Remixing Stereo Music with Score-Informed Source Separation*

John Woodruff
Music Technology,
School of Music
Northwestern University
Evanston, IL 60208, USA
j-woodruff@northwestern.edu

Bryan Pardo
Electrical Engineering and
Computer Science
Northwestern University
Evanston, IL 60208, USA
pardo@northwestern.edu

Roger Dannenberg
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213 USA
rbd@cs.cmu.edu

Abstract

Musicians and recording engineers are often interested in manipulating and processing individual instrumental parts within an existing recording to create a *remix* of the recording. When individual source tracks for a stereo mixture are unavailable, remixing is typically difficult or impossible, since one cannot isolate the individual parts. We describe a method of informed source separation that uses knowledge of the written score and spatial information from an anechoic, stereo mixture to isolate individual sound sources, allowing remixing of stereo mixtures without access to the original source tracks. This method is tested on a corpus of string quartet performances, artificially created using Bach four-part chorale harmonizations and sample violin, viola and cello recordings. System performance is compared in cases where the algorithm has knowledge of the score and those in which it operates blindly. The results show that source separation performance is markedly improved when the algorithm has access to a well-aligned score.

Keywords: source separation, score alignment, music.

1. Introduction

Musical *remixing* can be broadly defined as the process of manipulating and processing individual instrument parts within an existing recording. This could mean simply raising the level of a single instrument in a poorly balanced mixture, or completely reworking a piece of music through editing and applying effects to individual instruments. When individual source tracks for a stereo mixture are unavailable, remixing is typically difficult or impossible.

In order to remix existing recordings, one must perform *source separation*—separation of the audio mixture into its component sound sources. While perfect reconstruction of individual sources from a musical mixture is not currently possible in the general case, even imperfect isolation is useful for a number of purposes, including improved instrument identification and analysis within polyphonic recordings, structured audio coding and both the creative and restorative remixing applications described above.

In this paper we describe a method that performs source separation using information from the written score and spatial cues present in a stereo recording. The combination of these lets our method isolate individual musical parts in a corpus of four-part Bach chorale recordings so that audio effects, equalization and volumes can be altered on an instrument-by-instrument basis.

The remaining sections of this paper describe current research in source separation, our existing source separation method, our score alignment method, how we combine score information and spatial information to improve source separation, and experimental results. We also provide links to example remixes created using our approach at http://bryanpardo.com/papers/ismir2006.

2. Current Work in Source Separation

The difficulty of the source separation problem depends on the number of sources (instruments) in the recording and the number of sensors (microphones) used to make the recording. When the number of available audio channels (mixtures) equals or exceeds the number of individual sources (a quadraphonic recording of a trio, for example), one may use *Independent component analysis* (ICA) [7].

The source separation problem is considered *degenerate*, or under determined, when the number of sources exceeds the number of mixtures. Standard ICA algorithms are not effective in the degenerate case. Since millions of audio recordings exist in a stereo format (two-mixtures), but typically consist of more than two source signals, it should be clear why solving the degenerate source separation problem is of considerable interest to researchers.

Recent approaches to degenerate source separation of speech mixtures have exploited the sparsity of speech signals in the time-frequency domain. Speech is considered *sparse* because the vast majority of time-frequency frames in a speech signal have magnitude near zero. This is used to justify the assumption that at most one source signal (talker) has significant energy in any given time-frequency

*Published as: John Woodruff, Bryan Pardo, and Roger B. Dannenberg, "Remixing Stereo Music with Score-Informed Source Separation," in *Proceedings of the 7th International Conference on Music Information Retrieval,* Victoria, BC, Canada, October 2016, pp. 314-319.

frame (the signals are time-frequency disjoint). Given this assumption, a time-frequency masking approach can be used that exploits spatial cues from an anechoic, stereo recording to separate sources from a mixture [14].

