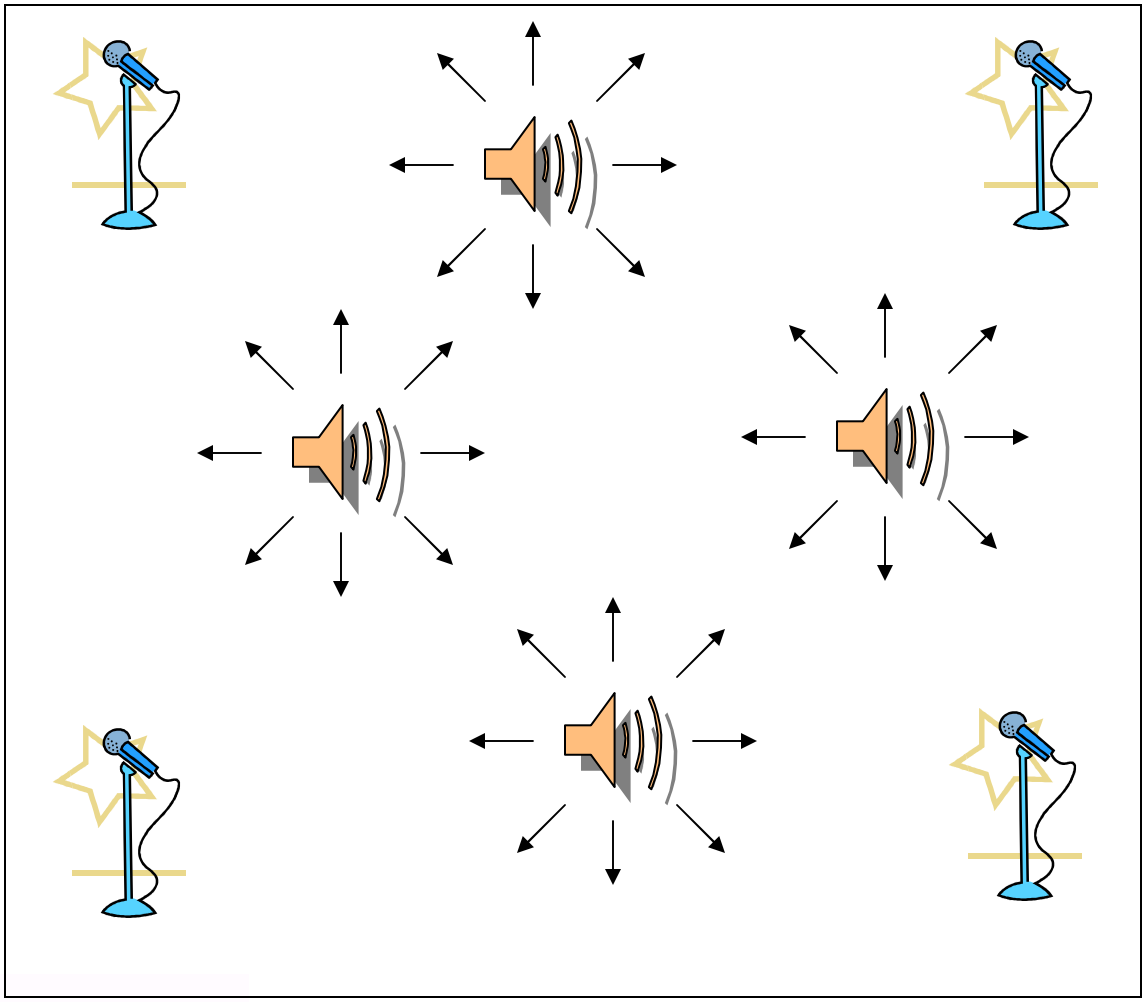
Blind Source Separation Using Repetitive Structure

R. Mitchell Parry & Irfan Essa
Computational Perception Lab
GVU Center / College of Computing
Georgia Institute of Technology
Atlanta, GA USA

