Unsupervised Single Channel Speech Separation Based on Optimized Subspace Separation

Belhedi Wiem, Ben Messaoud Mohamed anouar, Pejman Mowlaee, Bouzid Aicha

PII: S0167-6393(17)30127-9
DOI: 10.1016/j.specom.2017.11.010
Reference: SPECOM 2506

To appear in: *Speech Communication*

Received date: 6 April 2017
Revised date: 24 November 2017
Accepted date: 29 November 2017

Please cite this article as: Belhedi Wiem, Ben Messaoud Mohamed anouar, Pejman Mowlaee, Bouzid Aicha, Unsupervised Single Channel Speech Separation Based on Optimized Subspace Separation, *Speech Communication* (2017), doi: 10.1016/j.specom.2017.11.010

Unsupervised Single Channel Speech Separation Based on Optimized Subspace Separation

Belhedi Wiem*, Ben Messaoud Mohamed anouar*, Pejman Mowlaee**, and Bouzid Aicha*

*Université de Tunis El Manar, École Nationale d'Ingenieurs de Tunis, Laboratoire du Signal, Images et Technologies de l'Information, BP 37, Le Belvédère 1002, Tunis, Tunisie
(e-mail: bel.hedi.wiem@gmail.com)

**Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria.

Abstract

Single channel speech separation (SCSS) is widely used in many real-time applications such as preprocessing stage for speech recognition to control humanoid robots and in hearing aid. The performance of the separation is crucial for these applications. In this paper, we propose a new approach for unsupervised SCSS. The separation relies on an optimization of the subspace separation by decomposing the mixed signal into three estimates which are namely; the sparse subspace, the sub-sparse subspace and the low-rank subspace. Soft mask is used in the core of the proposed approach for the final decision. The proposed system generates two separated signals of different qualities and provided in two different channels. The channel classification is done using Fuzzy logic which requires two parameters. The first parameter is the quality of separated signal that we determine using a nonintrusive metric for speech quality and intelligibility. The second parameter is the gender of the speaker, determined using a proposed F_0 tracking algorithm. The evaluation results of the proposed approach are reported and compared to other state-of-art approaches. The proposed method on average achieves 67.9% improvement in PESQ, 59.5% improvement in signal-to-interference ratio (SIR) and 10.5% improvement in the target-related perceptual score (TPS) versus the benchmark methods.

Keywords: Unsupervised SCSS, Optimized Subspace Decomposition, Soft mask, Fuzzy Logic, Wavelet Transform, Multi-scale Product, Nonintrusive metric for speech quality and intelligibility.

1. Introduction

Speech separation refers to the technique that restores, from a mixture, a source whose characteristics are not given in advance. If we dispose of at least two observations, then it would be possible to use the spatial characteristics of the mixture by locating areas that belong to the same source. However, if the number of sources is smaller than the number of observations (the monaural case) then the problem is more difficult because the mixture is not reversible. In this paper we investigate the most challenging separation case: single channel separation in which the number of speakers (sources) is always higher than the number of channels (= 1).
Research in single channel speech separation (SCSS) has split in three main directions: Compu-