Improving Separation of Harmonic Sources with Iterative Estimation of Spatial Cues

NORTHWESTERN UNIVERSITY

JINYU. HAN[1], BRYAN. PARDO

The Interactive Audio Lab[2], Northwestern University, IL, USA

[1] jinyuhan@gmail.com [2] http://music.cs.northwestern.edu

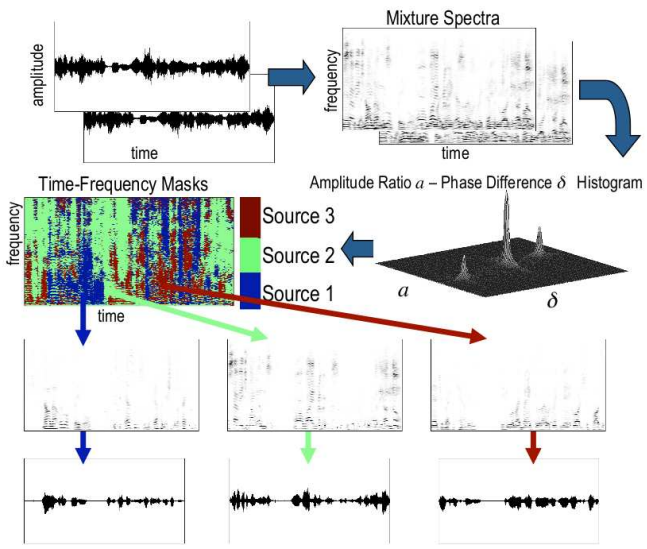
ABSTRACT

Spatial cues (cross-channel amplitude and phase difference coefficients) have been widely used in source separation of two-channel mixtures. However, as sources increasingly overlap in the time-frequency domain or the angle between sources decreases, these spatial cues become unreliable. We introduce an iterative method to re-estimate the spatial cues for mixtures of harmonic sources. Results on a set of three-source mixtures of musical instruments show this approach significantly improves separation performance of two existing time-frequency masking systems.

1 Introduction

DUET [1] and ASE [2] are are blind approaches to separating sources in two-channel anechoic acoustic mixtures. Both use spatial cues to estimate time-frequency masks for separating sources. Fig. 1 [a] illustrates how DUET uses spatial cues in stereo source separation. Spatial cues become obscured as the angle between sources decreases, hurting performance.

Figure 1 How to use Spatial Cues ?



[a] image by John Woodruff

2 System Description

Given a mixture, we estimate the sources as illustrated in Fig. 2):

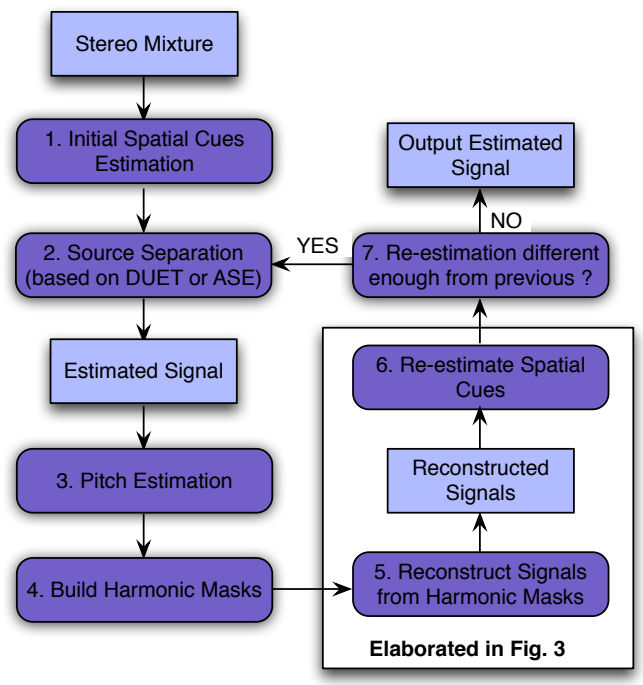
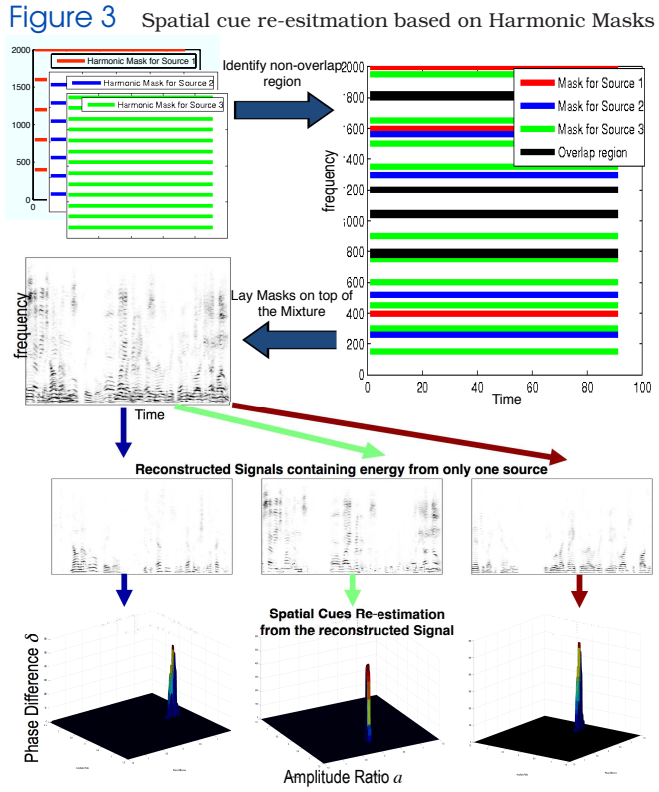