Blind Source Separation



Level difference
Delay
Reverberation

Simple Data Model

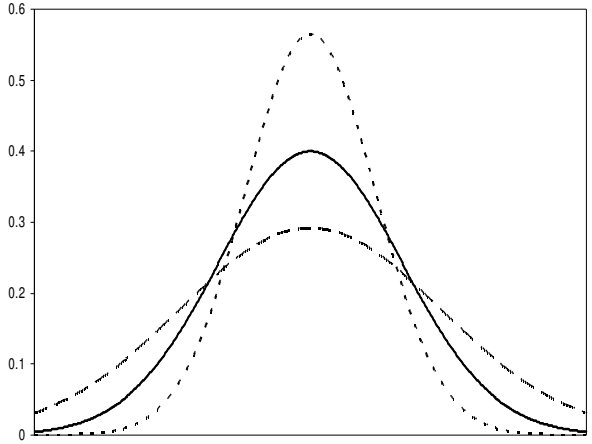
- Simplifying assumption:
  - No delay
  - No reverberation
- *M* linear mixtures, *N* sources
  - $$x_i(t) = \sum_{j=1}^{N} a_{ij} s_j(t)$$
  - $$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$
- Estimate $\mathbf{s}(t)$ or $\mathbf{A}$ given the mixtures.
  - $$\hat{\mathbf{s}}(t) = \mathbf{B}\mathbf{x}(t),\ \hat{\mathbf{A}} = \mathbf{B}^{-1}$$

Independent Component Analysis

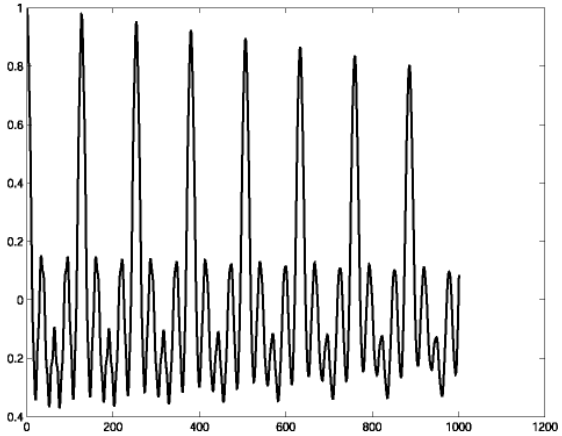
- Assume sources are statistically independent
  - $p(s_1, \ldots, s_N) = p(s_1) \ldots p(s_N)$
- General algorithm:
  - First, whiten mixtures with $\mathbf{W}$.
    - $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$
  - Then, find *unitary* matrix $\mathbf{U}$ that makes the estimated sources independent.
    - $\hat{\mathbf{s}} = \mathbf{U}^*\mathbf{z}(\mathbf{t}), \quad \hat{\mathbf{A}} = \mathbf{W}^{-1}\mathbf{U}$
- Independence is *not* enough
  - Need additional structure in sources

