

Representations of speech sound patterns in the speaker's brain: Insights from perception studies¹

Noël Nguyen

Laboratoire Parole et Langage, Aix-Marseille Université & CNRS, Aix-en-Provence, France

1 Introduction

Over the last few years, considerable advances have been made in our understanding of the processes that allow listeners to perceive and extract meaning from speech. To a significant extent, these advances have been facilitated by the development of instrumental techniques that have been routinely available to researchers for some time now and are used in conjunction with more standard experimental procedures. For example, eye-tracking (e.g., Allopenna et al., 1998; Creel et al., 2008; Dahan et al., 2008) and the tracking of hand movements (Spivey et al., 2005) now make it possible to continuously follow the dynamics of speech processing as the speech signal unfolds over time. Likewise, event-related brain potentials studies (Dehaene-Lambertz, 1997; Molfese et al., 2005; Phillips, 2001) and brain-imaging studies (Scott, 2003; Scott and Johnsrude, 2003) provide crucial new insight into the cerebral underpinnings of speech perception and comprehension (see ***Idsardi and Poeppel***, this volume, for a review). Another major development concerns the fact that speech perception studies now go well beyond the processing of speech sounds produced by an individual speaker in a laboratory setting, and extend to spoken language in the context of its primary site of occurrence (Local, 2003), that is, social interaction. Fragments of conversational speech are used as

material in perceptual tests for example, and more generally, rigorously-controlled experimental designs have been combined in a variety of innovative ways with large-scale investigations of spontaneous speech data². Yet another important evolution relates to the increasingly large number of studies of the contribution to speech perception of visual articulatory information associated with movements of the speaker's face (Massaro, 1998). These advances in speech perception research have contributed to establishing new links with speech technology, among other disciplines, and to triggering the development of automatic speech recognition systems whose design partly mirrors the way in which speech is processed by human listeners (Moore, 2007; Scharenborg, 2007).

In extending beyond the limits of its traditional domain, speech perception research has shed new light on what has consistently constituted one of the most important issues for laboratory phonology: the way in which speech sound patterns are represented in the speaker/listener's brain. Indeed, conjectures have been made for quite a long time — well before the inception of laboratory phonology itself — about what such representations might be. In the following, a short overview will be presented of what recent speech perception studies tell us about the form and function of these mental representations for speech sound patterns.

2 Central issues

Speech sounds are highly variable, yet, listeners seem to extract meaning from speech effortlessly and successfully in most circumstances. To account for this seemingly paradoxical fact of our everyday experience, theories and models of speech perception

and comprehension have to deal with two central problems. The first problem relates to how speech sound patterns are represented in the listener's memory. The second problem concerns the way in which access to such representations is achieved by the listener from the input speech signal. I will refer to these as the *representational* and the *mapping* problem, respectively. The solutions offered by speech perception models to these two problems are inevitably intertwined (**Hawkins**, this volume): for example, the mapping mechanism is likely to take a different form if the sound shape associated with each word in memory is specified as a function of the context of occurrence for that word, as opposed to being context-independent.

According to an approach that long prevailed (see Klatt, 1989, for a historical overview), speech perception involves retrieving invariant properties relative to distinctive features and phonemes, independent of the variability shown by the corresponding speech sounds both within and between speakers, and it is in terms of these invariant properties that words are represented in the listener's mental lexicon. In this approach, a clear demarcation is posited between the surface phonetic form of a word and the underlying phonemic representation for that word. Variations in the production of speech sounds attributable to inter-individual anatomical differences are assumed to be factored out at an early stage of perceptual processing by means of a speaker normalization procedure.

It is now generally considered that this approach does not offer a satisfactory characterization of a number of major properties that have been found to coexist in the speech perception system. One of these properties is the remarkable robustness to alterations in words' surface shapes caused by various phonological processes. To explain

how words can be successfully recognized in spite of these alterations, researchers have moved away from the traditional approach, to propose either more sophisticated mapping mechanisms, alternative lexical representations, or both. Another no less remarkable property is the sensitivity shown by listeners to indexical information about the speaker's individual and social identity. Such a phenomenon appears to be largely inconsistent with phoneme-based models of speech perception, and a major challenge today is understanding how sensitivity to speaker indexical information may combine with perceptual robustness despite phonological variation. Other recent studies have raised yet more questions for the traditional view by suggesting that so-called fine phonetic detail can be perceptually relevant for listeners. These properties now attributed to the speech perception system are successively discussed in the next sections. We start with the perceptual processing of phonological variation.

