UNDER-DETERMINED REVERBERANT SPEECH SEPARATION USING BINAURAL CUES AND BLIND SOURCE SEPARATION APPROACH

Atiyeh Alinaghi, Wenwu Wang and Philip JB Jackson

Centre for Vision, Speech and Signal Processing (CVSSP)
Department of Electronic Engineering (FEPS)
University of Surrey, Guildford GU2 7XH, UK

1. INTRODUCTION

Many speech separation algorithms cannot deal with underdetermined cases, when the number of sources is greater than that of microphones. The problem becomes even worse in reverberant environments, where the sensors receive not only the direct sounds from the sources, but also the reflections from the walls, ceiling and other surfaces in the room.In [1], we combined binaural cues, inspired by the human auditory system, and blind source separation techniques [2], based on statistical properties of the sources, to compensate for the weaknesses of the two methods and improve the performance of state-of-the-art algorithms. However, the experiments were limited and based on simulated data. Here we used real room impulse responses which were measured in 4 different rooms with different reverberation times. At each room various source/ interferer configurations are examined.

2. EXPERIMENTS AND RESULTS

Separation performance was tested for both 2-speaker and 3-speaker mixtures, varying the room reverberation time, $T_{60} = [0.32s, 0.47s, 0.68s, 0.89s]$ and angular displacement between the sources [3], in the range of $10^\circ - 90^\circ$ with the smallest resolution of 5°. The target source was always positioned at zero azimuth while the interfering sources were located at positive or negative azimuths. Thus, we were able to show a systematic improvement in the signal-to-distortion ratio (SDR) compared with a state-of-the-art baseline system [4].

By using the frequency-independent version of the model together with a garbage model, we obtained an average 0.8 dB enhancement over the best baseline [1]. This significant improvement is most consistent for the more-challenging configurations where the angular separation between sources is less than 45 degrees, and therefore offers a valuable benefit. A similar trend was observed for the 3-speaker case as shown in figure 1. It is also detectable that in some configurations (e.g. 45°), the random error (i.e. variance) of the results has been reduced which is also desirable.

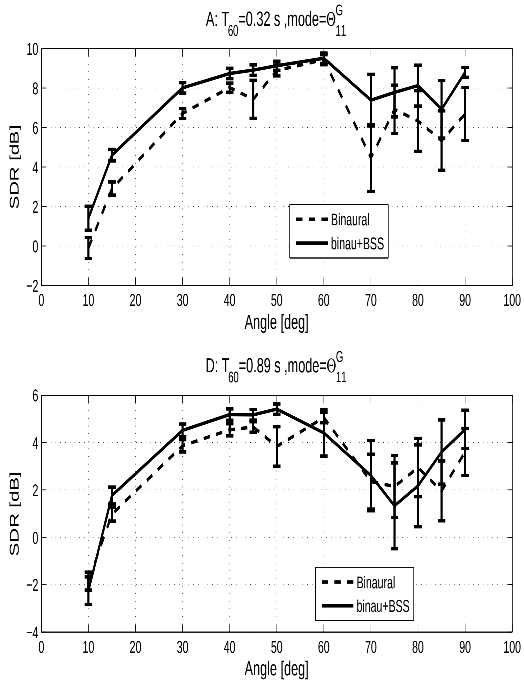

Fig. 1. SDRs of the recovered signals averaged over 15 mixtures (of 3-speaker) for each configuration with the error bar. The dashed line is for the baseline method using only the binaural cues and the solid line for the proposed algorithm where the parameters are frequency independent Θ_{11}^{G}

3. REFERENCES

[1] Atiyeh Alinaghi, Wenwu Wang, and Philip Jackson, “Integrating binaural cues and blind source separation method for separating reverberant speech mixtures,” in *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, May 2011.

[2] H. Sawada, S. Araki, and S. Makino, “Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment,” *IEEE Trans. Audio, Speech, and Language Processing*, 2011.

[3] Christopher Hummersone, *A psychoacoustic engineering approach to machine sound source separation in reverberant environments*, Ph.D. thesis, Music and Sound Recording, University of Surrey, UK, 2011.

[4] M. I. Mandel, R. J. Weiss, and D. P. W. Ellis, “Model-based expectation-maximization source separation and localization,” *IEEE transasction on audio, speech, and language processing*, vol. 18, no. 2, pp. 382–394, February 2010.