ICA: Additional Structure

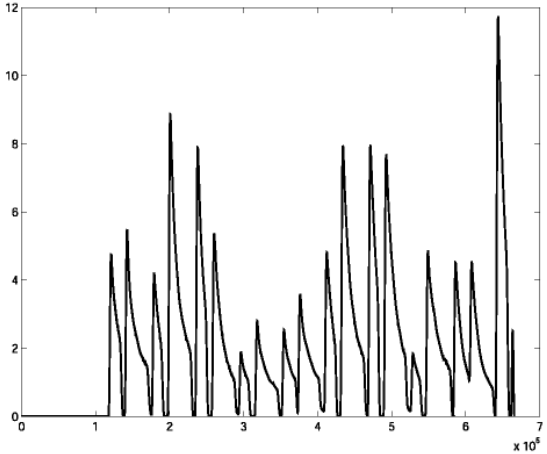
Non-Gaussian Distribution



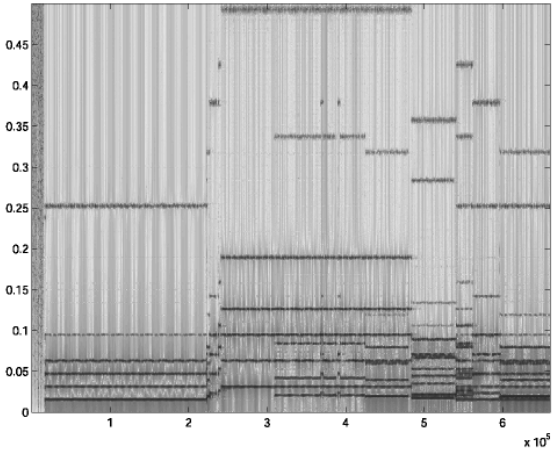
Correlated at Time Lags



Non-Stationary in Time



Non-Stationary in Time-Frequency



Time-Frequency Algorithm

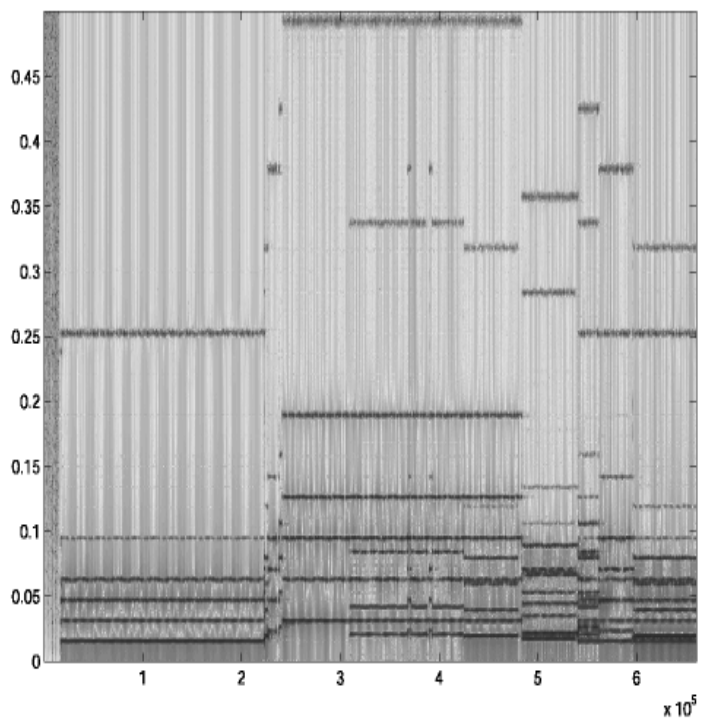
- Time-Frequency points with only one active source are autoterms.
- Correlation matrices between the original sources are diagonal at autoterm points.
- Use joint diagonalizer to find $\mathbf{U}$ that best diagonalizes these correlation matrices.

- $\hat{\mathbf{s}} = \mathbf{U}^{*}\mathbf{W}\mathbf{x}(\mathbf{t})$
- $\hat{\mathbf{A}} = \mathbf{W}^{-1}\mathbf{U}$

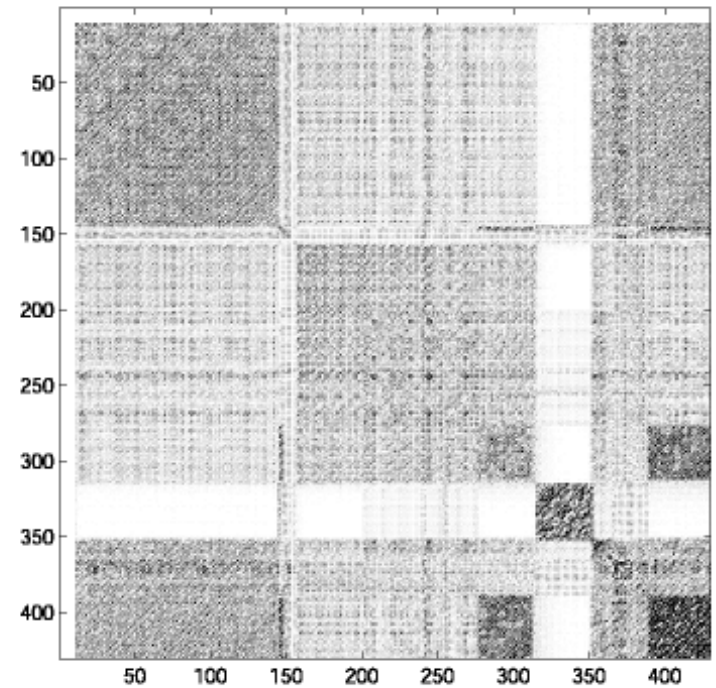
What about repetitive structure?

