Audiovisual Speech Processing

When we speak, we configure the vocal tract which shapes the visible motions of the face and the patterning of the audible speech acoustics. Similarly, we use these visible and audible behaviors to perceive speech. This book showcases a broad range of research investigating how these two types of signals are used in spoken communication, how they interact, and how they can be used to enhance the realistic synthesis and recognition of audible and visible speech. The volume begins by addressing two important questions about human audiovisual performance: how auditory and visual signals combine to access the mental lexicon, and where in the brain this and related processes take place. It then turns to the production and perception of multimodal speech, and how structures are coordinated within and across the two modalities. Finally, the book presents overviews and recent developments in machine-based speech recognition and synthesis of AV speech.

GÉRARD BAILLY is a senior CNRS Research Director at the Speech and Cognition Department, GIPSA-Lab, University of Grenoble where he is now Head of Department.

PASCAL PERRIER is a professor in the GIPSA-Lab at the University of Grenoble.

ERIC VATIKIOTIS-BATESON is Professor in the Department of Linguistics and Director of the Cognitive Systems program at the University of British Colombia.

Audiovisual Speech Processing

Edited by

Gérard Bailly, Pascal Perrier, and Eric Vatikiotis-Bateson

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781107006829

First published 2012

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Audiovisual speech processing / [edited by] G. Bailly, P. Perrier,
and E. Vatikiotis-Bateson.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-107-00682-9 (hardback)
I. Bailly, G. (Gérard) II. Perrier, Pascal. III. Vatikiotis-Bateson, Eric.
[DNLM: 1. Speech Perception. 2. Lipreading. 3. Phonetics.
4. Speech – physiology. 5. Visual Perception. WV 272]
616.85′5–dc23

2011053319

ISBN 978-1-107-00682-9 Hardback

Dedicated to Christian Benoît (1956–1998)

Contents

Figures

Tables

Contributors

VIRGINIE ATTINA MARCS Auditory Laboratories, University of Western Sydney – Australia. Dr. Virginie Attina did her doctoral research on the production and perception of Cued Speech at GIPSA-Lab, Grenoble, France. She then worked on the phonology of the deaf exposed to Cued Speech at the University of La Laguna, Spain. In 2007, she joined the Brain Dynamics and Cognition lab (U821, INSERM, Bron, France) to work on new interfaces in the study of cerebral activity. She is now a postdoctoral research fellow at MARCS Auditory Laboratories (UWS, Australia) working on tone languages with Professor D. Burnham. She is more generally interested in multimodal speech production, perception, and integration using a wide range of techniques (motion capture, eyetracking, and electrophysiology).

PIERRE BADIN Speech and Cognition Department, GIPSA-Lab, CNRS & Grenoble University – France. Pierre Badin is a senior CNRS Research Director at the Speech and Cognition Department, GIPSA-lab, Grenoble. Head of the "Vocal Tract Acoustics" team from 1990 to 2002, Associate Director of the Grenoble ICP from 2003 to 2006, he has been Deputy Head of the department since 2007. He has worked in the field of speech communication for more than thirty years. He gained international experience through extended research periods in Sweden, Japan, and UK, and is involved in a number of national and international projects. He is associate editor for speech at *Acta Acustica*, and a reviewer for many international journals. His current interest is speech production and articulatory modeling, with an emphasis on data acquisition, development of virtual talking heads for augmented speech, and speech inversion.

GÉRARD BAILLY Speech and Cognition Department, GIPSA-Lab, CNRS & Grenoble University – France. Gérard Bailly is a senior CNRS Research Director at the Speech and Cognition Department, GIPSA-lab, Grenoble where he is now Head of Department. He has worked in the field of speech communication for more than twenty-five years. He has supervised 20 PhD

Theses and authored 32 journal papers and over 200 book chapters and papers in major international conferences. He coedited *Talking Machines: Theories, Models and Designs* (1992) and *Improvements in Speech Synthesis* (2002). He is associate editor for the *Journal of Acoustics, Speech & Music Processing*, the *Journal of Multimodal User Interfaces* and a reviewer for many international journals. He is a founder member of the ISCA SynSIG and SproSIG special-interest groups. His current interest is multimodal and situated interaction with conversational agents using speech, facial expressions, head movements, and eye gaze.

