

DAVID B. PISONI & ROBERT E. REMEZ (eds.), *The handbook of speech perception*. Oxford: Blackwell, 2005. Pp. xi + 708. ISBN 0-631-22927-2 (hbk). doi:10.1017/S002510030800371X

Reviewed by **Ruben van de Vijver**

Universität Potsdam  
ruben@ling.uni-potsdam.de

Speech perception deals with the questions of how speech is perceived, how perception helps building up mental representations and how these representations help in processes such as spoken word recognition. *The handbook of speech perception* is another useful book in Blackwell's Handbook series. The editors, David B. Pisoni and Robert E. Remez, have done a good job of bringing together authors from the field of speech perception. The book provides us with a review of what the author thinks are the most pressing issues.

The handbook consists of six parts. The first part consists of five papers, all of which discuss seminal issues in speech perception. In the first paper, James R. Sawusch considers speech analysis as an aid to the study of speech perception, focusing on the need to ensure that speech analysis techniques are accurate models of human speech perception. Analysis informs synthesis, where more research needs to be done to convincingly synthesize the voices of women and children. In the next contribution, Remez asks how speech perception is really done. Speech is mostly studied without distortion or background noise, with controlled intelligibility and typically with only a single talker, conditions which are rare in real life. He concludes that the auditory correlates of speech are perceptually resolved into a coherent stream which is then analysed for their linguistic and indexical properties.

Next, Lawrence D. Rosenblum maintains that speech perception is multimodal, i.e. depends on visual and auditory input, and perhaps even haptic input. He cites four pieces of evidence: (i) visual speech perception helps hearing impaired listeners, children and adults under difficult hearing circumstances; (ii) multimodal information is integrated very early and true unimodal speech perception is rare; (iii) neurophysiological research finds that there are speech mechanisms that are unconcerned with modality; and (iv) speech is modality-neutral in the sense that its various sources – visual and auditory – are never separated. In the fourth chapter, Lynne E. Bernstein discusses whether there are cortical areas dedicated to speech and whether audiovisual speech processing relies on neuronal convergence of phonetic information. She reports that not all cortical areas that process speech are specialized for speech and that separate auditory and visual processing precedes the convergence of auditory and visual speech representations. The concluding chapter of this section, by Dennis L. Molfese, Alexandra P. Fonaryova Key, Mandy J. Maguire, Guy O. Dove & Victoria J. Molfese deals with Event Related Potentials (ERPs) (a time-locked portion of the ongoing electroencephalogram) and how they can be used to study speech perception. By comparing ERPs of different stimulus material, researchers can gain insight as to what happens in the brain and how these events are related temporally. They have used this technique to study the time course of speech perception. By comparing the ERPs of different groups of children, it is also possible to identify which are at risk of diminished language performance.

It is interesting to see how some of the conclusions of the chapters in this section unintentionally corroborate one another. Remez maintains that the auditory signal must first be organized perceptually before it can be analyzed as speech and the neurolinguistic chapters, and Bernstein concludes that the first step in speech perception is a general auditory analysis. There are, however, also contradicting conclusions. Rosenblum speculates that the auditory signal and the visual signal may not be separated from the beginning.

The seven chapters in the second section discuss how perception shapes linguistic representations and how linguistic representations shape perception. The first chapter

(Kenneth N. Stevens) proposes a six-step model of how humans extract words from running speech. In the first step, a general auditory analysis is performed. Speech is separated from non-speech (cf. Bernstein, above). Then landmarks which give clues to sound class are extracted, followed by a more detailed analysis. This information is used to estimate the distinctive feature bundles and syllable structure, which, in the next step, is used to access the lexicon. The hypothesized word is verified by a process of analysis by synthesis. This model explicitly states that there is no place for non-linguistic information (talker identity, speaking rate, emotional state of the speaker) in a lexical item. An entry in the lexicon consists of a string of segments and each segment is characterized by a bundle of binary distinctive features. In the next chapter, Edward Flemming concludes that speech perception influences the shape of phonological patterns, and that it does so by establishing constraints on the distinctiveness of contrasts. Lawrence Raphael's chapter then reviews the acoustic cues to the perception individual speech sounds. Researchers have moved away from the idea that a percept is caused by a single, invariant cue. Percepts are characterized by constellations of features which are analyzed statistically to influence listener responses. The following chapter by Rosalie Uchanski deals with clear speech: the speaking style a speaker adopts when he is speaking in difficult circumstances. It is unclear which characteristics of clear speech are especially helpful, but Uchanski concludes that it is nevertheless conducive to better understanding. More research is needed on the difference between clear speech and normal speech and what the consequences are of this difference for perception. In the fifth chapter of this section Jacqueline Vassière argues that it is very hard to study intonation since it is loaded with information, but progress has been made towards understanding the syntactic, informational, interactive, modal, attitudinal, emotional and indexal aspects of intonation. Anne Cutler reminds us in her chapter that lexical stress contributes to word recognition by constraining the candidate set, even though the extent to which this happens is not the same in all languages. What is needed is more laboratory research on the role of stress in lexical stress languages that are not related to Germanic or Romance languages. Next, Zinny Bond says that slips of the ear are far from random mistakes, instead showing that linguistic knowledge has a pervasive influence on what the listener thinks the speaker said. The last chapter (again by Kenneth Stevens) provides evidence for analysis by synthesis at the word level and 'above'.

