

R. Mitchell Parry
parry@cc.gatech.edu

Source Separation for Multichannel Music Audio

One fundamental difficulty with analyzing music audio is that it contains a mixture of sources. Pitch detection, instrument recognition, beat detection, segmentation, and transcription work best on single instrument audio. Source separation for music incorporates knowledge of musical signals and appears in several domains. Multipitch sound separation attempts to extract pitch from separate sources in mixtures (Klapuri; 2001). Voice separation techniques attempt to assign notes in a score to different instruments (Kilian and Hoos; 2002). Sound source localization attempts to position a single source spatially (Li and Levinson; 2003). Instrument recognition research attempts to classify sources (Martin; 1999). Two-source separation from stereo signals allows for greater than two sources as long as they do not overlap in time or frequency (Master; 2004).

Source separation also appears in statistically based techniques such as Independent Component Analysis (ICA) for Blind Source Separation (BSS) (Smaragdis; 2001). These techniques assume no knowledge about the sources except that at most one is Gaussian and that they are statistically independent. The setup includes N microphones in an environment and as many as N sources. Each source emanates from a unique spatial position. Therefore each microphone receives a unique mixture of signals. ICA algorithms estimate the independent components and their positions given the mixtures and the number of sources. ICA has been applied to stereo audio to separate the vocal from background sources (Feng et al.; 2002).

To our knowledge there has been no research that leverages the additional information provided by multichannel audio recordings such as DVD-Audio and Super Audio Compact Discs. DVD-Audio provides up to 6 channels of audio at 96 KHz sampling rate. This technical advance appears to substantially simplify the matter of separating sources. Editorial reviews claim that listening to an album that has been remixed for 6-channel audio is like hearing the recording for the first time. The greater spatial separation of instruments enhances our ability to distinguish and focus on a particular instrument. Psychoacoustical properties of multichannel audio may inform our method of analysis (Stuart; Psychoacoustics).

Although the new audio formats provide more information and greater separation potential, they also introduce new problems for instrument separation. For instance, ICA requires that the number of sources be known and unchanging. In general, we do not know how many sources to expect in music. Research in source number estimation (i.e. Aouada et al.; 2003) addresses this issue. By using small analysis windows we can expect that the number of sources is unlikely to change within a window. Also, in listening to these multichannel recordings it is evident that the rear speakers exhibit a slight echoing effect. Presumably, this is intended to simulate room acoustics in, for instance, a concert hall. Sound produced by an instrument arrives at a location in a room not only in a direct route, but also by bouncing off the walls in the room first. This is called the multipath problem. Blind deconvolution research addresses multipath issues with one source in

monophonic audio (Bell & Sejnowski; 1995). Multichannel blind deconvolution extends ICA techniques to simultaneously estimate the multipath filter for each source/channel pair (Lambert; 1996). Finally, the low frequency effect channel contains only low-frequency information. Even if emanates from an instrument present in other audio channels it does not share the same time-domain representation. However, ICA insists that there be one time-domain independent component that is simply mixed across multiple channels. Initially, ICA is poorly suited to incorporate information from the low frequency effect channel.

We propose to use a combination of techniques to separate sources from multichannel recordings. The highest goal is to separate all sources up to the number of channels in the audio. We can imagine separating more sources if they are separated in time. We expect that as the number of sources within a recording increase, the average time allocated each decreases. Because a vast majority of popular music contains fewer than 6 instruments, we are optimistic about the utility of such an approach.

The most daunting technical challenge will be reversing the effect of multipath issues with the rear channels. While a theoretical framework and implementation exist, solving for so many additional parameters than the standard ICA may require a greater number of samples to adequately train the model. Extra samples and a more complex model make this a much less desirable computational expense. Partial success may be achieved by employing only the three front audio channels for source separation. Of course, this provides a more incremental improvement to existing algorithms.

References:

- Aouada, S., Zoubir, A. M., and See, C. M. S. (2003). A Comparative Study on Source Number Estimation. In *Proceedings of the International Symposium on Signal Processing and its Applications*, Paris, France.
- Bell, A. J. & Sejnowski, T. J. (1995). An Information Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7, 1129-1159.
- Feng, Y., Zhuang, Y., Pan, Y. (2002). Popular Music Retrieval by Independent Component Analysis. In *Proceedings of the International Conference on Music Information Retrieval*, Paris, France.
- Kilian, J. & Hoos, H. H. (2002). Voice Separation — A Local Optimization Approach. In *Proceedings of the International Conference on Music Information Retrieval*, Paris, France.
- Klapuri, A. P. (2001). Multipitch Estimation and Sound Separation by the Spectral Smoothness Principle. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT.
- Lambert, R. H. (1996). *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*. Ph.D. dissertation, Electrical Engineering Department, University of Southern California.
- Li, D. & Levinson, E. (2003). A Bayes-Rule Based Hierarchical System for Binaural Sound Source Localization. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Hong Kong.
- Martin, K. D. (1999). *Sound-Source Recognition A Theory and Computational Model*. Ph.D. dissertation, Department of Electrical Engineering and Computer Science, MIT.
- Master, A. S. (2004). Bayesian Two Source Modeling for Separation of N Sources from Stereo Signals. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*. Montreal, Quebec.
- Smaragdis, P. (2001). *Redundancy Reduction for Computational Audition, a Unifying Approach*. Ph.D. dissertation, Media Arts and Sciences Department, MIT.
- Stuart, B. The Psychoacoustics of Multichannel Audio. White paper, Meridian Audio. Available on the Internet: http://www.meridian-audio.com/w_paper/multips3.pdf.