The linguistic basis of phonetic coherence*

Keith Johnson
UC Berkeley

"Phonetic coherence" refers to the fact that speech is not perceived acoustically. To illustrate this consider a spectrogram of the word "trip" in English (Figure 1). The acoustic intervals are silence, burst, voiceless /ɹ̥/, voiced /ɹ/, vowel /ɪ/, final /p/ closure, and burst. Yet people hear the segments /t/ /ɹ/ /ɪ/ and /p/. Some acoustic bits cohere with each other.

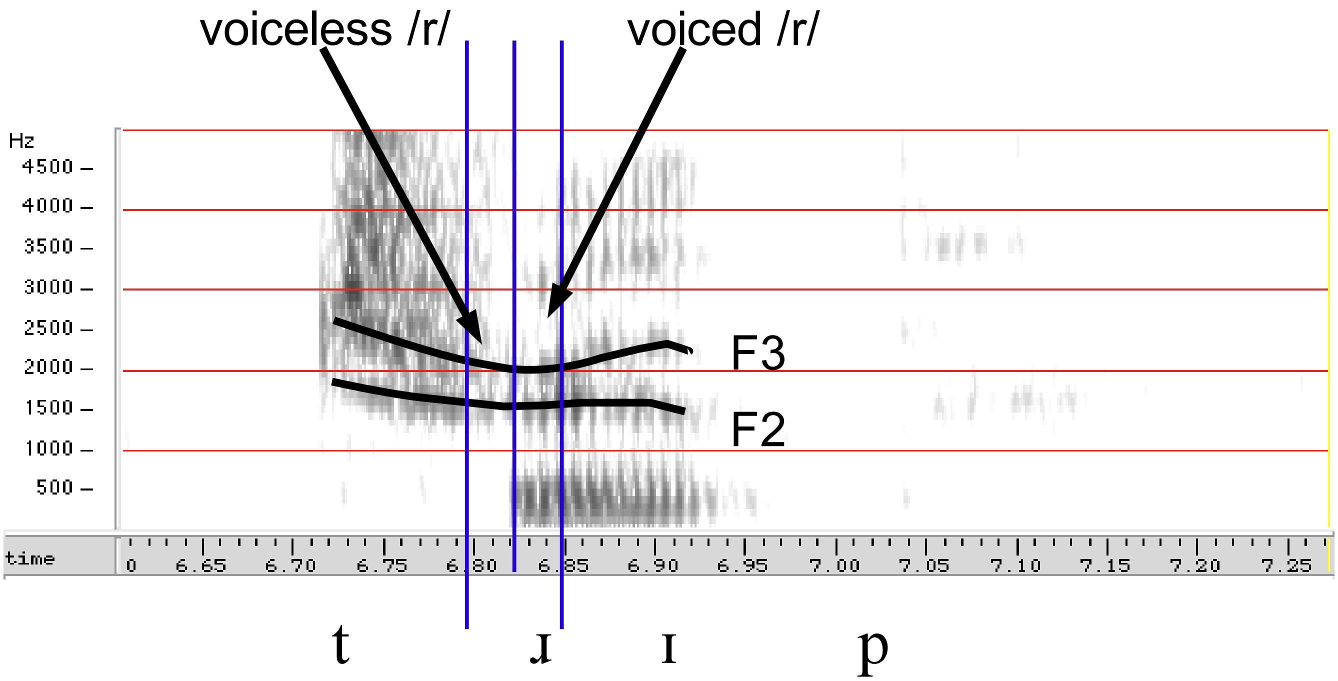

Figure 1. A spectrogram of the word "trip" with acoustic segments labelled.

One possible basis of phonetic coherence is that listeners hear acoustic "cues" like the voiceless and voiced portions of /r/ in terms of the vocal tract actions that produce them. Thus, the fact that there is a single rhotic gesture in "trip" causes the two rhotic acoustic bits to cohere into a single perceptual event.

It is fashionable in some circles to act as if experiencing the speech of others in terms of vocal tract actions is some sort of vaguely shamanistic approach to speech science. It has not always been so. For example, Abercrombie (1967, p. 97) discussing rhythm says,

> "There are important implications for perception here. Speech rhythm is experienced as a rhythm of movement. Obviously it is directly so experienced by

* Presented at the 11th Laboratory Phonology Conference. This paper is a response to Remez (2005) and Goldstein and Fowler (2003) – two important papers on phonetic coherence in speech perception.

> the speaker, but what of the hearer? We can say that he, too, is vicariously experiencing a rhythm of movement – he, in a sense, is a speaker also. ... We talk for convenience, about 'hearing' rhythm, but in fact we feel it, entering empathetically into the movements of the speaker, to which the sounds we hear are clues."

The attitude expressed in this quotation focussing on hearing as "vicarious experience" as "entering empathetically into the movements of the speaker" is key to an understanding of speech perception in terms of the gestural contents of speech, and also, I will suggest, of speech perception as a linguistic activity.

Another perception-based view of phonetic coherence relies on gestalt principles of grouping (Bregman, 1990) which are seen as causing acoustic bits to cohere with each other perceptually as auditory objects are formed (Diehl & Kluender, 1989). As with the gesture perception view, this is a view of phonetic coherence which depends on an autonomous speech perception mechanism wholly independent of language or linguistic knowledge.

I will argue in this paper that <u>phonetic coherence is a linguistic</u> <u>phenomenon</u> rather than only or primarily based in prelinguistic speech perception. In my view, the coherence of voiceless [ɹ̥] and voiced [ɹ] in "trip" is partly driven by the fact that English does not have a distinction between voiced and voiceless [ɹ̥]. We can easily imagine a language which does contrast them (as some dialects of English contrast voiced and voiceless [w] - "where" vs. "wear") and thus the sequence [ɹ̥ɹ] is perceived as just that, a sequence. That is, phonetic coherence in speech perception is guided by the listener's system of linguistic contrasts.

Linguistic Equivalences

The perceptual coherence models, whether relying on gesture perception or auditory scene analysis, argue for perceptual equivalences (such as /ɹ̥/ = /ɹ/ or equivalence of different cues for voicing- F1, duration, etc.) to explain phonetic coherence. I argue that linguistic equivalences are involved.

Linguistic communication is accomplished by the use of equivalences at a number of different levels. Among these, are phonetic equivalences which underlie phonetic coherence. However, before discussing phonetic equivalence it may be useful to illustrate the notion of linguistic equivalence with a few cases at different levels – this emphasizes that language makes extensive use of variation.

