

Editorial

Werner Verhelst

Department of Electronics and Information Processing, Vrije Universiteit Brussel, 1050 Brussel, Belgium
Email: wverhels@etro.vub.ac.be

Jürgen Herre

Fraunhofer Institute for Integrated Circuits (IIS), 91058 Erlangen, Germany
Email: hrr@iis.fhg.de

Gernot Kubin

Signal Processing and Speech Communication Laboratory, Graz University of Technology, 8010 Graz, Austria
Email: g.kubin@ieee.org

Hynek Hermansky

IDIAP Research Institute, 1920 Martigny, Switzerland
Email: hynek@idiap.ch

Søren Holdt Jensen

Department of Communication Technology, Institute of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7A, DK-9220 Aalborg, Denmark
Email: shj@kom.aau.dk

Anthropomorphic systems process signals “at the image of man.” They are designed to solve a problem in signal processing by imitation of the processes that accomplish the same task in humans. In the area of audio and speech processing, remarkable successes have been obtained by anthropomorphic systems: perceptual audio coding even caused a landslide in the music business.

At first sight, it could seem obvious that the performance of audio processing systems should benefit from taking into account the perceptual properties of human audition. For example, front ends that extract perceptually meaningful features currently show the best results in speech recognizers. However, their features are typically used for a stochastic optimization that is itself not anthropomorphic at all. Thus, it is not obvious why they should perform best, and perhaps the truly optimal features have not yet been found because, after all, “airplanes do not flap their wings.”

In general, we believe that there are several situations when an anthropomorphic approach may not be the best solution. First, its combination with nonanthropomorphic systems could result in a suboptimal overall performance (the quantization noise that was cleverly concealed by a perceptual audio coder could become unmasked by subsequent linear or nonlinear processing). Second, other approaches that are not anthropomorphic might be better adapted to the

technology that is chosen for the implementation (airplanes do not flap their wings because it is technically much more efficient to use jet engines for propulsion). Nevertheless, a lot can be learned from imitating natural systems that were optimized through natural selection. As such, anthropomorphic and, by extension, biomorphic systems can be considered to play an important role in the process of developing new technologies.

This special issue brings together a dozen papers from different areas of audio and speech processing that deal with aspects of anthropomorphic processing or in which an anthropomorphic or perceptual approach was taken.

The first of two papers on perceptual audio coding proposes a perceptual model for the specific distortion that is typically encountered in sinusoidal modelling, while the second paper introduces a novel parametric stereo coding technique based on binaural psychoacoustics. While these papers illustrate the use of human auditory perception for efficient audio coding, the three following papers present examples of efforts towards using different levels of neurophysiologic modelling directly for the representation and processing of audio signals: from a model for the adaptation behaviour in the chemical synapses between the inner hair cells and the auditory neurons, to a signal processing model for the early auditory system, and then a cortical audio representa-

tion for sound modification. In the last pair of audio papers, signal features that are based on our knowledge of the auditory system are used in conjunction with machine learning techniques, such as neural networks, to achieve more cognitive goals, such as audio source separation and classification.

A generally applicable technique that allows for discriminative training of hidden Markov models is introduced and applied on the confusable set of visemes for lip reading purposes in the first of five papers on speech processing. The next three of these papers all deal with the important problem of finding objective distortion measures for speech, and the last paper describes an articulatory speech synthesizer that, among other things, brought a better understanding of the Portuguese nasal vowels.

While the papers in this special issue can represent only a small sampling of anthropomorphic techniques in audio and speech processing, they are all very valuable in their own right and together, if nothing else, they show that anthropomorphic sound processing systems are invaluable in the form of computational models for human perception and that they can fuel our quest for further understanding of human nature and self-knowledge.

*Werner Verhelst
Jürgen Herre
Gernot Kubin
Hynek Hermansky
Søren Holdt Jensen*

Werner Verhelst obtained the Engineering degree, Burgerlijk Werktuigkundig Electrotechnisch Ingenieur, in 1980, and the Ph.D. degree in 1985, both from the Vrije Universiteit Brussel, Belgium. He specialised in digital speech and audio processing in general, and in speech and audio signal modification in particular. Verhelst also studied speech synthesis at the Institute for Perception Research, and audio signal modelling at the Katholieke Universiteit Leuven, Belgium. Since his graduation, he has been with the Vrije Universiteit Brussel where he is heading the Research Laboratory on Digital Speech and Audio Processing (DSSP) and teaching courses on digital signal processing and speech and audio processing.



