Source Separation for Multichannel Music Audio

R. Mitchell Parry

Computational Perception Lab

GVU Center / College of Computing

Georgia Institute of Technology

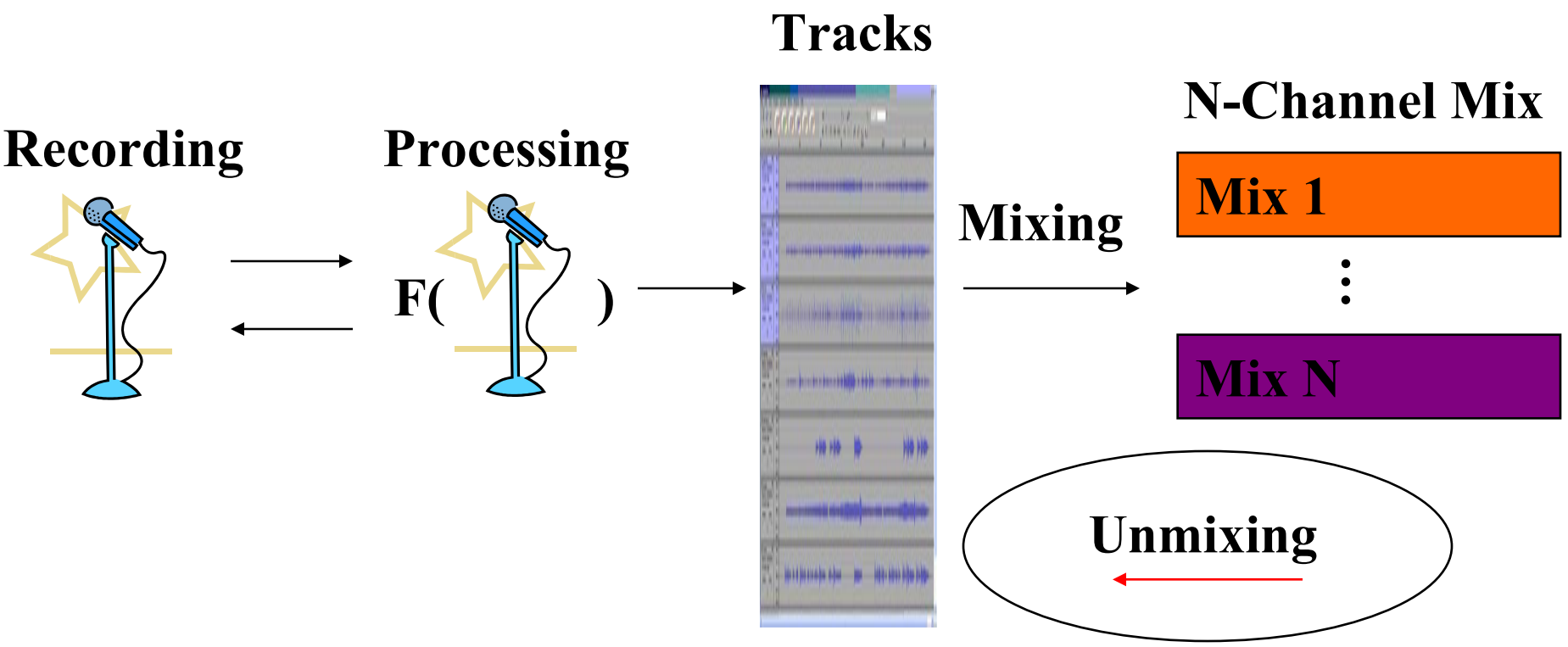
Goal

- Input: Multichannel mixed musical signals
- Output: Unmixed “instrument” source signals

Multichannel Audio

- DVD-Audio, Super Audio Compact Disc
- Up to 6 channels of 96 KHz sampling rate
- More than 1,000 titles and growing
- Better spatial separation of instruments
- Theoretically separate up to 6 instruments in a mix.

