Preface: Special Section: Advances in Speech, Music and Audio Signal processing (Articles 1–13)

K. C. Santosh¹ · Surekha Borra² · Amit Joshi³ · Nilanjan Dey⁴

Published online: 4 March 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Speech, music and audio signals are vital in communication (e.g. sharing information) as well as entertainment. Automatic processing of such signal processing reduces expert's and/or human's intervention. With the advances in system architectures, machine learning and deep learning techniques, smart processing is possible in various areas, such as speech synthesis, mining and recognition; human–machine interactions; and unstructured data retrieval. Similarly, music signals are of importance since they contain special structural characteristics, where feature representations and analysis could potentially be different.

This special issue is organized to promote and publish the state-of-art research related to speech, music and audio processing covering aspects from information acquisition, processing, analysis, synthesis, retrieval, storage, coding, privacy, security, automation and application. This special issue includes 13 articles.

In the first article, Himadri et al. proposed voice activity detection (VAD) system with an objective of reducing the computational overhead apart from elevating the recognition performance. The technique extracted the line spectral frequency-based features for extreme learning-based classification of vocal segments from audio clips of multifarious sources and obtained overall accuracy of 99.43%.

In the second article, Behraz et al. presented a novel onset detection methodology for a traditional Iranian musical instrument namely Tar. The pitch and energy features are extracted to detect the onsets and reaz, for precise separation between two adjacent notes.

In the third article, Yosra et al. described and evaluated a pre-processing technique which employs the concepts of steady-state suppression in the temporal domain and a priori knowledge of distributions of the powers and the durations of voiced and unvoiced phonemes for detection of the speech voiced segments and to enhance the speech intelligibility in reverberant spaces.

In the fourth article, Mohan et al. presented a low bit-rate speech coding method based on multicomponent amplitude and frequency modulated signal model, the Fourier–Bessel series expansion and the discrete energy separation algorithm for parametric representation of speech phonemes. The symmetric Itakura–Saito and the root-mean-square log-spectral distance measures are used for comparison of the original and reconstructed speech signals.

In the fifth article, Prashant et al. proposed a Mel scaled M-band wavelet filter bank structure which extracts robust acoustic features for speech recognition application, with a flexibility of frequency partition. The filter performance was analyzed using AMUAV corpus and VidTIMIT corpus, and the results indicated improvement in terms of word recognition accuracy at all SNR range (20–0 dB) over baseline (MFCC) features and dyadic features.

In the sixth article, Mohamed et al. proposed a VAD independent backward BSS crosstalk-resistant algorithm which makes use of input correlation properties for automatic enhancement of blind speech quality and to result effective noise reduction, low misalignments, high convergence rates and tracking capabilities.

In the seventh article, Sophiya et al. proposed Deep Multilayer Perceptron architecture for Apache Spark Audio Scene Classification based on Log Mel band features. The system was evaluated with TUT dataset (2017) and the results were compared with the parameters of the DNN baseline of DCASE 2017 challenge.

In the eighth article, Yash et al. introduced a monaural speech separation technique based on non-negative tucker

✉ K. C. Santosh
santosh.kc@usd.edu

¹ The University of South Dakota, Vermillion, SD, USA

² K.S. Institute of Technology, Bangalore, Karnataka, India

³ Global Knowledge Research Foundation, Chandkheda, Ahmedabad, India

⁴ Techno India College of Technology, Kolkata, West Bengal, India

decomposition considering the effect of sparsity regularization factor on TIMIT and noisex-92 datasets.

In the ninth article, Charu et al. provided a summary of the challenges and research directions in speech communication focusing mostly on VAD and background noise reduction techniques.

In the tenth article Arun et al. reported the investigation results of the effect of speech coding in the quality of features which included a variety of cepstral coefficients extracted from codecs such as G.711, G.729, G.722.2, enhanced voice services, mixed excitation linear prediction and a few codecs based on compressive sensing framework. The analysis also included the variation in the quality of extracted features with various bit-rates supported by enhanced voice services, G.722.2 and compressive sensing codecs.

Kewen et al. in the 11th article, proposed three algorithms to enhance quality of speech fragments under various conditions. The algorithms are designed to obtain the core eigen components by joint diagonalization of clean speech and noise covariance matrix.

Azzedine et al. in the 12th article studied the effects of the mean subtraction, variance normalization, and Autoregressive Moving Average (ARMA) filtering (MVA) normalization method on the ETSI advanced front-end features.

Amal et al. in the 13th article investigated the use of hidden Markov models (HMM) for extraction of suitable contextual features from modern Standard Arabic language particularities such as vowel quantity and gemination.

In this issue, the guest editors selected 13 research articles (with an acceptance rate of 28%) and confirmed they will be effective and valuable for multitude of readers/researchers. Note that technical standard and quality of published content are based on the strength of the submitted articles. We are grateful to the authors for their imperative research contributions to this issue and their patience during the revision stages. We take this opportunity to give our special thanks to the Editor-in-chief Amy Neustein, for all the support, and competence rendered to this special issue.

Guest editors

K.C. Santosh, *The University of South Dakota, SD, USA.*

Surekha Borra, *K.S. Institute of Technology, Bangalore, Karnataka, India.*

Amit Joshi, *Global Knowledge Research Foundation, India.*

Nilanjan Dey, *Techno India College of Technology, West Bengal, India.*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.