The Handbook of
Speech
Perception

Second Edition

Edited by

**Jennifer S. Pardo, Lynne C. Nygaard,
Robert E. Remez, and David B. Pisoni**

The Handbook of Speech Perception

Blackwell Handbooks in Linguistics

This outstanding multivolume series covers all the major subdisciplines within linguistics today and, when complete, will offer a comprehensive survey of linguistics as a whole.

The most recent publications in the series can be found below. To see the full list of titles available in the series, please visit www.wiley.com/go/linguistics-handbooks

The Handbook of
Speech Perception

Second Edition

Edited by

*Jennifer S. Pardo, Lynne C. Nygaard,
Robert E. Remez, and David B. Pisoni*

WILEY Blackwell

Contents

List of Contributors

Faisal M. Aljasser is an associate Professor of Applied Linguistics in the Department of English language and Translation, College of Arabic Language and Social Studies at Qassim University. He received his Ph.D. in Applied Linguistics from Newcastle University, UK in 2008. His research centers on the production and perception of Arabic as a native language and as a second language.

Melissa Baese-Berk is the David M. and Nancy L. Petrone Faculty Scholar and Associate Professor of Linguistics at the University of Oregon, where she directs the Speech Perception and Production Laboratory. She earned her PhD from Northwestern University in 2010. Her research focuses on speech perception and production, with special attention to speakers and listeners who do not share a native language with their interlocutor. Her work has been supported by grants from the *National Science Foundation*. Recent publications have appeared in *Cognition*, *Journal of the Acoustical Society of America*, *Journal of Phonetics*, and *Attention, Perception, and Psychophysics*.

Tessa Bent is Professor of Speech, Language and Hearing Sciences and director of the Speech Perception Laboratory at Indiana University. She received her Ph.D. in Linguistics from Northwestern University in 2005. Her research focuses on children's and adults' perception and representation of variable speech signals, with a focus on regional dialects and non-native accents. This research is currently supported by a grant from the National Science Foundation and has previously been supported by the National Institutes of Health. She is a Fellow of the Acoustical Society of America.

Sheila E. Blumstein is the Albert D. Mead Professor Emerita of Cognitive, Linguistic, and Psychological Sciences at Brown University. She received her Ph.D. in Linguistics from Harvard University in 1970. She spent her entire professional career from 1970 until 2018 at Brown University. Her research focuses on the neural basis of speech and language and the processes and mechanisms underlying speaking and understanding. She received a Claude Pepper Award from the National Institutes of Health and The Silver Medal in Speech Communication from the Acoustical Society of America and was elected Fellow to a number of professional societies.

Z. S. Bond, Professor Emerita, Ohio University, earned a Ph.D. in linguistics, with psychology and hearing and speech sciences as concentrations, from the Ohio State University. She has worked at the University of Alberta, Ohio University, Ohio State University and the Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base. Her research areas include phonetics, psychology of language, speech perception, and language contact. Currently she is analyzing the pronunciation of Latvian in recordings from WW I. She has published papers in various journals including *Journal of the Acoustical Society of America, Language and Speech*, *Perception and Psychophysics,* and *Journal of Phonetics*. She is a member of Acoustical Society of America, Association for the Advancement of Baltic Studies, Linguistic Society of America and a foreign member of Latvian Academy of Science.

Ann R. Bradlow is the Abraham Harris Professor of Linguistics and Associate Dean for Graduate Studies at Northwestern University. She received her PhD in Linguistics from Cornell University in 1993, and then completed postdoctoral fellowships in Psychology at Indiana University (1993-1996) and Hearing Science at Northwestern University (1996-1998). Over the past three decades, Bradlow has pursued an interdisciplinary research program in acoustic phonetics and speech perception with a focus on speech intelligibility under conditions of talker-, listener-, and situation-related variability. Her work has been supported by grants from the *National Institutes of Health*. Recent publications have appeared in *the Journal of the Acoustical Society of America*, *Journal of the Association for Laboratory Phonology, International Journal of Audiology, Applied Psycholinguistics, Journal of Phonetics, Language & Speech*, and *Bilingualism, Language, & Cognition*.

Susan Brady received her Ph.D. in Cognitive Psychology at the University of Connecticut in 1975 and presently is an Emerita Professor of Psychology at the University of Rhode Island. She has held additional positions at the University of Sussex, St. Andrews University, and Haskins Laboratories. Concentrating on topics in the field of reading, her research has focused primarily on the roles of speech perception and verbal working memory in individual differences in reading ability. Likewise, she has endeavored to translate the implications of the larger body of reading research for practice, and has conducted professional development projects for educators.

Axelle Calcus is an Assistant Professor of Psychology at Université Libre de Bruxelles (Brussels, Belgium). She received her PhD in Psychology in 2015, and worked as a postdoctoral researcher at Boston University (MS, United States), University College London (London, UK) and Ecole Normale Supérieure (Paris, France). Her main research interest focuses on the development of perception of speech in noise in children with and without hearing difficulties. Her work has been supported by awards from H2020 (European Commission). She is a member of the board of the Belgian association for audiology, and of the executive committee of the Belgian association for psychological sciences. Her recent work has been published in *Developmental Science* and *eLife*.