DENIS BEAUTEMPS Speech and Cognition Department, GIPSA-Lab, CNRS & Grenoble University – France. Denis Beautemps is a CNRS Researcher at the Speech and Cognition Department, GIPSA-lab, Grenoble. He has worked in the field of speech communication for more than ten years. He is now the head of the "Talking Machines, Conversational Agents, Face-to-Face interaction" team. His current interest is multimodal and supplemented speech, fusion of multimodal components. He is particularly studying the perception, production, and automatic processing of Cued Speech.

ATEF BEN YOUSSEF Speech and Cognition Department, GIPSA-Lab, CNRS & Grenoble University – France. Atef Ben Youssef is a PhD student in signal, image, speech, and telecoms at Polytechnic Institute of Grenoble, France. He has an MSc from Grenoble University III, France and Monastir University, Tunisia. He joined GIPSA-Lab in 2008. His thesis focused on talking heads for augmented speech communication and his research interests are in the area of speech analysis and include speech production, synthesis, recognition, and multimodal processing. He works on acoustic-to-articulatory speech inversion using statistical methods, e.g. acoustic phone recognition, articulatory phone synthesis.

LYNNE BERNSTEIN Department of Speech and Hearing Sciences, George Washington University – USA. Lynne E. Bernstein is a Professor in the Speech and Hearing Sciences Department of the George Washington University, Washington, DC. There, she leads the Communication Neuroscience Laboratory. She is also the Program Director for the Cognitive Neuroscience Program of the US National Science Foundation. For almost fifteen years, until the past year, her laboratory was located at the House Ear Institute in Los Angeles, California. Her research has focused on speech perception, lipreading, and multisensory integration in hearing and deaf adults.

JONAS BESKOW Centre for Speech Technology, KTH, Stockholm – Sweden Jonas Beskow is Associate Professor at the KTH Centre for Speech

Technology (Stockholm, Sweden). His research interests are in the areas of visual and acoustic speech processing and synthesis and embodied conversational agents, and application of audiovisual speech technology to facilitation of human communication. One example of this is the SynFace system for real-time lipreading support for hard-of-hearing persons. Jonas has also had a key role in the development of WaveSurfer, an open-source application used by speech and audio researchers worldwide. In 1998–1999 Jonas spent eighteen months at the Perceptual Science Lab at University of California Santa Cruz supported by a Fulbright grant. In 2006 he received the Chester Carlson Award from Xerox and the Royal Swedish Academy of Engineering Science.

CHRISTOPH BREGLER Vision Learning Graphics (VLG) Group, Courant Institute, New York University – USA Chris Bregler is an Associate Professor of Computer Science at NYU's Courant Institute. He received his MS and PhD in Computer Science from UC Berkeley in 1995 and 1998 and his Diplom from Karlsruhe University in 1993. Prior to NYU he was on the faculty at Stanford University and worked for several companies including Hewlett Packard, Interval, Disney Feature Animation, and LucasFilm's ILM. He founded the Stanford Movement Group and the NYU Movement Group, which does research in vision and graphics with a focus on motion capture, animation, interactive media, and applications to entertainment, art, and medicine. This resulted in numerous publications, patents, and awards from the National Science Foundation, Sloan Foundation, Packard Foundation, Office of Naval Research, Electronic Arts, Microsoft, and other sources. He was named Stanford Joyce Faculty Fellow and Terman Fellow in 1999. He received the Olympus Prize for achievements in computer vision and AI in 2002, and was named a Sloan Research Fellow in 2003. He was the chair for the SIGGRAPH 2004 Electronic Theater and Computer Animation Festival. At CVPR 2008 he was awarded the IEEE Longuet-Higgins Prize for "Fundamental Contributions in Computer Vision that have withstood the test of time."

