(Non-final version! Do not quote without permission of the author)

To appear in : *Laboratory Phonology 9*, Jennifer Cole and Jose-Ignacio Hualde (eds.).

Chapter X. Laboratory phonology and Socio-phonetics: Partners in a conversation whose time has come.

Marie K. Huffman

1. Introduction

The papers in this section illustrate the many insights that socio-phonetic research has to offer linguistics. Docherty surveys the potential benefits of more discourse between laboratory phonology and sociophonetics, emphasizing the inseparability of socio-phonetic and linguistic knowledge. Moreton and Thomas make elegant use of archival and modern recordings to test competing explanations of the vowel change often called Canadian Raising. Stuart-Smith demonstrates the fine manipulation of phonetic difference in social indexing, as seen in gender, age and class effects on Glaswegian /s/ acoustics. Warren, Hay and Thomas demonstrate that expectations about distinctiveness, based on speaker characteristics such as age, can influence perception as heavily as actual acoustic properties of the stimuli. Taken together, these papers demonstrate the astounding richness of phonetic detail which humans manipulate every day, in speech production and perception. Once we fully accept this richness, we have more to study, and more tools to use in developing improved models of speech representation and processing.

2. Validity of all forms of speech

As Docherty (this volume) points out, laboratory phonologists have concentrated heavily on averaged group speech characteristics. But averages are only a numerical representation of central tendency and do not necessarily reflect the relative importance that a particular

type of speech has, in mental representation, or in language use. We can gain different, but equally valid, insights about speech by looking at how language is variably used in different communication situations. Attending to patterns in all types of language use will help us better discern the structure of knowledge of sounds. For decades linguists have used language games as evidence of linguistic structure, such as the independence of segmental and prosodic representations. For example, Kenstowicz (1994) discusses the secret language Kishingelo, used by speakers of the Bantu language Sanga, in which the last two syllables of a word are interchanged, but not their length or tone: *múkwèètù-> mútùùkwè*. If this form of language use is a valid source of information about knowledge of sound, why not attend more systematically to even more common ways people manipulate language daily, such as changing register, making perceptual adjustments for a new conversational partner, or switching dialects? The entities manipulated in such activities must be real linguistic elements. If we accept this premise, then all language used is important and worthy of study. A number of important consequences follow from the conclusion.

First, traditional laboratory speech is a valid object of study. Careful citation forms represent part of a speaker's knowledge of the language, and they are used occasionally in real life situations (spelling bees, forms given in response to misunderstanding, etc). The difference between these situations and laboratory speech, however, is that in the latter, the speaker usually has no specific addressee in mind. Thus, it is not clear whether the aspects of speech that would be adjusted for an addressee are 1) not varied by the speaker (holding to default values), or 2) are varied, in random (or, worse, non random but unknown) ways. Whether to further vindicate laboratory speech, or to move beyond it, we need to know more about what sound properties are adjusted for a partner.

Another consequence of the conclusion that all language use matters, is that we must improve our empirical foundation by studying more kinds of speech. There has been a strong tendency in recent years for

laboratory phonologists to move toward addressing more "realistic" types of speech than citation forms. Whether it is using meaningful sentences rather than frame sentences, or analyzing telephone conversations between friends, our data base has definitely broadened. Attention to prosodic influences on segmental structure has helped drive this change, as we consider words and segments in a variety of phrasal and prominence relations. Easier access to recorded speech corpora has also enlarged our data base. While a laboratory phonologist may be concerned about the lack of prosodic or segmental control in some such corpora, they can nonetheless give us a more representative picture of the range of variation commonly manipulated by speaker/listeners (e.g., Byrd 1994, Dilley, Shattuck-Hufnagel and Ostendorf 1996, Cole et al. 2003, to name a few). We could also gain by considering seriously what we can learn from pre-existing interview data collected by field linguists and sociolinguists over the last several decades. Moreton and Thomas (this volume) demonstrate this elegantly in their study of the time course of changes in the /ai/ diphthong in Cleveland Ohio (which has an /ai/ alternation similar to Canadian Raising). They acquire data on pronunciation of speakers spanning 100 years by using 3 data sources: recordings from the Dictionary of American Regional English (Cassidy & Hall 1985-2002), civic organization recordings of local political figures, and and recordings they made in recent years. Their analysis shows that the differences in vowels before voiced and voiceless codas are not rooted in vowel length differences. What the paper leaves unresolved is a question for phonetics generally, which is a satisfactory explanation of formant differences before voiced and voiceless codas that cannot be explained by timing, or by voicing-related articulatory adjustments (on which, see also Hawkins and Nguyen (2004)). Thus, attention to socially and temporally varying properties of English vowels highlights core theoretical challenges for linguistic phonetics.

