
Recovering Audio Sources in a multi-path Environment

Lucas Parra, Paul Sajda

Air-coupled Acoustic Sensors Workshop

Report Documentation Page

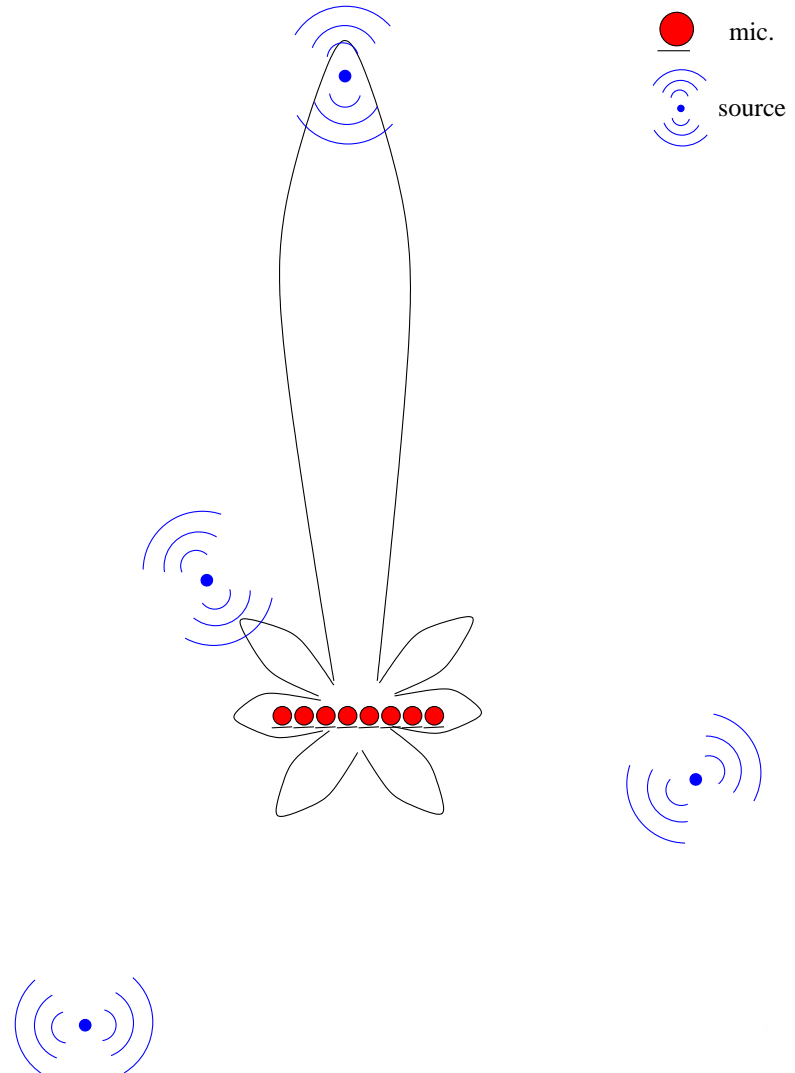
Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

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	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
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:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