N. MICHAEL BROOKE Lately with Department of Computer Science, University of Bath, Bath – UK. Lately Reader in Computing at the University of Bath and active in the field of audiovisual speech processing since 1978, N. Michael Brooke has been a visiting scientist at the MRC Institute of Hearing Research in Nottingham, at AT&T Bell Laboratories, Murray Hill, NJ, and at the Speech Unit of DRA Malvern. He was, with C. Benoît, a Co-Director of the NATO Advanced Study Institute meeting on Speech reading held at Bonas in 1995. He has lectured widely in the USA and Europe and has also been secretary, then chairman, of the Speech Group of the UK Institute of Acoustics. He retired in 1999.

VICKI BRUCE School of Psychology, Newcastle University, Newcastle upon Tyne – UK. Vicki Bruce graduated from Newnham College, University of Cambridge in 1974 with a BA in Natural Sciences and completed her PhD "Processing and remembering pictorial information" in 1977 at the MRC Applied Psychology Unit, supervised by Alan Baddeley. Bruce worked briefly as a demonstrator at Newcastle University before moving to the University of Nottingham as a Lecturer in 1978, where she was promoted to Reader in 1988 and Professor in 1990. In 1992 she moved to University of Stirling, where she was Deputy Principal for Research from 1995 until 2002. From 2002 to 2008 she was Vice Principal and Head of the College of Humanities and Social Science at the University of Edinburgh. In 2008 she became Head of the School of Psychology at Newcastle University. She is known for her work on human face perception and person memory, including face recognition and recall by eye-witnesses and gaze and other aspects of social cognition. She is also interested in visual cognition more generally.

DENIS BURNHAM MARCS Institute, University of Western Sydney – Australia. Following an honours degree (University of New England, Australia, 1974), and PhD (Monash University, 1980), Burnham was in the School of Psychology at the University of NSW for eighteen years before taking up the position of Inaugural Director of MARCS Auditory Laboratories, Sydney, Australia. MARCS Labs has over sixty members and specializes in speech and music research. Burnham's own research concerns five areas of speech perception and production: *Ontogenetic* studies of infants' and children's speech perception development; phonetic, attentional, and emotional aspects of *Special Speech Registers* to infants, foreigners, pets, lovers, and computers; *Cross-Language* studies on the relationship between speech perception and vocabulary, reading, and second language learning; *Auditory-Visual* speech perception studies with infants, children, and adults within and across languages; and *Lexical Tone*, an understudied but prevalent speech feature affording investigations of speech-music, and segmental-suprasegmental relationships. Burnham is President, Australasian Speech Science and Technology Association (ASSTA); Executive member, ARC Research Network on Human Communication Science (HCSNet); member of the International Advisory Council (IAC) and Interspeech Steering Committee (ISC) of the International Speech Communication Association (ISCA); and Co-founder/Vice-Chair, Auditory-Visual Speech Perception Association (AVISA). Burnham has been funded continuously by the Australian Research Council (ARC) and other grant bodies for the past twenty-six years, and has directed various large inter-disciplinary and international research projects. Currently he is lead

Chief Investigator (CI) on an ARC Discovery grant on tone perception (2009–2012); lead CI of an ARC and National Health and Medical Research Council Special Initiative, the $3.4M Thinking Head project (2005–2011) encompassing Embodied Conversational Agents and Human-Robot Interaction; and lead CI of the $1M ARC infrastructure project, the Big Australian Speech Corpus (Big ASC, 2010), in which 3 hours of speech data from 1000 people all around Australia are being collected.

RUTH CAMPBELL Deafness Cognition and Language (DCAL) Research Centre, UCL – UK. Ruth Campbell retired from University College London, where she was Professor of Communication Disorder from 1996 to 2008. Her previous appointments were at Goldsmiths College, University of London, and Oxford University. She trained as a cognitive psychologist and neuropsychologist and has had interests in the processes underlying speechreading for thirty years.