Repetitive Structure

Time-Frequency Representation



Self-Similarity Matrix =
Time-Time Representation



These two representations are related…

Time-Time Representation

- We design time-time and time-frequency representations to be related.
- Pseudo Wigner time-frequency rep.:

$$\mathbf{D}_{\mathbf{z}}(t,f) = \int h(\tau)\mathbf{z}(t+\frac{\tau}{2})\mathbf{z}^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau$$

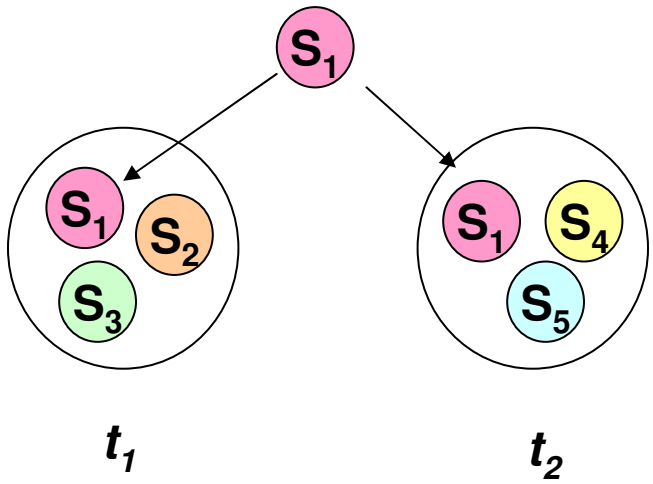
- Our analogous time-time representation:

$$\mathbf{S}_{\mathbf{z}}(t_1,t_2) = \int h(\tau)\mathbf{z}(t_1+\frac{\tau}{2})\mathbf{z}^*(t_2-\frac{\tau}{2})d\tau$$

So that the same algorithm applies…

New Time-Time Algorithm

- A time-time autoterm occurs when only one source is active at *both* points in time.



- Source correlation matrices at time-time autoterms are diagonal.
- Use joint diagonalizer to find $\mathbf{U}$ that best diagonalizes these correlation matrices.

Results

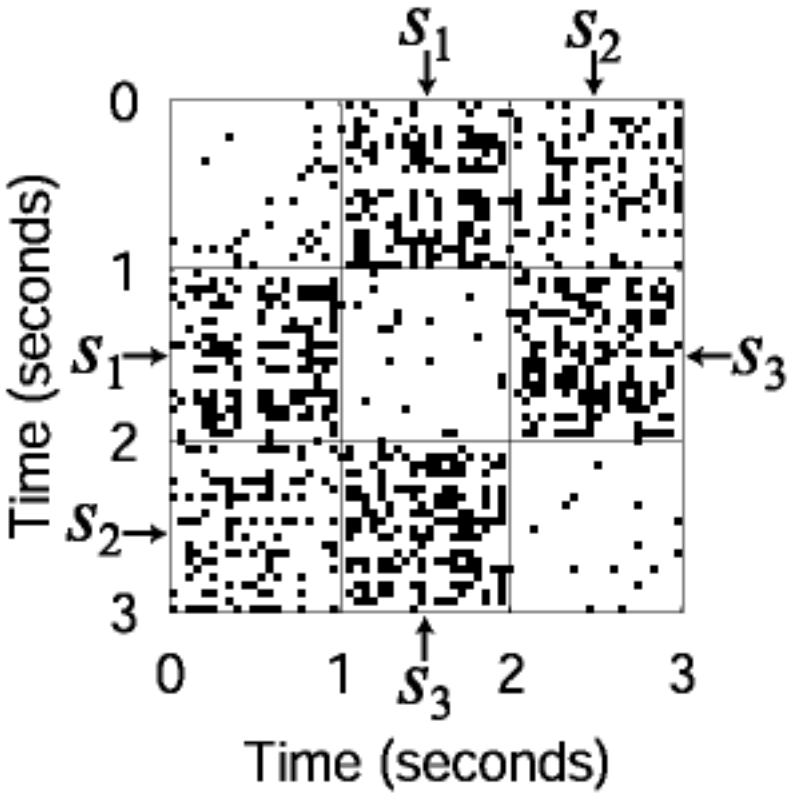
- Compare time-frequency and time-time separation based on estimated mixing matrices.
- Test cases:
  1. Synthetic sources varying in similarity
  2. Synthetic bell tower example
  3. Bass guitar and organ song excerpt
  4. Three clarinets playing the same note

