

Editorial

Intelligent Audio, Speech, and Music Processing Applications

Woon S. Gan,¹ Sen M. Kuo,² and John H. L. Hansen³

¹ *Digital Signal Processing Laboratory, School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798*

² *Department of Electrical Engineering, Northern Illinois University, Dekalb, IL 60115, USA*

³ *Center for Robust Speech Systems (CRSS), Department of Electrical Engineering, Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, Richardson, TX 75083, USA*

Correspondence should be addressed to Woon S. Gan, ewsgan@ntu.edu.sg

Received 3 November 2008; Accepted 3 November 2008

Copyright © 2008 Woon S. Gan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Future audio, speech, and music processing applications need innovative intelligent algorithms that allow interactive human/environmental interfaces with surrounding devices/systems in real-world settings to control, process, render, and playback/project sound signals for different platforms under a diverse range of listening environments. These intelligent audio, speech, and music processing applications create an environment that is sensitive, adaptive, and responsive to the presence of users. Three areas of research are considered in this special issue: analysis, communication, and interaction. Analysis covers both preprocessing of sound signals and extraction of information from the environment. Communication covers the transmission path/network, coding techniques, and conversion between spatial audio formats. The final area involves intelligent interaction with the audio/speech/music environment based on the users' location, signal information, and acoustical environment.

This special issue on intelligent audio, speech, and music processing (IASMP) consists of 13 papers that reflect a diverse range of disciplines in speech, audio, and music processing. These papers are grouped under analysis, communication, and interaction areas.

Under the analysis grouping, the first paper is "Phasor representation for narrowband active noise control systems" by Fu-Kun Chen et al. This paper uses signal phasors to analyze behavior of two-tap adaptive filters for canceling narrowband noise, and proposes a best signal basis to improve both the convergence speed and steady-state performance. The second paper is "On a method for improving impulsive sounds localization in hearing defenders" by Benny Sällberg et al. This study presents a new algorithm to enhance perceived directionality of active

hearing defenders used in police and military applications. The algorithm uses interaural level difference to enhance spatial information without increasing the impulse sound levels. "Auditory sparse representation for robust speaker recognition based on tensor structure," by Qiang Wu et al., looks into using a non-negative tensor principal component analysis for speech feature extraction. By encoding the speech in higher-order tensors, discriminative features can be extracted in the spectral-temporal domain to increase accuracy in speaker recognition in noisy environments. The next paper entitled "Towards an intelligent acoustic front-end for automatic speech recognition: built-in speaker normalization (BISN)," by Umit Yapanel and John Hansen, proposes a novel online vocal track length normalization algorithm entitled built-in speaker normalization. This algorithm unifies the nonlinear frequency warping function and speaker variability due to vocal tract length differences in the front-end of the automatic speech recognition and significantly reduces computational complexity. Significant word error-rate performance has also been achieved by this new algorithm for in-car and military noisy environments. The final paper in this grouping is on "Using SVM as back-end classifier for language identification" by Hongbin Suo et al. This paper describes an approach using support vector machines (SVMs) with radial basis function kernel for back-end classifier in language identification. Furthermore, a pairwise posterior probability estimation is used to calibrate the output of each classifier.

Under the communication area, we have the following papers. The first paper in this grouping is a "Frequency-domain adaptive algorithm for network echo cancellation in VoIP" by Shaw Lin et al. This paper introduces a new

frequency-domain adaptive algorithm for network echo cancellation. The proposed frequency-domain multidelay filtering algorithm has advantages of low complexity, low delay, and fast convergence which are particularly important for voice over internet protocol applications. The second paper in this area is entitled “Estimation of interchannel time difference in frequency subbands based on nonuniform discrete fourier transform” by Bo Qiu et al., which looks at the binaural cue coding in the latest MPEG Surround standard. A novel algorithm is proposed to estimate the interchannel time difference (ICTD) by using the nonuniform discrete Fourier transform (NDFT), and a sub-band coherence factor to determine whether interchannel time difference estimation needs to be performed. Subjective measurements show that NDFT-based ICTD schemes result in a very good performance for sound image width and audio quality. The third paper is on “Measurement combination for acoustic source localization in a room environment” by Pasi Pertil et al., which looks into a class of acoustic source localization methods. This method is based on a two-step approach that applies time delay estimation (TDE) function on the measurement data and combines these TDE functions to produce the spatial likelihood function (SLF). The intersection-based combination methods results in a better location of RMS error compared to union-based combination methods. The final paper in this grouping is “Beamforming under quantization errors in wireless binaural hearing aids” by Sriram Srinivasan et al. This last paper analyzes quantization error in a low bit-rate wireless communication link between left and right hearing aids for a binaural beamforming structure. The generalized sidelobe canceller is considered, and the effect of head shadow is incorporated into the experimental analysis.

In the final grouping of interaction, we have the following papers. The first paper is “Tango or waltz?: putting ballroom dance style into tempo detection” by Björn Schuller et al. This paper enhances a data-driven tempo detection algorithm by incorporating ballroom dance style and meter recognition, and tests its performance based on a large database containing about two thousands Latin dance music. The second paper is “On-line personalization of hearing instruments” by Alexander Ypma et al. In this paper, a linear mapping from acoustic features to tuning parameters is used in hearing aids. Efficient feature representations are selected using a sparse Bayesian approach. The online personalization on an experimental hearing aid is pitted against the default setting, and found to have superior performance. The third paper is on “Automatic music boundary detection using short segmental acoustic similarity in a music piece” by Yoshiaki Itoh et al. This paper proposes a new approach for detecting music boundaries or music/speech boundaries for musical video data. By using a new algorithm employing segmental continuous dynamic programming, music boundaries for both evaluation musical data and real broadcast music programs can be accurately detected. The final paper in this special issue is on “Real-time perceptual simulation of moving sources: application to the Leslie cabinet and 3D sound immersion” by R. Kronland-Martinet and T. Voinier. This last paper combines physical

and perceptual approaches to develop real-time model for a moving source, and applies it to two audio applications.

The key themes in all the papers submitted in this special issue focus on some form of intelligent, adaptive, automation, and human interaction. Intelligent audio, speech, and music processing applications will definitely becoming more pervasive; and newer algorithms, models, and methods will be needed to meet the demands from these applications. Such applications are clearly out of the laboratory and being employed in real, everyday environments. As such, it becomes imperative to incorporate the user, subject, and context to improve the overall human-machine experience. While a number of impressive strides have been made in these accepted papers, there are still many research challenges and unanswered questions. More research work is necessary to address these important and exciting areas. We hope that this diverse collection of articles in this special issue will help motivate new research and collaborative work and inspire new ideas in intelligent audio, speech, and music processing.

ACKNOWLEDGMENTS

The guest editors would like to extend their sincere gratitude and thanks to all authors and reviewers who have contributed to this special issue. They also would like to thank the editorial staff from Hindawi Publishing Corporation for assisting them in managing this special issue.

*Woon S. Gan
Sen M. Kuo
John H. L. Hansen*