For all of the forms of speech we study, we will improve our empirical foundation by attending to the personal and social factors that socio-phonetics has shown to correlate with phonetic differences.

These factors must be controlled, or purposefully varied. Many phonetic studies report simply that their subjects were college students between 18 and 30, and most are vague as to dialect of the subjects, though this seems to be improving. The papers in this section demonstrate the importance of dialect, age, class and gender. In addition, ethnicity, though not addressed in the research presented in this section, clearly needs serious consideration within laboratory phonology.

3. Manipulating social variables rather than controlling or ignoring them

Just as one of the most socially neutral types of speech (laboratory speech) is a valid object of study, so is one of the most complicated forms of speech; namely, that between partners who differ in one or more relevant socio-indexical factors. This is a common speech situation, as parents talk with children, adults with different dialects exchange information, or strangers waiting in long lines exchange theories about the explanations for the delay.

3.1 Investigating socially-driven variability in production

We know that one or more of participants in a conversation adjust their speech in response to differences between the participants, (e.g., Pardo (2004) on pronunciation, Bortfeld and Brennan (1997) on lexical choice) and in response to changes in common ground between participants (Brennan and Clark 1996). Sociolinguists have tried to minimize the effect of social differences by having two friends participate in a recorded conversation, or by recruiting a community member to interview subjects. An alternative approach, being taken by some linguists and psycholinguists, is to study variability directly—setting up controlled situations of difference between speakers, and assessing the ways people adapt their speech in response to characteristics of a partner. Two major factors driving speaker adaptation are communicative need—the speaker's assessment of the addressee's needs, or what went wrong in

communication, and the relative importance of the social differences between the speakers.

Taking communicative need first, if the speaker believes that the listener needs a particular kind of input, he may adjust his speech, as in the literature on speech intended for the hard of hearing (e.g., Picheny, Durlach and Braida 1986; Krause and Braida 2004), or speech intended for a more generalized recipient who has trouble understanding (as in much of the "clear" speech literature; e.g., Bradlow (2002); Bradlow and Bent (2002)). Recent work on speech addressed to computers suggests that when computers misunderstand a human speaker, the human makes adjustments similar to those found in human-human conversation (Oviatt 1996; Oviatt, MacEachern and Levow 1998; Stent, Huffman and Brennan 2006). Since we can control the behavior of the computer partner, we can use this forum to test subtler questions. For example, Stent, Huffman and Brennan report that when there is a localized misunderstanding (such as a speech recognizer "mishearing" one or more words in an utterance), the phonetic adjustments made (such as a slowing of speech rate) persist in the discourse, sometimes quite far beyond the problematic utterance. In addition, segmental adjustments (such as more careful pronunciation of consonants) are made not only to repeated forms of the misunderstood words, but also to forms preceding and following them. The speech rate effect may implicate global, or long term, adjustment of a general speech production variable such as clock rate. It would be interesting to see how this phenomenon interacts with, or is differentiated from, effects on clock rate found at prosodic domain boundaries (Byrd et al. 2000). The segmental "clarity" effect may be an instance of somewhat localized change in gestural specifications such as gesture stiffness, though more extreme changes, such as use of the full vowel for *a* rather than schwa, might be argued to be instances of planned lexical substitution with register shift. Thus, research on interactive speech, even with a non-human partner, can yield independent support for core theoretical concepts such as control parameters in speech production.

Another type of research on interactive speech focuses on how people adjust their speech in response to indirect contextual information about their partner's knowledge (e.g., Lockridge and Brennan 2002). Recent research documents how non-native speakers make such adjustments to resolve phonological ambiguity in a second language. Hwang (2006) found that Koreans speaking English to a native speaker were more likely to produce accurate forms of words with voiced coda stops (e.g., *bib*) when there was a competing form with voiceless stop (e.g., *bip*) in their partner's visual field. Adjustments included vowel length and consonant closure voicing. Future studies are needed to determine how information about the addressee's grammar affects the ways that speakers resolve potentially ambiguous reference. Possible manipulations include comparing speech to addressees who are native versus non-native speakers, or who have the same or different dialects as the speaker.