For example, gesture may be used in face to face conversation to disambiguate diectic pronouns. When a person, looking at an apple and an orange states "I like these better than those" the utterance is ambiguous. Does he (or she) prefer apples? Pointing first to the apple (while saying "these") and then the orange (while saying "those") disambiguates the statement. In this situation we have a linguistic equivalence - "I like these better than those" with pointing is equivalent to "I like apples better than oranges."

Words may also exhibit linguistic equivalence. For example, "children" and "kids" are equivalent in "The children were playing in the sandbox" and "the kids were playing in the sandbox" in the sense that they may refer to the same group of people in a social setting. Of course there may be subtle connotative differences between "children" a more formal word, and "kids" a less formal word, but in terms of the semantics of reference they are equivalent.

My main claim in this paper is that phonetic coherence is a result of linguistic equivalence. This idea is not new (see for example, Pierrehumbert, Beckman & Ladd, 2000), though I hope that the particular constellation of data or their presentation may constitute an original contribution of some small incremental value. A word or a phoneme is a set of phonetically coreferential forms, so that as different words may refer to the same concept, so also different phonetic realizations may "refer to" or signify the same linguistic item.

This many-to-one mapping between phonetic forms and lexical entities has significant implications for the theory of speech perception, so let's turn now to a consideration of some of the facts that suggest that phonetic coherence has a linguistic basis.

Phonological Equivalence

The paradox that we must face when it comes to the pronunciation of language is that on the one hand <u>speakers</u> have exquisitely fine control over the phonetic details of their speech, while on the other hand <u>listeners</u> tolerate a tremendous range of pronunciation variation.

I was most impressed by the fine phonetic control in examining the data published as Johnson, Ladefoged, and Lindau (1993). In looking at x-ray microbeam traces of articulator movement, I was observed that a speaker's repetitions of a phonological sequence were remarkably consistent – articulatory traces could be superimposed on each other and they overlapped almost entirely, while traces of the same sequences produced by different speakers were never so similar. That is, individual differences were much larger than within-speaker variability.

Recently, John Westbury and his colleagues at Wisconsin (Westbury et al., 1998) have quantified this observation in their study of American English /r/. Figure 2 shows in panel (a) that the tongue tip follows the same trajectory in repeated productions of the word “across” within-talker. Westbury et al. measured the “standard distances” (a multidimensional measure of variability) for these within-talker repetitions and found that the deviation was only about 1mm. Panel (b) shows that different talkers have quite different tongue shapes in the /r/ of “row”.

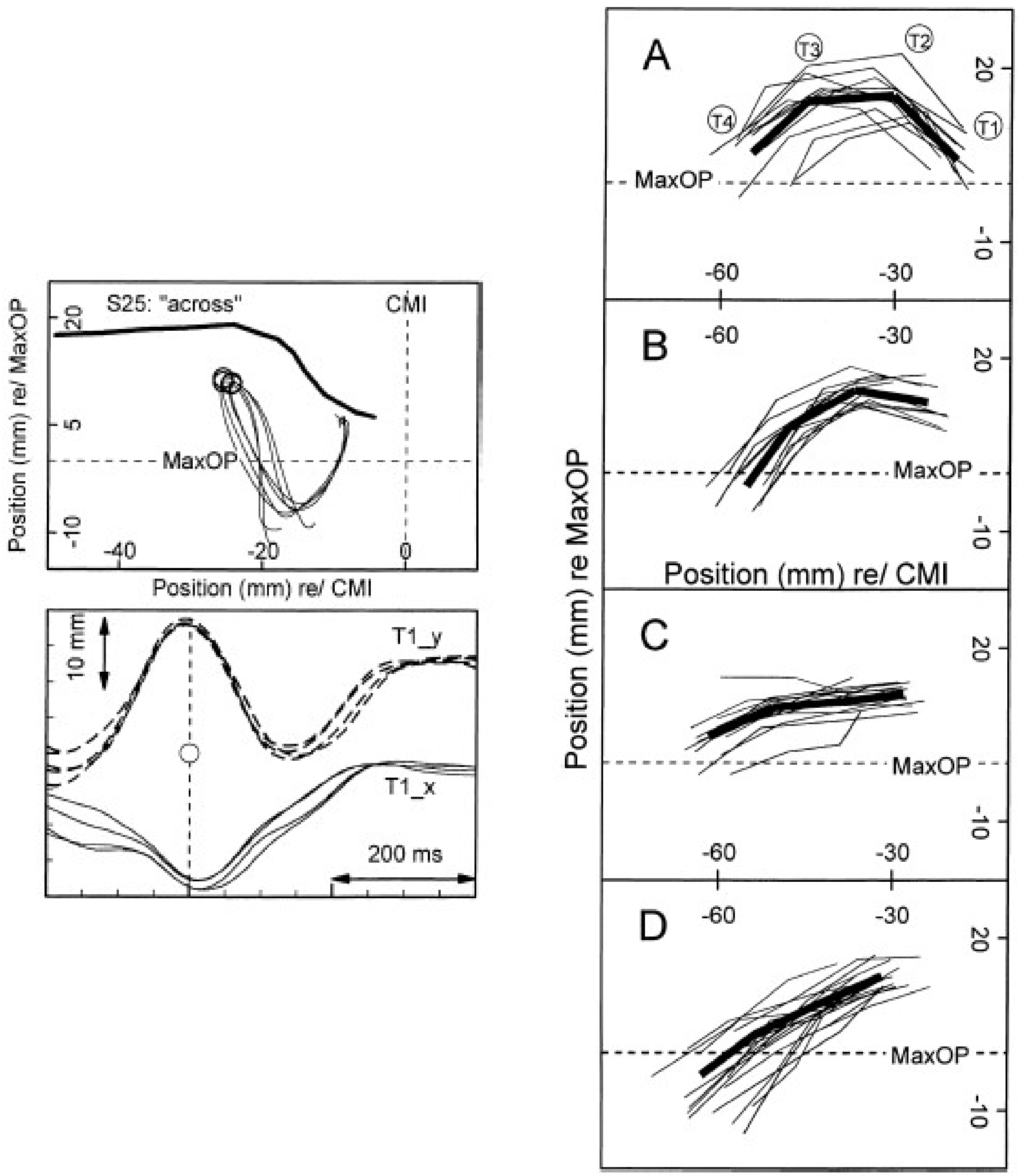

Figure 2. Data reproduced from Westbury et al. showing (a, left panel) within speaker consistency, and (b, right panel) between speaker differences in the articulation of American English /r/.