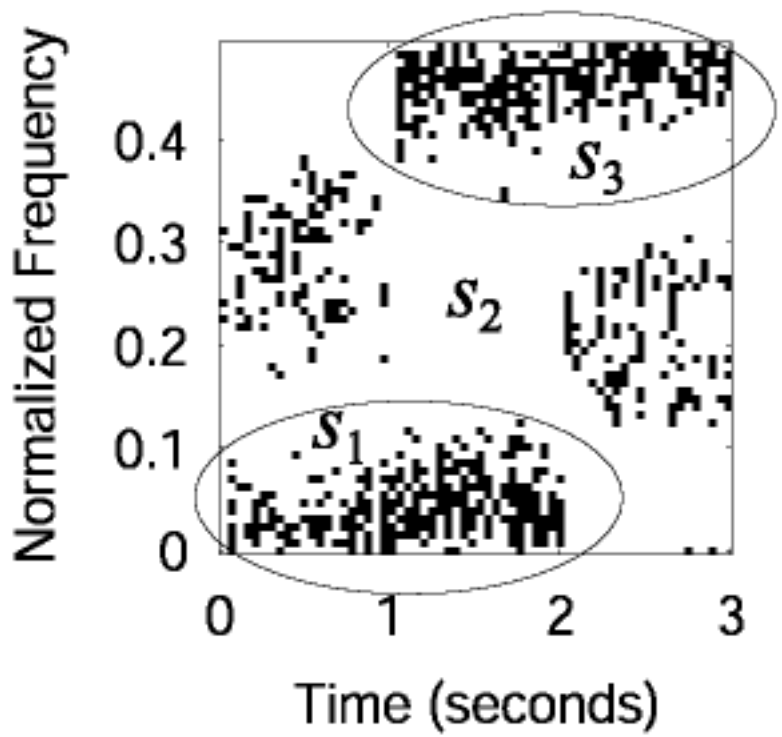
1: Synthetic Sources

- Sources, s_i, drawn from a Gaussian distribution and bandpass filtered around frequencies, f_i , and activated with pattern, a_i .
- $f_1 = 0.25 - \delta f \quad , \quad a_1 = [\texttt{on}, \texttt{on}, \texttt{off}]$
- $f_2 = 0.25 \quad , \quad a_2 = [\texttt{on}, \texttt{off}, \texttt{on}]$
- $f_3 = 0.25 + \delta f \quad , \quad a_3 = [\texttt{off}, \texttt{on}, \texttt{on}]$

1: Synthetic Source Autoterms

- $\delta f = 0.2$





1: Synthetic Source Error

δf	Time-Time Error (ISR)	Time-Frequency Error (ISR)
0.200	-20.9 dB	-20.6 dB
0.050	-14.2 dB	-13.1 dB
0.010	-11.2 dB	-8.9 dB
0.002	-10.8 dB	-8.7 dB
0.000	-10.8 dB	-8.7 dB

Georgia Institute of Technology

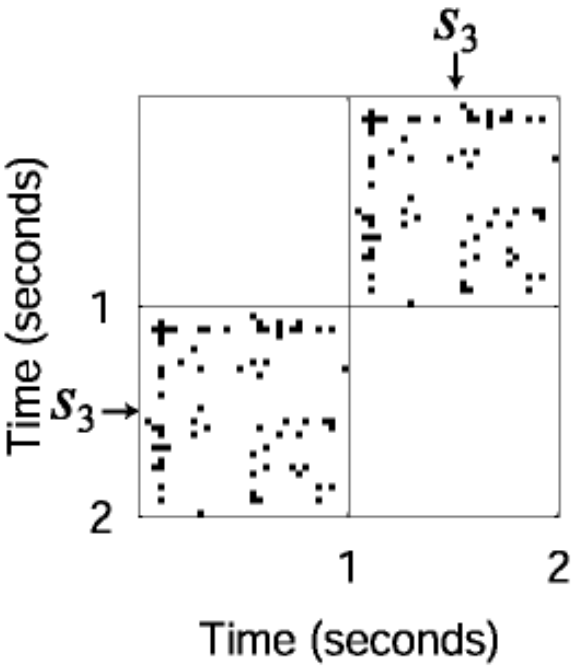
[gvu•center] computational perception lab