Recording / Mixing

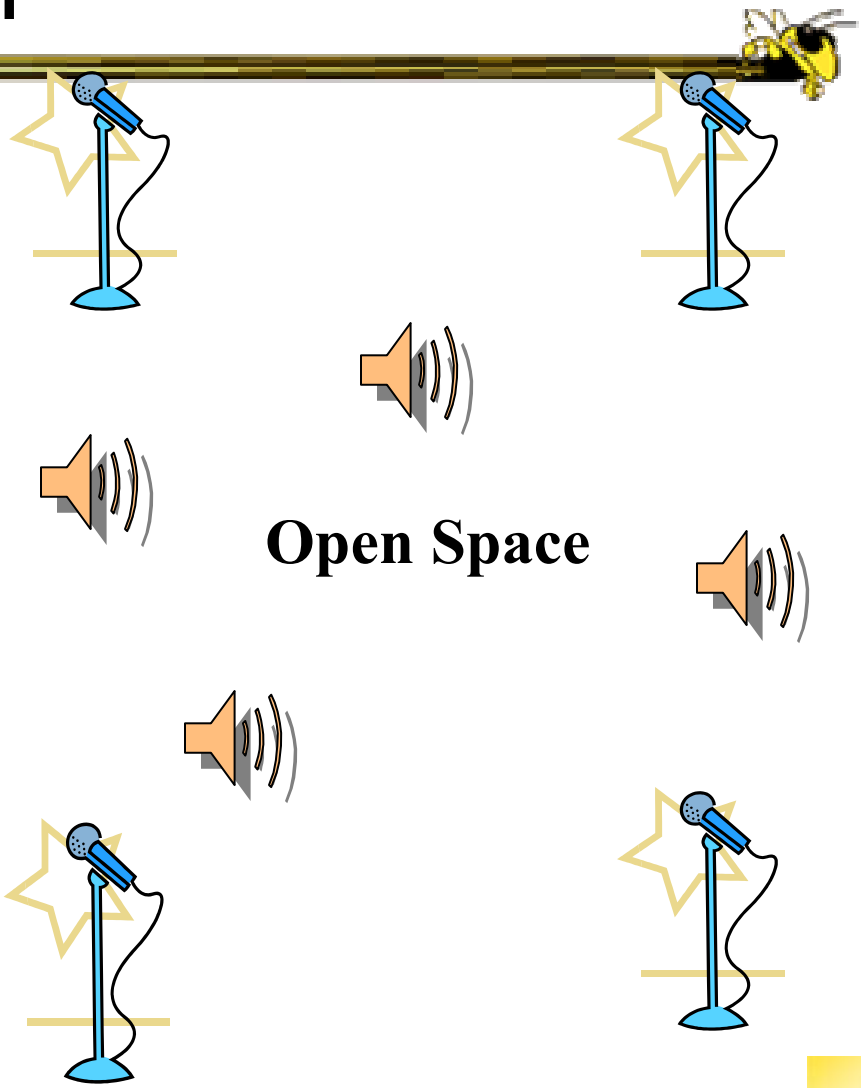


Applications …

- Enables common music analysis algorithms:
  - Pitch detection
  - Instrument recognition
  - Beat detection
  - Segmentation
  - Transcription
  - Structure analysis, ...
- Automatic sample extraction for synthesis / visualization applications.