3 Spoken word recognition in the face of phonological variation

Words can show substantial variations in their surface form under the influence of a variety of phonological phenomena such as assimilation or deletion (**Anttila**, this volume). A well-known example is provided by word-final coronals which take the place of articulation of a following labial or velar consonant in English³ among other languages (Nolan, 1992), as in *green boat* [gri:m bæʊt]. It has been a matter of much discussion how listeners can correctly identify words in which the final segment has undergone regressive place assimilation. Indeed, this phenomenon has come to form a key test for speech perception models whose domain of application extends much beyond the processing of

assimilation itself (see, among others, Darcy, 2003; Gaskell and Snoeren, 2008; **Lahiri**, this volume; Pitt, 2009; Ranbom and Connine, 2007).

In a series of studies (Eulitz and Lahiri, 2004; Friedrich et al., 2008; Lahiri and Marslen-Wilson, 1991, 1992; Lahiri and Reetz, 2002; see also Fitzpatrick and Wheeldon, 2000; Wheeldon and Waksler, 2004, as well as **Lahiri**, this volume), Lahiri and her colleagues have gathered both behavioral and brain-imaging data suggesting that assimilation does not have a disruptive effect on word recognition, and that listeners are in fact little or not sensitive to it. In the Featurally Underspecified Lexicon (FUL) model of word recognition proposed by Lahiri and colleagues, this is attributable to the fact that each word is associated in the mental lexicon with a highly abstract phonological representation, which is underspecified for certain features such as [coronal]. As a result, variations that a surface form may show with respect to these features do not prevent it from remaining consistent with the underlying phonological representation.⁴ According to the featural underspecification hypothesis, it is because of the way in which lexical representations are tailored that listeners are able to successfully recognize phonological variants of a given word, and this applies irrespective of the context in which these variants are encountered.

By contrast, Gaskell and his colleagues (e.g., Gaskell and Marslen-Wilson, 1996; Gaskell, 2003; Gaskell and Snoeren, 2008; Snoeren et al., 2009) have emphasized the role that context may play in the perceptual treatment of assimilation. Gaskell and colleagues' view is that listeners retrieve the underlying form of a word by means of an inference process that aims to reverse the effect that assimilation may have had on how

this word was produced. Studies conducted by the authors have revealed that this compensation-for-assimilation process is triggered in the context of a viable environment only, as opposed to an unviable one (in the case of word-final coronals in English, environmental viability refers to a subsequent word-initial labial or velar consonant). In contrast, in the context-independent and representational FUL approach, assimilatory variation is dealt with as the input word form is mapped onto the lexicon, the phonological inference account assumes that compensation for assimilation occurs at a prelexical level and may extend to novel words or nonwords.

Although assimilation is traditionally characterized as causing a categorical change in the value of a segmental feature, studies on the phonetic realization of segments that occur in a context appropriate for assimilation have found a variety of patterns from non-assimilated forms through cases of partial assimilation to fully assimilated forms (e.g., Browman and Goldstein, 1990; Ellis and Hardcastle, 2002). When assimilation is incomplete, cues to the underlying identity of the target segment are still available to listeners in the speech signal (e.g., see Snoeren et al., 2008; Surprenant and Goldstein, 1998). In yet a different model of the perception of assimilation, namely the feature cue-parsing model developed by Gow (2001, 2002, 2003), listeners are tuned to these fine-grained acoustic cues, which provide them with information about both the (partially) assimilated segment and the assimilating segment. There is evidence showing for example that when a word-final alveolar is assimilated to the following velar, as may be the case in *lead covered*, differences in F_1 and F_2 frequency extending throughout the preceding vowel and possibly even further towards the beginning of the word can be

found, relative to a /g/-final word such as *leg* in *leg covered* (Nolan, 1992; Local, 2003). Such differences may contribute to explaining why assimilated alveolars are perceptually recoverable from speech (Wright and Kerswill, 1989). This is consistent with recent work showing more generally that perceptually relevant cues to the identity of a given segment are spread over an interval that can extend well outside the segment's most prominent boundaries in the speech signal (e.g., Hawkins and Nguyen, 2004; West, 1999). Because these cues are both numerous and largely distributed in the temporal domain, the feature cue-parsing model contends that listeners are in most circumstances able to directly identify segments from speech, and that neither underspecified lexical representations nor phonological inference mechanisms are necessary for word recognition.