Jürgen Herre joined the Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany, in 1989. Since then he has been involved in the development of perceptual coding algorithms for high-quality audio, including the well-known ISO/MPEG-Audio Layer III coder (aka "MP3"). In 1995, Dr. Herre joined Bell Laboratories for a postdoc term working on the development of MPEG-2 advanced audio coding (AAC).



Since the end of 1996, he has returned to Fraunhofer to work on the development of advanced multimedia technologies including MPEG-4, MPEG-7, and secure delivery of audiovisual content. Currently he is the Chief Scientist for the audio/multimedia activities at Fraunhofer Institute for Integrated Circuits (IIS), Erlangen. Dr. Herre is a Fellow of the Audio Engineering Society, Cochair of the AES Technical Committee on Coding of Audio Signals, and Vice Chair of the AES Technical Council. He also served as an Associate Editor of the IEEE Transactions on Speech and Audio Processing and is an active member of the MPEG audio subgroup.

Gernot Kubin was born in Vienna, Austria, on June 24, 1960. He received the Dipl.-Ing. degree in 1982, and Dr. Techn. degree (sub auspiciis praesidentis) in 1990, both in electrical engineering from TU Vienna. He has been a Professor of nonlinear signal processing and the Head of the Signal Processing and Speech Communication Laboratory (SPSC), Graz University of Technology, Austria, since September 2000. Earlier international appointments include CERN, Geneva, Switzerland, 1980; TU Vienna, from 1983 to 2000; Erwin Schroedinger Fellow at Philips Natuurkundig Laboratorium, Eindhoven, The Netherlands, 1985; AT&T Bell Labs, Murray Hill, USA, from 1992 to 1993, and 1995; KTH, Stockholm, Sweden, 1998; Vienna Telecommunications Research Centre (FTW) from 1999 up to date as Key Researcher and Member of the Board; Global IP Sound, Sweden and USA, from 2000 to 2001 as a Scientific Consultant; Christian Doppler Laboratory for Nonlinear Signal Processing from 2002 up to date as the Founding Director. Dr. Kubin is a Member of the Board of the Austrian Acoustics Association and Vice Chair for the European COST Action 277, Nonlinear Speech Processing. He has authored or coauthored over ninety peer-reviewed publications and three patents.



Hynek Hermansky works at the IDIAP Research Institute, Martigny, Switzerland. He has been working in speech processing for over 30 years, previously as a Research Fellow at the University of Tokyo, a Research Engineer at Panasonic Technologies, Santa Barbara, California, a Senior Member of the research staff at US WEST Advanced Technologies, and a Professor and Director of the Center for Information Processing, OHSU, Portland, Oregon. He is a Fellow of the IEEE for the "invention and development of perceptually-based speech processing methods," a Member of the Board of the International Speech Communication Association, and a Member of the Editorial Boards of Speech Communication and of Phonetica. He holds 5 US patents and authored or coauthored over 130 papers in reviewed journals and conference proceedings. He holds a Dr. Eng. degree from the University of Tokyo, and Dipl.-Ing. degree from Brno University of Technology, Czech Republic. His main research interests are in acoustic processing for speech and speaker recognition.



Søren Holdt Jensen received the M.S. degree in electrical engineering from Aalborg University, Denmark, in 1988, and the Ph.D. degree from the Technical University of Denmark, in 1995. He has been with the Telecommunications Laboratory of Telecom Denmark, the Electronics Institute of the Technical University of Denmark, the Scientific Computing Group of the Danish Computing Center for Research and Education (UNI-C), the Electrical Engineering Department of Katholieke Universiteit Leuven, Belgium, the Center for PersonKommunikation (CPK) of Aalborg University, and is currently an Associate Professor in the Department of Communication Technology, Aalborg University. His research activities are in digital signal processing, communication signal processing, and speech and audio processing. Dr. Jensen is a Member of the Editorial Board of the EURASIP Journal on Applied Signal Processing, and a former Chairman of the IEEE Denmark Section and the IEEE Denmark Section's Signal Processing Chapter.