The four chapters in the next section deal with the perception of individual features of speech. In the first, Cynthia G. Clopper & David B. Pisoni point out that we do not fully understand how listeners perceive and encode dialect variation but that the more we understand about this, the better our model of spoken language processing will be. One important question that is not mentioned is whether dialect variation is part of a lexical representation. This would be a natural extension of models of spoken word recognition in which items are instances of experienced words. In the next chapter, Jody Kreiman, Diana Vanlancker-Sidtis & Bruce R. Gerratt show that voice quality may reflect sarcasm and irony, and may thus cause the meaning of what one is saying to be the opposite of what it appears to mean, thus performing a linguistic function. Even so, it is not obvious what should be measured when measuring vocal quality. Keith Johnson next asks how it is possible to effortlessly extract a message out of a signal which includes much inter- and intra-speaker variation. Most answers involve some kind of normalization: the signal is stripped of any unnecessary variation and what remains is an abstract, invariant signal. An alternative to all of these methods is to say that the variance is part of the representation. In the last chapter of this part, Lynne C. Nygaard discusses the relationship between linguistic and non-linguistic information in spoken language processing. Non-linguistic information influences language processing, but the effect may be different from the role of linguistic information. Nygaard asks how the delicate balance of these factors used by humans can be modelled.

The seven chapters of the fourth part focus on listeners whose experience is different from the experience of monolingual adults. Derek M. Houston deals with speech perception in infants. Infants have surprising speech perception abilities and our knowledge about these abilities provides us with new knowledge of how they support word learning. Even

though young children can distinguish very fine-grained differences, this ability seems to disappear once they begin learning words. Another focus of the chapter is speech perception in deaf children who receive cochlear implants. Research on these infants can inform us about the importance of early sensory experience. In the next chapter, speech perception in childhood is discussed by Amanda C. Walley. This area is not as thoroughly researched as speech perception in infants, but as it involves the continuing growth and refinement of the vocabulary, the changing interaction between phonetic and lexical levels of processing, developing phonological awareness and the emergence of literacy, it is obviously an important domain of study. There are strong links between phonological awareness and beginning reading ability. Moreover, it appears to be the case that words that children are familiar with (early acquisition or high frequency) are better recognized in gating and word repetition tasks when there are few similar words to choose from. More difficulty in recognition and gating has been found for tasks involving phonological awareness and similarity classification. Mitchell S. Sommers next says that the single most influential factor to explain deterioration of spoken word recognition over time is age-related reduction of auditory sensitivity. In the fourth chapter in this section, David Pisoni discusses speech perception in children with cochlear implants. Children who have received a cochlear implant can be roughly divided into 'successful' and 'unsuccessful'. The difference between these groups is that children who do well were previously immersed in an oral-only environment and the others were immersed in an environment where spoken language was supported by signs. Another group of special listeners are people with focal brain injury, discussed by William Badecker. Their lesion site together with the pattern of the deficit may give us an indication of where the particular deficit is located in the brain. The various patterns of deficits suggest that the auditory route to the lexical meaning includes several processing stages. Núria Sebastián-Gallés next discusses cross-language speech perception. Speech illusions are quite common when one listens to a foreign language. Sometimes, a difference cannot be heard, or something is heard that isn't there, or a listener changes one sound to another. It has been shown that a range of factors, segmental and also suprasegmental, causes these illusions. The final chapter in this section, by Susan Ellis Weismer, deals with speech perception in specific language impairment. Children with specific language impairment (SLI) display a significant language disorder without a clear pathological cause. It has become clear that children with SLI do not form a homogenous group, leading to a hypothesis that there are various subgroups with each characterized by their own underlying factors. Children with SLI seem to exhibit deficits in nonword repetitions, usually taken as a measure for phonological working memory. The nature of lexical representations presents several quandaries. The first is that while children are very sensitive to fine distinctions in speech, they still seem to have great trouble distinguishing a word that is a minimally different variant of a word they have just learned. An explanation for this would necessarily assume a certain degree of abstract analysis, I think. Another mystery concerns the illusions we experience when listening to a foreign language. Such illusions imply, to my mind, (mis)analyzing the signal and not just categorization upon retrieval. It shows that we need to learn more about what part of the experienced instance is stored, how this is learned and how experience shapes the analysis of the signal and retrieval of lexical items.

