Recovering Audio Sources in a multi-path Environment

Lucas Parra, Paul Sajda
Air-coupled Acoustic Sensors Workshop
1. REPORT DATE  
25 AUG 1999

2. REPORT TYPE  
N/A

3. DATES COVERED  
-

4. TITLE AND SUBTITLE  
Recovering audio sources in a multi-path environment

5a. CONTRACT NUMBER  
-

5b. GRANT NUMBER  
-

5c. PROGRAM ELEMENT NUMBER  
-

5d. PROJECT NUMBER  
-

5e. TASK NUMBER  
-

5f. WORK UNIT NUMBER  
-

6. AUTHOR(S)  
Sarnoff Corporation

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  
Sarnoff Corporation

8. PERFORMING ORGANIZATION REPORT NUMBER  
-

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)  
-

10. SPONSOR/MONITOR’S ACRONYM(S)  
-

11. SPONSOR/MONITOR’S REPORT NUMBER(S)  
-

12. DISTRIBUTION/AVAILABILITY STATEMENT  
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES  
DARPA, Air-Coupled Acoustic Microsensors Workshop held on August 24 and 25, 1999 in Crystal City, VA., The original document contains color images.

14. ABSTRACT  
-

15. SUBJECT TERMS  
-

16. SECURITY CLASSIFICATION OF:  
<table>
<thead>
<tr>
<th>a. REPORT</th>
<th>b. ABSTRACT</th>
<th>c. THIS PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
</tr>
</tbody>
</table>

17. LIMITATION OF ABSTRACT  
UU

18. NUMBER OF PAGES  
10

19a. NAME OF RESPONSIBLE PERSON  
-

Standard Form 298 (Rev. 8-98)  
Prescribed by ANSI Std Z39-18
Geometric (Adaptive) Beamforming

- **What it is**
  - microphone array with fixed geometric configuration
  - adaptive algorithms to steer and adjust beam pattern
- **Typical Applications**
  - Attenuation of jammers
  - source localization
- **Problems/Issues**
  - Requires known/fixed array configuration
  - Cannot handle multiple sources
  - Signal leakage, reverberation
Statistical (Blind) Beamforming

- **What it is**
  - Multiple sensors at arbitrary locations
  - adaptive algorithm to recover independent/decorrelated source signals

- **Typical Applications**
  - simultaneous recovery of multiple sources
  - jammer attenuation under reverberation, and target signal leakage

- **Problems/Issues**
  - requires low noise sensors
  - computational complexity
Recovering Speech from Simultaneous Recording
(Statistical Beamforming Demo)

... isolating individual speakers with multiple microphones ...

- Instantaneous mixture - corresponds to environment with no reverberation and known time delays.
- Solution can’t assume knowledge of speaker/source location - requires a “blind” algorithm.
- Demo 1 - linear mixing of 10 speakers.
- Demo 2 - “real-world” deconvolution with 2 speakers.
- Sarnoff algorithm exploits non-stationarity of speech signal, performing multiple decorrelation across time to compute a matrix of “unmixing” FIR filters.
Acoustic signals $x(t)$ recorded simultaneously in a reverberant environment $A(\tau)$ can be described as sums of differently convolved sources $s(t)$.

$$x(t) = \sum_{\tau = 0}^{P} A(\tau)s(t-\tau) + n(t)$$

with $\dim(x) \geq \dim(s)$
Context on Blind Source Separation

**PCA:**

\[ x = R s \]
\[ R : s = R^T x \quad <s_i s_j> = \delta_{ij} \lambda_i \]

**ICA:**

\[ x = A s \]
\[ A : s = A^{-1} x \quad <s^n_i s^m_j> = \delta_{ij} \lambda_i^{n+m} \]
\[ W : s = W x \quad <s_i(0) s_j(t)> = \delta_{ij} \lambda_i(t) \]

**BSS:**

\[ x = A \otimes s(t) \]
\[ A : s(t) = A^{-1} \otimes x(t) \quad <s^n_i(0) s^m_j(t)> = \delta_{ij} \lambda_{inm}(t) \]
\[ W : s(t) = W \otimes x(t) \quad <s_i(t) s_j(t')> = \delta_{ij} \lambda_i(t,t') \]
Approach - Use Non-stationarity

Measure time dependent second order statistic

\[
\overline{R} \left( \omega, t \right) = \frac{1}{N} \sum_{n=0}^{N-1} x(\omega, t + nT)x^H(\omega, t + nT)
\]

Where \( x(\omega, t) \) are the frequency components of frame \( [x(t), \ldots, x(t + T)] \)
Experimental Setup: Speaker with Interfering Source

Reverberant environment (small office room). The interfering signal was a competing speaker or music.
Left:  $a = b = 50^\circ$, $c = 50^\circ$, $6^\circ$.
Right: $b = 30^\circ$, $c = 60^\circ$, $\alpha = 45^\circ$, $180^\circ$. 
Multiple Microphone Performance

Performance of multiple microphones in simulated room (small office) for separating a speaker from music background. Microphone distance 2m.

Separation Performance of two sources in simulated room

Separation Performance of two sources in simulated room

- SIR improvement in dB
- Number of microphones
- Q=512
- Q=2048

9 Parra/08/24/1999
Speech Recognition Improvement

Word error rate (WER) of ViaVoice (IBM) on a short text (Wallstreet Journal article of 760 words length) before and after source separation. The result is contrasted to clean recording with no interfering source.

up to 50% reduction in word error rate
for IBM Viavoice