Cynthia G. Clopper is Professor of Linguistics at Ohio State University and a Fellow of the Acoustical Society of America. She received her Ph.D. in Linguistics and Cognitive Science from Indiana University and held post-doctoral positions in Psychology at Indiana University and in Linguistics at Northwestern University before joining the faculty at Ohio State. Her major areas of expertise are phonetics, speech perception, sociophonetics, and laboratory phonology. Dr. Clopper's current research projects examine the effects of geographic mobility and linguistic experience on cross-dialect lexical processing, the relationships between linguistic and indexical sources of variation in speech processing, and regional prosodic variation in American English.

Anja-Xiaoxing Cui is a postdoctoral fellow at the University of British Columbia and visiting professor of systematic musicology at Osnabrück University. She studied psychology and piano performance before receiving her Ph.D. in Cognitive Neuroscience from Queen's University in 2019. Her research centers on auditory processing and the interactions of music and learning, and has been supported by NSERC and the Social Sciences and Humanities Research Council of Canada. Anja received additional support through the German Academic Exchange Service and the German Academic Scholarship Foundation.

Anne Cutler is Distinguished Professor at the MARCS Institute, Western Sydney University, Australia. She studied languages and psychology in Melbourne, Berlin and Bonn, took a PhD in psycholinguistics at the University of Texas, held positions at MIT, Sussex University and the MRC Applied Psychology Unit (Cambridge, UK), and then from 1993 to 2013 was Director and Comprehension Group head at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. Her research concerns listening to spoken language, and in particular how native language experience tailors speech decoding processes. She is an elected member of national academies in Europe, the US and Australia.

Josh Dorsi is a Postdoctoral Scholar in the Neurology Department of the Penn State College of Medicine. He received his Ph.D. in Psychology from the University of California, Riverside in 2019. His research investigates the role of multisensory and lexical information in supporting speech perception, as well as the role of crossmodal correspondences in speech and language pathologies. Some recent publications of this work have appeared in *Attention, Perception, & Psychophysics*; *The Quarterly Journal of Experimental Psychology;* and the *Journal of Cognitive Psychology*.

Cynthia R. Hunter is Assistant Professor of Speech-Language-Hearing and Director of the *Speech Perception, Cognition, and Hearing Laboratory* at the University of Kansas. She earned her Ph.D. in Psychology from the StateUniversity of New York at Buffalo in 2016, and completed postdoctoral fellowships at Indiana University and the Indiana Clinical and Translational Sciences Institute. Her research centers on the neural and cognitive factors that allow individuals with and without hearing loss to understand speech in adverse listening conditions.

Her recent workhas appeared in *Ear and Hearing*, *Journal of Speech, Language, and Hearing Researc*h, *Neuropsychologia*, and *Brain and Language*.

Sara Incera is an Assistant Professor of Psychology and Director of the *Multilingual Laboratory* at Eastern Kentucky University. She received her Ph.D. in Psychology from Cleveland State University in 2016, Conor McLennan was her Ph.D. advisor. Her research interests include foreign accents, bilingualism, and language development across the lifespan. Her most recent work has focused on the relationships between language and emotion. Her articles have been published in *Cognition & Emotion, Journal of Psycholinguistic Research, Mind & Language, Aging, Neuropsychology, & Cognition, International Journal of Bilingualism, Acta Psychologica*, and *Bilingualism: Language & Cognition*.

Alexandra Jesse is an Associate Professor of Psychological and Brain Sciences and Director of the Language, Intersensory Perception, and Speech Laboratory at the University of Massachusetts Amherst. After receiving her Ph.D. in Psychology from the University of California Santa Cruz in 2005, she held a research position at the Max Planck Institute for Psycholinguistics until 2010. Her research focuses on speech perception, particularly on audiovisual speech and aging, and has been supported by the National Institutes of Health, the German Research Foundation, and the Netherlands Organization for Scientific Research. Some recent publications have appeared in Cognition, Journal of Experimental Psychology: Learning, Memory and Cognition, and Biological Psychology.

Keith Johnson is Professor of Linguistics and Chair of the Department of Linguistics at UC Berkeley. He received his PhD in Linguistics from Ohio State University in 1988 and held research positions in Psychology at Indiana University, in Linguistics at UCLA, and in Speech and Hearing Science at University of Alabama, Birmingham, and academic positions at Ohio State University, and Berkeley. His research is on perceptual processes involved in compensating for phonetic talker differences.

Steven Lamontagne is a PhD candidate at Queen's University and a visiting doctoral scholar at Harvard's McLean Hospital. In 2017, he received his MSc in Cognitive Neuroscience at Queen's University, where he studied interactions between reward and stress circuitry using animal models. His current research, which is supported by NSERC, centers on the neurophysiological correlates of reward learning and cognitive control in people with treatment-resistant major depressive disorder. Some recent publications of his research have appeared in *Psychopharmacology*, *Physiology & Behavior,* and *Behavioural Brain Research*.

Susannah V. Levi is an Associate Professor of Communicative Sciences and Disorders and Director of the Acoustic Phonetics and Perception Lab at New York University. She received her Ph.D. in Linguistics from the University of Washington in 2004. She completed a postdoctoral fellowship in the Speech Research Lab at Indiana University with David Pisoni. Her research focuses on the relationship between linguistic and speaker information during of spoken language processing.

Her research has been supported by grants from the National Science Foundation and the National Institutes of Health. Some recent publications have appeared in the Journal of Speech, Language, and Hearing Research, Journal of the Acoustical Society of American, and Cognitive Science.