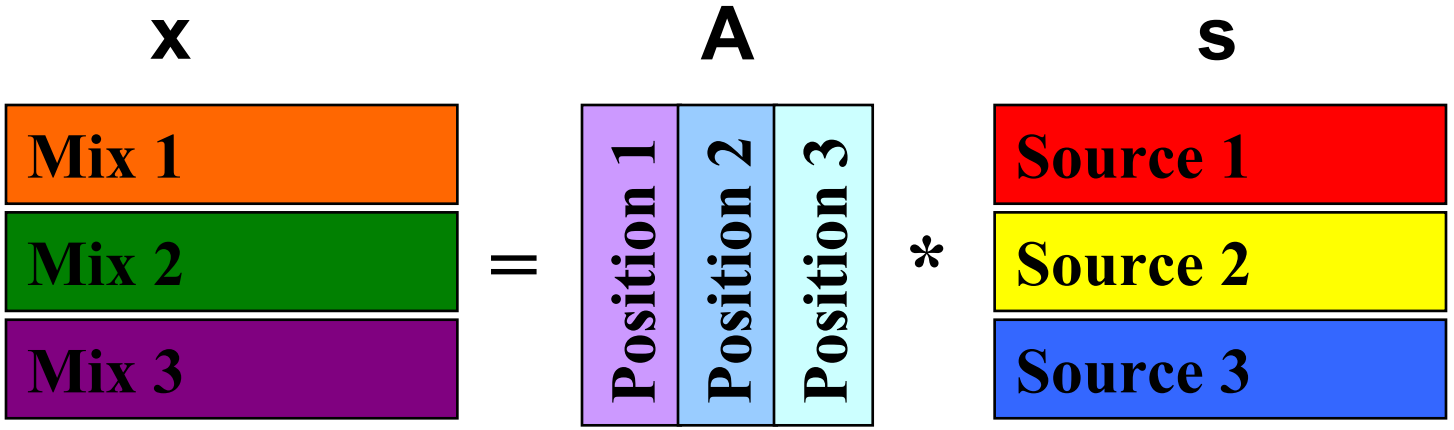
Source Separation

- Blind Source Separation (BSS) / Independent Component Analysis (ICA)
  - *N* recordings (mixtures)
  - *M* sources
  - *M* <= *N*
- Smaragdis, 2001



Blind Source Separation

- x = A*s



Challenges

- Number of sources
  - unknown, changing, > number of channels
- Moving sources
- Multiple paths
- Low frequency effects channel
- Evaluation

Unknown Number of Sources

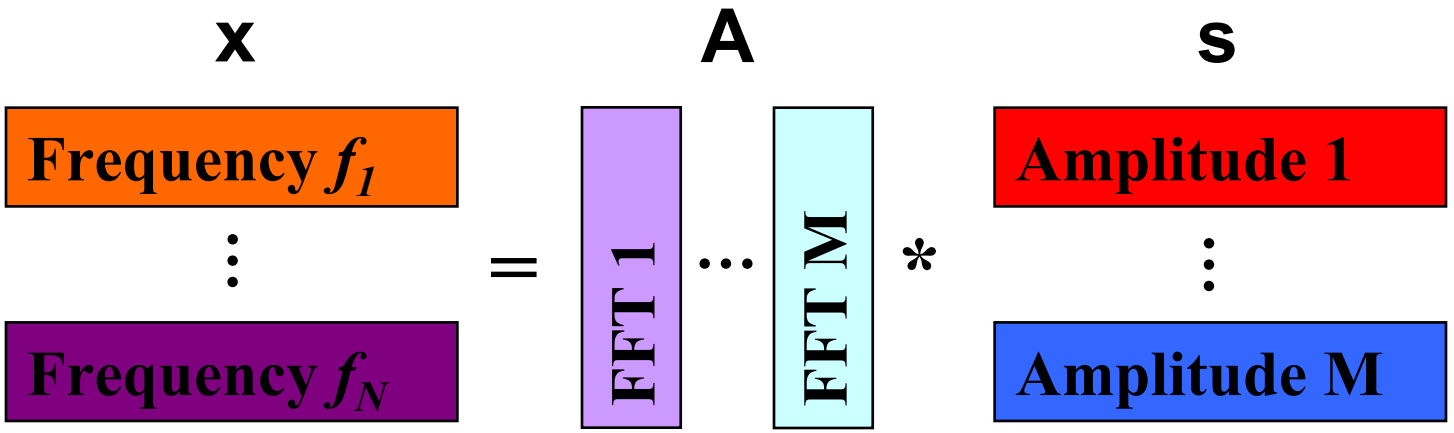
- ICA algorithms are parameterized by the number of sources present
- Source number estimation algorithms exist
  - Aouada et al. 2003
- Estimate sources, then ICA

Changing Number of Sources

- ICA assumes the number of sources does not change
- Separate sources and positions in small analysis windows
- Cluster source positions and combine

More Sources than Channels

- Solvable if not simultaneous
- Otherwise, move to frequency domain
  - x = Spectrogram
  - A = FFT of sources
  - s = Amplitude envelope of sources