MARIE-AGNÈS CATHIARD Center for Research on the Imaginary (CRI), Grenoble University – France. Marie-Agnès Cathiard, Doctor in Cognitive Psychology of Mendès-France University (Grenoble), is Maître de Conférences in Phonetics and Cognition at Stendhal University (Grenoble). Her main interests are in visual and audiovisual speech perception and in the production and perception of French Cued Speech. Her recent projects at the Center for Research on the Imaginary (CRI Stendhal) deal with body-part illusions in speech (face and hand), within the cognitive framework concerned with limb and body phantom phenomena. International publications in: *Speech Communication* (Ed.), *Revue Parole*, *AMSE-Journals*, *Behavioral and Brain Sciences*, *NeuroImage*, *Perception & Psychophysics*; and in books: *Phonetics, Phonology, and Cognition* (2002); *Lecture Notes in Artificial Intelligence* (2006); *Emergence of Linguistic Abilities* (Cambridge, 2008); *Speech Motor Control* (2010); *Cued Speech* (2010).

RASHID CLARK Rashid Clark is currently an independent contractor, and was a part of the Perceptual Science Laboratory at the University of California, Santa Cruz from 1998–2002, contributing as a tools developer. He currently works in the gaming industry as a software engineer. His more general interests include merging the technical and the aesthetic, and currently develops the internet cartoon series, *The Cosmospolitans*. Rashid received his BA in Cybernetics from the University of California, Los Angeles in 1995.

MICHAEL M. COHEN Michael M. Cohen is a Principal Investigator on a SBIR research grant iGlasses: an automatic wearable speech supplement in

face-to-face communication and classroom situations, awarded to Animated Speech Corporation, now TeachTown. He was previously a research associate in the Program in Experimental Psychology at the University of California – Santa Cruz. His research interests include speech perception and production, speechreading, information integration, learning, and computer facial animation. He received a BS in Computer Science and Psychology (1975) and an MS in Psychology (1979) from UW-Madison, and a PhD in Experimental Psychology (1984) from UC-Santa Cruz.

TONY EZZAT MIT – CBCL, McGovern Institute – USA. Tony Ezzat is a Research Affiliate at the Center for Biological and Computational Learning at MIT, as well as a Principal Scientist at Kayak.com. His research interests span a wide variety of topics including speech processing, auditory neuroscience, computer vision, computer graphics, and machine learning. His Thesis work on audiovisual speech synthesis received the Christian Benoit Prize in 2002.

GADI GEIGER MIT – CBCL, McGovern Institute – USA. Gadi Geiger is a Research Scientist at the Center for Biological and Computational Learning at MIT. His research is on visual and auditory perception. Over the last two decades his research has focused on visual and auditory perceptual aspects in dyslexia.

RAFAËL LABOISSIÈRE Space and Action, U864, INSERM/Claude Bernard University, Lyon 1 – France. Dr Rafael Laboissière was awarded his PhD by the Institut National Polytechnique de Grenoble in 1992 since when he has been a CNRS researcher in France. He worked on biomechanical and neurophysiological models of speech production at the Institut de la Communication Parlée, Grenoble till 2001. Between 2001 and 2005 he was the head of the Sensorimotor Coordination team at the Max Planck Institute for Human Cognitive and Brain Sciences in Munich, Germany. His current research interests involve human motor control and multisensory integration.

KAREN LANDER Department of Psychology, University of Manchester – UK. Dr. Karen Lander received her PhD from the University of Stirling in 1999. She then worked as a Research Fellow for one year, at the University of Stirling, on an ESRC grant awarded to Professor Vicki Bruce and herself. She has been a lecturer at the Department of Psychology, University of Manchester, since January 2001. She was promoted to Senior Lecturer in 2008, and has worked extensively on the role of movement in the recognition and learning of faces and more generally in face perception.

HÉLÈNE LOEVENBRÜCK Speech and Cognition Department, GIPSA-Lab, CNRS & Grenoble University – France. Hélène Loevenbruck is currently CNRS Research Associate at the Speech and Cognition Department of GIPSA-Lab. She received an engineering training in electronics, signal processing, and computer science and obtained an MSc in cognitive sciences in 1992. After her PhD at the former Institut de la Communication Parlée on articulatory control, she was assistant lecturer in computer science and obtained a second MSc in phonetics. In 1997–98 she spent a post-doctoral year at the Department of Linguistics of the Ohio State University in the United States. She was awarded a bronze medal from the CNRS in 2006 for her works on the neural correlates of vocal pointing. Her research areas include (i) the acoustic, articulatory, perceptual, and neural correlates of prosody, (ii) language development (prosody, phonology, crosslinguistic differences), (iii) auditory verbal hallucination in schizophrenia (EMG and fMRI), (iv) murmured, silent, and inner speech (NAM, EMG, ultrasound), (v) self/other speech perception.

