warren, Paul, Jennifer Hay & Brynmor Thomas. 2007. The locus of sound-change effects in recognition and perception. In Jennifer Cole & Jose I. Hualde (eds.), *Laboratory Phonology* 9, 87–112. New York: Mouton de Gruyter.
- Watson, David & Auke Tellegen. 1985. Toward a consensual structure of mood. *Psychological Bulletin* 98. 219–235.
- Wells, John C. 1982. *Accents of English*. Cambridge: University of Cambridge Press.
- Wright, Richard A. 2004. Factors of lexical competition in vowel articulation. In John J. Local, Richard Ogden & Rosalyn Temple (eds.), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.
- Yaeger-Dror, Malcah. 1994. Phonetic evidence for sound change in Quebec French. In P. Keating (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*. 267–292. Cambridge, MA: Cambridge University Press.