

Lori L. Holt • Jonathan E. Peelle
Allison B. Coffin • Arthur N. Popper
Richard R. Fay
Editors

Speech Perception



Editors

Lori L. Holt
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA, USA

Jonathan E. Peelle
Department of Otolaryngology
Washington University in St. Louis
Saint Louis, MO, USA

Allison B. Coffin
Integrative Physiology and Neuroscience
Washington State University
Vancouver, WA, USA

Arthur N. Popper
Department of Biology
University of Maryland
Silver Spring, MD, USA

Richard R. Fay
Department of Psychology
Loyola University Chicago
Chicago, IL, USA

ISSN 0947-2657 ISSN 2197-1897 (electronic)
Springer Handbook of Auditory Research
ISBN 978-3-030-81541-7 ISBN 978-3-030-81542-4 (eBook)
<https://doi.org/10.1007/978-3-030-81542-4>

© Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Chapter 1

The Auditory Cognitive Neuroscience of Speech Perception in Context



Lori L. Holt and Jonathan E. Peelle

Abstract Speech is undeniably significant as a conspecific human communication signal, and it is also perhaps the most ubiquitous class of acoustic signals encountered by the human auditory system. However, historically there was little integration between speech research and the field of auditory neuroscience. Much of this divide can be traced back to the Motor Theory of speech perception, which framed speech not as an auditory process but as one grounded in motor gestures. Recent decades have seen a marked shift in perspective, with mutual interest from researchers in understanding both how neuroscientific principles can be used to study speech perception and, conversely, how speech as a complex acoustic stimulus can advance auditory neuroscience. This introductory chapter reviews this historical context for the modern field of auditory cognitive neuroscience before placing the remaining chapters of the book in context. A number of important themes emerge: methodological improvements, particularly in human brain imaging; the ability to study more natural speech (stories and conversations, rather than isolated stimuli); an appreciation for ways in which different listeners (e.g., of different ages or hearing levels) perceive speech; and incorporation of regions outside traditional auditory and language networks into our neuroanatomical frameworks for speech perception. Evolving techniques, theories, and approaches have provided unprecedented progress in understanding speech perception. These opportunities challenge researchers to ask new questions and to fully integrate speech perception into auditory neuroscience.

Keywords Hearing · Brain · Cognition · Language · Neuroimaging · Neurosciences · Motor Theory

L. L. Holt (✉)

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: loriholt@cmu.edu

J. E. Peelle

Department of Otolaryngology, Washington University in St. Louis, St. Louis, MO, USA
e-mail: jpeelle@wustl.edu

1.1 Speech Perception Research: A Historical Perspective

In many circumstances, speech is crucial to guiding human behavior. The focus of the current book is on auditory speech: acoustically complex changes in sound waves uttered by a talker with the intention of conveying information to a listener. Whether catching up on a favorite television series, taking in the news on public radio, chatting with a friend at a café, or listening to a colleague describe her latest idea, we consume a daily perceptual diet of acoustically complex speech that originates from diverse talkers and blends with distinct acoustic backgrounds. A classic and oft-cited analysis attributes 70–80% of the workday to communication, with about 55% of this time devoted to speech listening (Klemmer and Snyder 1972). Even our own voice provides us with rich input; systematic recordings of natural conversations indicate that we utter an average of about 16,000 words each day (Mehl et al. 2007). Although changes in technology and culture over the years may affect the specifics of these estimates, speech is undeniably significant as a conspecific human communication signal, and it is also perhaps the most ubiquitous class of acoustic signals encountered by the human auditory system. It may seem surprising through a modern perspective, then, that interdisciplinary efforts linking auditory neuroscience and speech perception were not always appreciated. At least some of the reasons for this historical divide can be traced back to the progression of early theoretical ideas about speech perception.

When researchers began investigating speech perception in earnest in the 1950s and 1960s (reviewed in Diehl et al. 2004; Samuel 2011), their landmark research resulted in the discovery of a list of perceptual phenomena that appeared to be present for speech perception, but not for perception of other auditory signals (Cooper et al. 1951; Liberman 1957). This work provided the foundation for what is known about how acoustic cues map to linguistic units like phonemes and revealed the complexity of this relationship (Peterson and Barney 1952; Delattre et al. 1955). Evidence emerged that acoustic information relevant to perceiving phonemes like those that differentiate *bear* from *pear* was categorical and context-dependent – not invariant – and further that it smeared across adjacent phonemes; speech was not as simple as an acoustic alphabet (Fowler 2001). A theory took shape from these observations that had an incredibly strong influence on the course of research spanning many decades.