Tonal music makes extensive use of multiple simultaneous instruments, playing consonant intervals (such as unisons, octaves and perfect fifths). When two harmonic sources form a consonant interval, their fundamental frequencies are related by a ratio that results in significant overlap between the *harmonics* (regions of high-energy at integer multiples of the fundamental frequency) of one source and those of another. Non-harmonic instruments, such as percussion instruments, further complicate the problem due to their wide-band (noisy) spectral characteristics. Thus, instrument signals frequently overlap in both time and frequency, rendering approaches that assume time-frequency disjoint sources ineffective.

To deal more effectively with overlapping source signals, researchers have introduced assumptions about the structure of the sound sources. In the single-mixture (monophonic) domain, Virtanen and Klapuri [9, 10] assume source signals are harmonic, allowing multi-pitch estimation of the polyphonic mixture to determine frequency regions in which source signals overlap. By assuming that the signals have a smoothly decaying overtone series as a function of frequency, source amplitudes in the overlapping frequency regions can be estimated.

Every and Szymanski [1] use a single mixture and prior knowledge of instrument pitches to determine regions of source signal overlap. They linearly interpolate between known harmonics in cases where multiple sources overlap and achieve separation through spectral-filtering of the mixture.

If the number of audio channels equals or exceeds the number of sound sources, Viste and Evangelista [11] show they can perform iterative source separation by minimizing the variance of the temporal envelopes of each source's individual harmonics. While this method does very well in situations where two sources overlap and can potentially deal with reverberant recordings, it cannot be applied in the degenerate case.

Vincent [8] approaches demixing stereo recordings with two or more instruments by incorporating grouping rules from computational auditory scene analysis [6], spatial cues and time-frequency source signal priors to cast the demixing problem into a Bayesian estimation framework. This is done to let the system handle reverberant recordings, but requires significant prior knowledge of each source signal in the mixture and is not suited to the remixing applications described in the introduction.

While remixing is possible when source separation can be achieved, researchers have also approached remixing without attempting to fully isolate each sound source. Methods for isolating percussion instruments from the rest of a stereo recording for remixing purposes are proposed in [2, 15]. While the overarching goal of this work is related to our own, our effort has been focused on isolating and remixing harmonic instruments in music recordings, requiring distinctly different processing techniques.

We previously introduced the ASE method to perform separation of stereo, anechoic mixtures of any number of harmonic, monophonic sources [13]. This approach requires no prior information about the sources, but can deal effectively with mixtures that contain significant source overlap. ASE can accurately resolve situations where two sources overlap, and uses this information to resolve regions of recordings where three or more sources are simultaneously active. We describe the details of this method in the next section.

3. ASE Source Separation

The Active Source Estimation (ASE) source separation approach assumes an anechoic, stereo (two-channel) mixture of harmonic sound sources. The two mixture channels are modeled as follows,

$$X_1(\tau,\omega) = \sum_{n=1}^{N} S_n(\tau,\omega) \quad (1)$$

$$X_2(\tau,\omega) = \sum_{n=1}^{N} a_n e^{-j\omega\delta_n} S_n(\tau,\omega) \quad (2)$$

where $X_1(\tau,\omega)$ and $X_2(\tau,\omega)$ represent the left and right mixtures in the time-frequency domain, with time frame τ and frequency bin ω. Here, $S_n(\tau,\omega)$ is the nth source signal, a_n is the cross-channel amplitude scaling and δ_n is the cross-channel time-shift associated with source n. We call a_n and δ_n the *mixing parameters* of source n.

ASE takes a three-step approach to source separation. In the first stage, common amplitude and phase differences between the two mixtures are assumed to result from the differing spatial locations of the individual sources. The most common cross-channel scaling and time-shift factors are thus associated with the individual sources as the mixing parameters a_n and δ_n, and used to identify the time-frequency frames of the mixture that result from only one source [5, 13, 14].

The energy from these single-source frames is distributed to create initial source estimates while time-frequency frames that do not match the mixing parameters of any of the sources are left in the mixtures for later processing.