David Lutes received his M.Sc. in Cognitive Neuroscience at Queen's University in 2019, where he used virtual reality devices to study the impact that various image characteristics have on the brain's ability to effectively fuse separate images in binocular vision. To further his interest in the applications of virtual reality, David is continuing his education into video game development, as well as public health and neuroscience.

Conor T. McLennan is a Professor, Chair of the Department of Psychology, and Director of the Language Research Laboratory at Cleveland State University. He received his Ph.D. from the University at Buffalo in 2003. His research interests include language perception, bilingualism, cognitive aging, and other topics in language, memory, and perception. His research has been supported by the *National Science Foundation* and the *National Institutes of Health*, and has been published in a variety of journals, including *Aging, Neuropsychology, & Cognition, Attention, Perception, & Psychophysics, Cognition & Emotion, Journal of Experimental Psychology: Learning, Memory, & Cognition*, and *Language & Speech*.

K. G. Munhall is a professor in the Department of Psychology at Queen's University. He received his Ph.D. in psychology from McGill University in 1984. His research focuses on sensorimotor processing in speech production, audiovisual speech perception, and perceptual and cognitive factors in conversational interaction. His work has been supported by grants from the National Institute on Deafness and Other Communication Disorders, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canadian Institutes of Health Research (CIHR). Some recent publications of his work have appeared in *Journal of the Acoustical Society of America, Experimental Brain Research, Multisensory Research,* and *Attention, Perception, & Psychophysics*.

Emily B. Myers is an Associate Professor of Speech, Language, and Hearing Sciences and Psychological Sciences at the University of Connecticut. She received her PhD from Brown University in 2005. Her work focuses on the processes that allow a listener to map the speech signal to meaning, how these processes are instantiated in the brain, and how the system breaks down in cases of language disorder. Her work has been funded by the National Institutes of Health and the National Science Foundation.

Susan Nittrouer received her PhD from the City University of New York in Speech and Hearing Science. After a post-doctoral fellowship at Haskins Laboratories she worked at Boys Town National Research Hospital, Utah State University, and the Ohio State University. Currently she is Professor of Speech, Language, and Hearing Sciences at the University of Florida. Her research focuses on the intersection between auditory and language development, and on the challenges encountered by

children with risk factors for developmental language delays, including hearing loss, poverty, or conditions leading to dyslexia. Susan's goal is to develop more effective interventions for these children.

Lynne C. Nygaard is Professor of Psychology and Director of the Center for Mind, Brain, and Culture, and the Speech and Language Communication Laboratory at Emory University, USA. Her research on the perceptual, cognitive, biological, and social underpinnings of human spoken communication has appeared in many journals, including *Psychological Science, Brain and Language*, and *Cognitive Science*.

Ellen O'Donoghue is a Ph.D. Candidate at The University of Iowa, in the Department of Psychological and Brain Sciences. She received her M.Sc. in Cognitive Psychology from Queen's University in 2018. Her research concerns the fundamental mechanisms that support learning and categorization across species, with particular emphasis on humans and pigeons.

Jennifer S. Pardo is Professor of Psychology and Director of the *Speech Communication Laboratory* at Montclair State University. She received her Ph.D. in Cognitive Psychology from Yale University in 2000, and has held academic positions at Barnard College, Wesleyan University, and The New School for Social Research. Her research centers on the production and perception of spoken language in conversational interaction and on understanding variation and convergence in phonetic form, and has been supported by grants from the *National Science Foundation* and the *National Institutes of Health*. Some recent publications of this work have appeared in *Journal of Memory & Language*, *Journal of Phonetics*, *Language & Speech,* and *Attention, Perception, & Psychophysics*.

Oiwi Parker Jones is a Hugh Price Fellow at Jesus College, University of Oxford. He did his doctoral research in Oxford on NLP with a focus on the application of machine learning to endangered languages. From there he trained as an imaging and computational neuroscientist at University College London and Oxford. His primary interest is in the development of a neural speech prosthetic. This includes basic research on speech and language in the brain, including work on clinical populations. His papers have been published in journals like *Science* and *Brain* and at machine learning conferences like *NeurIPS*, *ICLR*, and*ICML*.

David B. Pisoni is Distinguished Professor of Psychological and Brain Sciences and Chancellor's Professor of Cognitive Science at Indiana University, Bloomington, USA, and Professor in the Department of Otolaryngology at Indiana University School of Medicine, Indianapolis, USA. He has made significant contributions in basic, applied, and clinical research in areas of speech perception, production, synthesis, and spoken language processing.

Lawrence J. Raphael is Professor Emeritus of both the Graduate School of CUNY and Adelphi University. He was a research associate at Haskins Laboratories for 26 years. His research interests include speech perception, speech acoustics and the physiology of the speech mechanism. His research has been published in a variety of scholarly journals. He is a co-author of Speech Science Primer, 6th edition and

co-editor of *The Biographical Dictionary of the Phonetic Sciences*, *Language and Cognition* and *Producing Speech*. Professor Raphael is a Fellow of the New York Academy of Sciences.

Robert E. Remez is Professor of Psychology at Barnard College, Columbia University, USA, and Chair of the Columbia University Seminar on Language and Cognition. His research has been published in many scientific and technical journals, including *American Psychologist*, *Developmental Psychology*, *Ear and Hearing*, *Experimental Aging Research*, *Journal of Cognitive Neuroscience*, and *Journal of Experimental Psychology*.