1.1.1 Motor Theory

Alvin Liberman and his colleagues at the Haskins Laboratories became convinced that perceived phonemes and features have a more nearly one-to-one relationship to articulation than to speech acoustics, and this gave rise to the *Motor Theory* of speech perception (Liberman 1957; Liberman et al. 1967). This Motor Theory took as a first principle that speech signals, by virtue of being human vocalizations

providing entry to language, engage human-specific processing entirely distinct from processing other sounds (Liberman et al. 1967; Liberman and Mattingly 1985). In the strongest form of Motor Theory, speech was posited to be perceived as a motor object, not an acoustic one. Specifically, the objects of speech perception were proposed to be the intended phonetic articulatory gestures of the speaker represented as the invariant motor commands that called upon articulator movements to speak. Thus, the Motor Theory imagined the invariant motor commands to be a common currency linking speaking and listening. Crucially, the theory argued that this perceptual-motor relationship did not emerge as a learned association by virtue of having been both a speaker and a listener. Instead, the link was posited to be innately specified as a human-specific mode of perception as part of a larger specialization for language with an adaptive advantage provisioned by the “common currency” to automatically translate from sound to articulatory gesture. Of course, from this perspective, it made very little sense to study the auditory system to understand speech perception, or to study speech perception to understand auditory perception of complex signals. The two were simply distinct systems.

Although Motor Theory was extremely influential in early speech perception research, it was not without controversy. Intellectual debates raged in the 1980s and 1990s. By the early 2000s, weaker versions of Motor Theory were proposed (Galantucci et al. 2006) to accommodate empirical observations that systematically ticked off the list of phenomena purported to differentiate perception of speech from perception of other sounds by demonstrating that under the right conditions, speech and nonspeech perceptual phenomena align (Diehl et al. 2004). In the end, considering nonspeech perception in richer contexts that drew upon attention, learning, and cognitive control demonstrated that the hallmarks of speech perception could often be replicated in nonspeech signals when listeners were afforded the right expertise or listening context (Holt and Lotto 2010; Heald and Nusbaum 2014). Categorical perception provides an example (Harnad 1987). Perhaps the best-known pattern of speech perception, categorical perception refers to the observation that speech sounds gradually changing in their acoustics tend to be perceived categorically, with a sharp boundary in how they are labeled rather than a gradual, graded change in perception that mirrors the acoustics. Further, when listeners discriminate pairs of stimuli drawn from a series of speech sounds, the resulting discrimination function is discontinuous. It is nearly perfect for stimuli that lie on opposite sides of the sharp identification boundary, whereas it is very poor for pairs of stimuli that are equally acoustically distinct but fall on the same side of the identification boundary. Categorical perception was thought to be a peculiarity of speech perception, not evident for nonspeech sounds (Liberman et al. 1957). However, later research demonstrated that categorical perception could emerge for nonspeech sounds when listeners trained to apply category labels to them (Mirman et al. 2004).

Further in contrast to the predictions of Motor Theory, research demonstrated that speech and nonspeech acoustics interacted strongly in perception providing more evidence for a shared substrate (Lotto and Kluender 1998; Holt 2005). Moreover, nonhuman animal listeners (who lack a human speech motor system) were found to exhibit some of the very speech perception behaviors that were

thought to differentiate speech from nonspeech perception, including categorical perception (Kuhl and Miller 1978; Kluender et al. 1987), and context effects in perception of speech (Lotto et al. 1997). Finally, damage to the motor speech areas (e.g., in Broca's aphasia) did not produce the speech perception deficits that would be predicted by Motor Theory (Moineau et al. 2005; Hickok 2009). The overall weight of the empirical evidence did not side with the elegant, parsimonious predictions of the Motor Theory.

As evidence contrasting with predictions of the Motor Theory accumulated, the lively – often impassioned – debates regarding the objects of speech perception ultimately moved the field forward. But, there were casualties. The field lost decades of opportunity for realizing the reciprocal benefits of studying the human auditory system in alignment with human speech perception and aligning it with interpretative frameworks from nonhuman animal auditory research. More, it was denied the broader enterprise of understanding the human auditory system using one of the richest, most complex perceptual challenges: speech.