In the second step, ASE estimates the number of sources that are active in each remaining time-frequency frame of the mixtures. It does this by pitch-tracking the partial source estimates from the first step. These fundamental frequency estimates, combined with simple harmonic models (estimated from the first stage), let the system identify which sources are likely to contribute

energy to the remaining time-frequency frames in the stereo mixture. It is during this stage that the pitch information from the score can be utilized. We discuss the implementation of score knowledge into this stage of the algorithm in section 4.

Given our mixture models, the problem when two sources are active in a given time-frequency frame is even determined, allowing us to solve for the appropriate source energies in these frames. This requires solving the system of equations provided by (1) and (2) under the assumption that only two sources are active. If we denote the active sources in a particular time-frequency frame, (τ,ω), by $S_g(\tau,\omega)$ and $S_k(\tau,\omega)$, the following equations result,

$$S_k(\tau,\omega) = X_1(\tau,\omega) - S_g(\tau,\omega) \quad (3)$$

$$S_g(\tau,\omega) = \frac{X_2(\tau,\omega) - a_k e^{-j\omega\delta_k} X_1(\tau,\omega)}{a_g e^{-j\omega\delta_g} - a_k e^{-j\omega\delta_k}} \quad (4)$$

By determining the appropriate source energy in each of these frames, more complete source estimates are created, leaving only those time-frequency frames that contain energy from three or more sources.

In mixture frames that have energy from three or more sources, the problem is under determined and the relative energy contribution of each source cannot be solved for directly. In this case, a third step is taken. The system models the amplitude variation of the harmonics in each source estimate. These models are used in conjunction with the mixture energy to predict the relative strength of the sources in the remaining time-frequency frames.

This approach lets ASE separate mixtures containing time-frequency frames in which multiple harmonic sources are active without prior knowledge of source characteristics. This method is, however, susceptible to errors in pitch tracking. Inaccurate estimates of fundamental frequency will result in the system making mistakes about which sources contributed energy to individual time-frequency frames, causing source separation to fail. In order to improve the reliability of the fundamental frequency estimates, it is helpful to use information from the musical score, when available.

4. Score Alignment

A key ingredient in this approach to source separation is the labeling of audio with symbolic pitches. Labeling *could* be done manually, but it is much easier to start with a symbolic, machine-readable score, e.g. MIDI. In practice, MIDI files can often be found on the Web. While a MIDI file encodes basic rhythmic timing, it lacks any information about expressive timing and tempo in the audio recording. Alignment can recover this information.

Polyphonic alignment is performed by converting the MIDI file and the audio file into a *chromagram* representation [12]. A chromagram is a sequence of *chroma vectors*, 12-element vectors representing the total spectral energy corresponding to each of the 12 pitch classes (C, C#, D, …, B). The chroma vector is chosen because it captures harmonic and melodic information, which is shared by audio and MIDI, and it tends to be insensitive to amplitude and timbral differences, which often do not to match very well between audio and MIDI [4].

Audio data is divided into 125 ms frames, each of which is converted to a chroma vector. For MIDI data, chroma vectors are estimated by summing all the matching pitch classes sounding during that frame, weighted by the key velocity and duration (0.125, or less if the note begins or ends during the frame).

Audio and MIDI chroma vectors are normalized to have a mean of 0 and a standard deviation of 1. The next step uses dynamic time warping to find the best time alignment, using Euclidean distance between chroma vectors.

Finally, the rough alignment, which is a path quantized to points along a 125 ms grid, is smoothed at each point by finding the best fit to the nearest 7 points, using linear regression. The resulting points define a sampled function that can be linearly interpolated to map between MIDI file time and audio file time.