Work on the role of phonological variation in word recognition has recently turned to another potentially influential factor, namely the listener's degree of exposure to this variation. It is well known that lexical frequency has an important function in both production and comprehension, e.g. high-frequency words are recognized more rapidly than low-frequency words (see Jurafsky, 2003, for a review). There is now evidence that frequency effects apply to the phonological variants for a given word. For example, Connine (2004) showed that American English listeners are biased towards perceiving the word *pretty* (as opposed to the nonword *bretty*) to a greater extent when presented with a speech sound sequence that contains the more frequently produced voiced alveolar flap ([p.ɹɹi]), compared with the less frequent voiceless alveolar stop ([p.ɹti]) in intervocalic position. According to Connine and her colleagues (Connine, 2004; Connine et al., 2008; Connine and Pinnow, 2006; Ranbom and Connine, 2007), this is inconsistent with a

model of word recognition in which *pretty* has a single underlying phonological representation with a voiceless alveolar stop, which is recovered by the listener from the flapped variant. Rather, Connine and colleagues have argued that the main phonological variants of a word are jointly stored in the listener's mental lexicon and that each of these variants has a strength in memory that reflects the listener's frequency of exposure to that particular form. In Connine and colleagues' proposal, phonological variation is therefore directly encoded in the lexicon, in the form of a repertoire of alternative phonological representations for each word, contrary to the featural underspecification theory (in which the representation for each word is unique) as well as both the inference process and the feature cue-parsing theory (which assume that phonological variation is factored out at a prelexical stage of processing).

As we have seen, the models of speech comprehension reviewed above diverge from one another in quite a large measure with respect to the characterization of how phonological variation is dealt with by the listener. One major bone of contention relates to how much of this variation is incorporated into the lexicon, and how much is abstracted away during lexical access. In that respect, an opposition has arisen between the abstractionist viewpoint, as embodied by the FUL model for example, and the exemplar-based viewpoint, according to which each word is associated in the lexicon with a list of exemplars that each reflect a particular context-dependent realization for this word. Studies on indexical effects in speech perception, to which we now turn, have allowed further advances in this debate.

4 Indexical effects in speech perception

Speech contains a large variety of properties which relate to the speaker's physical, psychological, and social characteristics, and which have been referred to as indexical properties (Abercrombie, 1967). As listeners, we are highly sensitive to these properties. For example, we are able to recognize the correspondence that may exist between a specific phonetic pattern and the speaker's social category (Foulkes and Docherty, 2006; Labov, 1966; Stuart-Smith, 2007). Little or no attention was paid to indexical properties in traditional models of speech perception (§ 2), which assumed that listeners concentrate on the most prominent acoustic cues relative to phonemic contrasts as a first step towards extracting meaning from speech. The role of indexical properties in the recognition of the speaker's individual and social identity was undisputed, but these properties were assumed to be processed independently of phonemic cues. However, studies over the last twenty years have revealed that speech sounds are processed differently by listeners depending on the speaker's perceived gender (Johnson et al., 1999; Strand, 2000), age and social class (Hay et al., 2006b), idiolect (e.g., Norris et al., 2003), and dialect (Dahan et al., 2008; Evans and Iverson, 2004; Hay et al., 2006a; Niedzielski, 1999), and that indexical properties have a significant impact on spoken word recognition itself. These findings have contributed to reshaping our conceptions on the listener's mental representations for speech patterns.