1.1.2 Speech Perception from an Auditory Perspective

Like most pervasive aspects of our lives, it is easy to take speech for granted. We are so adept at speech perception that it hardly seems a major accomplishment. However, the ease with which we perceive speech belies the complexity of the perceptual, cognitive, and neural mechanisms involved and the rich opportunities for advancing understanding of *general* human auditory perceptual abilities by studying the specific perceptual challenges introduced by speech. The fundamental units of speech that carry information may exist for mere moments. These units are complex and may be signaled by a dozen or more variable acoustic dimensions even for simple distinctions that change meaning, like *bear* from *pear*.

Complicating matters further, acoustic speech is often mixed with considerable noise, and even overlapping speech from other talkers. Yet, from this fleeting and complex acoustic signal, we are able to apprehend the linguistic message of the speaker as well as information about her gender, age, region of origin, identity, and emotional state (Kraus et al. 2019). Speech thus provides a rich testbed for understanding general principles of auditory processing, and for observations of auditory processing directed at other nonspeech acoustic signals to inform how we understand the mechanisms available to speech perception. As a complex, ecologically significant acoustic signal, speech presents challenging perceptual dilemmas spanning sensory encoding, prediction, attention, learning, memory, and integration with multimodal sensory inputs as well as other important sensory, perceptual, and cognitive issues. There is much to be gained by investigating the human auditory system through the lens of speech perception.

1.1.3 *Speech Perception Today*

Contemporary research is realizing this promise. The field of speech perception has radically shifted to embrace these reciprocal benefits, with a methodological toolbox equipped to support the endeavor. With the advent of noninvasive functional neuroimaging using hemodynamic (Evans and McGettigan 2017; Peelle 2017) and electrophysiological (Wöstmann et al. 2017) approaches, and the application of invasive neurosurgical approaches to speech perception (Leonard and Chang 2016), there is unprecedented opportunity to examine the human brain's response to speech. Accelerating benefits, auditory science more generally has developed a nascent appreciation for the cognitive aspects of auditory processing, the field of auditory cognitive neuroscience has begun to develop traction, and general cognitive and perceptual mechanisms are increasingly understood to play a role in speech communication (Pichora-Fuller et al. 2016; Peelle 2018).

At the same time, theoretical models of speech originating from cognitive science have greatly informed neurobiological approaches to understanding speech perception. Early cognitive models of speech and the human behavioral research that tested them provided evidence of hierarchically organized levels of representation whereby speech signals activate acquired representations for lower-level phonetic features, categories, and words (McClelland and Elman 1986; Norris 1999), and there is interactive processing across levels (Elman and McClelland 1988) that is modulated by attention (Mirman et al. 2008), the history of experienced that shaped the acquired representations (Kronrod et al. 2016), and online adaptation to short-term input regularities (Norris et al. 2003; Kraljic et al. 2008).

Yet, there remains much to be learned from cognitive science and behavioral approaches; indeed, the very nature of speech representations is actively under debate (Samuel 2020). Nonetheless, at this point in time, general auditory mechanisms, whether described at the cognitive or neurobiological level, are so systematically integrated into accounts of speech perception that early career researchers will likely find it most unusual to learn that the literature raged for decades about whether this was appropriate. In this volume of the *Springer Handbook of Auditory Research*, we showcase these advances at a truly exciting time for research.

1.2 The Auditory Cognitive Neuroscience of Speech Perception

This book is organized such that interested readers can dip into individual chapters of interest, or read the book cover to cover. Although it would be impossible to review the auditory cognitive neuroscience of speech perception in its entirety in a single volume, the chapters included here survey a broad range of theoretical perspectives, methodological approaches, and listening contexts that highlight current successes, challenges, and controversies in the field.

In Chap. 2, Bharath Chandrasekaran, Rachel Tessmer, and G. Nike Gnanateja provide an overview of the subcortical processing of speech sounds. This perspective is important, in part, because it is possible to develop a “cortical bias” in understanding how the brain processes speech, particularly in the context of its role in language. Nonetheless, as Chandrasekaran and colleagues review, there are important subcortical contributions to speech perception. Rather than simply relaying acoustic information to higher-order centers of the auditory system, contemporary research reveals substantial cortical-subcortical interactions in speech processing. There is significant bottom-up as well as top-down processing, a theme that recurs across this book’s chapters. Chandrasekaran, Tessmer, and Gnanateja guide readers through a thorough review of cortical and subcortical anatomy and physiology to situate discussion of the role of subcortical processing in extraction, encoding, and experience-dependent modulation of incoming speech.

