World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:10, No:05, 2016

Speaker Identification by Atomic Decomposition of Learned Features Using Computational Auditory Scene Analysis Principals in Noisy Environments

Authors : Thomas Bryan, Veton Kepuska, Ivica Kostanic

Abstract : Speaker recognition is performed in high Additive White Gaussian Noise (AWGN) environments using principals of Computational Auditory Scene Analysis (CASA). CASA methods often classify sounds from images in the time-frequency (T-F) plane using spectrograms or cochleargrams as the image. In this paper atomic decomposition implemented by matching pursuit performs a transform from time series speech signals to the T-F plane. The atomic decomposition creates a sparsely populated T-F vector in &ldquo;weight space&rdquo; where each populated T-F position contains an amplitude weight. The weight space vector along with the atomic dictionary represents a denoised, compressed version of the original signal. The arraignment or of the atomic indices in the T-F vector are used for classification. Unsupervised feature learning implemented by a sparse autoencoder learns a single dictionary of basis features from a collection of envelope samples from all speakers. The approach is demonstrated using pairs of speakers from the TIMIT data set. Pairs of speakers are selected randomly from a single district. Each speak has 10 sentences. Two are used for training and 8 for testing. Atomic index probabilities are created for each training sentence and also for each test sentence. Classification is performed by finding the lowest Euclidean distance between then probabilities from the training sentences and the test sentences. Training is done at a 30dB Signal-to-Noise Ratio (SNR). Testing is performed at SNR&rsquo;s of 0 dB, 5 dB, 10 dB and 30dB. The algorithm has a baseline classification accuracy of ~93% averaged over 10 pairs of speakers from the TIMIT data set. The baseline accuracy is attributable to short sequences of training and test data as well as the overall simplicity of the classification algorithm. The accuracy is not affected by AWGN and produces ~93% accuracy at 0dB SNR.

Keywords : time-frequency plane, atomic decomposition, envelope sampling, Gabor atoms, matching pursuit, sparse dictionary learning, sparse autoencoder

Conference Title : ICMCSSE 2016 : International Conference on Mathematical, Computational and Statistical Sciences and Engineering

Conference Location : Amsterdam, Netherlands

Conference Dates : May 12-13, 2016

Open Science Index, Mathematical and Computational Sciences Vol:10, No:05, 2016 publications.waset.org/abstracts/51069.pdf