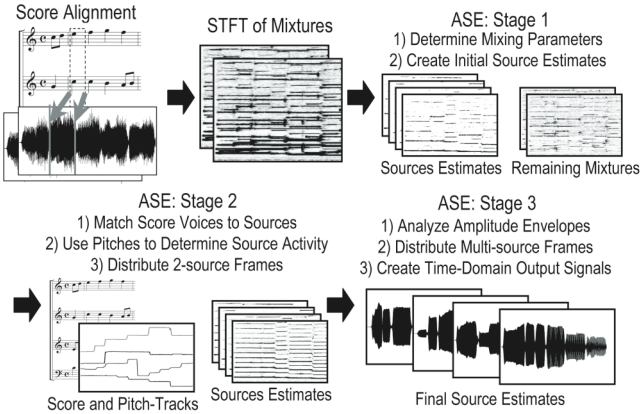

Figure 1: Illustration of the combined Score Alignment and ASE Source Separation algorithm. Score alignment is carried out prior to ASE separation. Score knowledge is incorporated during stage 2 of ASE.

5. The Combined System

To incorporate knowledge of the score into the source separation algorithm, we must accomplish three primary tasks. First, we must correctly associate individual *voices* (instrument parts) contained in the score with the mixing parameters of each source. Second, the pitches in a score give only a rough estimate of the actual fundamental frequencies present in the mixture. The system must refine the pitch estimates provided by the score in order to achieve accurate separation. Third, further timing alignment between performance and score must be performed within the combined system, since the score-alignment can be expected to make note onset timing errors on the order of 60 ms.

To account for these issues we incorporate the information in the score after the first stage of the ASE algorithm. This allows us to use fundamental frequency and amplitude envelope characteristics of the initial signals to solve all three problems. The next two sections will describe the techniques employed in more detail. An illustration of the overall system is provided in Figure 1.

5.1 Associating Scored Voices with Initial Source Estimates

The first stage of the ASE algorithm distributes time-frequency frames of the mixture that contain energy from a single source. For the score to be of use in the second and third stages of energy distribution, we must determine which voice in the score is associated with a particular source estimate. Since ASE is already estimating the fundamental frequency of each source, we take a simple approach to establishing this relationship. We compare the fundamental frequency estimates of each source to each pitch track provided by the score. We calculate the number of time frames in which the fundamental frequency is at most half a *semitone*, or within roughly ±3%, from the frequency associated with the pitch in the score. For each source estimate established in stage one of ASE, we store the pitch track from the score that has the most time frames in common with the source's fundamental frequency estimate. We take care to ensure that each source is associated with a different voice in the score by giving priority to sources that have higher similarity ratings to a pitch track in the score. Using this simple method, the system correctly associated 99% of the score voices with source estimates on our testing corpus.

5.2 Refinement of Pitch and Timing Information Provided by the Score

The pitch-tracks provided by the score are useful to ASE in the determination of where sources are likely to have high amounts of energy in time-frequency space. However, using the score without frequency refinement would often cause the algorithm to miss the true fundamental frequencies of the sources in the mixture, since intonation variation and vibrato are not typically represented in scores. Also, even the time-aligned score can have errors of up to 60 ms, depending on the amount of expressive timing variation and asynchrony between performers. Without timing refinement, perceptually important signal features like note onsets can be lost.

To refine the frequency of the pitch tracks provided by the score, we again turn to the sources' fundamental frequency estimates calculated in stage two of ASE. Since we have determined which pitch track is most similar to each source's fundamental frequency estimate, we simply use the fundamental frequency estimate provided by ASE in all time frames in which it is within half a semitone of the pitch track frequency from the score.

To refine the timing information, we calculate the amplitude envelopes of the sources from the stage one estimates, set an amplitude threshold to determine when the source is active, and record all frames that transition from below to above the threshold as possible note onsets. We then allow the note onsets provided by the score to be altered to the possible onsets calculated above if the estimated onsets are within 3 time frames of the onset in the score. Again, we chose 3 frames because the difference between the centers of consecutive time windows is 23 ms and the score alignment had a maximal error of roughly 60 ms on this corpus.