Turning to social differences between speakers, we are aware of register shifts and even dialect changes that speakers make when conversing with different partners. Stuart-Smith (this volume) demonstrates how speakers manifest age, gender and class in a laboratory setting (word list reading). A perfect companion study would compare behavior of these same subjects when speaking to a partner who differs in one or more socio-linguistically relevant characteristics. Brennan, Huffman and Stent (2006) have used a laboratory setting to study phonetic adjustment to conversational partners differing in age, status, and dialect. Students with certain Long Island English pronunciation features (especially r-dropping) played a card game with a fellow student from Long Island (who deliberately used a strong Long Island/New York accent, including r-dropping) and then with a professor who speaks a dialect more like textbook English, but specifically lacking r-dropping. The students used less, or less extreme r-dropping when playing with the professor, as might be expected. Since we have observed of our students making quite dramatic dialect switches when speaking to faculty

versus students, perhaps the most surprising result here is that in this instance the switch was not perfectly clean. This may be due to the fact that the subjects had already associated the task with their own dialect pronunciation, from the first session with the Long Island dialect speaker. This account would thus offer support for the notion that we store extensive experiential information with our representations of the sound characteristics of words or segments, and that these experiences affect our choices in speech production. Alternatively, it may be that the relaxed conversational style of the professor partner led subjects to use a more casual register, in which, by self report, they usually use more r-dropped forms[1]. These results, while preliminary, suggest the merit of doing additional laboratory studies which further separate the influence of age, status and dialect in adaptation.

3.2 Investigating socially-driven variability in perception

For all forms of speech, the essential question in speech perception is the same: how the variability that listeners encounter is manipulated to arrive at a phonological analysis, and ultimately, meaning. Research has shown that when we perceive vowels, we adjust our expectations based not only on presumed vocal tract size of the speaker (Nordstrom and Lindblom 1975), phrasal phonetic context (Ladefoged and Broadbent 1957) and phrasal fundamental frequency (Johnson 1990), but also on properties of the speaker, including language and gender (Johnson 1997) and age, class and related expectations of patterns of variability (Warren, Hay and Thomas, this volume). Warren, Hay and Thomas argue that knowledge of socially-indexed variability is represented in the lexicon, partly through the association of varying phonemic shapes with the same lexical item. For example, in New Zealand English, there is a vowel raising process is underway, such that a word like *fair*, may be heard either as [feə] or [fiə]. They propose that both phonetic forms are included in the set of stored possible phonemic shapes for the word. This general approach is in line with other proposals of exemplar type representations of words (e.g., Johnson 1997; Pierrehumbert

2002; Hawkins 2003; Pitt and Johnson 2003). However, the results show that speakers/listeners have much more fully elaborated knowledge, which needs to be modeled. Their perceptual and acoustic analyses indicate that hearing a speaker who is older (and thus of the generation that usually maintains *fear* and *fair* as distinct) biases listeners to hear tokens of these two words as more distinct, even if they are not phonetically more different. It is not clear that these results can be handled by a model like that sketched in their paper, where a pre-lexical processor assigns a phonemic categorization independently of the lexicon.

The independent status of the envisioned pre-lexical processor is also called into question by cases of perceptual learning, an abstract form of listener adaptation, in which lexically based expectations of meaning-to-sound correspondence influence sound categorization. Norris, McQueen and Cutler (2003) have shown that Dutch listeners can be trained to accept variants of a target phoneme that are quite distinct from those they can normally be expected to have heard. After a training phase, in which they heard a fricative intermediate between f and s on a series of f-final words, subjects accepted more items along an [ɛf] to [ɛs] continuum as "f" than prior to training. Kraljic and Samuel (2005) have shown that similar perceptual learning effects can last as long as 25 minutes after training, which they take to be indicative of change to stored representations. The test items for phoneme categorization after perceptual learning training were nonsense words, and the task was to label the sounds as "S" or "SH", so it would appear that the lexicon is not involved in this perception task. In this case, then, is the pre-lexical processor involved at all? If so, the results suggest that the pre-lexical processor's boundaries were affected by the training phase, via mediation from the lexicon. If this is the case, then the location of experiential information in the lexicon (only), and thus its neat separation from a lexical pre-processor, is called into question. Either there is bi-directional movement of information between the pre-lexical processor and the lexicon, or the division must be abandoned. Future research is needed to assess more carefully the influence that

lexical information can have on phonemic categorization. As indicated by Stuart-Smith's, and Warren, Hay and Thomas' results, attention to class and age-related expectations could be fruitful.