An early demonstration of the effects of indexical properties on word recognition was provided by Mullenix et al. (1989). These authors showed that response speed and

accuracy in a word recognition task both decrease when listeners are presented with words produced by multiple speakers rather than a single speaker. The fact that it is more difficult for listeners to identify words originating from more than one speaker may be accounted for in a way consistent with traditional speech perception models, if one supposes that decreased performance in the multiple-speaker condition is due to the greater amount of computing resources consumed by a speaker normalization mechanism that comes into play prior to lexical access. However, an alternative explanation has been proposed that relies on the assumption that speaker-specific phonetic characteristics are encoded by listeners in long-term memory as spoken words are being processed (e.g., Bradlow et al., 1999; Pisoni, 1993). Empirical evidence from experimental studies has accumulated in support of this proposal. For example, Palmeri et al. (1993) found that it is easier for listeners to recognize that a word has already been presented to them (i.e. an “old” word as opposed to a “new” one) when both tokens of that word were produced by the same speaker rather than by different speakers. Likewise, Goldinger (1996) showed that prior exposure to a word facilitates later recognition of that word to a greater extent when the speaker is the same as opposed to different across the two repetitions. On this account, the Mullenix et al. (1989) effect is attributable to the fact that memory encoding of speaker-specific phonetic characteristics takes more time and resources when listeners are exposed to voices from a larger variety of speakers.

The finding that indexical properties come into play in word recognition has lent strong support to so-called exemplar-based models of speech perception and understanding (Coleman, 2002; Goldinger, 1996, 1998; Johnson, 1997). While these

models draw on a long-established line of research in cognitive psychology, their introduction into the field of speech perception is relatively new and they stand in stark contrast to the traditional phoneme-based approach. Exemplar models take the view that, for each encountered token of the word, an exemplar forms in the listener's memory that includes all the perceptual and contextual details specific to that token. These include sensory-motor, semantic and pragmatic characteristics, but also indexical information about the speaker's identity and the situation of occurrence, to mention but a few properties. Exemplars are therefore deeply anchored within their context of occurrence in the largest possible sense and this has drastic implications for how spoken language may be represented in the brain (Bybee and McClelland, 2005). In non-analytic models such as Johnson's (1997; 2005) XMOD, exemplars have no internal structure, and are conceived as unanalyzed auditory representations associated with whole words. More abstract representations associated with words' sound shapes are assumed to exist, but only as the result of a pattern formation process in the online processing of speech. In this view, "abstract phonological structure is a fleeting phenomenon—emerging and disappearing as words are recognized" (Johnson, 1997), as opposed to being more permanently stored in memory, as is assumed in abstractionist models.

Central to the exemplar approach is the assumption that indexical properties are integral to how words are represented in the mental lexicon, along with lexically-contrastive phonetic properties. More generally, talker-specific information and linguistic information are viewed as being processed in an integrated fashion by the listener (Nygaard, 2005). However, experimental evidence suggests that listeners are little

sensitive to at least some aspects of the talker's voice in spoken word recognition. For example, artificially-produced variations in words' overall acoustic amplitude do not affect response accuracy in a word recognition (Sommers et al., 1994) or word monitoring (Magnuson and Nusbaum, 2007) task, and overall amplitude does not seem to be retained in long-term memory as a perceptually-relevant aspect of the words' surface forms (Bradlow et al., 1999; Church and Schacter, 1994). Thus, listeners seem to employ a mechanism that filters out overall amplitude and potentially other acoustic characteristics prior to the long-term storage of surface forms. More generally, representations for words in memory must be, to a certain extent at least, abstract, since it is estimated that the auditory trace of speech fades away after about 400 ms (Pardo and Remez, 2006). Recent repetition-priming studies (Luce and McLennan, 2005; McLennan et al., 2003; McLennan and Luce, 2005) indicate that both abstract phonological representations and talker-specific exemplars may in fact coexist in memory and come into play at different stages in spoken word recognition. Specifically, early processing would be dominated by abstract representations, whereas talker-dependent information would exert an influence at a later stage of processing.

5 Beyond the lexicon: Perceptual relevance of fine phonetic detail

In recent years, research has increasingly focused on the listener's sensitivity to properties of the speech signal that are generically referred to as "fine phonetic detail" (FPD; see **Hawkins**, this volume). This research suggests that FPD has a significant impact on speech perception and understanding, at least in some circumstances. FPD includes

allophonic variation, sometimes specific to certain words or classes of words (Pierrehumbert, 2002). Fine phonetic detail is designated as such in the sense that it is to be distinguished from the local and most perceptually prominent cues associated with phonemic contrasts in the speech signal. Crucially, FPD “rarely, if ever, is a major contributor to distinguishing the citation form of lexical items from one another. In other words, FPD is subphonemic phonetic variability that contributes to phonological or other contrasts that distinguish meanings, but not necessarily lexical items” (Hawkins, 2010).

