Reconstructing Completely Overlapped Notes from Musical Mixtures

Jinyu Han & Bryan Pardo

Northwestern University, EECS Department, Evanston, IL, USA.
jinyuhan@gmail.com • *pardo@northwestern.edu* • *http://music.cs.northwestern.edu/*

Introduction

Systems that perform source separation on recordings of music often encounter situations where a harmonic sound is almost completely overlapped in time and frequency (e.g. when two instruments play one octave apart). *Common Amplitude Modulation* is an effective method to resolve the problem of overlapped harmonics. It, however, relies on non-overlapped harmonics from the same note being available. We propose a musical sound separation system for monaural recordings that explicitly deals with the "completely overlapped" notes problem, based on *Harmonic Temporal Envelope Similarity*. We learn a harmonic envelope model for each instrument from the non-overlapped harmonics of notes of the same instrument, wherever they occur in the recording. This model is used to reconstruct the harmonic envelopes for overlapped harmonics.

Review

▶ Common Amplitude Modulation (CAM)

CAM assumes the harmonic amplitude envelopes from the same note are correlated. CAM holds most of the time for the first few strong harmonics but fails to hold for those with weak energy.

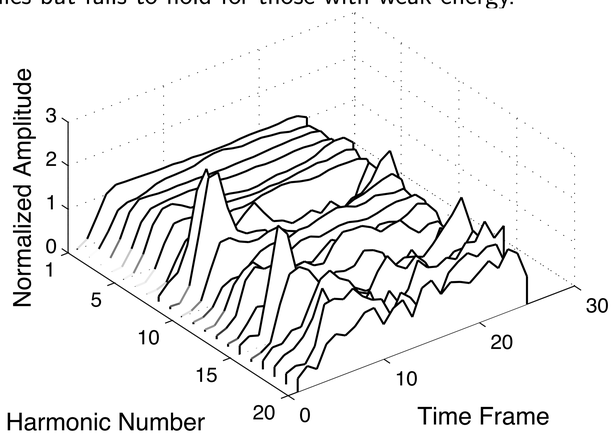

Figure: First 20 harmonic envelopes of a F4 note played by a clarinet. Amplitude envelopes are normalized in volume so that the shapes are easier to see. The first 8 harmonics do share the same general modulation trend, while the envelopes of the remaining harmonics have little correlation with each other.

Contributions

▶ Harmonic Temporal Envelope Similarity (HTES)

Instead of using CAM to estimate harmonic envelopes, We utilize the property that notes played by the same instrument within a short period of time (e.g. within musical phrase boundaries) have similar harmonic envelopes. We call this *Harmonic Temporal Envelope Similarity*.

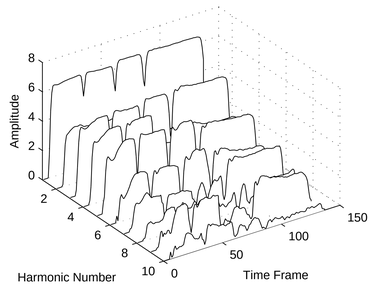

Figure: The first 10 harmonics of four consecutive notes of 400Hz, 375Hz, 300Hz and 330Hz played by a clarinet

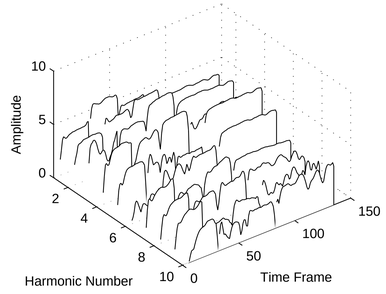

Figure: The first 10 harmonics of four consecutive notes of 132Hz, 147Hz, 197Hz and 100Hz played by a bassoon

▶ Proposed Method

In a mixture containing several instruments, we learn a harmonic envelope model for each instrument from the non-overlapped harmonics of notes played by that source throughout the recording. We consider this as a curve-fitting problem and fit the envelopes of the non-overlapped harmonics using a polynomial function. To recover a completely overlapped note from the mixture, this model is applied to reconstruct the harmonic envelope for each overlapped harmonic.