3. Additional directions for future research

If we open up everything for study, it is a vast field indeed. Here I will briefly discuss some existing research threads which might be productively extended as areas of collaboration between laboratory phonology and socio-phonetics.

One of the most compelling areas for additional research is socio-phonetic patterning in languages other than English. What types of variability are used to signal social-indexical information in other language communities? Are there cross-linguistic trends that give insight into the structural content of sounds? For example, are certain features more often manipulated to signal register differences? The second language acquisition literature provides some discussion of non-English socio-linguistic patterning. For example, Beebe (1987) discusses variants of "r" in Thai that are conditioned by formality, and consequences of transfer in pronunciation of English "r". Initial "r" in Thai is trilled in the most formal situations and flapped in less formal but still careful speech. Beebe found that Thai's speaking English used more trilled r's (and thus fewer approximant r's) in list reading than in interviews. Thus, the more formal situation did not result in more care to pronunciation, and greater accuracy, but actually meant less accuracy, due to first language interference. The Thai data follow English and other languages in which phonetic differences accompanying register differences seem to involve fortition/lenition. In articulatory phonological terms, we might attribute lenition in less formal situations to lessened gesture stiffness, or to the consequences of accommodating more overlapped, but contradictory, gestures for adjacent sounds. Further research, including research on more languages, may clarify this question.

Another important research area is cross-linguistic investigation of the phonetic changes used to signal factors such as class, ethnicity and age. An example research question, raised implicitly in Stuart-Smith's paper (this volume), is whether certain sounds serve as better "hosts" for signaling socio-linguistic information. Stuart-Smith suggests that [s] shows more such variation than [ʃ] in Glaswegian because [s] is longer, and since the constriction portion is information bearing for fricatives, /s/ therefore can carry a higher informational load. It seems equally likely that paradigmatic pressures in a language may limit the extent to which they may be varied to signal socio-linguistic information. For example, English /s/ and /ʃ/ are fairly distinct acoustically to start with, leaving some room for socio-linguistically motivated variation in at least one of them, without endangering their distinctiveness. However, in a language in which there are additional fricatives in the alveo-palatal region (such as /s/ allophones [ɕ ɕʷ sʷ] of Korean), /s/ may not be a good candidate for signaling socio-phonetic information, whatever its length characteristics. The contextually determined allophones presumably co-opt the available "variation" space. Beyond the influence of segment inventories, we can also hypothesize that the nature of the articulatory:acoustic mapping for different types of sounds determines how much they can be varied for socio-linguistic purposes. That is, sounds participating in contrasts that involve more "quantal" properties (e.g., Stevens 1989) may not be good candidate hosts for socio-phonetic information, because small articulatory changes will have little acoustic effect, while past a certain point, articulatory changes will result in a big effect, and clearly a different phoneme. These types of questions can only be answered by analyzing data on socio-phonetic variation in a wide range of languages.

4. Concluding Remarks

The papers in this section are only a sample of the interesting and theoretically informative work being done in socio-phonetics. The theoretical issues and future research areas sketched here are just the beginning of what I believe will be a long and productive conversation between laboratory phonologists and socio-phoneticists. The cross-pollenation of ideas and methodology brought about by active collaboration between researchers in these areas should significantly advance our efforts to develop detailed models of mental representations of sound structure which have true predictive power.

Notes

[1] In fact, many students from Long Island equate r-dropping with casual register, and are surprised that for someone from another region, say California, r-dropping isn't possible in casual speech.

Acknowledgements

Thanks to Giorgios Tserdanelis and Susan Brennan for comments on a draft of this paper, preparation of which paper was supported in part by the National Science Foundation under Grant No. 0325188.

References

Beebe, Leslie 1987 Sociolinguistic variation and style shifting. In *Interlanguage Phonology: the acquisition of a second language sound system*. Georgette Ioup, and Steven Weinberger (eds.), 378-388. Cambridge, Mass.: Newbury House.

Bortfeld, Heather and Susan Brennan1997 Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes,* 23: 119-147

Bradlow, Ann 2002 Confluent talker- and listener-oriented forces in clear speech production. *Papers in Laboratory Phonology VII*, Carlos Gussenhoven and Natasha Warner, (eds.), 241-273. Berlin/New York: Mouton de Gruyter.