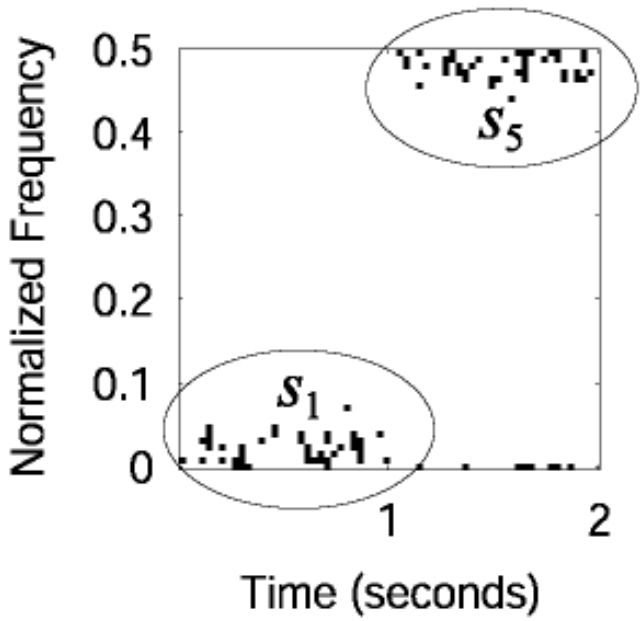
2: Synthetic Bell Tower

- Sources drawn from a Gaussian distribution and filtered using different activation patterns:

$$
\begin{aligned}
f_1 &= 0.05 & a_1 = a_2 &= [\texttt{on}, \texttt{off}] \\
f_2 &= 0.15 & a_3 &= [\texttt{on}, \texttt{on}] \\
f_3 &= 0.25 & a_4 = a_5 &= [\texttt{off}, \texttt{on}] \\
f_4 &= 0.35 \\
f_5 &= 0.45
\end{aligned}
$$