Figure 2 System Overview

Given the harmonic masks built from pitch estimation, we re-estimate the spatial cues as illustrated in Fig. 3):

Figure 3 Spatial cue re-esitmation based on Harmonic Masks



3 Evaluation

The instrument recordings used in the testing mixtures are individual notes played by horn, bass clarinet and oboe, taken from the instrument samples made available by the University of Iowa [a]. Five systems are compared in our experiments. They are DUET with ground truth pitch information for spatial cues refinement (IDEAL), unmodified DUET [1], DUET with spatial cues iterative refinement (DUET+ITER), unmodified ASE [2] and ASE with spatial cues iterative refinement (ASE+ITER). Ground truth pitches are the pitches estimated from the original recordings of the isolated sources, prior to mixing.

3.1 Experiment I

Performance results [b] when the angle between instruments is 20^o are shown in Fig. 4. We created 30 mixtures, each of a block-chord major triad with one note per instrument. The median value of separation performance for each method is labeled with arrow text. Iterative spatial cues refinement improves DUET's median performance by 5.2 dB and ASE's median performance by 2.5 dB. Furthermore, our proposed system (DUET+ITER) performs as well as the system using ground truth pitches.

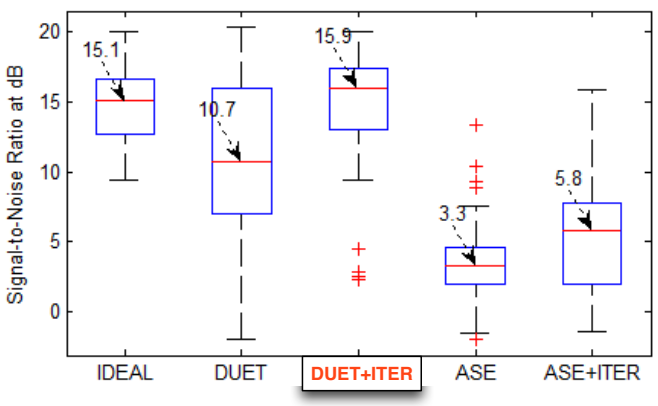


Figure 4 Performance of different methods at mixing angle 20^o

[a] http://theremin.music.uiowa.edu/

[b] Audio Examples could be found in http://music.cs.northwestern.edu/lab/research.php?project=separation

3.2 Experiment II

Performance results from mixtures created using different mixing angle are shown in Fig. 5. In this figure, each data point indicates an average result for 30 mixtures. The proposed system (DUET+ITER) consistently outperformed the existing systems' performance (excluding the system using ground truth pitches) for nearly all the mixing angles above 18^o. The iterative spatial cues estimation improves DUET or ASE when the sources are close to each other (In Figure 2, this is the case when the mixing angle is between 18^o and 30^o).

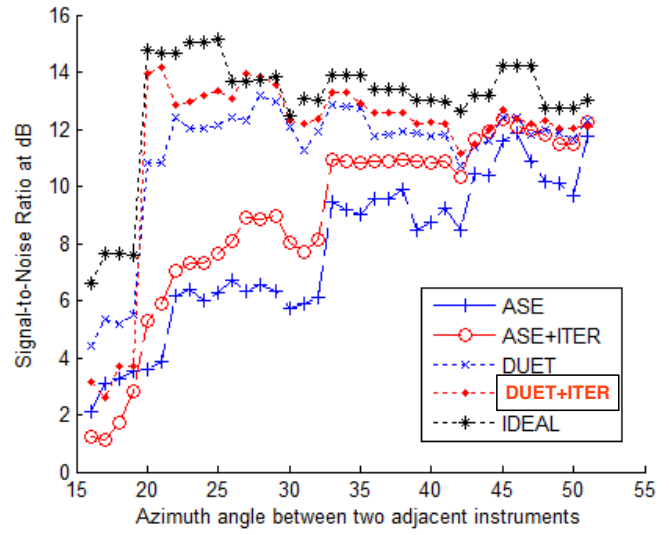


Figure 5 Signal to Noise Ratio of different methods with the mixing angle changing from 15^o and 50^o

4 Conclusions

We have proposed a method for improved source separation of anechoic two-channel mixtures of harmonic sound sources. We use an existing source separation system to do the initial estimate and improve the results by incorporating the pitch and energy distribution information to further refine the spatial cues. Results on a database of three-instrument mixtures show this approach improves both the DUET and the ASE source separation systems, especially as the angle between two adjacent instruments falls below 40 degree.

Acknowledgements

We would like to thank the reviewers for their helpful comments and suggestions; This work is supported by National Science Foundation Career Award (grant 0643752).

References

[1] O. Yilmaz and S. Rickard: "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, Vol. 52, pp. 1830- 1847, 2004.

[2] J. Woodruff and B. Pardo: "Using pitch, amplitude modulation and spatial cues for separation of harmonic instruments from stereo music recordings," *EURASIP Journal on Applied Signal Processing.*, Vol. 2007, 2007.

More references are listed in the paper