Interestingly, Misty Azara and I found a similar pattern of within and between talker variation in an as yet unpublished study of speech produced by identical twins. Twin's vowel formant trajectories were remarkably similar to each other and different from other speakers. For example, figure 3 shows productions of /o/ by four speakers – two sets of twins. Although we did find differences between twins suggestive of idiosyncratic style differences, the degree of similarity between twins was very striking.

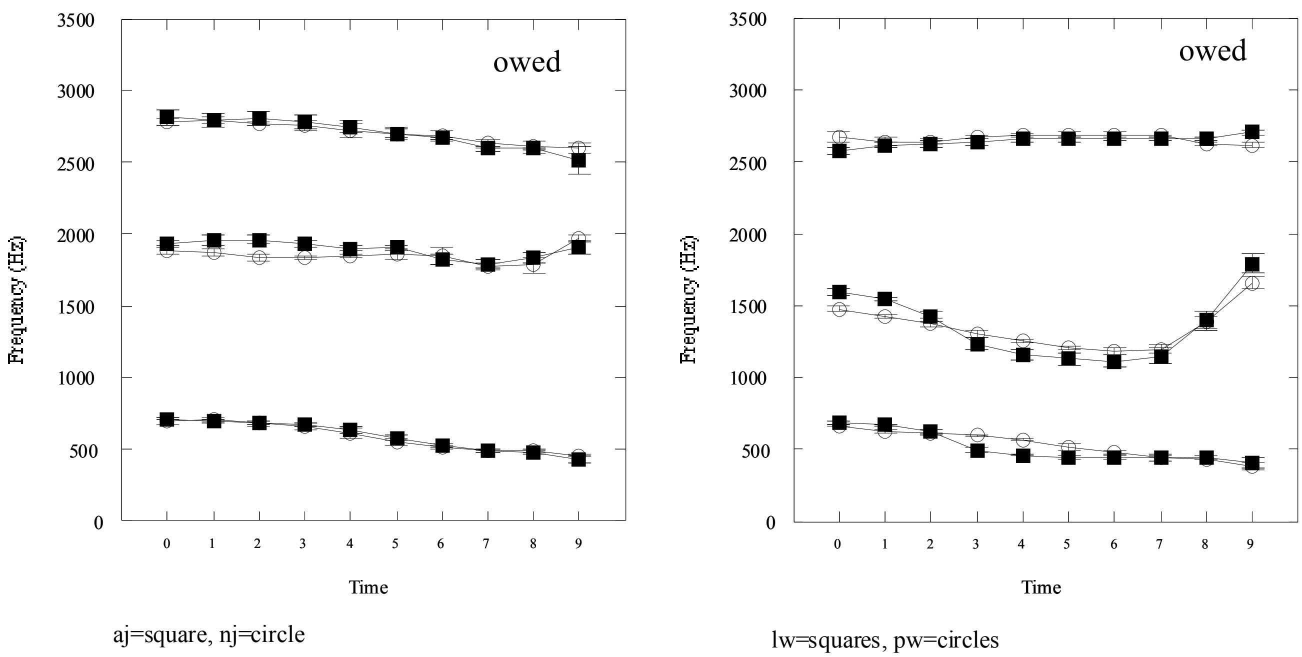

Figure 3. Vowel formants of twins' productions of American English /o/ in "owed". (a, left panel) twins AJ and NJ, (b, right panel) twins PW and LW.

Now, it is true that there is a substantial amount of within-speaker phonetic variation (e.g. Johnson, 2004). If we assume, for example, that all [s] articulations are the same in the sense that they result from the same motoric plan, then contextual and stylistic variation poses a substantial problem for my argument that speakers control the fine phonetic details in speech. For example, Perkell et al. (2000) suggest that the acoustic targets for speech motor control are broad swaths in two dimensional time/frequency space (following Keating, 1988, 1990) – a notion that is antithetical to the lack of variation noted in the matched speaking conditions discussed above. However, Browman & Goldstein (1986) articulatory phonology describes how within-talker phonetic variation across context and rate may emerge from invariant motor plans. Provisionally, I accept this account of within-talker within-word variation.

Returning then to the point of this section. The difference between talkers' and listeners' tolerance of variation is striking. Where talkers exhibit very restricted and lawful variation, listeners accept a huge amount of phonetic variation. For example, when I lived in the US state of Alabama (known for its “southern” accent), I learned to understand a new dialect of American English that I had never heard before – to any substantial extent. Many people have had this experience, and I think that the pattern is fairly common. At first exposure to a new dialect you may (or not!) understand much of what a person is saying, but there are certainly some points of confusion, some definite decreased phonetic intelligibility.[1] Gradually, with continued exposure to the unusual dialect, you adapt to the new phonetic patterns and intelligibility increases as a result of “perceptual learning” (Greenspan et al., 1988) So, although there are certainly exceptional individuals who can imitate different dialects convincingly (though the phonetic basis of even this ability is slim, Zetterholm, 2000), such skilled mimicry is the exception that proves the rule – the phonetic range we command as listeners exceeds our range as talkers.

This fact – that <u>listeners tolerate a larger range of variation than they control as speakers</u> – suggests that phonetic equivalence is at work in speech perception. Listeners do not require physical (gestural or auditory) uniformity in their linguistic categories, the vowel of “I” is the same linguistic element whether it is pronounced [ɑɪ̯] as in California or [ɑː] as in Alabama.

Gestures and mouths.

A second fact suggesting that phonetic coherence is not based on perception of gestures is that people's mouths differ from each other. Men and women have different ratios of pharynx and mouth length (Hagiwara, 1995) and therefore constriction between tongue body and soft palate impacts different locations in the standing waves of vocal tract resonances for men and women. That is, if men and women produce the same gesture – aiming at the same location and degree of maximum oral constriction then they should produce acoustically distinct sounds. Paradoxically, to produce the acoustic output that speakers do actually produce (say the lowest F3 possible for a given vocal tract) their gestures must differ.

This observation, <u>the gestural nonidentity of speech in different vocal tracts</u>, is related to the first fact, listener toleration of variance, but from a gestural point of view. The variation of speech across talkers is not simply a matter of acoustic variation – people also make different speech gestures while producing the “same” speech sounds. Westbury et al.'s (1998) data on /r/ production shown in figure 2 above (as well as more recent data from Mielke et al., 2007) is one indication of the range of gestural nonidentity found in speech production.

Language effects in speech perception.

The idea that we are considering is that phonetic coherence is a product, to a substantial

1 This has been found even for nonstereotypical male and female voices in a known dialect (Strand, 2000).

degree, of linguistic perception. In this conception, listeners impose a linguistic “frame” on the phonetic substance of speech – interpreting the speech “signal” in terms of the “signs” provided by language.