Lawrence D. Rosenblum is a Professor of Psychology at the University of California, Riverside. He studies multisensory speech and talker perception as well as ecological acoustics. His research has been supported by grants from the National Science Foundation, National Institutes of Health, and the National Federation of the Blind. He is the author of numerous publications including the book See What I'm Saying: The Extraordinary Powers of our Five Senses. His research has been featured in Scientific American, The New York Times, and The Economist.

Jan W. H. Schnupp is a sensory neuroscientist with a long standing interest in the processing of auditory information by the central nervous system. He received his DPhil from the University of Oxford in 1996, and he held visiting and faculty positions at the University of WIsconsin, the Italian Institute of Technology and the University of Oxford before taking up a professorship at the City University of Hong Kong. His research interests range widely, from central representations of auditory space to pitch and timbre, temporal predictive coding and auditory pattern learning. His work has been funded by the Wellcome Trust, BBSRC, MRC, and the UGC and HMRF of Hong Kong. He has published over 80 papers in numerous neuroscience and general science journals and he coauthored the textbook "Auditory Neuroscience".

Diana Van Lancker Sidtis (formerly Van Lancker) is Professor Emeritus of Communicative Sciences and Disorders at New York University, where she served as Chair from 1999-2002; Associate Director of the Brain and Behavior Laboratory at the Nathan Kline Institute, Orangeburg, NY; and a certified and licensed speech-language pathologist (from Cal State LA). Her education includes an MA from the University of Chicago, PhD from Brown University, and an NIH Postdoctoral Fellowship at Northwestern University. Dr. Sidtis has continued to mentor students and perform research in speech science, voice studies, and neurolinguistics. She is author of over 100 scientific papers and review chapters, and coauthor, with Jody Kreiman, of *Foundations of Voice Studies*, Wiley-Blackwell. Her second book, *Foundations of Familiar Language*, is scheduled to appear in 2021.

Matthias J. Sjerps received his Ph.D. in Cognitive Psychology at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. He has held postdoc positions at the Max Planck Institute, The Radboud University of Nijmegen,

and at the University of California at Berkeley. His main research line has been centered on the perception of speech sounds, with a specific focus on how listeners resolve variability in speech sounds. His work has been supported by grants from the *European Committee (Marie curie grant)* and Max Planck Gesellschaft. Some recent publications of this work have appeared in *Nature Communications*, *Journal of Phonetics*, *Journal of Experimental Psychology: Human Perception and Performance*. Since 2019 he is working as a researcher for the Dutch Inspectorate of Education, focusing on methods of risk-assessment of schools and school-boards.

Rajka Smiljanic is Professor of Linguistics and Director of UT Sound Lab at the University of Texas at Austin. She received her Ph.D. from the Linguistics Department at the University of Illinois Urbana-Champaign, after which she worked as a Research Associate in the Linguistics Department at Northwestern University. Her work is concentrated in the areas of experimental phonetics, cross-language and second language speech production and perception, clear speech, and intelligibility variation. Her recent work appeared in the *Journal of the Acoustical Society of America*, *Journal of Speech, Language, and Hearing Research*, and *Journal of Phonetics*. She was elected Fellow of the Acoustical Society of America in 2018 and is currently serving as a Chair of the Speech Communication Technical Committee.

Mitchell S. Sommers is Professor of Psychological and Brain Sciences at Washington University in St. Louis. He received his PhD in Psychology from the University of Michigan and worked as a postdoctoral Fellow at Indiana University. His work focuses on changes in hearing and speech perception in older adults and individuals with dementia of the Alzheimer's type. His work has been published in Ear & Hearing, Journal of the Acoustical Society of America, and Journal of Memory and Language, among others. He received a career development award from the Brookdale Foundation and his work has been supported by NIH, NSF, and the Pfeifer Foundation.

Christina Y. Tzeng is Assistant Professor of Psychology at San José State University. She received her Ph.D. in Psychology from Emory University in 2016. Her research explores the cognitive mechanisms that underlie perceptual learning of variation in spoken language and has been supported by the American Psychological Association. She has published her research in journals such as *Cognitive Science*, *Journal of Experimental Psychology: Human Perception and Performance, and Psychonomic Bulletin and Review*.

Michael S. Vitevitch is Professor of Psychology and Director of the *Spoken Language Laboratory* at the University of Kansas. He received his Ph.D. in Cognitive Psychology from the University at Buffalo in 1997, and was an NIH post-doctoral trainee at Indiana University before taking an academic position at the University of Kansas in 2001. His research uses speech errors, auditory illusions, and the mathematical tools of network science to examine the processes and representations that are involved in the perception and production of spoken language. His work has been supported by grants from the *National Institutes of Health,* and has been published in Psychology journals such as *Journal of Experimental Psychology:*

General, *Cognitive Psychology,* and *Psychological Science,* as well as journals in other disciplines such as *Journal of Speech, Language, and Hearing Research,* and *Entropy*.