2: Synthetic Bell Tower Results



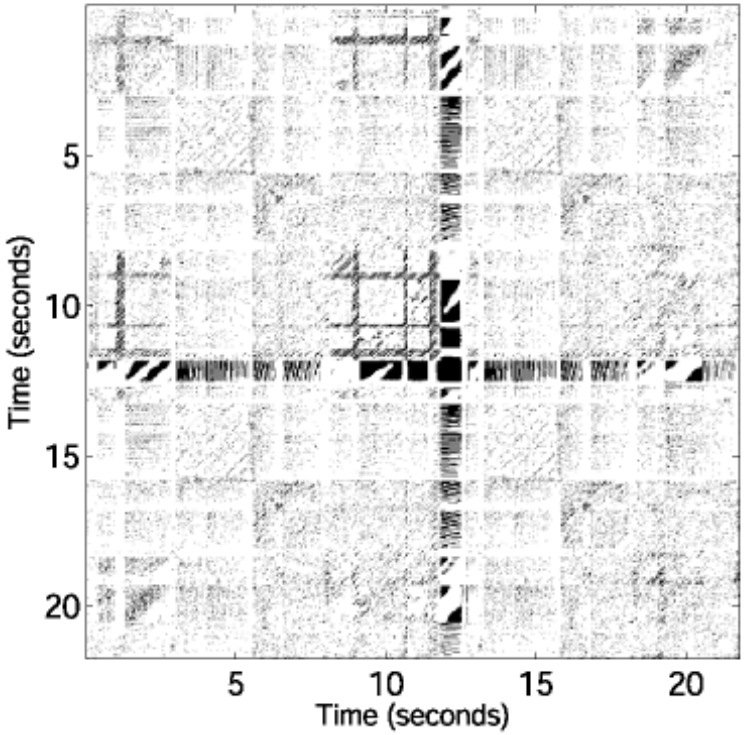


Error (ISR) = -17.5 dB

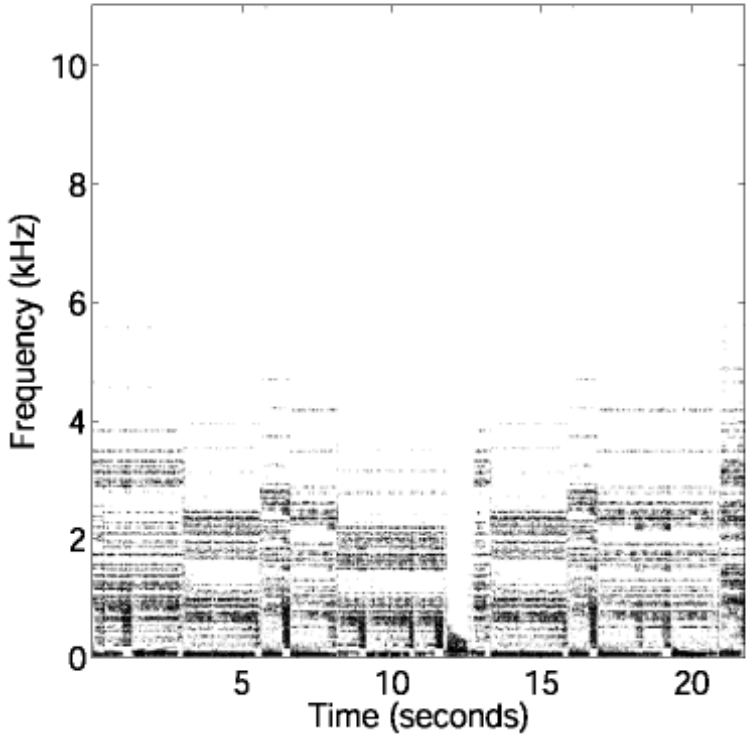
No s_3 autoterms with which to estimate $\hat{s}_3$.