Recent studies on the role of FPD in spoken word recognition have provided evidence that perceptually-relevant allophonic variation includes vowel-consonant acoustic transitions (e.g., Marslen-Wilson and Warren, 1994), within-category variations in voice onset time (Allen and Miller, 2004; Andruski et al., 1994; Ju and Luce, 2006; McMurray et al., 2009), long-domain resonance effects associated with liquids (West, 1999), and graded assimilation of place of articulation in word-final coronals (e.g., Gaskell, 2003).⁵ To a certain extent, however, the fact that listeners are sensitive to allophonic variation was established much earlier. For example, studies conducted in the 1970s and 1980s consistently showed that coarticulation between neighboring segments provides listeners with perceptually-relevant cues to segment identity (and by extension to word recognition). A well-known example is regressive vowel-to-vowel coarticulation in English, which allows the identity of the second vowel to be partly predictable from the acoustic cues associated with it in the first vowel (Martin and Bunnell, 1982). Between these early studies and more recent research on fine phonetic detail in speech perception, however, a change in the focus of interest has occurred. Whereas the former centered on

the role of coarticulation in phoneme identification, the latter extends the potential influence of fine phonetic detail to higher levels of processing, ranging from lexical access and syntactic parsing to the interpretation of conversational interaction. Central to this line of research is the assumption that information contained in fine phonetic detail can percolate up to the lexical level and above, contrary to an approach to speech perception in which access to meaning from the speech signal is accomplished through the mediation of a sequence of abstract infra-lexical units from which fine phonetic detail is left out.

FPD-oriented research takes the view that the central place typically attributed to the lexicon in theories of speech understanding has led to an overemphasis on short-domain phonetic properties related to phonemic contrasts. It is argued that shifting the focus away from the lexicon allows us to see fine phonetic properties associated with phonological contrasts that are spread over long temporal windows, and/or which perform functions other than lexical differentiation. Thus, the phonetic exponents of phonological contrasts (including fine phonetic detail) have been shown to differ in function words as opposed to content words, a phenomenon attributed to the fact that function words and content words form two different systems of contrastivity, with a restricted inventory and less variation in phonological structure for function words (Local, 2003). Likewise, it is now established that fine phonetic detail is related in systematic ways to the time course of conversational interaction, and in particular to patterns of turn-taking and cooperative exchange of information in a conversation (e.g., Local, 2003; Ogden, 2006; Plug, 2005; Szczepek-Reed, 2006).

The phonetics of conversational interaction (e.g., Couper-Kuhlen and Ford, 2004)

is an area in which evidence for the role of FPD in speech perception is growing. In the course of such an interaction, the behavior of each talker can evolve with respect to that of the other talker in two opposite directions: it may become more similar to the other talker's behavior (a phenomenon referred to as convergence, or imitation) or more dissimilar. Convergence effects have been shown to be systematic and recurrent, and manifest themselves under many different forms, which include posture (e.g., Shockley et al., 2003), head movements and facial expressions (e.g., Estow et al., 2007; Sato and Yoshikawa, 2007) and, as regards speech, vocal intensity (Natale, 1975), pitch curve (Bosshardt et al., 1997; Gregory et al., 1993) and rate of speech (Giles et al., 1991). These phenomena may facilitate conversational exchange by contributing to setting a common ground between speakers (Giles et al., 1991). Recent studies (e.g., Pardo, 2006) have indeed shown that perceived similarity in pronunciation between talkers increases over the course of the interaction and persists beyond its conclusion. Conversational interaction therefore seems to have long-term effects on the pronounced form of words, and this may be taken as indicating that words are stored in memory in a form that is highly dependent on their context of occurrence. More specifically, it now appears that the representations associated with words in the mental lexicon for each talker may dynamically evolve during conversation under the influence of the other talker's speech patterns, and retain the traces of that influence once the conversation has ended.

