

Editorial

Signal Processing Platforms and Algorithms for Real-Life Communications and Listening to Digital Audio

Alexander Petrovsky,¹ Wanggen Wan,² Manuel Rosa-Zurera,³ and Alexey Karpov⁴

¹*Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus*

²*Shanghai University, Shanghai, China*

³*University of Alcalá, Madrid, Spain*

⁴*St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russia*

Correspondence should be addressed to Alexander Petrovsky; palex@bsuir.by

Received 2 April 2017; Accepted 2 April 2017; Published 19 July 2017

Copyright © 2017 Alexander Petrovsky 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.

Design of modern electronic communication systems involves diversified scientific areas including algorithms, architectures, and hardware development. Variety of existent multimedia devices gives rise to development of platform-dependent signal processing algorithms. Their integration into existent digital environment is an urgent problem for application engineers. Considering a wide range of applications including hearing aids, real-life communications, and listening to digital audio, the following research areas are of particular importance: advanced time-frequency representations, audio user interfaces, audio and speech enhancement, assisted listening, and perception and phonation modeling.

This special issue aims at publishing papers presenting novel methodologies and techniques (including theoretical methods, algorithms, and software) correspondent to these research areas. It includes high-quality papers dealing with applications in speech recognition, emotion recognition in speech signals, or informed source separation in orchestral recordings. The problems addressed in the accepted papers are among the top trends in the signal processing and communications research community.

Human interaction with computers through voice-user interfaces is the issue one of the papers deals with. It is based on Automatic Speech Recognition, and nowadays the objective is to solve the problem of spontaneous speech recognition. Spontaneous speech is characterized by hesitations, disfluencies, and changes that convey information about the

speaker. The paper “Experiments on Detection of Voiced Hesitations in Russian Spontaneous Speech,” by V. Verkhodanova and V. Shapranov, addresses the issue of voiced hesitations (filled pauses and sound lengthenings) detection in Russian spontaneous speech by utilizing different machine learning techniques: from grid search and gradient descent in rule-based approaches to such data-driven ones as ELM and SVM based on the automatically extracted acoustic features. Experimental results on the mixed and quality diverse corpus of spontaneous Russian speech indicate the efficiency of the techniques for the task in question, with SVM outperforming other methods.

The need for understanding business trends, ensuring public security, and improving the quality of customer service has caused a sustained development of speech analytics systems which transform speech data into a measurable and searchable index of words, phrases, and paralinguistic markers. Keyword spotting technology makes a substantial part of such systems. In the article entitled “A Russian Keyword Spotting System Based on Large Vocabulary Continuous Speech Recognition and Linguistic Knowledge” by V. Smirnov et al., the authors present an automatic system for keyword spotting in continuous speech. This system uses high-level linguistic knowledge and models for Russian speech and language; it has been implemented as software and applied in real-life telecommunication tasks for continuous speech processing.

In recent years, more attention is paid to the study of emotion recognition. Speech, as one of the most important ways of communication in human daily life, contains rich emotional information. Speech emotion recognition because of its wide application significance and research value in intelligence and naturalness of human-computer interaction aspects has got more and more attention from the researchers in recent years. The authors Z. Cairong et al. of the article "A Novel DBN Feature Fusion Model for Cross-Corpus Speech Emotion Recognition" present an automatic system for speaker emotion recognition by speech analysis. This system uses Deep Belief Nets for feature fusion and selection; it has been studied in cross-corpus experiments for emotion recognition tasks using emotional Chinese and German speech databases.

Speech intelligibility and speech recognition are important and trending topics of research in various fields of science: Linguistics, Medicine, Electrical Engineering, and Information Technology. Speech recognition process is investigated from different sides, as only an integrated approach could lead to a better understanding of this process. One of these research areas is intelligibility improvement of synthesized speech in noise, which has demanded much attention in recent years. In the article entitled "Crosslinguistic Intelligibility of Russian and German Speech in Noisy Environment" by R. Potapova and M. Grigorieva, the authors present quantitative results of experimental research on speech perception and intelligibility of spoken utterances and words by different listeners in various noisy conditions. Multiple experiments have been made using Russian and German speech data with addition of white and pink acoustic noises with different intensity and signal-to-noise ratios.

Audio source-separation is also a challenging task when we have sources corresponding to different instrument sections, which are strongly correlated in time and frequency. Without any previous knowledge, it is difficult to separate two sections which play, for instance, consonant notes simultaneously. One way to tackle this problem is to introduce into the separation framework information about the characteristics of the signals, such as a well-aligned score. The paper entitled "Score-Informed Source Separation for Multichannel Orchestral Recordings" by M. Miron et al. proposes and evaluates a system for score-informed audio source separation for multichannel orchestral recordings. The given article aims at adapting and extending score-informed audio source separation to the inherent complexity of orchestral music. This scenario involves not only challenges, like changes in dynamics and tempo, a large variety of instruments, high reverberance, and simultaneous melodic lines, but also opportunities as multichannel recordings. Results show that it is possible to align the original score with the audio of the performance and separate the sources corresponding to the instrument sections. In addition, authors derive applications, which allow for multiperspective audio enhancement, as acoustic scene rendering, and integrate them into an online repository which allows the distribution of the generated audio content.

Acknowledgments

We thank all the authors who made submissions to this special issue and the reviewers for their support and detailed reviews in making this special issue possible.

*Alexander Petrovsky
Wanggen Wan
Manuel Rosa-Zurera
Alexey Karpov*



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