JUERGEN LUETTIN Robert Bosch GmbH, Corporate Research (CR/AEA5) – Germany. Dr. Juergen Luettin was awarded his PhD by the University of Sheffield, UK in 1997. From 1997 to 2000 he headed the computer vision group at IDIAP (Dalle Molle Institute for Perceptual Artificial Intelligence), Switzerland, acquired several EU and national research projects in speech and image processing and has been an invited scientist at Johns Hopkins University, USA. From 2000 to 2002 he headed the pattern recognition group at Ascom AG, Switzerland and was responsible for research and development in video surveillance products. In 2002 he joined the video-based driver assistance systems division at Robert Bosch GmbH, Stuttgart. His current work is on electromobility and system development and he is also a consultant and instructor for systems and requirements engineering. Dr. Luettin has published over forty scientific articles.

MAIRÉAD MACSWEENEY Institute of Cognitive Neuroscience, University College London – UK. Dr. Mairéad MacSweeney is Wellcome Trust Career Development Research Fellow, based at the Institute of Cognitive Neuroscience, University College London. She trained as a psychologist and has a particular interest in the neurobiological basis of language processing in people born profoundly deaf. This includes sign language processing, reading, and speechreading.

DOMINIC W. MASSARO Department of Psychology, University of California – USA. Dominic W. Massaro is Professor of Psychology and Computer Engineering at the University of California, Santa Cruz. He is best known for his fuzzy logical model of perception, and more recently, for his

development of the computer-animated talking head Baldi. Massaro is Director of the Perceptual Science Laboratory, past president of the Society for Computers in Psychology, book review editor for the *American Journal of Psychology*, founding Chair of UCSC's Digital Arts and New Media program, and was founding co-editor of the interdisciplinary journal *Interpreting*. He has been a Guggenheim Fellow, a University of Wisconsin Romnes Fellow, a James McKeen Cattell Fellow, an NIMH Fellow, and in 2006 was recognized as a Tech Museum Award Laureate. Massaro received his BA in Psychology from the University of California, Los Angeles in 1965, and completed his PhD in Mathematical Psychology at the University of Massachusetts, Amherst in 1968. After his postdoctoral work at the University of California, San Diego, he was Professor of Psychology at the University of Wisconsin, Madison from 1970 to 1979, before moving to UCSC where he has remained since. Massaro's research focuses on applying an information processing approach to the study of language, perception, memory, cognition, and decision making. In collaboration with Gregg Oden, he developed the fuzzy logical model of perception, which stresses the integration of multiple sources of information when modeling perception. Stemming from this early work, Massaro established a research program demonstrating the importance of information from the face in speech perception. As part of this program, Massaro, along with researcher Michael Cohen, developed the computer-animated talking head known as Baldi. The Baldi technology is special in its extraordinary accuracy, and has been expanded to speak in numerous languages. In recent years, Massaro has become more involved with applied research, using his talking head technology to benefit language learners, including those facing learning challenges such as deafness and autism. For this work, he was named a 2006 Tech Microsoft Education Award Laureate by the Tech Museum of Innovation.

IAIN MATTHEWS Disney Research, Pittsburgh – USA. Iain Matthews is a Senior Research Scientist at Disney Research, Pittsburgh, where he leads the computer vision group. He also holds an adjunct appointment at The Robotics Institute, Carnegie Mellon University. Prior to joining Disney he spent two years at Weta Digital working on the facial motion capture system for James Cameron's *Avatar*, and the upcoming Spielberg/Jackson movie *Tin Tin*. Between 1999–2007 he was post-doc then systems faculty at Carnegie Mellon University, working on real-time, high fidelity, non-rigid face tracking. He obtained his PhD in Audio-Visual Speech Recognition from the University of East Anglia in 1998.