6. Experimental Results

In previous work we tested the efficacy of the ASE source separation algorithm in isolation [13]. In this work, we were interested in measuring the possible improvement of source separation when score information is available. To this end, we created a corpus of 100 stereo mixtures of four-part Bach chorales and tested our source separation under four conditions: blind (no score), an un-aligned initial score, a machine-aligned score and ideal score. This section describes the result in detail.

6.1 Corpus of Scores and Audio Mixtures

Our test corpus consisted of 100 typical Bach soprano-alto-tenor-bass four-part chorale harmonizations. For each harmonization, we randomly chose a four second segment, typically equating to about one or two measures in the music. For each segment of the harmonization chosen, we created three MIDI versions. The first version was an unaltered representation of the selected segment of the harmonization. We call this the *original score*.

From each original score, we created the second MIDI version by randomly altering the tempo of each piece between 71% and 140% of the original tempo, with the average deviation being roughly 20%. This version, the *ideal score*, was used to generate the audio mixture. Although a typical interpretive performance of a piece of music would likely include tempo variation throughout the duration of the piece, our scored segments were only a measure or two long, so we felt that a simple tempo scaling was a reasonable simulation of a performance of the harmonization segment.

For each notated instrument part in the ideal score we created an audio file using recorded samples of violin (soprano and alto part), viola (tenor) and cello (bass). The samples used were from a commercial instrument sample library, *Xsample Professional Sound Libraries, Volume 41: Solo Strings*. These individual audio recordings (one for each instrument part in the score) were then combined to create a stereo audio mixture of each chorale harmonization. We created mixtures in this way in order to measure the difference between the ideal (the pre-mix individual signals) and the source estimates extracted from each mixture.

We then performed score following on each audio mixture, aligning the original score to the mixture. The output of the score follower was a MIDI file that had been time-altered to match the timing of the audio mixture. This is the *aligned score*.

6.2 The Experiment

For each audio mixture we performed source separation four times: once with no score (the standard ASE algorithm), once with the ideal score, once with the aligned score, and once with the original score. For this experiment, we used a window length of 186 ms and a 163 ms overlap between time frames in the time-frequency analysis of the mixture.

Since we were interested in testing the system improvement when incorporating score knowledge, we must note that one key aspect of the separation algorithm was not tested in the presented data. During the first stage, ASE uses the approach presented in [14] to determine each source's mixing parameters (a_j,δ_j). Since the experiment in this paper was designed to measure how score knowledge improves the second stage of demixing (separation of overlapped harmonics), we wanted all variation in results to be due to the use of score information and used known values for the mixing parameters.

6.3 The Error Measure

The performance of the algorithm was measured by its ability to achieve complete isolation of the individual sources. For each source estimate created from the mixture, we calculate the *Signal-to-Distortion Ratio* (SDR) as shown in Equation 5 [3]. Here, s is the original source signal, and $\hat{s}$ is the source estimate provided by the algorithm.

$$SDR = 10\log_{10}\left(\frac{\left|\langle \hat{s}, s\rangle\right|^2}{\left|\langle \hat{s}, \hat{s}\rangle\right|^2 - \left|\langle \hat{s}, s\rangle\right|^2}\right) \quad (5)$$

A high SDR represents a strong correlation between the estimated and original signal, with little noticeable distortion. Through informal listening tests, we feel that an SDR under 3 or 4 dB results from estimated signals that are similar to the original sources, but with very noticeable interference or artifacts due to demixing errors. Signals with an SDR above 6 dB are better and can be sufficient for many remixing applications. Signals with an SDR above 8 or 9 dB may still contain audible artifacts when isolated, but these artifacts are easily masked when recombined with other instruments from the original recording.

6.4 Results

We found that using knowledge of the score greatly improved the performance of the source separation algorithm. Without score knowledge, the fundamental frequency estimation in stage 2 of ASE was accurate (within half a semitone) in an average of 69.4% of a source signal's time frames. Using the aligned and refined scores increased this accuracy to 92.9%. The increased accuracy of the fundamental frequency estimates resulted in improved separation performance in 78.25% of the separated signals. The SDR improvement between the median *blind* and median *aligned score* performance was 1.7 dB.