In Chap. 3, Yulia Oganian, Neal P. Fox, and Edward F. Chang review contributions of human intracranial recordings to our understanding of speech perception, focusing on the superior temporal gyrus. Although electrophysiology using nonhuman animal models has long played a role in understanding speech perception (Palmer and Shamma 2004; Quam et al. 2017), there are inherent limitations in how much we can learn from species that do not, themselves, use speech to communicate. Oganian, Fox, and Chang provide readers with an overview of empirical findings and the computational tools that have been essential in revealing speech perception in human auditory cortex. Supported in equal parts by the availability of human intracranial data collected in the context of human neurosurgery and advanced computational approaches to analyzing these data, the past two decades have seen an incredibly rapid expansion of our understanding of how auditory regions of the superior temporal gyrus represent speech information. In harmony with empirical literature and theoretical models reviewed by other chapters in this book, these discoveries include demonstrations that the neural representation of speech is nonlinear, and not a faithful representation of the input. Rather, it enhances behaviorally relevant information and is influenced strongly by top-down processing.

In Chap. 4, Sarah Tune and Jonas Obleser explore the role of neural oscillations, the rhythmic or repetitive patterns of neural activity in the central nervous system that generally arise from feedback connections among neurons that result in synchronization of firing patterns, in speech perception. Tune and Obleser provide an introduction to the key characteristics of neural oscillations, as well as their origins and the functions they are thought to support. The authors argue that neural oscillations, studied extensively across sensory and cognitive domains, provide a parsimonious connection of speech perception to broader strategies for sensory, perceptual, and cognitive processing by the brain. Whereas the authors caution against the allure of ascribing distinct oscillations to specific functions, they also present a case for why understanding neural oscillations more generally will allow researchers to relate the complex dynamics of speech perception to neural dynamics. Finally, in linking to other book chapters, Chap. 4 critically examines evidence for the role neural oscillations may play in the perceptual analysis of continuous speech, from analysis of the sounds of speech to sentence-level comprehension.

In Chap. 5, Laura Gwilliams and Matthew H. Davis introduce an information-based approach to speech communication, grounded in statistical properties of speech content and the linguistic information conveyed by speech. Information-based frameworks for spoken communication have a long history in the field, and have recently found new utility in cognitive neuroscience. The chapter provides an overview of the evidence that the neural processing of speech is influenced by linguistic structure of a language – the morpheme and word-level statistical properties of the information conveyed by the acoustic speech signal. The authors situate these findings in information theoretic measures entropy and surprisal, demonstrating their value in understanding neural responses to speech. The authors argue that modeling the *information* content of the speech signal helps to explain the interface between sensory information conveyed by speech and how that interacts with listeners' sensitivity to the statistically structured patterns of linguistic input learned through years of experience. Importantly, information-based approaches can be applied at different levels of analysis (phonemes, words, sentences, and so on), providing a common currency for comparing responses at each of these levels.

In Chap. 6, Stephen C. Van Hedger and Ingrid S. Johnsrude explore how listeners understand speech in adverse listening conditions. They cover a range of challenges listeners might encounter, including background noise, competing talkers, an unfamiliar talker, and more. They provide a systematic review of behavioral and neurobiological evidence demonstrating that even minor challenges to listening demand interactions across perceptual, cognitive, and linguistic processes. The authors make the case that abstract knowledge and context are particularly important when the acoustic speech input is degraded and that although listeners likely draw upon multiple mechanisms to cope with the diversity of adverse listening conditions, the processes generally appear to be attentionally demanding. They describe evidence for the involvement of the cingulo-opercular network – especially anterior insula – in directing the cognitive effort involved in speech perception under adverse listening conditions. Finally, the chapter highlights the importance of the interaction of various listening contexts with individual differences in the cognitive resources available to speech perception, a theme that appears also in Chap. 9 (Rogers and Peelle).