Seung Yun Yang, Ph.D., CCC-SLP, is an assistant professor in the department of Communication, Arts, Sciences, and Disorders. She is also a member of the Brain and Behavior Laboratory at the Nathan Kline Institute for Psychiatric Research in Orangeburg, New York. She received her doctorate from the Department of Communicative Sciences and Disorders at New York University. Her research primarily focuses on understanding the neural bases of nonliteral language and on understanding how prosody is conveyed and understood in the context of spoken language. Her research works have been published in peer-reviewed journals such as Journal of American Speech, Language, and Hearing Research and Clinical Linguistics & Phonetics.

Romi Zäske is a researcher at the University Hospital Jena and the Friedrich Schiller University of Jena, Germany. She received her Ph.D. from the Friedrich Schiller University of Jena in 2010, and has conducted research projects at Glasgow University, UK, and at the University of Bern, Switzerland. Her research centers on the cognitive and neuronal mechanisms subserving human voice perception and memory, including individual differences, and has been supported by grants of the *Deutsche Forschungsgemeinschaft (DFG)*. Some recent publications of her work have appeared in *Royal Society Open Science, Behavior Research Methods, Attention, Perception, & Psychophysics, Cortex,* and *Journal of Neuroscience*.

Foreword to the Second Edition

Two remarkable developments have taken hold since the publication of the first edition of *The Handbook of Speech Perception* in 2006. The first is directly connected to the study of speech perception and stands as a testament to the maturity and vitality of this relatively new field of research. The second, though removed from the study of speech perception, provides a timely pointer to the central theme of this book. Both of these developments are so overbearing that they simply cannot go without notice as I write this preface in the last quarter of 2020. They also help us to see how the complex landscape of speech perception research intersects with some of the most challenging and exciting scientific frontiers of our time.

The first of these developments is the appearance of virtual assistants such as Apple's Siri, Amazon's Alexa, and Microsoft's Cortana. While it is a well-worn cliché to mark time by technological developments, the rapid adoption of these speech technologies over the past decade is hard to ignore when thinking about speech communication. The domain of speech perception now includes both humans and machines as both talkers and listeners. What exactly does machine speech recognition have to do with the body of research presented in the chapters of this handbook, all of which address *human* speech perception? These speaker-hearer machines certainly do not perceive speech in a human-like way; Siri, Alexa, and Cortana do not sense speech as do human ears and eyes, their machine learning algorithms do not result in neurocognitive representations of linguistic properties, and they are not participants in the relationships and social meanings encoded in the indexical properties of speech. In his preface to the first edition of this handbook, Michael Studdert-Kennedy noted that "alphabetic writing and reading have no independent biological base; they are, at least in origin, parasitic on spoken language." Studdert-Kennedy went on to suggest that, "speech production and perception, writing and reading, form an *intricate biocultural nexus*" (my italics). With the invention of virtual assistants, spoken language once again participates in a symbiotic relationship with a new medium of verbal communication. Within this complex and evolving ecology of spoken–written–digital language, the study of human speech perception continues to reveal, in increasing detail, the contours of this biocultural nexus. Immersion into this field of inquiry, made so accessible by the carefully selected and recently updated collection of chapters in this handbook, is so stimulating precisely because it illuminates the milestones that mark the path to, through, and beyond this nexus.

The second development that cannot go without mention as this updated handbook goes to the printing press in 2020 is the startling spread of the Covid-19 virus through human communities across the globe. Speech sounds, words, ideas, and (unfortunately) viruses are all transmitted from person to person through the air that we breathe during social interactions. With the covering of visible speech gestures by virus-blocking masks, the awkward turn-taking of video conferencing tools with single-track audio channels, and the social distancing that protects us from the Covid-19 virus, this pandemic is a constant reminder of the multimodal nature of speech perception and of the centrality of in-person social interaction for seamless speech communication. The arrangement of the first three major sections of this handbook – *I: Sensing Speech*, *II: Perception of Linguistic Properties*, and *III: Perception of Indexical Properties* – provides the scaffold for an understanding of speech perception as far more than perception of a particular auditory signal. Instead, the chapters in these sections, along with the applications and theories covered in the remaining two sections – *IV: Speech Perception by Special Listeners*, and *V: Theoretical Perspectives* – develop the overarching argument that the observation, measurement, and modeling of speech perception must be conducted from a vantage point that encompasses its broad cognitive and social context. This central point is brought home in the final chapter by David Pisoni, one of the founders of the field and editors of this handbook:

> ". . . hearing and speech perception do not function as independent autonomous streams of information or discrete processing operations that take place in isolation from the structure and functioning of the whole information-processing system. While it is clear that the early stages of speech recognition in listeners with normal hearing are heavily dependent on the initial encoding and registration of highly detailed sensory information, audibility and the sensory processing of speech is only half of the story".

The chapters in this handbook provide a superbly sign-posted map of the full story.

Any compendium of knowledge on a particular topic represents a body of knowledge that developed in a specific time and place. The contributors to this handbook cover several generations of researchers spread over many academic disciplines working primarily on both sides of the North Atlantic Ocean. Yet, the scientific study of speech perception as presented in this outstanding handbook is still relatively young and localized. Perhaps one of the lasting lessons of the current pandemic is that we are all even more connected than we thought. New ideas and new ways of knowing can circulate as extensively, though maybe not quite as quickly, as a virus. This bodes well for the future of speech perception research.