One of the most direct ways to test this conception of speech perception is in cross-linguistic studies. For example, Boomershine et al. (2007) tested the perception of [ð] [d] and [ɾ] by American English and Spanish-speaking listeners. The phonology of English suggests that [ð] will be different from [d] and [ɾ] which are allophonically related, while the phonology of Spanish suggests that [ð] should be similar to [d] because they are allophonically related while flap [ɾ] is the odd one out as an allophone of trilled /r/. Listeners' judgments of perceptual similarity follow the linguistic system for both stimuli presented for difference rating judgments (Figure 4) and in speeded discrimination (Figure 5).

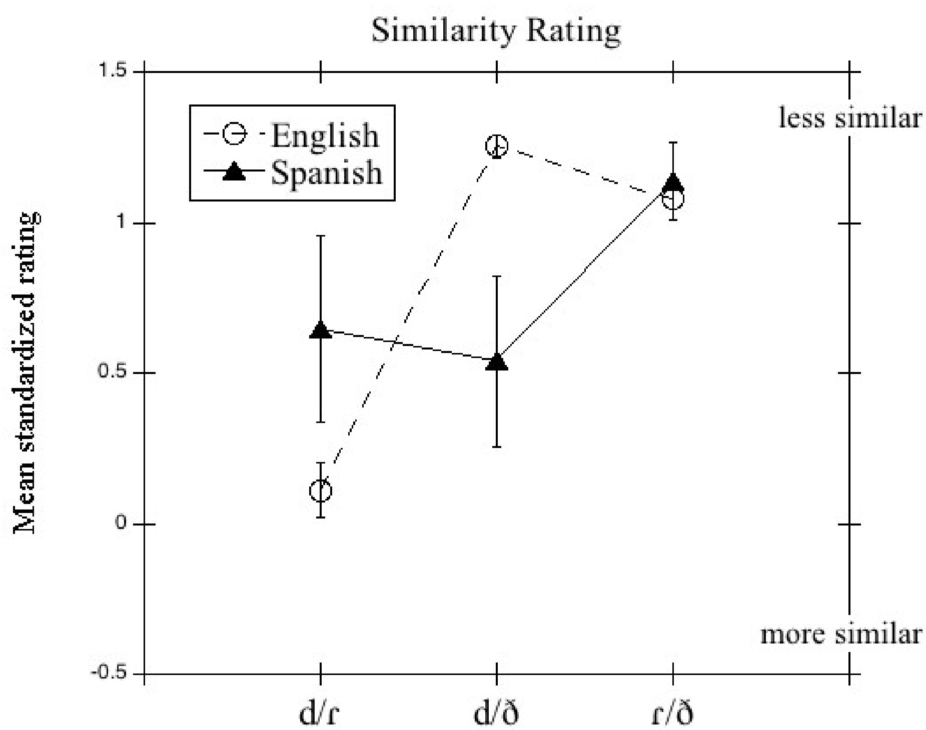

Figure 4. Figure 4. Normalized difference rating judgments for English speaking and Spanish speaking listeners (from Boomershine et al.,2007).

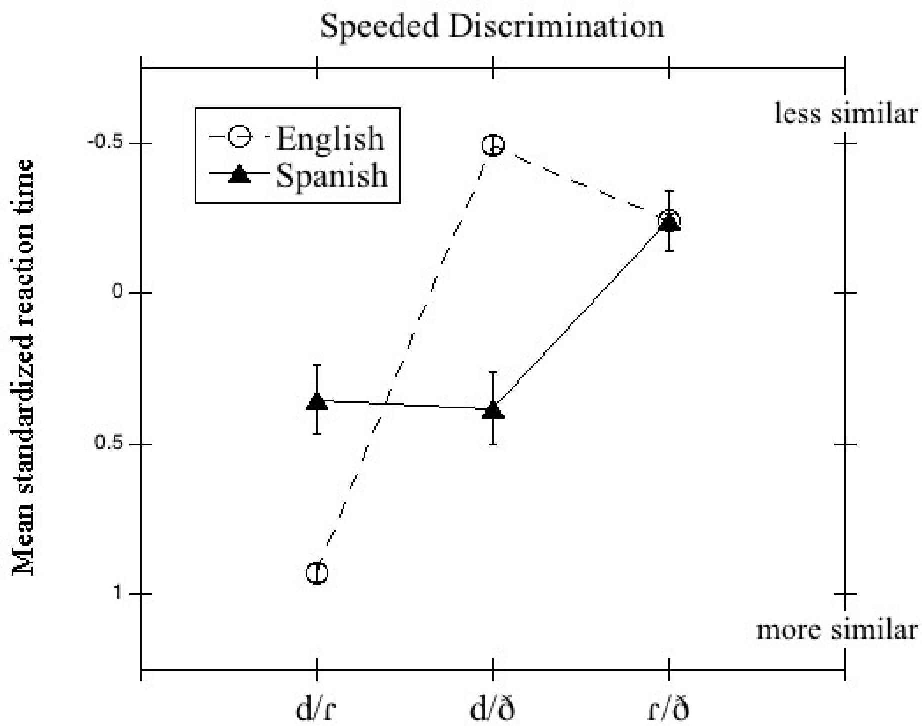

Figure 5. Normalized reaction time in speeded discrimination for English speaking and Spanish speaking listeners (from Boomershine et al., 2007).

The language effect in this experiment is that the listener's interpretation of the phonetic difference between [ð] and [d] (for example) is influenced by the status of that contrast in his/her native language. That is, phonetic perceptual space is warped by linguistic experience.

So, we have three arguments suggesting that phonetic perception is linguistic in nature. First, listeners tolerate a larger range of variation than they control as speakers. Second, gestures for the same speech sound produced in different vocal tracts are nonidentical. And third, phonetic perceptual space is warped by linguistic experience. There are other phenomena that are consistent with the idea that phonetic coherence has a linguistic basis, some of which will be touched on later in this paper, but these three are sufficient to establish the basic credibility of my point. In the next section we will look back at some historical considerations and in the final sections conclude with a look forward toward the future of phonetics.

Phonological Learning

The motor theory view (Liberman & Mattingly, 1985) was that in speech perception listeners recover the articulatory "intension" of the speaker, so what is "heard" is articulatory in nature. Crucial in Liberman's view is the idea that the recovery of articulatory intension is guided by an innate cognitive module. Children do not need to learn to recover articulations because the speech perception organ is equipped to do this.[2] Liberman (1996) rejected his earlier (Liberman, Harris, Eimas, Lisker & Bastian, 1961) account of learning in speech perception and the language-specificity of perception remained a key problem for the motor theory. When I asked him about it in the late 1990's he said "of course the module is 'tuned' by experience." This undiscussed tuning mechanism significantly weakened the motor theory.