In Chap. 7, Shruti Ullas, Milene Bonte, Elia Formisano, and Jean Vroomen review evidence that the mappings from acoustics to phonetic categories representing the speech sounds of a native language are flexible, rather than fixed. It has long been observed that context can resolve ambiguous speech acoustics (see Chap. 6, Van Hedger and Johnsrude). The movement of the speaker's lips, the context of the sound in a familiar word, and adjacent speech sounds each can provide contextual support to resolve ambiguity in the mapping from acoustics to phonetic categories. The chapter reviews studies that demonstrate that when listeners experience repeated instances of this contextual resolution, longer-lasting perceptual learning or recalibration can occur such that perception of the ambiguous speech acoustics is shifted even when contextual support is no longer available. The chapter also reviews a rich literature that has developed to investigate this adaptive plasticity in speech

perception and relates these investigations to theories of speech perception and neuroimaging data that inform its neural underpinnings.

In Chap. 8, Judit Gervain reviews the development of speech perception. Before we are native speakers, we are native listeners – infants begin learning about the patterns of speech in their native language even before birth and, by their first birthday, exhibit substantial experience-dependent reorganization of auditory processing of speech that accommodates the sound patterns of the native language(s). In this way, examination of speech perception across early development provides a window into experience-dependent auditory processing. The chapter reviews the major milestones of the development of speech perception, beginning prenatally and continuing through the first year of life and into the toddler years when word learning and bootstrapping of grammar by the prosodic properties of speech become apparent. The review makes clear that the developing brain orchestrates acquisition of spoken language in parallel across multiple levels of representation that ultimately support speech perception in the native language(s).

Finally, in Chap. 9, Chad S. Rogers and Jonathan E. Peelle discuss interactions between audition and cognition in hearing loss and aging. Earlier chapters (Chap. 5, Gwilliams and Davis, and Chap. 6, in particular, Van Hedger and Johnsrude) make the case that speech perception involves a distributed network of processes, including cognitive processes that vary rather substantially across individuals. Rogers and Peelle highlight the central role of cognitive processes in speech perception among older adults with hearing loss. The chapter reviews age-related changes in both hearing and cognition and describes converging evidence demonstrating their interplay – the evidence indicates that when confronted with acoustically challenging speech, cognitive effort is required. Individual differences in hearing and cognitive abilities determine the cognitive demand of a listener in a particular listening context, and therefore the cognitive and neural resources that contribute to speech perception.

1.3 Common Threads and Future Directions

Although each chapter covers its own specific topic, there are also a number of important themes that cut across chapters that are important to highlight.

A clear shift in the field has occurred with the widespread availability of functional neuroimaging and electrophysiological measurements to study how the brain processes speech. Every chapter in this volume reviews neural data collected from human listeners that were simply unavailable during earlier eras of speech perception research. In fact, reading this volume alongside an earlier volume on speech perception in the *Springer Handbook of Auditory Research* series provides an excellent bird's-eye view of how the field has evolved as new approaches to examining the human brain became ubiquitous (Greenberg and Ainsworth 2004). Keeping pace with advances in data collection methods, there have been increasingly sophisticated approaches to modeling data that incorporate acoustic or linguistic features

and permit extraction of neural signatures of specific aspects of the speech signal (Chap. 3, Oganian, Fox, and Chang; Chap. 4, Tune and Obleser; Chap. 5, Gwilliams and Davis). A challenge introduced by this wealth of approaches is to keep sight of the value of integrating what we learn across different methods, levels of analysis, time domains, and populations to advance deeper understanding. Just as important, it will be crucial for the field to not only address “old” questions with these new techniques but also to reconceptualize speech in the context of distributed processing across an interactive brain, and to start asking questions from this new perspective.

Supported by methodological advancements, there is now also an increasing use of more “natural” speech signals, including movies and short stories, to study speech perception (Chap. 3, Oganian, Fox, and Chang; Chap. 4, Tune and Obleser; Chap. 5, Gwilliams and Davis). Of course, most of our everyday communication does not happen listening to isolated phonemes, words, or sentences over headphones while lying in an MRI scanner, and the move toward ever-more natural speech is a positive one. At the same time, almost by definition, these natural stimuli are not well controlled for various acoustic or linguistic features of interest. Thus, the strongest claims will likely need to be backed by converging evidence from both “traditional” experimental paradigms (offering tight control over experimental conditions) and naturalistic listening (verifying real-world applicability).