The fifth part of the *Handbook* consists of two chapters. In the first, Paul Luce & Conor McLennan suggest that indexical and allophonic variation present challenges for current models of perception, since it is clear that these two factors play a role in the behavior of subjects in all kinds of psycholinguistic tests. The next chapter, by Edward Auer & Paul Luce, deals with the role of probabilistic phonotactics in spoken word recognition. The authors suggest that frequently-occurring units are favored during recognition and that active units compete with one another. It seems to me that addressing the issues that have been put forward in these two chapters will influence the theory of lexical selection.

The final part of the book consists of two chapters, the first, by Carol Fowler & Bruno Galantucci, defending the point of view that what is perceived are gestures, coordinated

actions of two or more articulators. The evidence for this claim is weak in the speech domain, but becomes stronger when the pervasiveness of perception–production links in many other domains of cognition are considered. The second chapter, by Timothy Gentner & Gregory Ball, makes the point that while language is special, and while the existence of shared neural structures hardly makes a good case for comparing human speech and birdsong, we have already learned a lot about human speech by studying songbirds. Songbirds parse a song into motifs that capture the relevant variation in a song and they represent these dynamically, perhaps comparable to the way in which representations of articulatory gestures that underlie single speech sounds are represented by humans.

The editors state in the introduction to the *Handbook* that each chapter should not only summarize the state of the art but should also indicate issues that need more research. I think they have met this goal. The literature on speech perception is vast and it would be easy to get lost in it. Yet all chapters are well focused. Both the authors and the editors deserve praise for having achieved this. The book has shown the rise of theories of lexical items as experienced instances of words. These theories address quandaries that have occupied researchers for a long time, such as how children extract units from the speech stream and how the form of the input (child-directed speech as opposed to adult-directed speech) helps or hinders the child. The *Handbook of speech perception* offers a wealth of data, theories and ideas to apply to this question and others.

MARK TATHAM & KATHERINE MORTON, *Speech production and perception*. New York: Palgrave Macmillan, 2006. Pp. xviii + 326. ISBN: 1403917337 (hbk).

doi:10.1017/S0025100308003721

Reviewed by **Stefan A. Frisch & Sylvie M. Wodzinski**

Department of Communication Sciences and Disorders

University of South Florida

*frisch@cas.usf.edu*

In the conclusion to this book, the authors state: ‘Our objective has been to identify key areas of speech production and perception where we believe theoretical work needs to concentrate’ (p. 299). The book certainly accomplishes this objective. The foremost area identified by the authors, Mark Tatham & Katherine Morton (T&M), is in the representation of expressive content in speech. T&M also highlight a lack of good integration between work on prosody and work on segmental phenomena, such as coarticulation. There are two other domains where the book makes a significant contribution. One is in the definition of terms and explanation of the historical development of phonetic theory. The other is in the description of the TM model of speech production and perception, which provides some inroads into tackling the areas of theoretical development for representing expressive content in descriptions of speech and in theoretical development of the representation of language sound structure in the mind.

The first chapter provides a very nice overview of the perspective of classical phonetics and introduces phonetic terminology. The reader is also immediately confronted with one of the strengths and one of the weaknesses of the book. T&M are very thorough and explicit in elucidating terminology, and perhaps the biggest strength of the book is in the definition of terms throughout. For the reader who takes the time to carefully examine what T&M say and how they say it, there is much insight to be gained into theoretical issues in phonetic science. There are numerous books on phonetics, but in most cases the focus of these is on data and phonetic description. The discussion of issues and presentation of terminology in