To put this into a bit of context. Concurrent with the development of motor theory Chomsky and Halle (1968) building on work by Trubetskoy and Jakobson, were developing generative phonology. Their book SPE responded to the typological "universals" found in the phonological inventories and phonologies of the world's languages (cf. the Stanford Universals project) by asserting that infants are endowed with a Language Acquisition Device which provides an innate universal language independent phonetic module. This Universal Phonetics maps distinctive feature combinations to movements of the mouth in a deterministic way – regardless of language. Liberman's motor theory is (or was in its strong form) the perceptual module implementing the inverse of universal phonetics. This is a great model except for the "tuning" part.

During the 1980's and 1990's the UCLA phonetics lab, under the direction of Peter Ladefoged, was engaged in a project to document and highlight the language-specificity of phonetics (Disner, 1983; Nartey, 1982, Ladefoged, 1984). This was in direct opposition to the SPE conception of universal phonetics. Ladefoged knew, because of a mass of phonetic research showing it, that sounds vary from language to language.

Consider, for example Jonas Nartey's cross-linguistic study of fricative noises. Figure 6 shows a typical result. Averaging critical band spectra from the same number of male and female speakers of Navajo and Zuni, Nartey found that the [ʃ] of Zuni has a lower spectral center of gravity than does the [ʃ] of Navajo. This is probably an unremarkable finding to most of us now. Why would we think that [ʃ] necessarily has to be the same in all languages, after all? But in an era of debate about universal phonetics (even though Labov, Yeager, & Steiner, 1972 had amply documented the arbitrariness and specificity of phonetic details of dialect variation) it was important to document the cross-linguistic arbitrariness of phonetics. Now of

2 Goldstein and Fowler (2006) cite infant face imitation behavior implying that no gesture learning needs to takes place in language acquisition.

course this language-specific phonetics is so a part of our shared understanding of phonology that, for example, Kuhl et al. (1992) assumed that Swedish and English babies learn phonetically different [i] vowel "perceptual magnets".

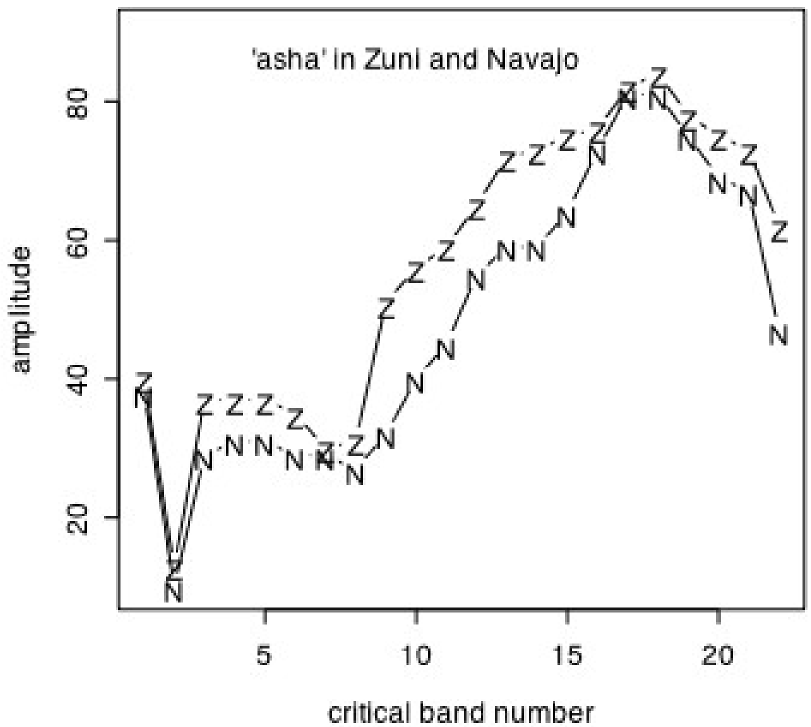

Figure 6. Average spectrum of [ʃ] in Zuni and Navajo. The figure is drawn from data published by Jonas Nartey (1982).

Findings such as these on the language specificity of phonetics indicate that universal phonetics is not a viable concept. That there is no innate universal mapping between distinctive features and phonetics. In response to this finding, it became common to speak of the "phonetics phonology interface" - a language-specific mapping between universal but now contentless distinctive features and the observable phonetic properties of language.

Emergence

From the nineties to the present decade, researchers have had to ask themselves – 'what are these mysterious contentless distinctive features?' This has lead to serious consideration of the notion of phonological emergence (Cole and Iskarous, 2001; Jusczyk, 1992, 1997; Lacerda, 2003; Lindblom, 1992; Ohala, 1990; Plaut & Kello, 1998). In this view, children build phonology during acquisition rather than tune it. Although a complete tabula rosa view of phonology acquisition is not seriously tenable, it is possible to posit a humanistic emergence in which the construction of phonology is aided by certain innate capacities.

One of these is a capacity to imitate other humans using a rough visual-to-articulatory mapping. This capacity is likely not tied to speech alone (consider waving, smiling, etc.) and is probably much too coarse-grained to account for much in speech acquisition other than providing a starting point for acoustic imitation. For example, the McGurk effect (McGurk & McDonald, 1976) is weak for children from 3-5 years old illustrating that the visual-to-articulatory mapping is coarse-grained, yet the importance of this acquisition bootstrap is illustrated by the fact that blind children experience language delay as compared with sighted children (Mills, 1987).

Another key innate capacity, which is again not likely to be speech specific, is the capacity to form cognitive categories through intersecting streams of processing (Edelman, 1987; 2004). The capacity to accumulate knowledge of phonological categories has been investigated in terms of "exemplar-based" memory systems (Goldinger, 1998; Johnson, 1997, 2007), but whatever mechanism turns out to be correct, the main observation here is that when we assume that language learners have a basic mechanism for building phonological categories, rich with variability and cross-cutting semantic, morphological, and social information, then we no longer need to assume that children are endowed with a universal phonetics. The theory comes into better alignment with the cross-linguistic data on the language-specificity of phonetics.

The significance of the emergent phonology view is that phonology is now private. Because, I as a speaker of language have constructed my phonology on the basis of the quirks of my own production and perception anatomy, and the quirks of my own experience with the languages that I speak, I have a phonology that is not identical with that of the other speakers of my languages (compare this with Goldstein and Fowler's, 2003, view that speech gestures are the public expression of language).