Figure 2 shows notched box-plots of the SDR over all trials for the three score knowledge scenarios. Each box represents the performance on 400 signals, four for each chorale harmonization. The lower and upper lines of each box show 25th and 75th percentiles of the sample. The line in the middle of each box is the sample median. The lines extending above and below the box show the extent of the rest of the sample, excluding outliers. Outliers are defined as points further from the sample median than 1.5 times the interquartile range (the overall height of the box) and are indicated by plus signs. The notches in each box show the 95% confidence interval around the median. Since the notches in the box-plot for the blind case and the aligned score do not overlap, we conclude, with 95% confidence, that use of the aligned score provides significant performance improvement.

While knowledge of the score can improve the algorithm's performance, a misaligned score can actually degrade separation. In comparing the blind algorithm performance to the performance with the original score (the non-aligned score), the median SDR decreased by 3.51 dB with 79.25% of the cases performing worse when the algorithm had knowledge of the misaligned score. This result emphasizes the necessity of score alignment if one is to incorporate score knowledge into a signal separation algorithm.

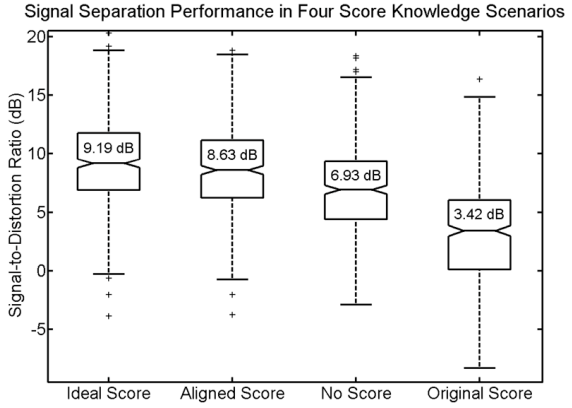

Figure 2: Performance results over all mixtures, compared between score knowledge conditions.

Direct comparisons between this system and other musical separation or remixing systems are difficult because of the lack of commonality between source signal assumptions, mixture assumptions, and testing data. Considering the variety of approaches to the problem, the most suitable algorithm for a given mixture depends primarily on how well suited the algorithm's assumptions

and required *a priori* knowledge are to the given source signals and mixing process.

6.5 Example Remixes

To illustrate the effectiveness of the combined algorithm for score-informed source, we created a number of example remixes that are accessible here:

http://bryanpardo.com/papers/ismir2006

We have included examples of amplification and attenuation of individual instruments in the mixture, and also some in which we have applied reverberation and other effects processing to individual instruments. A final example mimics a complete reworking of a piece of music by applying editing, looping and effects to the individual instrument parts.

We also provide audio examples of isolated source estimates. One of the benefits of using source separation for musical remixing is that although isolated source estimates may contain audible artifacts or interference, these distortions are due to other sources in the recording and are often masked when the signal is recombined into a remix. Effective manipulation of level and instrument timbre is possible even at relatively low SDR levels. To illustrate this, we provide examples of remixes using source estimates at various SDR levels.

7. Conclusions and Future Work

Musicians, recording engineers and composers often desire remixing of fully notated musical pieces. We have presented a method combining score alignment and source separation to achieve such a task for anechoic, stereo recordings. We have discussed the implementation of a musical source separation algorithm incorporating score knowledge and found this yields a notable improvement in separation performance.

Our results make it clear that the fundamental frequency stage of the ASE algorithm can be inaccurate. In future work, we plan to explore more robust methods of fundamental frequency estimation, which our results show will improve overall performance.