6 Conclusion: Towards hybrid models of speech perception and understanding

Experimental evidence is now available that provides support for the role of both detailed phonetic characteristics and abstract phonological categories in speech perception. This has opened the way towards the development of new models of speech perception and understanding that aim to bridge the gap between the exemplar-based and abstractionist approaches. For example, Tuller and her colleagues (Tuller et al., 1994; Tuller, 2004; see also Nguyen et al., 2009) have proposed a model that uses concepts from the theory of nonlinear dynamical systems to account for the mechanisms involved in the categorization of speech sounds, and according to which there are two complementary aspects to speech perception. On the one hand, speech perception is assumed to be a highly context-dependent process sensitive to the detailed acoustic structure of the speech input. On the other hand, it is viewed as a non-linear dynamical system characterized by a limited number of stable states, or attractors, which allow the system to perform a discretization of perceptual space and which are associated with abstract perceptual categories. The recent development of so-called hybrid models (Hawkins, 2003, 2010; Luce and McLennan, 2005; McLennan and Luce, 2005; Pierrehumbert, 2006) is also governed by the assumption that detailed phonetic properties and abstract phonological categories combine with each other in the representations associated with words in memory.

Notes

¹Thanks are due to Cheryl Frenck-Mestre, Pauline Welby, one anonymous reviewer, and the Editors, for helpful comments on an earlier version of this section.

²In a different but complementary perspective, **Holt**, this chapter, shows how the growing focus on ecological validity leads studies using laboratory speech to be increasingly combined with work on more natural speech communication.

³Recent work by Dilley and Pitt (2007), however, showed that regressive place assimilation in word-final coronals in conversational speech in English is less frequent than has been previously assumed.

⁴For example, a so-called no-mismatch relationship will be said to exist between the surface form [g.i:m] and the phonological representation for *green*, because the coronal place of articulation of the final consonant in *green* is assumed to be underspecified in the lexicon. Thus, listeners are expected to interpret both [g.i:m] and [g.i:n] as referring to the word *green*.

⁵The studies cited here were conducted on either American or British English.

References

- Abercrombie, D. (1967). *Elements of General Phonetics*. Aldine, Chicago, IL, USA.
- Allen, J. and Miller, J. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 115:3171–3183.
- Allopenna, P., Magnuson, J., and Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38:419–439.
- Andruski, J., Blumstein, S., and Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52:163–187.
- Bosshardt, H.-G., Sappok, C., Knipschild, M., and Hölscher, C. (1997). Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers. *Journal of Psycholinguistic Research*, 26:425–448.
- Bradlow, A., Nygaard, L., and Pisoni, D. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, 61:206–219.
- Browman, C. and Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In Kingston, J. and Beckman, M., editors, *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, pages 341–376. Cambridge University Press, Cambridge, UK.

Bybee, J. and McClelland, J. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22:381–410.

Church, B. and Schacter, D. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:521–533.

Coleman, J. (2002). Phonetic representations in the mental lexicon. In Durand, J. and Laks, B., editors, *Phonetics, Phonology, and Cognition*, pages 96–130. Oxford University Press, Oxford, UK.

Connine, C. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, 11:1084–1089.

Connine, C. and Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review*, 23:235–245.

Connine, C. M., Ranbom, L. J., and Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, 70(3):403–411.

Couper-Kuhlen, E. and Ford, C., editors (2004). *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*. John Benjamins, Amsterdam, The

Netherlands.

Creel, S., Aslin, R., and Tanenhaus, M. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 106(2):633–664.

Dahan, D., Drucker, S., and Scarborough, R. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108(3):710–718.

Darcy, I. (2003). *Assimilation phonologique et reconnaissance des mots*. PhD thesis, École des hautes études en sciences sociales, Paris, France.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, 8:919–924.

Dilley, L. and Pitt, M. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America*, 122:2340–2353.

Ellis, L. and Hardcastle, W. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. *Journal of Phonetics*, 30:373–396.

Estow, S., Jamieson, J., and Yates, J. (2007). Self-monitoring and mimicry of positive and negative social behaviors. *Journal of Research in Personality*, 41:425–433.

Eulitz, C. and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive*

Neuroscience, 16:577–583.

Evans, B. and Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*, 115:352–361.