Ann R. Bradlow
Northwestern University

Foreword to the First Edition

Historically, the study of audition has lagged behind the study of vision, partly, no doubt, because seeing is our first sense, hearing our second. But beyond this, and perhaps more importantly, instruments for acoustic control and analysis demand a more advanced technology than their optic counterparts: having a sustained natural source of light, but not of sound, we had lenses and prisms long before we had sound generators and oscilloscopes. For speech, moreover, early work revealed that its key perceptual dimensions are not those of the waveform as it impinges on the ear (amplitude, time), but those of its time-varying Fourier transform, as it might appear at the output of the cochlea (frequency, amplitude, time). So it was only with the invention of instruments for analysis and synthesis of running speech that the systematic study of speech perception could begin: the sound spectrograph of R. K. Potter and his colleagues at Bell Telephone Laboratories in New Jersey during World War II, the Pattern Playback of Franklin Cooper at Haskins Laboratories in New York, a few years later. With these devices and their successors, speech research could finally address the first task of all perceptual study: definition of the stimulus, that is, of the physical conditions under which perception occurs.

Yet, a reader unfamiliar with the byways of modern cognitive psychology who chances on this volume may be surprised that speech perception, as a distinct field of study, even exists. Is the topic not subsumed under general auditory perception? Is speech not one of many complex acoustic signals to which we are exposed, and do we not, after all, simply hear it? It is, of course, and we do. But due partly to the peculiar structure of the speech signal and the way it is produced, partly to the peculiar equivalence relation between speaker and hearer, we also do very much more.

To get a sense of how odd speech is, consider writing and reading. Speech is unique among systems of animal communication in being amenable to transduction into an alternative perceptuomotor modality. The more or less continuously varying acoustic signal of an utterance in any spoken language can be transcribed as a visual string of discrete alphabetic symbols, and can then be reproduced from that string by a reader. How we effect the transforms from analog signal to discrete

message, and back again, and the nature of the percept that mediates these transforms are central problems of speech research.

Notice that without the alphabet as a means of notation, linguistics itself, as a field of study, would not exist. But the alphabet is not merely a convenient means of representing language; it is also the primary objective evidence for our intuition that we speak (and language achieves its productivity) by combining a few dozen discrete phonetic elements to form an infinite variety of words and sentences. Thus, the alphabet, recent though it is in human history, is not a secondary, purely cultural aspect of language. The inventors of the alphabet brought into consciousness previously unexploited segmental properties of speech and language, much as, say, the inventors of the bicycle discovered previously unexploited cyclic properties of human locomotion. The biological nature and evolutionary origins of the discrete phonetic categories represented by the alphabet are among many questions on which the study of speech perception may throw light.

To perceive speech is not merely to recognize the holistic auditory patterns of isolated words or phrases, as a bonobo or some other clever animal might do; it is to parse words from a spoken stream, and segments from a spoken word, at a rate of several scores of words per minute. Notice that this is not a matter of picking up information about an objective environment, about banging doors, passing cars, or even crying infants; it is a matter of hearers recognizing sound patterns coded by a conspecific speaker into an acoustic signal according to the rules of a natural language. Speech perception, unlike general auditory perception, is intrinsically and ineradicably intersubjective, mediated by the shared code of speaker and hearer.

Curiously, however, the discrete linguistic events that we hear (segments, syllables, words) cannot be reliably traced in either an oscillogram or a spectrogram. In a general way, their absence has been understood for many years as due to their manner of production: extensive temporal and spectral overlap, even across word boundaries, among the gestures that form neighboring phonetic segments. Yet, how a hearer separates the more or less continuous flow into discrete elements is still far from understood. The lack of an adequate perceptual model of the process may be one reason why automatic speech recognition, despite half a century of research, is still well below human levels of performance.

The ear's natural ease with the dynamic spectro-temporal patterns of speech contrasts with the eye's difficulties: oscillograms are impossible, spectrograms formidably hard, to read – unless one already knows what they say. On the other hand, the eye's ease with the static linear string of alphabetic symbols contrasts with the ear's difficulties: the ear has limited powers of temporal resolution, and no one has ever devised an acoustic alphabet more efficient than Morse code, for which professional rates of perception are less than a tenth of either normal speech or normal reading. Thus, properties of speech that lend themselves to hearing (exactly what they are, we still do not know) are obstacles to the eye, while properties of writing that lend themselves to sight are obstacles to the ear.

Beyond the immediate sensory qualities of speech, a transcript omits much else that is essential to the full message. Most obvious is prosody, the systematic variations in pitch, loudness, duration, tempo, and rhythm across words, phrases, and

sentences that convey a speaker's intentions, attitudes, and feelings. What a transcript leaves out, readers put back in, as best they can. Some readers are so good at this that they become professional actors.

Certain prosodic qualities may be peculiar to a speaker's dialect or idiolect, of which the peculiar segmental properties are also omitted from a standard transcript. What role, if any, these and other indexical properties (specifying a speaker's sex, age, social status, person, and so on) may play in the perception of linguistic structure remains to be seen. I note only that, despite their unbounded diversity within a given language, all dialects and idiolects converge on a single phonology and writing system. Moreover, and remarkably, all normal speakers of a language can, in principle if not in fact, understand language through the artificial medium of print as quickly and efficiently as through the natural medium of speech.