The approach of the ASE method is to first create initial signal estimates, which can be analyzed to assist with demixing more difficult mixture regions. Our future work will examine more sophisticated analysis methods and signal modelling to leverage learned structural information concerning the sources in conjunction with spatial information present in stereo recordings. We feel that the development of an effective separation algorithm requires the exploitation of simultaneous signal features, and the ability to assess the reliability of these features at a given time. Relying heavily on cross-channel amplitude and timing differences is unrealistic in reverberant or studio-produced recordings. We believe, however, that additional source characteristics can be learned from corrupted or mixed signals, which will allow systems such as ASE to degrade more gracefully as the recording process or environment becomes more challenging.

References

[1] M. Every and J. Szymanski. "A Spectral-Filtering Approach to Music Signal Separation", *in DAFx 04 Seventh Int. Conf. on Digital Audio Effects Proc*., 2004, pp. 197-200.

[2] O. Gillet and G. Richard. "Extraction and Remixing of Drum Tracks from Polyphonic Music Signals", in *WASPAA 05 IEEE Workshop on App. of Signal Proc. and to Audio and Acoustics Proc.*, 2005, pp. 315-318.

[3] R. Gribonval, L. Benaroya, E. Vincent, C. Févotte. "Proposals for Performance Measurement in Source Separation", in *ICA 03 Fourth Int. Symp. on Ind. Comp. Analysis and Blind Signal Sep. Proc.*, 2003, pp. 763-768.

[4] N. Hu, R. Dannenberg, G. Tzanetakis. "Polyphonic Audio Matching and Alignment for Music Retrieval", in *WASPAA 03 IEEE Workshop on App. of Signal Proc. to Audio and Acoustics Proc.*, 2003, pp. 185-188.

[5] A. Master. "Sound Source Separation of N Sources from Stereo Signals via Fitting to N Models Each Lacking One Source", Stanford University, CCRMA Technical Report, 2002.

[6] D.F. Rosenthal, and H.G. Okuno. *Computational Auditory Scene Analysis*. Lawrence Eribaum Associates, 1998.

[7] J.V. Stone. *Independent Component Analysis: A Tutorial Introduction*, Cambridge, Mass.: MIT Press, 2004.

[8] E. Vincent. "Musical Source Separation Using Time-Frequency Priors", *IEEE Trans. on Audio, Speech and Language Proc.*, vol. 14, no. 1, pp. 91-98, 2006.

[9] T. Virtanen, and A. Klapuri. "Separation of Harmonic Sounds using Multipitch Analysis and Iterative Parameter Estimation", in *WASPAA 01 IEEE Workshop on App. of Signal Proc. to Audio and Acoustics Proc.*, 2001, pp. 83-86.

[10] T. Virtanen, and A. Klapuri. "Separation of Harmonic Sounds using Linear Models for the Overtone Series", in *ICASSP 02 IEEE Int. Conf. on Acoustics, Speech and Signal Processing Proc.*, 2002, pp. 1757-1760.

[11] H. Viste and G. Evangelista. "A Method for Separation of Overlapping Partials Based on Similarity of Temporal Envelopes in Multi-Channel Mixtures", *IEEE Trans. on Audio, Speech and Language Proc*., in press.

[12] G.H. Wakefield. "Mathematical Representation of Joint Time-Chroma Distributions", in *SPIE 99 Int. Symp. on Opt. Sci., Eng., and Instr. Proc.*, 1999, pp. 637-645.

[13] J. Woodruff and B. Pardo. "Active Source Estimation for Improved Source Separation", Northwestern University, EECS Dept. Technical Report, NWU-EECS-06-01, 2006

[14] O. Yilmaz and S. Rickard. "Blind Separation of Speech Mixtures via Time-Frequency Masking", *IEEE Transactions on Signal Processing*, vol. 52, no. 7, 2004, pp. 1830-1847.

[15] K. Yoshii, M. Goto, H. Okuno. "Inter:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums", in *EWIMT 05 2nd Euro. Workshop on the Int. of Knowledge, Semantic and Digital Media Tech. Proc*., 2005, pp. 205-212.