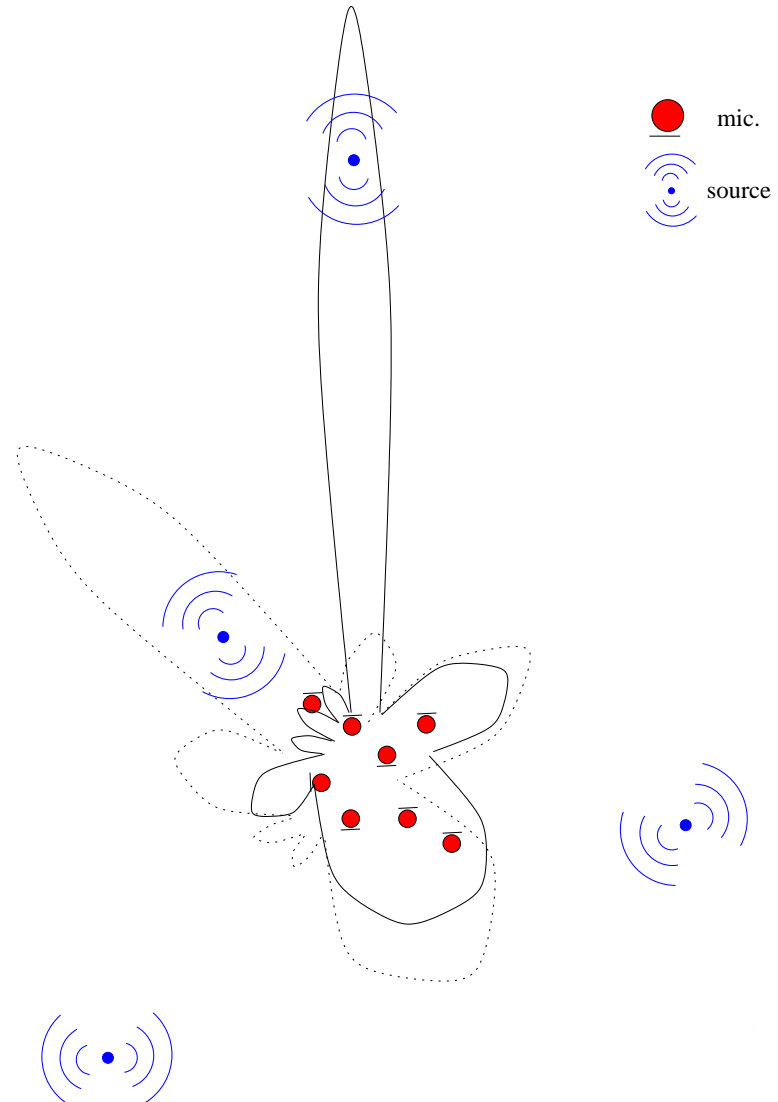
Geometric (Adaptive) Beamforming

- 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 know/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.

The Problem: Convolutional Mixture



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:

$$\mathbf{x} = \mathbf{R}s$$

$$\mathbf{R} : s = \mathbf{R}^T \mathbf{x}$$

$$\langle s_i s_j \rangle = \delta_{ij} \lambda_i$$

ICA:

$$\mathbf{x} = \mathbf{A}s$$

$$\mathbf{A} : s = \mathbf{A}^{-1} \mathbf{x}$$

$$\mathbf{W} : s = \mathbf{W} \mathbf{x}$$

$$\langle s_i^n s_j^m \rangle = \delta_{ij} \lambda_i^{n+m}$$

$$\langle s_i(0) s_j(t) \rangle = \delta_{ij} \lambda_i(t)$$

BSS:

$$\mathbf{x} = \mathbf{A} \otimes s(t)$$

$$\mathbf{A} : s(t) = \mathbf{A}^{-1} \otimes \mathbf{x}(t)$$

$$\mathbf{W} : s(t) = \mathbf{W} \otimes \mathbf{x}(t)$$

$$\langle s_i^n(0) s_j^m(t) \rangle = \delta_{ij} \lambda_{inm}(t)$$

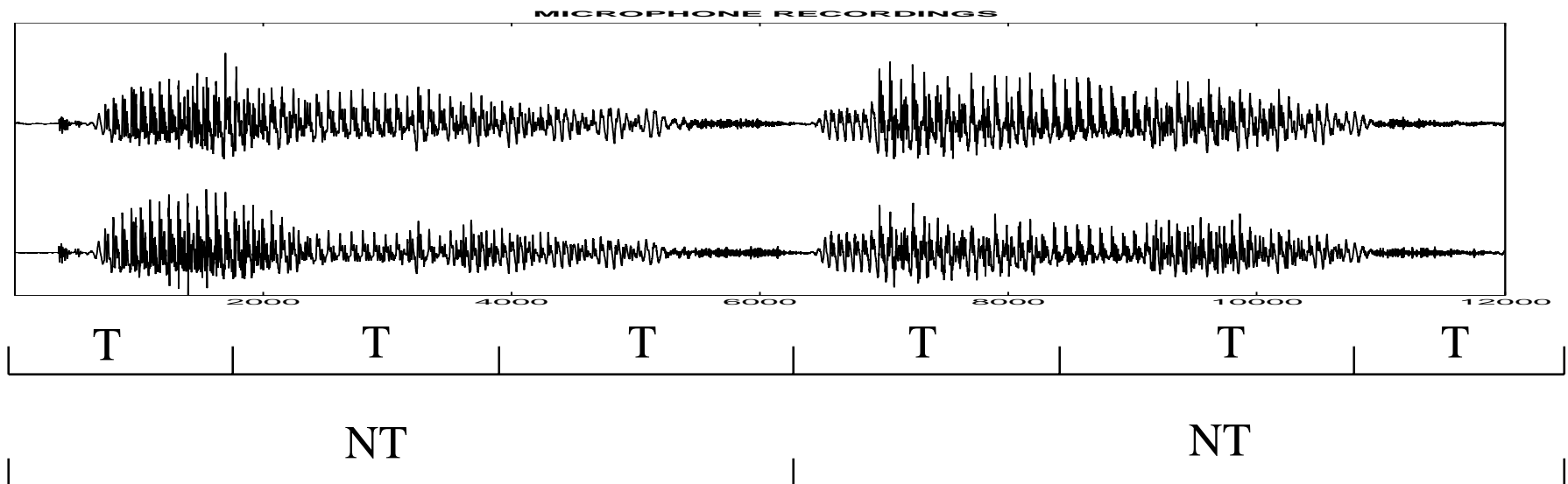
$$\langle s_i(t) s_j(t') \rangle = \delta_{ij} \lambda_i(t, t')$$

Approach - Use Non-stationarity

Measure time dependent second order statistic

$$\bar{\mathbf{R}}_{\div}(\omega, t) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}(\omega, t + nT) \mathbf{x}^H(\omega, t + nT)$$

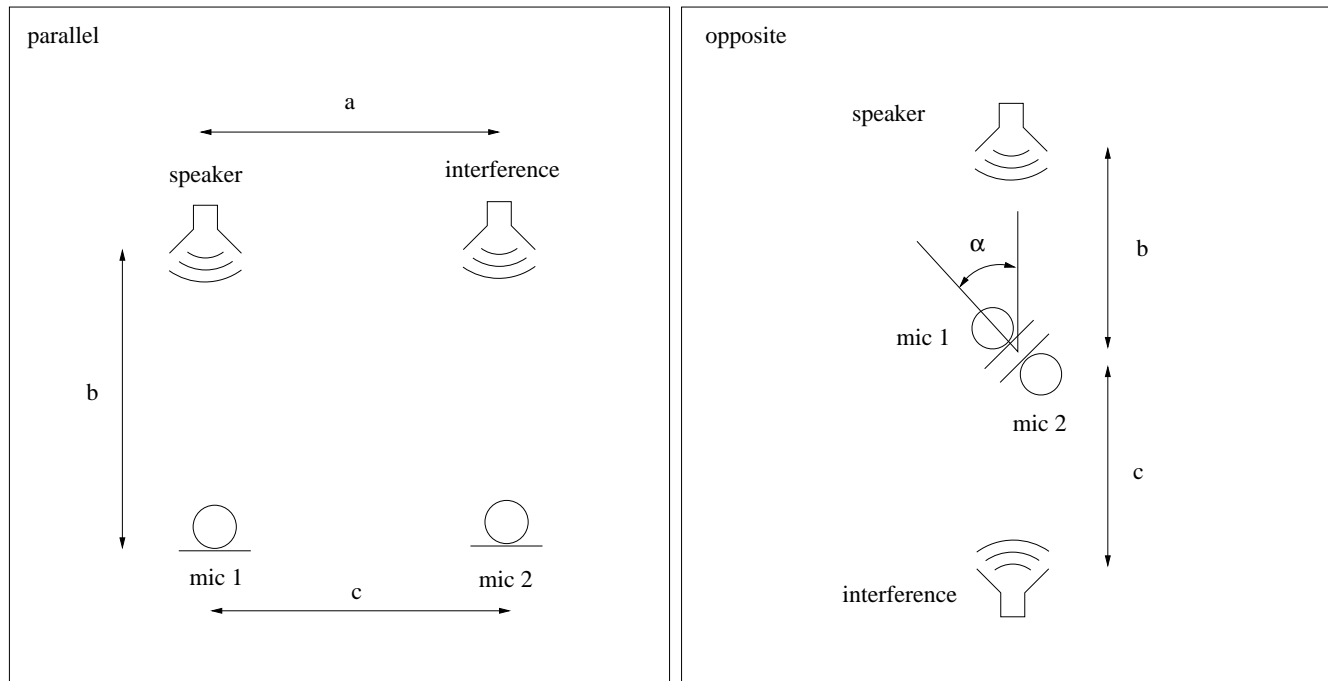
Where $\mathbf{x}(\omega, t)$ are the frequency components of frame $[\mathbf{x}(t), \dots, \mathbf{x}(t + T)]$