KEVIN MUNHALL Departments of Psychology and Otolaryngology, Queen's University, Kingston, Canada Kevin Munhall is the director of the Queen's

Biological Communication Centre. He did his undergraduate studies at the University of Waterloo and his graduate work at McGill University under the supervision of David Ostry. He was a post-doctoral fellow at Haskins Laboratories from 1984–86 and following that he was a Research Scientist at Haskins. Munhall taught at York University in Toronto and the University of Western Ontario before taking up a position at Queen's University in 1990. His research has focused on coordination in articulation, auditory feedback in speech production and audiovisual speech communication.

CHALAPATHY NETI Human Language Technologies Department, IBM Thomas J. Watson, Research Center – USA. Chalapathy Neti is currently the Associate Director and Global Leader, Healthcare Transformation, at IBM Research. Previous to this, he was an Executive Architect in the Information Agenda organization, IBM Software Group where he consulted with healthcare institutions to develop an information management strategy for improved outcomes and efficiency. Prior to his assignment in Software Group, Chalapathy Neti was a Senior Manager for Information Analysis and Interaction technologies at IBM Research. In this role, he managed several research groups involved in developing text, image, and video analysis technologies and its application to bio-medical informatics, medical imaging, and real-time decision intelligence (e.g. customer intelligence, fraud detection, and video surveillance. Before taking on the senior management role, he held various senior technical and management positions including the CTO of IBM's Digital Media business, manager of audiovisual speech technologies and technical roles in rich media analysis (speech, audio, and video) and mining. He has been with IBM since 1990. Chalapathy Neti received his PhD degree in Biomedical Engineering from the Johns Hopkins University (1990), and BS degree from Indian Institute of Technology, Kanpur (1980). He has more than twenty years of advanced R&D experience and has authored over fifty articles (conference and journal) in various fields related to bio-medical informatics, medical imaging, speech and video analysis, computational neuroscience and VLSI design. He has sixteen patents and several pending. He is an active member of IEEE, a former member of IEEE Multimedia Signal Processing technical committee (2001–2004), an associate editor of *IEEE transactions on Multimedia* (2002–2005) and a guest editor for the *IBM Systems Journal*. He is currently a member of the Executive Review Board of MI3C (Medical Imaging Informatics Innovation Center – a collaboration between IBM and Mayo Clinic).

PASCAL PERRIER Speech and Cognition Department, GIPSA-Lab, CNRS & Grenoble University – France. Pascal Perrier was awarded a PhD in Electronic Systems by the Institut National Polytechnique de Grenoble in

1982 and an HDR (Accreditation in Research Supervision) in Speech Communication in 1990. He is currently a Professor at Grenoble INP, where he teaches digital signal processing, digital communication, and speech processing. He is a member of the research laboratory GIPSA-Lab, where he was the head of the research group on Acoustics Aeroacoustics, Biomechanics and Control from 2005 to 2009. He co-authored 50 peer-reviewed international journal articles or book chapters, and around 100 papers in international conference proceedings. He is co-editor of the book *Some Aspects of Speech and the Brain* published in 2009. He has established collaborative projects with labs in different parts of the world, in particular the Centre for General Linguistics (ZAS) in Berlin, the Speech Communication group at MIT in Cambridge, USA, the Phonetic Laboratory at UQAM in Montréal, Canada, the JAIST in Ishikawa, Japan, and the MAGIC lab at UBC in Vancouver, Canada. He was Chair of AFCP (Association Francophone de la Communication Parlée), Special Interest Group of ISCA, from 2005 to 2006. His research interests include speech motor control, biomechanical modeling of speech articulators, and the interaction between physical and linguistic constraints in speech production.