The Nature of the Linguistic Frame

I have been arguing that phonetic coherence emerges from the interaction between the speech signal and the hearer's knowledge of language. It is important to explore, however briefly, the nature of the linguistic knowledge involved. In cross-linguistic studies of speech perception it is common to find perceptual influences resulting from the inventory of phonemic contrasts, or as Boomershine et al. found, the phonological organization of allophones in the listener's native language.

However, recent research in our laboratory (Palmer, 2007) suggests that the linguistic "frame" that provides phonetic coherence may be much more specific than merely the phonemes and allophones of a language. Palmer identified “reduced” and “full” variants of several words in the Buckeye corpus of conversational speech (Pitt et al., 2001), as illustrated in table 1. Reduced forms always had at least one vowel deletion, and hence one less syllable than the full forms. She cross-spliced tokens (with pitch smoothing) so that full variants were presented sometimes in other “full” contexts and sometimes in “reduced” contexts, and “reduced” variants were also inserted into both types of context. Listeners then performed a word spotting task with these words as targets.

Table 1. Example stimuli used by Palmer (2007).

Full 1: Because nobody is more similar [sɪ.mɪ.lar] to you than that sibling

Full 2: Is extraordinarily similar [sa.ma.lar] um, so what she says might

Reduced 3: Is often so similar [sɪm.lər] to where I’m coming from

Reduced 4: We were able to communicate in, uh, very similar [sm̩.lər] fashions

It was not surprising that full forms could be spotted more quickly than reduced forms in the full context (figure 7) because the full forms contained more phonological information than the reduced forms. Interestingly though, this advantage disappeared in the reduced contexts. Here, reduced forms were spotted slightly (nonsignificantly) faster than the full forms. That is, the match between context and target word was key to fast reaction time.

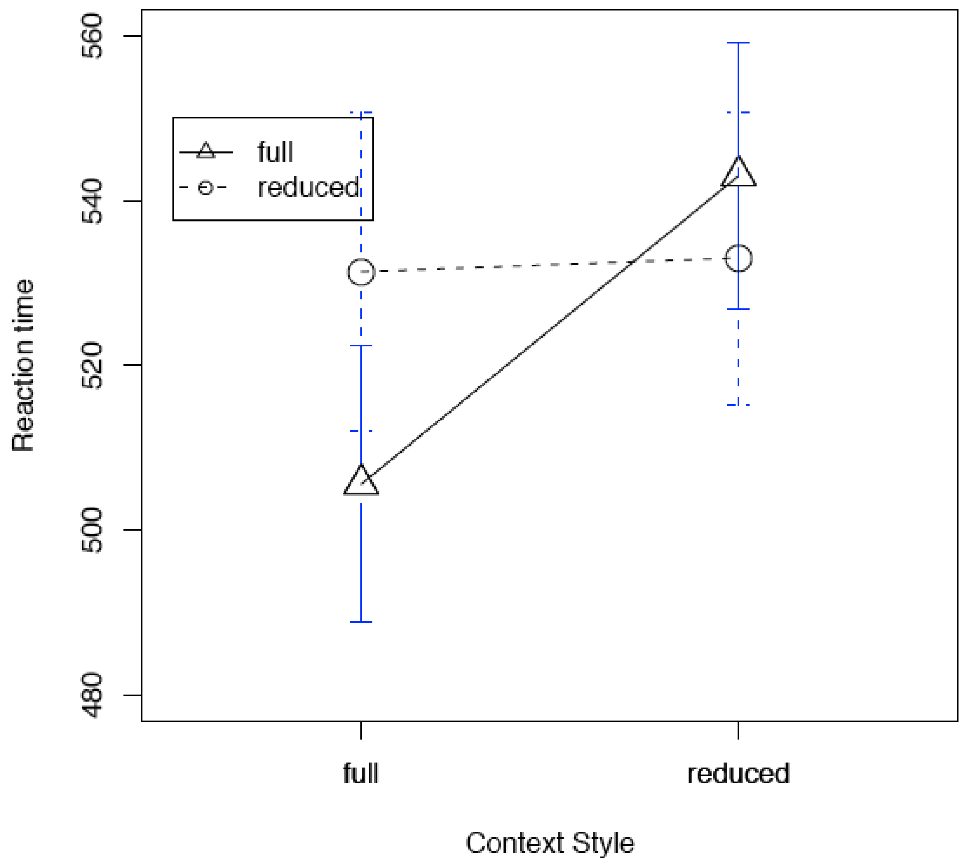

Figure 7. Lexical decision reaction times for full and reduced variants in full and reduced contexts (Palmer, 2007).

Here we have an example of how language affects perceptual expectations, with knowledge of language guiding the perceptual parsing of an inherently messy and ambiguous physical (acoustic or gestural) signal. Palmer's results show that the linguistic frame is closely tied with experience so that the expected forms of words adaptively match the speech rate or style of speech of the contexts in which they are heard.

Another recent study from our lab (Lau, 2008) illustrates that talkers also adjust to context. In her study, Lau had talker/listener pairs read words to each other in four different intelligibility test conditions as shown in figure 8. Lane & Tranel (1971) suggested that the Lombard effect, when talkers speak more loudly and clearly in the presence of noise is an indication of an active compensation for communicative difficulty because “the speaker does not change his voice level to communicate better with himself, but rather with others” (p. 692). However, this active talker view of the Lombard effect is suspect because it occurs even when the speaker is merely speaking into a microphone in the lab with no one else present. Lau’s study sought to determine whether talkers compensate when the listener is in noise as they do when the noise is present for the talker alone or for both the listener and talker. In the present context, considering phonetic coherence, this study helps to establish the linguistic frame for phonetic coherence. If talkers produce a version of the Lombard effect when only the listener is experiencing noise, we would conclude that phonetic coherence is dependent partly on an active computation on the part of the talker of the communicative needs of the listener.