Another dimension along which our understanding of speech perception is broadening relates to the people doing the listening. There is increasing realization that the challenges (and, hopefully, successes) of speech perception depend not only on the acoustic properties of the speech signal but on the auditory, linguistic, and cognitive abilities of individual listeners (highlighted in Chap. 6, Van Hedger and Johnsrude; Chap. 8, Gervain; and Chap. 9, Rogers and Peelle). The ways in which different listeners perceive speech are important not only to ensure generalizability of our theoretical approaches but also to test specific hypotheses. For example, if we have a hypothesis about how acoustic clarity affects speech perception, then studying speech perception in hearing-impaired listeners is one way to empirically test our claim. Considering speech perception across the lifespan, listeners with different abilities, and a variety of listening environments will ensure that the field converges on robust mechanistic accounts that accommodate the true demands on speech perception.

Neuroanatomically, there is still much focus on core auditory regions including the hindbrain and midbrain (Chap. 2, Chandrasekaran, Tessler, and Gnanateja) and superior temporal gyrus (Chap. 3, Oganian, Fox, and Chang). However, there is also an increasing appreciation for speech as a whole-brain activity. For example, the fact that regions outside traditional speech and language networks are engaged during adverse listening situations (Chap. 6, Van Hedger and Johnsrude; Chap. 9, Rogers and Peelle) highlights the systems-level interactions required for speech perception (at least under some circumstances). Recognition of these “extra-auditory” brain regions as crucial to speech perception goes hand in hand with the developing appreciation that learning, attention, and cognitive control are crucial components to any full theoretical account of speech perception.

In this regard, speech perception offers a rich testbed for cognitive science and cognitive neuroscience, more broadly. For example, although the Motor Theory did not hold up to empirical scrutiny, there remains important work to be done in understanding the nuanced interactions between speech perception and speech production. Future work also will be needed to blur the arbitrary lines that have traditionally been drawn between perception, learning, attention, and cognition – even outside of speech perception. Speech presents a model case for making progress in this regard; even “online” speech *perception* engages learning (Chap. 7, Ullas, Bonte, Formisano, and Vroomen), and attention (Chap. 6, Van Hedger and Johnsrude), and cognitive processing (Chap. 9, Rogers and Peelle). Similarly, given the intimate connection of speech input with distinct levels of language processing (phonemes, words, etc.), speech provides an ideal model for advancing general understanding of the interplay of hierarchical levels of representation and of predictive models in neural processing (Chap. 5, Gwilliams and Davis).

1.4 Summary

In summary, evolving techniques have provided unprecedented access to neural data, and theoretical perspectives of speech perception are making more and more contact with auditory neuroscience. These opportunities challenge researchers to ask questions that continue to further our understanding of speech perception in new and useful ways. It is an exciting time to be studying speech perception.

Acknowledgments This work was supported in part by grants R01 DC014281, R21 DC016086, R21 DC015884, and R56 AG059265 from the US National Institutes of Health to JEP as well as R01DC017734, R03HD099382, and R21DC019217 from the US National Institutes of Health and BCS1950054 and BCS1655126 from the US National Science Foundation to LLH.

Compliance with Ethics Requirements Lori L. Holt declares that she has no conflict of interest. Jonathan E. Peelle declares that he has no conflict of interest.

References

- Cooper FS, Liberman AM, Borst JM (1951) The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proc Natl Acad Sci U S A* 37:318–325
- Delattre PC, Liberman AM, Cooper FS (1955) Acoustic loci and transitional cues for consonants. *J Acoust Soc Am* 27:769–773
- Diehl RL, Lotto AJ, Holt LL (2004) Speech perception. *Annu Rev Psychol* 55:149–179
- Elman JL, McClelland JL (1988) Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes. *J Mem Lang* 27:143–165
- Evans S, McGettigan C (2017) Comprehending auditory speech: previous and potential contributions of functional MRI. *Lang Cogn Neurosci* 32:829–846
- Fowler CA (2001) Obituary: Alvin M. Liberman (1917–2000). *Am Psychol* 56:1164–1165