Fitzpatrick, J. and Wheeldon, L. (2000). Phonology and phonetics in psycholinguistics models of speech perception. In Burton-Roberts, N., Carr, P., and Docherty, G., editors, *Phonological Knowledge: Conceptual and Empirical Issues*, pages 131–160. Oxford University Press, Oxford, UK.

Foulkes, P. and Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34:409–438.

Friedrich, C. K., Lahiri, A., and Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6):1545–1559.

Gaskell, M. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31:447–463.

Gaskell, M. and Marslen-Wilson, W. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22:144–158.

Gaskell, M. G. and Snoeren, N. D. (2008). The impact of strong assimilation on the

perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6):1632–1647.

Giles, H., Coupland, N., and Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In Giles, H., Coupland, N., and Coupland, J., editors, *Contexts of Accommodation: Developments in Applied Sociolinguistics*, pages 1–68. Cambridge University Press, Cambridge, UK.

Goldinger, S. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22:1166–1183.

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105:251–279.

Gow, D. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45:133–159.

Gow, D. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28(1):163–179.

Gow, D. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65:575–590.

Gregory, S., Webster, S., and Huang, G. (1993). Voice pitch and amplitude convergence

as a metric of quality in dyadic interviews. *Language and Communication*, 13:195–217.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31:373–405.

Hawkins, S. (2010). Phonetic variation as communicative system: Perception of the particular and the abstract. In Fougeron, C., D’Imperio, M., Kühnert, B., and Vallée, N., editors, *Papers in Laboratory Phonology X*. Mouton de Gruyter, Berlin. to appear.

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

Hay, J., Nolan, A., and Drager, K. (2006a). From *fish* to *feesh*: Exemplar priming in speech perception. *The Linguistic Review*, 23:351—379.

Hay, J., Warren, P., and Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34(4):458—484.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics*, 50:101–113.

Johnson, K. (2005). Decisions and mechanisms in exemplar-based phonology. *UC Berkeley Phonology Lab Annual Report*, pages 289–311.

Johnson, K., Strand, E., and D’Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4):359–384.

Ju, M. and Luce, P. A. (2006). Representational specificity of within-category phonetic variation in the long-term mental lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1):120–138.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In Bod, R., Hay, J., and Jannedy, S., editors, *Probabilistic Linguistics*, pages 39–95. MIT Press, Cambridge, MA, USA.

Klatt, D. (1989). Review of selected models of speech perception. In Marslen-Wilson, W., editor, *Lexical Representation and Process*, pages 169–226. MIT Press, Cambridge, MA, USA.

Labov, W. (1966). *The Social Stratification of English in New York City*. Center for Applied Linguistics, Washington D.C., USA.

Lahiri, A. and Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38:245–294.

Lahiri, A. and Marslen-Wilson, W. (1992). Lexical processing and phonological representation. In Docherty, G. and Ladd, D., editors, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pages 229–254. Cambridge University Press, Cambridge, UK.

Lahiri, A. and Reetz, H. (2002). Underspecified recognition. In Gussenhoven, C. and Warner, N., editors, *Papers in Laboratory Phonology VII*, pages 637–675. Mouton de

Gruyter, Berlin, Germany.

Local, J. (2003). Variable domains and variable relevance: Interpreting phonetic exponents. *Journal of Phonetics*, 31:321–339.

Luce, P. and McLennan, C. (2005). Spoken word recognition: The challenge of variation. In Pisoni, D. and Remez, R., editors, *The Handbook of Speech Perception*, pages 591–609. Blackwell, Oxford, UK.

Magnuson, J. S. and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2):391–409.

Marslen-Wilson, W. and Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101:653–675.

Martin, J. and Bunnell, H. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 8:473–488.

Massaro, D. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. MIT Press, Cambridge, MA, USA.

McLennan, C. and Luce, P. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning*,

Memory and Cognition, 31:306–321.

McLennan, C., Luce, P., and Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29:539–553.

McMurray, B., Tanenhaus, M., and Aslin, R. (2009). Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, 60(1):65–91.

Molfese, D., Fonaryova Key, A., Maguire, M., Dove, G., and Molfese, V. (2005). Event-related evoked potentials (ERPs) in speech perception. In Pisoni, D. and Remez, R., editors, *The Handbook of Speech Perception*, pages 99–120. Blackwell, Malden, MA, USA.