Alphabetic writing and reading have no independent biological base; they are, at least in origin, parasitic on spoken language. I have dwelt on them here because the human capacity for literacy throws the biological oddity of speech into relief. Speech production and perception, writing and reading, form an intricate biocultural nexus at the heart of modem western culture. Thanks to over 50 years of research, superbly reviewed in all its diversity in this substantial handbook, speech perception offers the student and researcher a ready path into this nexus.

<div style="text-align: right">

Michael Studdert-Kennedy
Haskins Laboratories
New Haven, Connecticut

</div>

Preface

The Second Edition of the *Handbook of Speech Perception* presents a collection of essays on the research and theory that have guided our understanding of human speech perception. From their origins in psychoacoustic assessment of phonetics for telecommunication systems, the concerns of research have broadened with the growth of cognitive science and neuroscience. Now truly interdisciplinary in span, studies of speech perception include basic research on the perception of linguistic form while encompassing investigations of multisensory speech perception, speech perception with sensory prostheses, speech perception across the life span, speech perception in neuropathological disorders, as well as the study of the interchange of linguistic, paralinguistic, and indexical attributes of speech. Empirical practice has often turned to speech as a way to assess the potential of a new idea, making speech perception an intellectual crossroad for the subfields that compose contemporary behavioral neuroscience. This intellectual and scientific convergence is also reflected in the topics, large and small, that are represented here. The Second Edition, specifically, showcases new concerns, presents new understanding of lines of classic investigation, and offers a critical assay of technical and theoretical developments across the field of research.

Editors face many decisions in composing a handbook, one that can be useful for student and researcher alike. Early in our discussions, we understood that we would not be creating a comprehensive review of method and theory in research on speech perception. For one reason, technical methods and technical problems evolve rapidly as researchers explore one or another opportunity. For another, the *Annual Reviews* already exist and can satisfactorily offer a snapshot of a field at a particular instant. Aiming higher, we asked each of the contributors to articulate a point of view to introduce the reader to the major issues and findings in the field. The result is a broad-ranging and authoritative collection of essays offering perspectives on exactly the critical questions that are likely to move a rapidly changing field of research.

The twenty-five chapters are organized into five sections. Each essay provides an informed and critical exposition of a topic central to understanding, including: (1) a synthesis of current research and debate; (2) a narrative comprising clear examples and findings from the research literature and the author's own research program; and (3) a forward look toward anticipated developments in the field.

In Part I, *Sensing Speech*, four chapters cover a wide range of foundational issues in the field. Robert Remez discusses the perceptual organization of speech and how it differs from other auditory signals; Lawrence Rosenblum and Josh Dorsi present an argument and evidence for the primacy of multimodal speech perception; Jan Schnupp and Oiwi Parker-Jones describe the representation of speech in the brain; and Kevin Munhall and colleagues explain the role of perception in controlling speech production.

In Part II, *Perception of Linguistic Properties*, eight chapters survey major topics in human speech perception. Shelia Blumstein describes the role of linguistic features in speech perception and lexical access; Keith Johnson and Matthias Sjerps discuss perceptual accommodation of differences between individual talkers; Rajka Smilianic examines the differences between casual and clear speech; Conor McLennan and Sara Incera Burkert present a critical appraisal of specificity effects in spoken word identification; Anne Cutler and Alexandra Jesse discuss the role of lexical stress in the perception of spoken words; Zinny Bond considers speech misperception in an essay on slips of the ear; Michael Vitevitch and Faisal Aljasser assess the contribution of phonotactic knowledge to speech perception; and Diana Van Lancker-Sidtis and Sun-Yeung Yang discuss the implications of the use of formulaic speech.

Part III is devoted to the *Perception of Indexical Properties*, those aspects of the speech of individual talkers that make them identifiable. The five chapters in this section include a discussion of the perception of dialectal variation, by Cynthia Clopper; the resolution of the spoken signals of individual identity, by Diana Van Lancker-Sidtis and Romy Zäske; the integration of linguistic and non-linguistic properties of speech, by Lynne Nygaard and Christina Tzeng; an essay on perceptual learning of accented speech, by Tessa Bent and Melissa Baese-Berk; and an appraisal of the ability of children to notice indexical properties of speech, by Suzanne Levi.

In Part IV, the handbook considers *Speech Perception by Special Listeners*. Susan Nittrouer examines speech perception by children; Mitchell Sommers describes accounts of audiovisual speech perception in older adults; Cynthia Hunter and David Pisoni consider speech perception in prelingually deaf children when a cochlear implant is used; and Emily Myers examines the perception of speech following focal brain injury.

Part V includes four essays each offering a *Theoretical Perspective* on a new or classic concern of the field. Lawrence Raphael provides a detailed retrospective on the acoustic cues to segmental phonetic perception; Jennifer Pardo and Robert Remez offer a review of the influential idea that perception of speech relies on the dynamics of the production of speech; Susan Brady and Axelle Calcus consider the relation between reading and speech perception; and David Pisoni provides a review of the emerging field of cognitive audiology.

The scope of the topics encompassed in the *Handbook of Speech Perception* reflects the wide-ranging research community that studies speech perception. This includes neighboring fields: audiology, speech and hearing sciences, behavioral neuroscience, cognitive science, computer science and electrical engineering,

linguistics, physiology and biophysics, otology, and experimental psychology. The chapters are accessible to nonspecialists while also engaging to specialists. While the *Handbook of Speech Perception* takes a place among the many excellent companion volumes in the Wiley Blackwell series on language and linguistics, the collection is unique in its emphasis on the specific concerns of the perception of spoken language.