TOMASO A. POGGIO Massachusetts Institute of Technology, McGovern Institute for Brain Research, Brain and Cognitive Sciences Department – USA. Tomaso A. Poggio is the Eugene McDermott Professor at the Department of Brain and Cognitive Sciences; Co-Director, Center for Biological and Computational Learning (CBCL); member of the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT; and a member of the faculty of the McGovern Institute for Brain Research at MIT. Previous to his appointments at MIT, he spent ten years at the Max Planck Institut für Biologische Kybernetik, Tübingen, Germany. Professor Poggio's current research focuses on mathematical theories of learning and on computational models of brain function with the goal to understand human intelligence and to build intelligent machines that can mimic human performance. Professor Poggio is a Founding Fellow of AAAI, a Foreign Member of the Italian Academy of Sciences, a Fellow of the American Academy of Arts and Sciences and a member of the American Association for the Advancement of Science (AAAS). He was awarded the Otto-Hahn-Medaille Award of the Max-Planck-Society, the Max Planck Research Award (with M. Fahle), from the Alexander von Humboldt Foundation, the MIT 50K Entrepreneurship Competition Award, the Laurea Honoris Causa from the University of Pavia in 2000 (Volta Bicentennial), the 2003 Gabor Award, and the 2009 Okawa Prize.

GERASIMOS POTAMIANOS Institute of Informatics and Telecommunications, NCSR "Demokritos" – Greece. Dr. Gerasimos Potamianos was awarded his

PhD in Electrical and Computer Engineering from Johns Hopkins University, in Baltimore, Maryland in 1994. Since then, he has worked in the US at the Center for Language and Speech Processing at Johns Hopkins, at AT&T Labs – Research, and at the IBM T. J. Watson Research Center, and in Greece at the Institute of Computer Science (ICS) at FORTH, Crete. He is currently Research Director at the National Center for Scientific Research, "Demokritos" in Athens, Greece. His interests span the areas of multimodal speech processing with applications to human–computer interaction and ambient intelligence, with particular emphasis on audiovisual speech processing, automatic speech recognition, multimedia signal processing and fusion, as well as multimodal scene analysis. He has published over ninety articles in these areas and holds seven patents.

ROBERT E. REMEZ Department of Psychology, Barnard College, Columbia University – USA. Robert E. Remez is Professor of Psychology at Barnard College of Columbia University. He is Chair of the Columbia University Seminar on Language & Cognition. He has been an Associate Editor of the journals *Perception & Psychophysics* and *Journal of Experimental Psychology: Human Perception and Performance*, and he is co-editor, with David Pisoni, of the *Handbook of Speech Perception*. A Fellow of the Acoustical Society of America, the Association for Psychological Science, the American Association for the Advancement of Science, and the American Psychological Association, his research reports have appeared in the *American Psychologist*, *Developmental Psychology*, *Ear & Hearing*, *Journal of Experimental Psychology*, *Journal of Phonetics*, *Journal of the Acoustical Society*, *Memory & Cognition*, *Attention*, *Perception & Psychophysics*, *Psychological Review*, *Psychological Science*, *Psychonomic Bulletin & Review*, *Scandinavian Journal of Psychology, Science,* and *Speech Communication*.

LIONEL REVÉRET Jean Kuntzman Laboratory (LJK), INRIA – Grenoble University – France. Lionel Revéret is an INRIA researcher. He is with the EVASION project (formerly iMAGIS), in the Rhone-Alpes Research Unit of the INRIA located near Grenoble. He was awarded his PhD by the Institute of Spoken Communication (ICP) of the INPG University in Grenoble in 1999. His doctoral thesis dealt with issues in analysis and synthesis of lip motion during speech production, combining video tracking, and 3D modeling. Part of this doctoral research was spent as a visiting scholar at the ATR-HIP laboratories in Japan. He has since been involved in a France Télécom research project to extend visual speech analysis and synthesis to the whole face, using 3D textured model of the head and articulatory control (project MOTHER). From July 2000 to September 2001 as a Post-Doctoral Researcher at the Computational Perception Laboratory (CPL) at

GeorgiaTech he worked with Professor Irfan Essa on combining studies on visual speech and facial expressions. He is currently working on video-based motion capture, in particular in the measurement of animal motion and plant motion under wind effect.

CHRISTOPHE SAVARIAUX Speech and Cognition Department, GIPSA-Lab, CNRS & Grenoble University – France. Christophe Savariaux is a CNRS research engineer in charge of the Stendhal experimentation platform. This platform is dedicated to the acquisition of multimodal speech data using different techniques, such as multiple cameras, electromagnetic articulography, and aerodynamic/physiologic sensors. Christophe Savariaux works in collaboration with the teams of the Speech and Cognition Department for which he has developed different software toolkits for data visualization and analysis. His current interest is speech production, in particular the recovery of phonetic contrasts after temporary or permanent impairment of speech articulators.