Moore, R. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication*, 49:418–435.

Mullenix, J., Pisoni, D., and Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85:365–378.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32:790–804.

Nguyen, N., Wauquier, S., and Tuller, B. (2009). The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories. In Pellegrino, F., Marsico, E., Chitoran, I., and Coupé, C., editors, *Approaches to Phonological*

Complexity. Mouton de Gruyter, Berlin, Germany. in press.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18:62–85.

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In Docherty, G. and Ladd, D., editors, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pages 261–280. Cambridge University Press, Cambridge, UK.

Norris, D., McQueen, J., and Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2):204–238.

Nygaard, L. (2005). Perceptual integration of linguistic and nonlinguistic properties of speech. In Pisoni, D. and Remez, R., editors, *Handbook of Speech Perception*, pages 390–414. Blackwell, Oxford, UK.

Ogden, R. (2006). Phonetics and social action in agreements and disagreements. *Journal of Pragmatics*, 38:1752–1775.

Palmeri, T., Goldinger, S., and Pisoni, D. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19:309–328.

Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119:2382–2393.

Pardo, J. S. and Remez, R. E. (2006). The perception of speech. In Traxler, M. and

Gernsbacher, M., editors, *The Handbook of Psycholinguistics, Second Edition*, pages 201–248. Elsevier, Cambridge, MA, USA.

Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science*, 25:711–731.

Pierrehumbert, J. (2002). Word-specific phonetics. In Gussenhoven, C. and Warner, N., editors, *Papers in Laboratory Phonology VII*, pages 101–140. Mouton de Gruyter, Berlin, Germany.

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, 34:516–530.

Pisoni, D. (1993). Long-term memory in speech perception — Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, 13:109–125.

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language*, 61(1):19–36.

Plug, L. (2005). From words to actions: The phonetics of *Eigenlijk* in two communicative contexts. *Phonetica*, 62:131–145.

Ranbom, L. and Connine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57:273–298.

Sato, W. and Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, 104:1–18.

- Scharenborg, O. (2007). Reaching over the gap: A review of efforts to link human and automatic speech recognition research. *Speech Communication*, 49:336–347.
- Scott, S. (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication*, 41:23–34.
- Scott, S. and Johnsrude, I. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26:100–107.
- Shockley, K., Santana, M.-V., and Fowler, C. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29:326–332.
- Snoeren, N. D., Gaskell, M. G., and Betta, A. M. D. (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2):542–549.
- Snoeren, N. D., Segui, J., and Hallé, P. A. (2008). Perceptual processing of partially and fully assimilated words in French. *Journal of Experimental Psychology: Human Perception and Performance*, 34(1):193–204.
- Sommers, M., Nygaard, L., and Pisoni, D. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, 96:1314–1324.
- Spivey, M., Grosjean, M., and Knoblich, G. (2005). Continuous attraction toward

phonological competitors. *Proceedings of the National Academy of Sciences*, 102(29):10393–10398.

Strand, E. (2000). *Gender Stereotype Effects in Speech Processing*. PhD thesis, The Ohio State University, Columbus, OH, USA.

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in Glaswegian. In Cole, J. and Hualde, I., editors, *Papers in Laboratory Phonology 9*, pages 65–86. Mouton de Gruyter, Berlin, Germany.

Surprenant, A. and Goldstein, L. (1998). The perception of speech gestures. *Journal of the Acoustical Society of America*, 104:518–529.

Szczepek-Reed, B. (2006). *Prosodic Orientation in English Conversation*. Palgrave Macmillan, Houndmills, Basingstoke, UK.

Tuller, B. (2004). Categorization and learning in speech perception as dynamical processes. In Riley, M. and Van Orden, G., editors, *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*. National Science Foundation. URL: www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp.

Tuller, B., Case, P., Ding, M., and Kelso, J. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 20:3–16.

West, P. (1999). Perception of distributed coarticulatory properties in English /l/ and

/ɹ/. *Journal of Phonetics*, 27:405–426.

Wheeldon, L. and Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language*, 90(1-3):401–412.

Wright, S. and Kerswill, P. (1989). Electropalatography in the analysis of connected speech processes. *Clinical Linguistics & Phonetics*, 3:49–57.