3: Real Musical Signal Results

- Bass guitar and organ



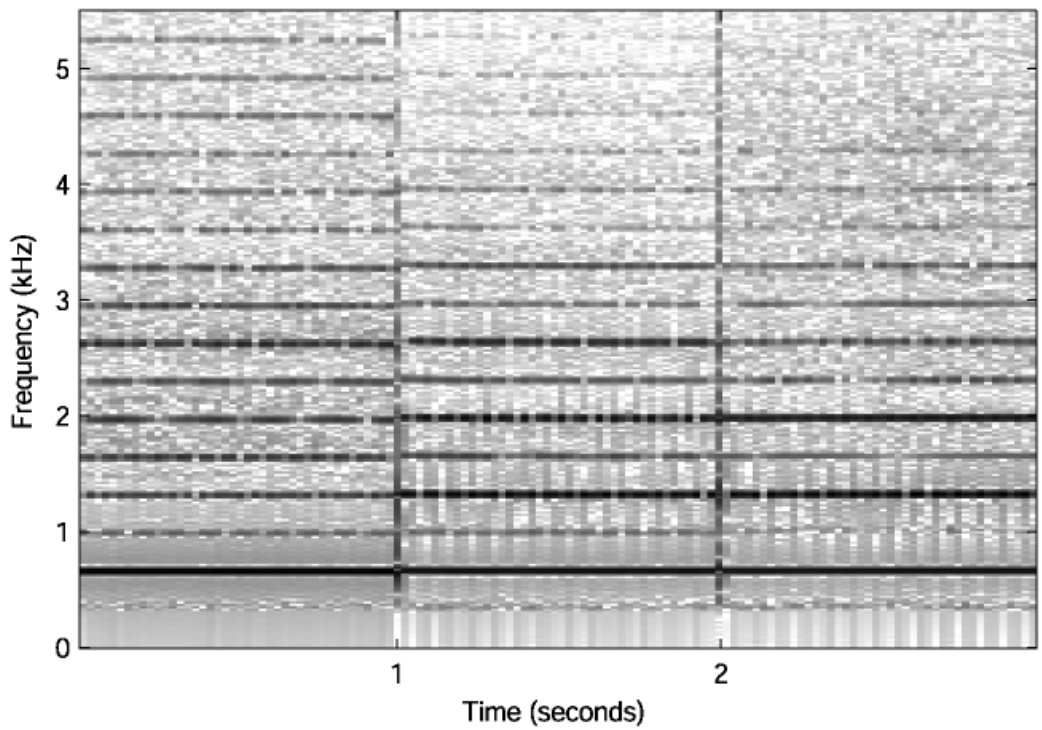
Error (ISR) = -19.2 dB



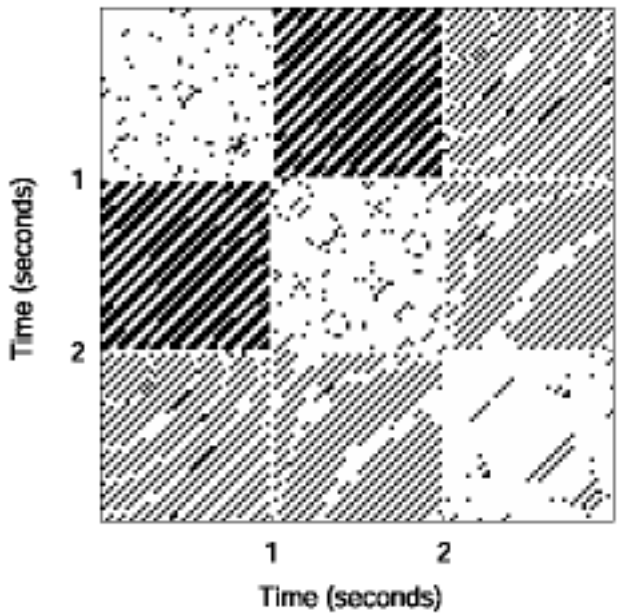
Error (ISR) = -18.0 dB

4: Three Clarinets, Same Note

- Bass, B♭, and E♭ clarinet

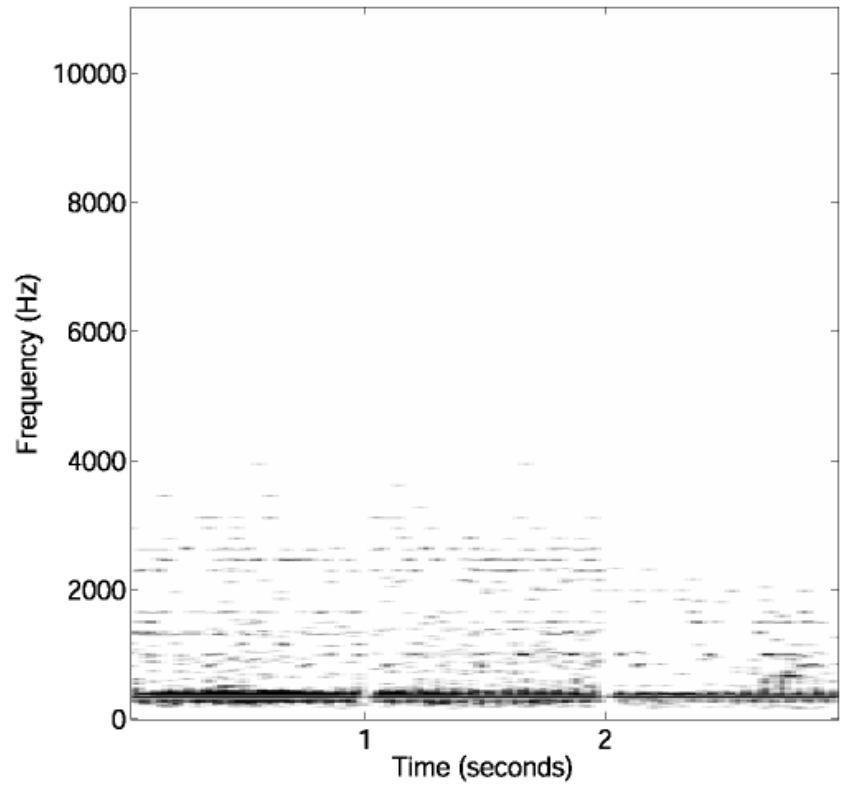


4: Clarinet Recordings Results



Error (ISR) = -13.1 dB

- Original
- Mix
- TF sep.
- TT sep.



Error (ISR) = -7.0 dB

Conclusion

- Time-Time separation leverages repetitive structure in sources.
- When sources are highly overlapping in time and frequency, time-time separation can still separate sources.
- Future work includes combining these approaches to leverage repetitive structure in the time-frequency domain.

Questions?