JEAN-LUC SCHWARTZ Speech and Cognition Department, GIPSA-Lab, CNRS & Grenoble University – France. Dr. Jean-Luc Schwartz is a senior CNRS Research Director at the Speech and Cognition Department, GIPSA-Lab, Grenoble. He was in charge of ICP (Institut de la Communication Parlée) from 2003 to 2006. His main areas of research involve speech and cognition, and more precisely auditory modeling, audiovisual speech perception, perceptuo-motor interactions, speech robotics, and the emergence of language. He has been involved in many national and European projects, and responsible for a number of them (for example, within the CNRS program ROBEA, within the "Complex Systems in Human and Social Sciences" program, and within the European Science Foundation "Origin of Man, Language and Languages" program). He has coordinated a number of special issues of journals such as *Speech Communication*, *Interaction Studies* and *Primatologie*, and organized several international workshops on Audiovisual Speech Processing, Language Emergence, or Face to Face Communication. He has authored or co-authored more than fifty publications in international journals (including *JASA*, *Speech Communication*, *Computer Speech and Language*, *IEEE Transactions on Speech and Audio Processing*, *Interaction Studies*, *Journal of Phonetics*, *Behavioural and Brain Research*, *Perception and Psychophysics*, *Cognition*, *NeuroImage*), and about thirty book chapters.

SIMON D. SCOTT Centre of Innovation & Technology (CTI), Asia Pacific University College of Innovation & Technology – Malaysia. Professor Dr Simon David Scott is the Chief Technologist of the Centre of Innovation & Technology (CTI), Asia Pacific University College of Innovation &

Technology. Under his direction, CTI (www.cti.my) has been awarded over RM4 million in research funding from the Malaysian Ministry of Science, Technology and Innovation and has produced numerous commercial products, winning various national and international awards including eight APICTA and two PIKOM Product of the Year awards. Previous positions include CTO of Guardware Ltd., Visiting Research Fellow at British Telecom's BTexact Asian Research Centre, and Research Officer at the University of Bath, UK.

KAORU SEKIYAMA Faculty of Letters, Kumamoto University – Japan. Kaoru Sekiyama was awarded her PhD by Osaka City University, Osaka, Japan. She spent some time at the City University of New York and Massachusetts Institute of Technology to collect data for her cross-linguistic studies while she was an assistant professor at Kanazawa University (Kanazawa, Japan). After working at Future University (Hakodate, Japan), she has been Professor of Cognitive Psychology at Kumamoto University (Kumamoto, Japan), since 2006. Her research interests include auditory-visual speech perception, experience-based modification of cognition, development, aging, brain plasticity, adaptation to reversed vision, body schema, visuomotor coordination, and crossmodal perception. Concerning speech perception, she has studied developmental and cross-linguistic examinations of auditory-visual speech perception with behavioral and neural measures.

MALCOM SLANEY Yahoo! Research, Santa Clara, USA. Malcolm Slaney is a principal scientist at Yahoo! Research Laboratory working on all types of multimedia data. He received his PhD from Purdue University for his work on computed imaging. He is a co-author, with A. C. Kak, of the IEEE book *Principles of Computerized Tomographic Imaging*. This book was republished by SIAM in their "Classics in Applied Mathematics" series. He is co-editor, with Steven Greenberg, of the book *Computational Models of Auditory Function*. Before Yahoo!, Dr. Slaney worked at Bell Laboratory, Schlumberger Palo Alto Research, Apple Computer, Interval Research, and IBM's Almaden Research Center. He is also a (consulting) Professor at Stanford's CCRMA where he organizes and teaches the Hearing Seminar. His research interests include auditory modeling and perception, multimedia analysis and synthesis, compressed-domain processing, similarity and search, and machine learning. For the last several years he has led the auditory group at the Telluride Neuromorphic Workshop. He is a Fellow of the IEEE.

MARIJA TABAIN Linguistics Program, La Trobe University, Melbourne – Australia. Marija Tabain is senior lecturer in phonetics at La Trobe