t_1

t_2

Experimental Setup: Speaker with Interfering Source



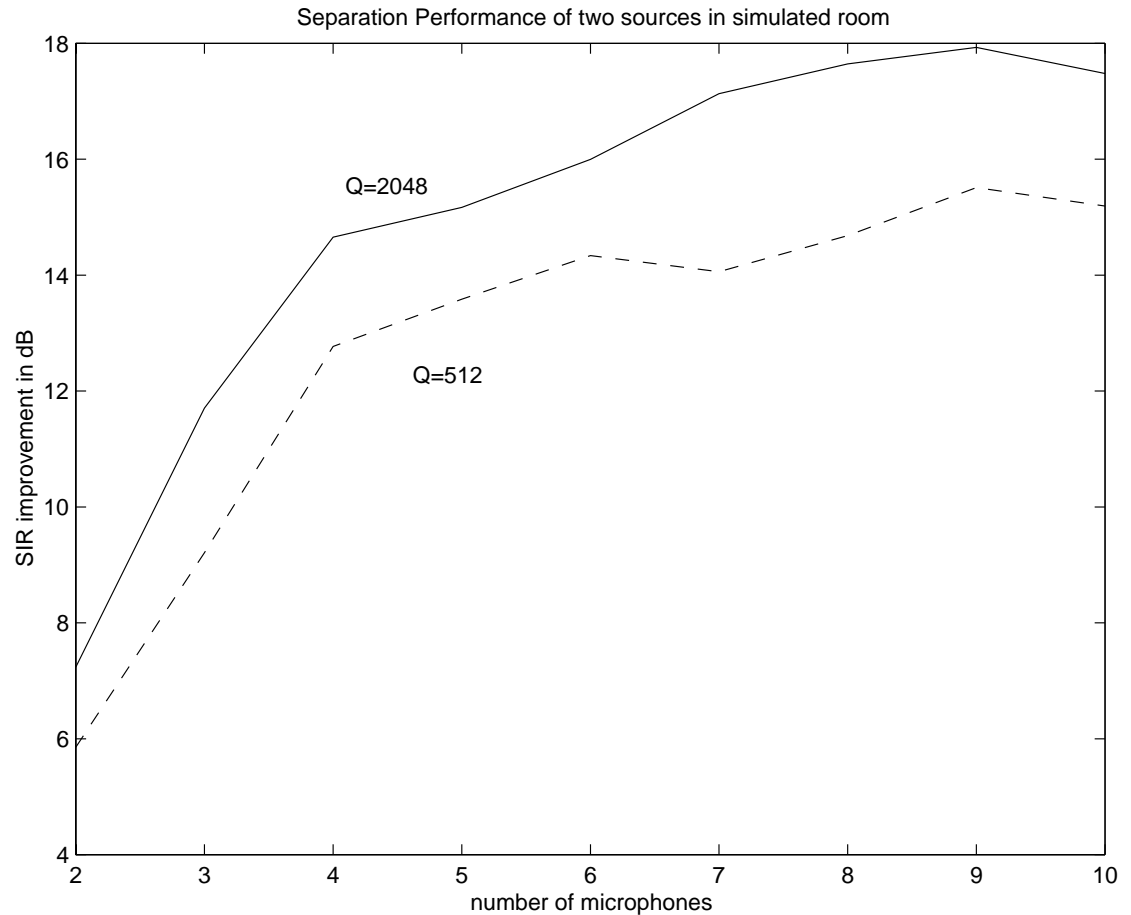
Reverberant environment (small office room). The interfering signal was a competing speaker or music.

