Abstract

The problem of separation of audio sources recorded in a real world situation is well established in modern literature. The method to solve this problem is Blind Speech Separation (BSS). The recording environment is usually modeled as convolutive (i.e. number of speech sources should be equal to or less than number of microphone arrays). In this paper, we propose a new frequency domain approach to convolutive blind speech separation. Matrix Diagonalization method is applied on cross power spectral density matrices of the microphone inputs to determine the mixing system at each frequency bin up to a permutation ambiguity. Then, we propose an efficient algorithm to resolve permutation ambiguity, where we group vectors of estimated frequency responses into clusters in such a way that each cluster contains frequency responses associated with the same source. The inverse of the mixing system is then used to find the separate sources. The performance of the proposed algorithm is demonstrated
by experiments conducted in real reverberant rooms.

Reference


Index Terms

Computer Science
Digital Signal Processing

Key words

Cross-Power Spectral Density Matrix
Diagonalization
Blind Speech Separation
Permutation ambiguity
Cluster