Evaluation

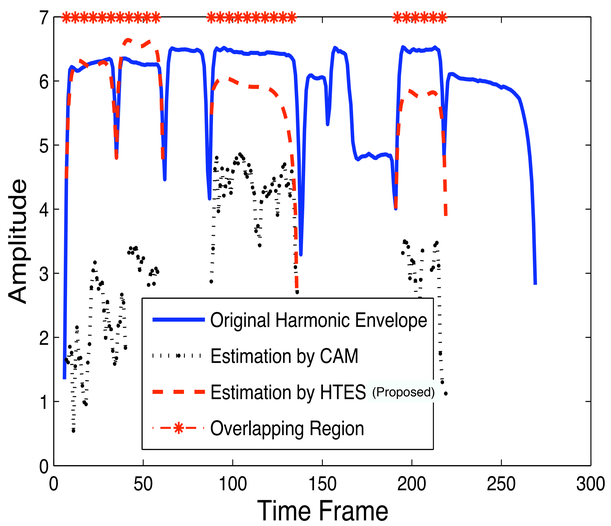

Figure: Original envelopes of the first harmonic are from nine notes played by clarinet. Four notes of 398.4Hz, 397.7Hz, 296.6Hz and 293.3Hz are completely overlapped with four bassoon notes of 132.6Hz, 198.2Hz, 98.2Hz and 146.9Hz.

The estimation based on *CAM* (dotted line) is very unstable and different from the original envelope. The dashed line is the envelope estimated by utilizing the note model from the third and last note of the example.

Our proposed model using HTES produces much better envelope estimates for the completely overlapped notes than the *CAM* does.

▶ Experimental Setup

- 10 Bach chorale recordings of about 330 seconds of audio.
- Mixtures of two instruments with one instrument (bassoon) playing the bass line and the other playing the alto (clarinet or trumpet) or soprano line (violin) of a Bach chorale.
- The proposed system was comapred to a musical sound separation system (denoted LWW) based on *CAM*

▶ Experimental Results

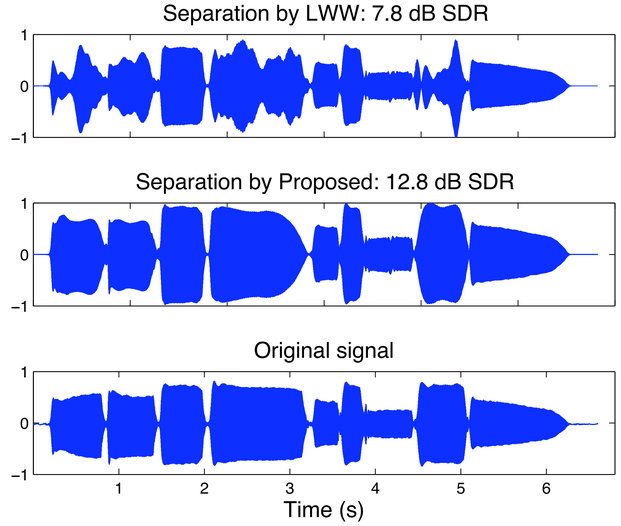

Figure: Separation of a clarinet from a 6.5 seconds mixture of clarinet and bassoon

Mixtures	SDR		SAR		SIR	
	Proposed	LWW	Proposed	LWW	Proposed	LWW
Clarinet	**12.3**	10.4	**12.3**	10.4	42.3	41.9
Trumpet	**10.7**	9.6	**10.8**	9.6	41.4	39.7
Violin	6.3	6.4	6.4	6.4	44.1	43.7

Table: Performance of the proposed system and the LWW. Numbers in bold indicate the difference between proposed method and LWW are statistically significant.

Conclusion

We proposed a monaural musical sound separation system that explicitly deals with completely overlapped notes in music recordings. Our approach is based on Harmonic Temporal Envelope Similarity, a new assumption on instrumental harmonic envelope we observed from real audio data. Experimental results show the proposed method achieves better separation performance than a state-of-art monaural music separation system that only exploits CAM.

This work was supported by NSF grant numbers IIS-0643752.