Bradlow, Ann and Tessa Bent 2002 The Clear Speech Effect for Non-Native Listeners. *Journal of the Acoustical Society of America* 112 (1): 272-284.

Brennan, Susan and Herbert Clark 1996 Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 6: 1482-1493.

Byrd, Dani 1994 Relations of sex and dialect to reduction. *Speech Communication* 15: 39-54.

Byrd, Dani, Abigail Kaun, Shrikanth Narayanan and Elliot Saltzman. 2000 Phrasal signatures in articulation. In *Papers in Laboratory Phonology V*: *Acquisition and the Lexicon*, Michael Broe and Janet Pierrehumbert, (eds.), 70-87. Cambridge: Cambridge University Press.

Cassidy, Frederic and Joan Hall (eds.) 1985-2002 *The Dictionary of American Regional English, Volumes I-IV*. Cambridge, Mass.: Harvard University Press.

Cole, Jennifer, Hansook Choi, Heejin Kim and Mark Hasegawa-Johnson 2003 The effect of accent on the acoustic cues to stop voicing in Radio News speech. *Proceedings of the XVth International Congress of Phonetic Sciences*, 2665-2668.

Dilley, Laura, Stephanie Shattuck-Hufnagel and Mari Ostendorf 1996 Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24: 423-444.

Hawkins, Sarah 2003 Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31: 373-405 .

Hawkins , Sarah and Noël Nguyen 2004 Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics* 32: 199-231.

Hwang, Jiwon 2006 Phonetic ambiguities in non-native speech and disambiguation. Ms. Department of Linguistics, Stony Brook University.

Johnson, Keith 1990 Contrast and normalization in vowel perception. *Journal of Phonetics* 18: 229-254.

Johnson, Keith 1997 Speech perception without speaker normalization: An exemplar model. In *Talker Variability in Speech Processing*, Keith Johnson & John Mullennix (eds.), 145-165. San Diego: Academic Press.

Kenstowicz, Michael 1994 *Phonology in generative grammar*. Cambridge, Mass.: Blackwell.

Kraljic, Tanya and Arthur Samuel 2005 Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* 51: 141-178.

Krause, Jean and Lou Braida 2004 Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America* 115 (1): 362-378.

Ladefoged, Peter, and Donald Broadbent 1957 Information conveyed by vowels. *Journal of the Acoustical Society of America* 29(1): 99-104.

Lockridge, Calion and Susan Brennan 2002 Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review* 9: 550-557.

Nordstrom, P and Bjorn Lindblom 1975 A Normalization Procedure for Vowel Formant Data. in *Proceedings of the International Congress of Phonetic Sciences*, Leeds, England.

Norris, Dennis, James McQueen and Anne Cutler 2003 Perceptual learning in speech. *Cognitive Psychology* 47: 204-238.

Oviatt, Sharon 1996 User-centered design of spoken language and mulitmodal interfaces. *Institute of Electrical and Electronics Engineers Multimedia*, 3: 26-35.

Oviatt, Sharon, Margaret MacEachern, and Gena-Anne Levow 1998 Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication* 24 (2): 87-110.

Pardo, Jennifer 2004 Acoustic-Phonetic convergence among interacting talkers. *Journal of the Acoustical Society of America* 115 (5,2): 2608.

Picheny, Michael, Nathaniel Durlach and Lou Braida 1986 Speaking clearly for the hard of hearing II: Acoustic Characteristics of clear and conversational speech. *Journal of Speech & Hearing Research*, 29(4): 434-446.

Pierrehumbert, Janet 2002 Word-specific phonetics in *Laboratory Phonology VII* Carlos Gussenhoven and Natasha Warner (eds.), 101-140. Berlin/New York: Mouton de Gruyter.

Pitt, Mark and Keith Johnson 2003 Using pronunciation data as a starting point in modeling word recognition. *Presented at the 15th International Congress of Phonetic Sciences, Barcelona.*

Stent, Amanda, Marie Huffman and Susan Brennan 2006 Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. Under review *Language and Speech*.

Stevens, Kenneth 1989 On the quantal nature of speech. *Journal of Phonetics* 17: 3-45.

Index entries regular: secret language, register, dialect, laboratory speech, age, class, gender, misunderstanding, conversation, adaptation, communicative need, r-dropping, perceptual learning,

Index entries Languages: Thai, Sanga, Korean