Left: $a = b = 50?$, $c = 50?$, $6?$.

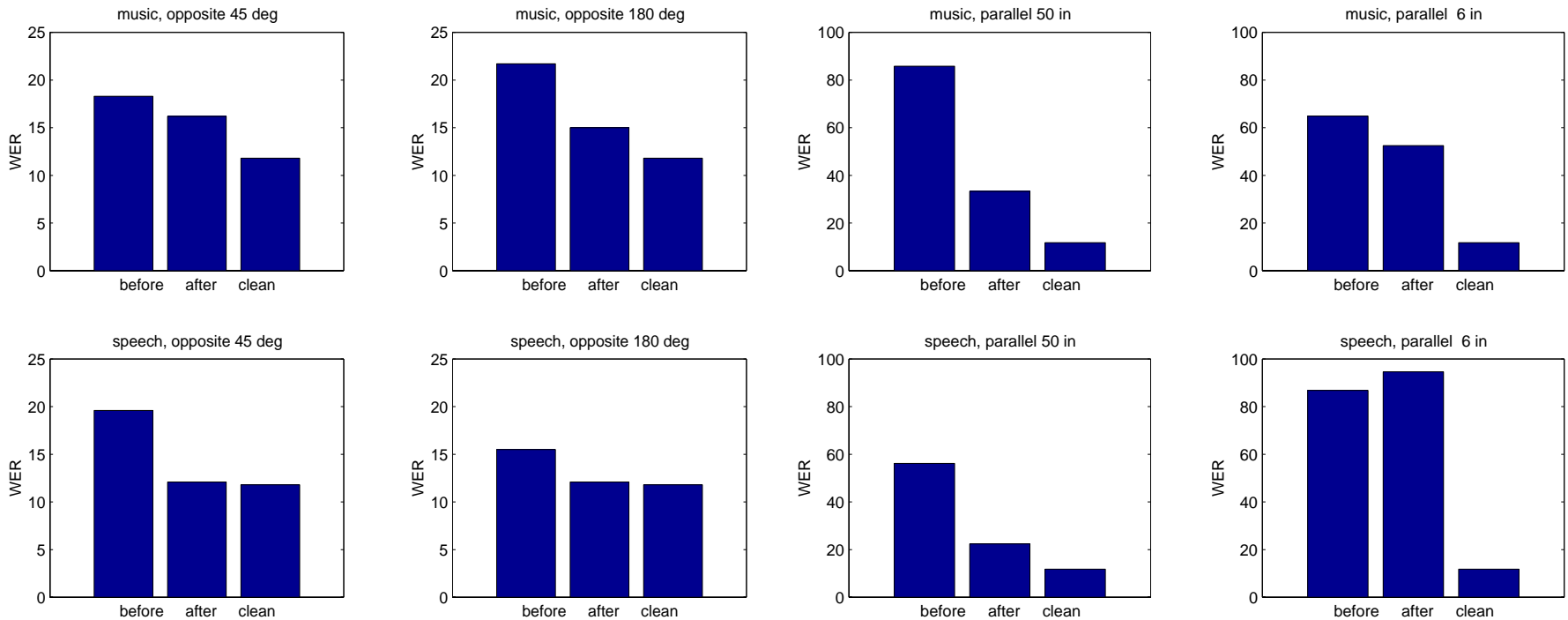
Right: $b = 30?$, $c = 60?$, $\alpha = 45^\circ$, 180° .

Multiple Microphone Performance

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



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