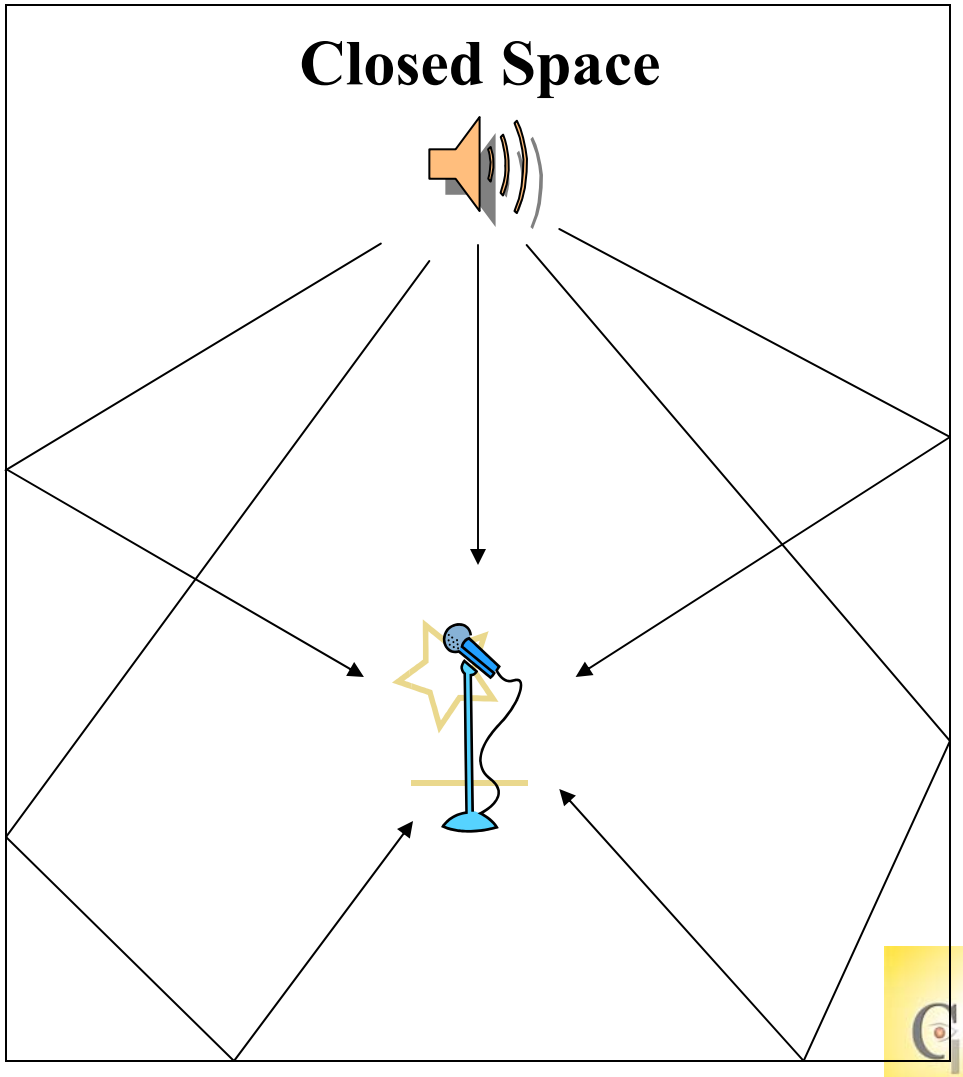


Moving Sources

- Spiraling effect
  - Tracking algorithms assume small change in position over time.
- Soloist always front and center
  - Cluster on timbre rather than position

Multiple Paths

- Concert Hall effect
- Impulse response of room



Multiple Paths

- Time delayed attenuated signals contribute to direct path
- For one source, one channel
  - Blind deconvolution (i.e. Bell & Sejnowski; 1995)
- For multiple source, multichannel
  - Multichannel blind deconvolution (Lambert; 1996)
  - ‘A’ is FIR matrix instead of scalar matrix

Low Frequency Channel

- Low frequency effect channel does not contain full source.
- ICA attempts to find components that spread across the channels
- Reduces to multiple path problem
  - FIR filter represents low-pass filter instead of impulse response of room.

Evaluation

- Evaluation
  - Mixtures of sinusoids
  - Mixtures of known sources
    - Anyone have multi-track recordings to share?
  - Mixtures with unknown sources (?)
    - DVD-Audio

Questions / Comments

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.