- Galantucci B, Fowler CA, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychon Bull Rev* 13:361–377
- Greenberg S, Ainsworth WA (2004) *Speech processing in the auditory system: an overview*. Springer, New York
- Harnad S (1987) *Categorical perception: The groundwork of cognition*. Cambridge University Press, Cambridge
- Heald S, Nusbaum HC (2014) Speech perception as an active cognitive process. *Front Syst Neurosci* 8:35
- Hickok G (2009) Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J Cogn Neurosci* 21:1229–1243
- Holt LL (2005) Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychol Sci* 16:305–312
- Holt LL, Lotto AJ (2010) Speech perception as categorization. *Atten Percept Psychophys* 72:1218–1227
- Klemmer ET, Snyder FW (1972) Measurement of time spent communicating. *J Commun* 22:142–158
- Kluender KR, Diehl RL, Killeen PR (1987) Japanese quail can learn phonetic categories. *Science* 237:1195–1197
- Kraljic T, Samuel AG, Brennan SE (2008) First impressions and last resorts: how listeners adjust to speaker variability. *Psychol Sci* 19:332–338
- Kraus MJ, Torrez B, Park JW, Ghayebi F (2019) Evidence for the reproduction of social class in brief speech. *Proc Natl Acad Sci USA* 116:22998–23003
- Kronrod Y, Coppess E, Feldman NH (2016) A unified account of categorical effects in phonetic perception. *Psychon Bull Rev* 23:1681–1712
- Kuhl PK, Miller JD (1978) Speech perception by the chinchilla: identification function for synthetic VOT stimuli. *J Acoust Soc Am* 63:905–917
- Leonard MK, Chang EF (2016) Direct cortical neurophysiology of speech perception. In: Hickok G, Small SL (eds) *Neurobiology of language*. Academic Press, London, pp 479–489
- Lieberman AM (1957) Some results of research on speech perception. *J Acoust Soc Am* 29:117–123
- Lieberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21:1–36
- Lieberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol* 54:358–368
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431–461
- Lotto AJ, Kluender KR (1998) General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Percept Psychophys* 60:602–619
- Lotto AJ, Kluender KR, Holt LL (1997) Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *J Acoust Soc Am* 102:1135–1140
- McClelland JL, Elman JL (1986) The TRACE model of speech perception. *Cogn Psychol* 18:1–86
- Mehl MR, Vazire S, Ramírez-Esparza N et al (2007) Are women really more talkative than men? *Science* 317:82
- Mirman D, Holt LL, McClelland JL (2004) Categorization and discrimination of nonspeech sounds: differences between steady-state and rapidly-changing acoustic cues. *J Acoust Soc Am* 116:1198–1207
- Mirman D, McClelland JL, Holt LL, Magnuson JS (2008) Effects of attention on the strength of lexical influences on speech perception: behavioral experiments and computational mechanisms. *Cogn Sci* 32:398–417
- Moineau S, Dronkers NF, Bates E (2005) Exploring the processing continuum of single-word comprehension in aphasia. *J Speech Lang Hear Res* 48:884–896
- Norris D (1999) The merge model: speech perception is bottom-up. *J Acoust Soc Am* 106:2295–2295
- Norris D, McQueen JM, Cutler A (2003) Perceptual learning in speech. *Cogn Psychol* 47:204–238

- Palmer A, Shamma S (2004) Physiological representations of speech. In: Greenberg S, Ainsworth WA (eds) *Speech processing in the auditory system: an overview*. Springer, New York
- Peelle JE (2017) Optical neuroimaging of spoken language. *Lang Cogn Neurosci* 32:847–854
- Peelle JE (2018) Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear* 39:204–214
- Peterson GE, Barney HL (1952) Control methods used in a study of the vowels. *J Acoust Soc Am* 24:175–184
- Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BW, Humes LE, Lemke U, Lunner T, Matthen M, Mackersie CL, Naylor G, Phillips NA, Richter M, Rudner M, Sommers MS, Tremblay KL, Wingfield A (2016) Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear* 37:5S–27S
- Quam RM, Ramsier MA, Fay RR, Popper AN (2017) *Primate hearing and communication*. Springer, Cham
- Samuel AG (2011) Speech perception. *Annu Rev Psychol* 62:49–72
- Samuel AG (2020) Psycholinguists should resist the allure of linguistic units as perceptual units. *J Mem Lang* 111:104070
- Wöstmann M, Fiedler L, Obleser J (2017) Tracking the signal, cracking the code: speech and speech comprehension in non-invasive human electrophysiology. *Lang Cogn Neurosci* 32:855–869