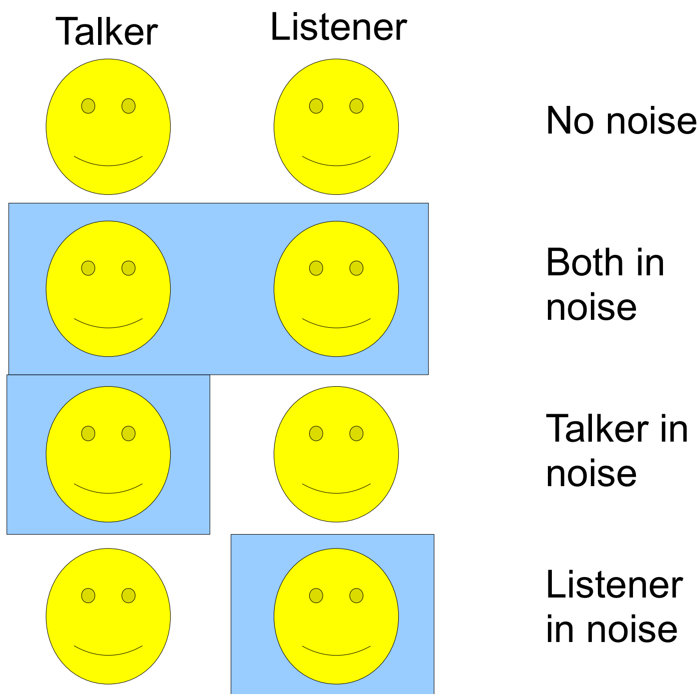

Figure 8. The four speaking conditions in Lau's (2008) study.

On average, talkers spoke more loudly in the both in noise and talker in noise conditions (figure 9). The condition of particular interest for the active talker hypothesis is the listener in noise condition. Here, the talker was speaking with no noise in his/her own earphones but was aware – via an intercom system with the listener – that the listener had a high level of background noise. The amplitude data of figure 9 suggest that talkers did speak louder in the listener in noise condition, but not as loudly as in the talker in noise or both in noise conditions. This partially supports Lane & Tranel's (1971) interpretation of the Lombard effect.

The average vowel spaces in the four conditions lead to a similar conclusion (Figure 10). All three of the noise conditions showed vowel space expansion relative to the no noise condition. In these data, it isn't obvious that the listener in noise condition resulted in a smaller effect than the talker in noise conditions. Lau (2008) also found increased durations of vowels in the three noise conditions compared with the no noise condition. So the result is clear, talkers respond to noise even when they themselves are not experiencing it, but they know that their listener is.

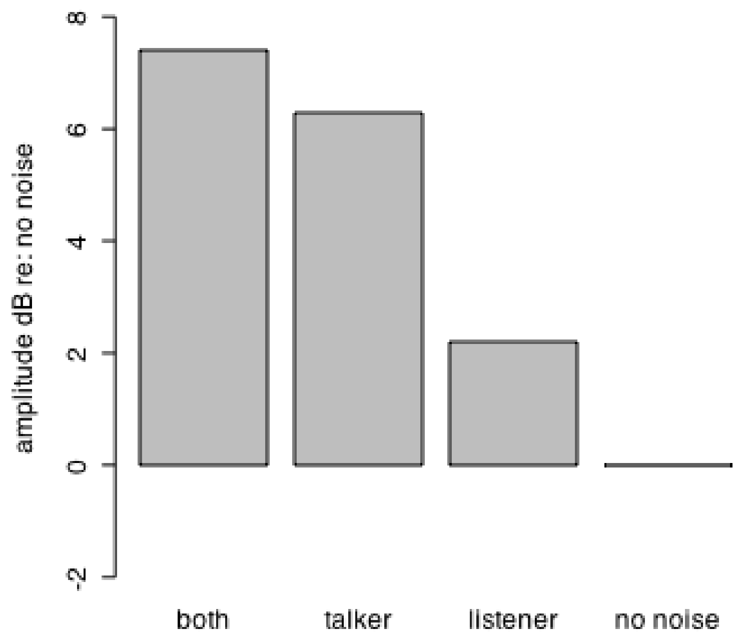

Figure 9. Average amplitude for the four conditions of Lau's (2008) experiment relative to the average amplitude in the no noise condition.

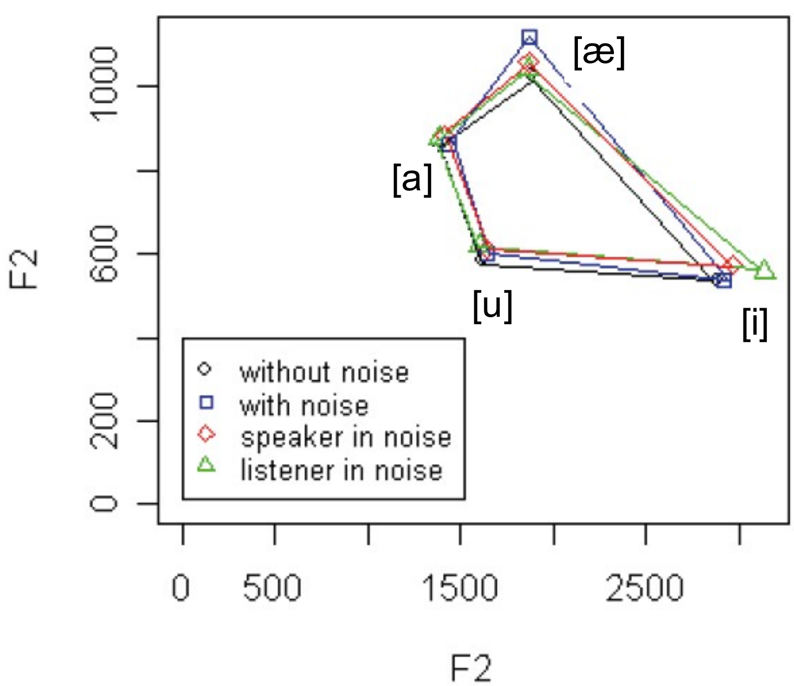

Figure 10. Average vowel spaces for the four conditions of Lau's (2008) experiment.

These two studies, Palmer's on auditory word recognition, and Lau's on the communicative basis of the Lombard effect, illustrate that the linguistic knowledge that forms a basis for phonetic coherence is context specific. Listeners' expectations vary depending on context in which a word occurs and talkers' performance also depends on the speaking situation.

Conclusion

Drawing on evidence from (1) listener's tolerance for wide phonetic variability in speech perception, (2) the gestural nonidentity of speech sounds, and (3) the language specificity of linguistic phonetics, I have argued that phonetic coherence is based in the linguistic equivalence rather than a (hypothetical) gestural or auditory equivalence of speech sounds. Further, consideration of listener's and talker's sensitivity to context suggests that the linguistic frame of reference for speech is highly adaptable and is sensitive to the contexts in which speech is produced.

References

Abercrombie, D. (1967) *Elements of General Phonetics*. Chicago, Aldine Publishing Co.

Boomershine A., Hall, K.C., Hume, E., & Johnson, K. (in press). The impact of allophony vs. contrast on speech perception. In P. Avery, E. Dresher & K. Rice (eds.), *Contrast in Phonology: Perception and Acquisition*. New York, Mouton de Gruyter.