If the advent of a handbook can be viewed as a sign of growth and maturity of a discipline, the appearance of this Second Edition is evidence of the longevity of research interest in spoken language. This new edition of the *Handbook of Speech Perception* brings the diverse field together for the researcher who, while focusing on a specific aspect of speech perception, might desire a clearer understanding of the aims, methods, and prospects for advances across the field. In addition to the critical survey of developments across a wide range of research on human speech perception, we also anticipate the *Handbook* facilitating the development of multi-disciplinary research on speech perception.

We cannot conclude without acknowledging the many individuals on whose creativity, knowledge, and cooperation this endeavor depended, namely, the authors whose essays compose the Handbook of Speech Perception. A venture of this scope cannot succeed without the conscientious care of a publisher to protect the project, and we have received the benefit of this attention from many people at Wiley, originating with Tanya McMullin who was instrumental at the start of the project, Angela Cohen, Rachel Greenberg, and Clelia Petracca.

With our sincere thanks,
Jennifer S. Pardo
Bedford, New York

Lynne C. Nygaard
Atlanta, Georgia

Robert E. Remez
New York, New York

David B. Pisoni
Bloomington, Indiana

Part I   Sensing Speech

1 Perceptual Organization of Speech

ROBERT E. REMEZ

Barnard College, Columbia University, United States

How does a perceiver resolve the linguistic properties of an utterance? This question has motivated many investigations within the study of speech perception and a great variety of explanations. In a retrospective summary over 30 years ago, Klatt (1989) reviewed a large sample of theoretical descriptions of the perceiver's ability to project the sensory effects of speech, exhibiting inexhaustible variety, into a finite and small number of linguistically defined attributes, whether features, phones, phonemes, syllables, or words. While he noted many distinctions between the accounts, with few exceptions they exhibited a common feature. Each presumed that perception begins with a speech signal, well composed and fit to analyze. This common premise shared by otherwise divergent explanations of perception obliges the models to admit severe and unintended constraints on their applicability. To exist within the limits set by this simplifying assumption, the models apply implicitly to a world in which speech is the only sound; moreover, only a single talker ever speaks at once. Although this designation is easily met in laboratory samples, it is safe to say that it is rare *in vivo*. Moreover, in their exclusive devotion to the perception of speech the models are tacitly modular (Fodor, 1983), even those that deny it.

Despite the consequences of this dedication of perceptual models to speech and speech alone, there has been a plausible and convenient way to persist in invoking the simplifying assumption. This fundamental premise survives intact if a preliminary process of perceptual organization finds a speech signal, follows its patterned variation amid the effects of other sound sources, and delivers it whole and ready to analyze for linguistic properties. The indifference to the conditions imposed by the common perspective reflects an apparent consensus at the time that the perceptual organization of speech is simple, automatic, and accomplished by generic means. However, despite the rapidly established perceptual coherence of the

constituents of a speech signal, the perceptual organization of speech cannot be reduced to the available and well-established principles of auditory perceptual organization.

Perceptual organization and the gestalt legacy

A generic auditory model of organization

The dominant contemporary account of auditory perceptual organization has been auditory scene analysis (Bregman, 1990). This theory of the resolution of auditory sensation into streams, each issuing from a distinct source, developed empirically in the cognitive era, though its intellectual roots run deep. The gestalt psychologist Wertheimer (1923/1938) established the basic premises of the account in a legendary article, the contents of which are roughly known to all students of introductory psychology. In visible and audible examples, Wertheimer described the coalescence of elementary figures into groups and contours, arguing that sensory experience is organized in patterns, and is not registered as a mere spatter of individual receptor states. By considering a series of hypothetical cases, and without knowing the sensory physiology that would not be described for decades (Mountcastle, 1998), he justified organizing principles of *similarity*, *proximity*, *closure*, *symmetry*, *common fate*, *continuity*, *set*, and *habit*. Hindsight suggests that Wertheimer framed the problem astutely, or so it now seems given our contemporary understanding of the functions of the sensory periphery that integrate the action of visual and auditory receptors (Hochberg, 1974).

Setting the indefinitely elastic principle of habit aside, the simple gestalt-derived criteria of grouping are arguably reducible to two functions: (1) to compose an inventory of sensory elements; and (2) to create contours or groups on the principle that like binds to like. Whether groups occur due to the spectral composition of auditory elements, their common onset or offset, proximity in frequency, symmetry of rate of change in an auditory dimension, harmonic relationship, the interpolation of brief gaps, and so on, each is readily understood as a case in which similarity between a set of auditory sensory elements promotes grouping automatically. A group composed according to these functions forms a sensory contour or perceptual stream. It is a small but necessary extrapolation to assert that an auditory contour consists of elements originating from a single source of sound, and therefore that perceptual organization parses sensory experience into concurrent streams, each issuing from a different sound-producing event (Bregman & Pinker, 1978).

In a series of ongoing experiments, researchers adopted Wertheimer's auditory conjectures, and calibrated the resolution of auditory streams by virtue of the historic principles and their derived corollaries. For example, Bregman and Campbell (1971) reported that auditory streams formed when a sequence of 100 ms tones differing in frequency was presented to listeners. According to a procedure that has become standard, the series of brief tones was presented repetitively to