Bregman, A. S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA, MIT Press.

Browman, C. P. and Goldstein, L. (1986) Towards an articulatory phonology. *Phonology Yearbook* 3: 219-252.

Chomsky, N. and Halle, M. (1968) *The Sound Pattern of English*. New York, Harper and Row.

Cole, J. and K. Iskarous. (2001) Effects of vowel context on consonant place identification: implications for a theory of phonologization. In E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*, pp. 103-122. New York, Academic Press.

Diehl, R.L., & Kluender, K.R. (1989) On the objects of speech perception. *Ecological Psychology*, **1**(2), 121-144.

Disner, S. (1983) Vowel quality: the relation between universal and language-specific factors. *UCLA Working Papers in Phonetics* **58**, 1-158.

Edelman, G.M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*. New York, Basic Books.

Edelman, G.M. (2004) *Wider than the Sky: The Phenomenal Gift of Consciousness*. New Haven, Yale University Press.

Goldstein, L. M. & Fowler, C. (2003) Articulatory phonology: a phonology for public language use. In Schiller, N.O. & Meyer, A.S. (eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. pp. 159-207. New York, Mouton de Gruyter.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *J. Exp. Psychol. Learn Mem. Cogn.*, **14**(3), 421-433.

Hagiwara, R. (1995) Acoustic realizations of American /r/ as produced by women and men. *UCLA Working Papers in Phonetics* 90, pp. 1-187, 1995.

Johnson, K. (2007) Decisions and Mechanisms in Exemplar-based Phonology. In Sole, M.J., Beddor, P. & Ohala, M. (eds) *Experimental Approaches to Phonology. In Honor of John Ohala*. Oxford University Press. pp. 25-40.

Johnson, K. (2006) Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* **34**, 485-499.

Johnson, K. (2004) Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.) *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*. Tokyo, Japan: The National International Institute for Japanese Language. pp. 29-54.

Johnson, K., Ladefoged, P., & Lindau, M. (1993) Individual differences in vowel production. *Journal of* *the Acoustical Society of America* **94**(2), 701- 714.

Jusczyk, P. (1992), Developing phonological categories from the speech signal. In Ferguson, C., Menn, L., and Stoel-Gammon, C. (Eds.), *Phonological development: Models, research, implications*, York Press, Maryland, 17-64.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge: MIT Press.

Keating, P.A. (1990) The window model of coarticulation: articulatory evidence. In J. Kingston & M. Beckman (eds.) *Papers in Laboratory Phonology I* , Cambridge University Press, pp. 451-470.

Keating, P.A. (1988) Underspecification in phonetics. *Phonology* **5 (**2), 275-292.

Kuhl, P. Williams, K., Lacerda, F., Stevens, K. and Lindblom, B. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255**, 606-608.

Labov, Yeager & Steiner (1972) *Quantitative Study of Sound Change in Progress*.

Lacerda, F. (2003) Phonology: An emergent consequence of memory constraints and sensory input. *Reading and Writing* **16**(1-2), 41-59.

Ladefoged, P. (1984) ‘Out of chaos comes order’: Physiological, biological, and structural patterns in phonetics. In M.P.R. Van den Broeke and A. Cohen (eds.) *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dordrecht: Foris Publications (pp. 83-95.

Lane, H.L. and Tranel, B. (1971). “The Lombard sign and the role of hearing in speech,” *J. Speech Hear. Res.* **14**, 677-709.

Liberman, A.M. (1996) *Speech, A Special Code*. Cambridge, MA: MIT Press.

Liberman, A.M.; Harris, K.S.; Eimas, P.; Lisker, L. & Bastian, J. (1961) An effect of learning on speech perception: Discrimination of durations of silence with and without phonemic significance. *Language and Speech* **4** (4), 175-195.

Liberman, A.M., & Mattingly, I.G. (1985) The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C.A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development* (pp. 131- 163). Timonium, Maryland: York Press.

McGurk, H., & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature*, 264, 746–748.

Mielke, J., Baker, A., and Archangeli, D. (2007) Variability and homogeneity in American English /r/ allophony and /s/ retraction.

Mills, A. (1987) The development of phonology in the blind child. In Dodd, B. and Campbell (eds.) *Hearing by Eye: the Psychology of Lip-reading.* London: Lawrence Erlbaum Associates Ltd.

Nartey, J.N.A. (1982) On Fricative Phones and Phonemes: Measuring the Phonetic Differences Within and Between Languages. *UCLA Working Papers in Phonetics*, **55**.

Ohala, J. (1990) The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, pp. 258-275. Cambridge: CUP.

Palmer, J.S. (2007) The Role of Context in Lexical Recognition. *UC Berkeley Phonology Lab Annual Report*, 2007, 127-146.

Perkell, J.S., Guenther, F.H., Lane, H. ... (2000) A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics* **28**, 233-372.

Pierrehumbert, J. Beckman, M. E., & Ladd, D. R. (2000). Conceptual foundations of phonology as a laboratory science. In Noel Burton-Roberts, Phil Carr & Gerry Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Cambridge University Press.

Pitt, M.A., Johnson, K., Hume, E., Kiesling, S., Raymond, W. (2005) The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication* **45** (1), 89-95.

Plaut, D.C. and Kello, C.T. (1998) The Emergence of Phonology from the Interplay of Speech Comprehension and Production: A Distributed Connectionist Approach. In B. MacWhinney (ed.), *The Emergence of Language*, Erlbaum, Mahweh, NJ.

Remez, R. E. (2005). The perceptual organization of speech. In D. B. Pisoni and R. E. Remez (Eds.), *The Handbook of Speech Perception*, (pp. 28-50). Oxford: Blackwell.

Strand, E. (2000) *Gender stereotype effects in speech processing*. PhD Dissertation, Ohio State University.

Summers, W. Van; David B. Pisoni, Robert H. Bernacki, Robert I. Pedlow, and Michael A. Stokes (1988). “Effects of noise on speech production: Acoustic and perceptual analyses,” *J. Acoust. Soc. Am.* **84**(3), 917-928.

Westbury, J., Hashi, M., & Lindstom, M. (1998) Differences among speakers in lingual articulation for American English /r/. *Speech Communication* **26**, 203-226.

Zetterholm, E. (2000). An impersonator and some of his voices. A phonetic study of human voice flexibility. *Proceedings Fonetik 2000, The Swedish Phonetics Conference*: 145-148, Skövde, May 24-26 2000.