Multi-source TDOA estimation using SNR-based angular spectra

Charles Blandin, Emmanuel Vincent and Alexey Ozerov

INRIA, Centre de Rennes - Bretagne Atlantique, France

emmanuel.vincent@inria.fr

Overall approach

We aim to find the *time differences of arrival* (TDOAs) τ_n of several sound sources from a two-channel mixture signal.

We consider the *angular spectrum* approach illustrated by GCC-PHAT, MUSIC or cSCT [4], which consists of

- defining a local angular spectrum $\phi_{tf}(\tau)$ describing sound direction in each time-frequency bin (t, f),
- summing it over the time-frequency plane

$$\phi(\tau) = \sum_{tf} \phi_{tf}(\tau), \quad (1)$$

- picking the J largest peaks of $\phi(\tau)$.

Current methods essentially assign the same weight to all time-frequency bins, whether they result from one or several sources.

SNR-based angular spectra

Following advances in source localization for instantaneous mixtures [1], we define $\phi_{tf}(\tau)$ as the *signal-to-noise ratio* between the signal power along direction τ *vs.* other directions.

From now on, we consider a single time-frequency bin (t, f) and omit its indices.

SNR estimation by beamforming

Beamforming can be used to estimate the power along direction τ and subtract it from the total power.

Denoting by $\mathbf{d}_\tau = [1, e^{-2i\pi f\tau}]^T$ the *steering vector* associated with TDOA τ and by $\widehat{\mathbf{\Phi}}_{\mathbf{xx}}$ the *covariance matrix* of the input signal, we get

$$\phi_{\mathrm{DS}}(\tau) = \frac{\mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}} \mathbf{d}_\tau}{2\mathrm{tr}(\widehat{\mathbf{\Phi}}_{\mathbf{xx}}) - \mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}} \mathbf{d}_\tau} \quad (2)$$

$$\phi_{\mathrm{MVDR}}(\tau) = \frac{(\mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}}^{-1} \mathbf{d}_\tau)^{-1}}{\frac{1}{2}\mathrm{tr}(\widehat{\mathbf{\Phi}}_{\mathbf{xx}}) - (\mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}}^{-1} \mathbf{d}_\tau)^{-1}} \quad (3)$$

Minimum variance distortionless response (MVDR) beamforming provides enhanced peaks compared to delay-and-sum (DS) beamforming, but overestimates the SNR at low frequencies.

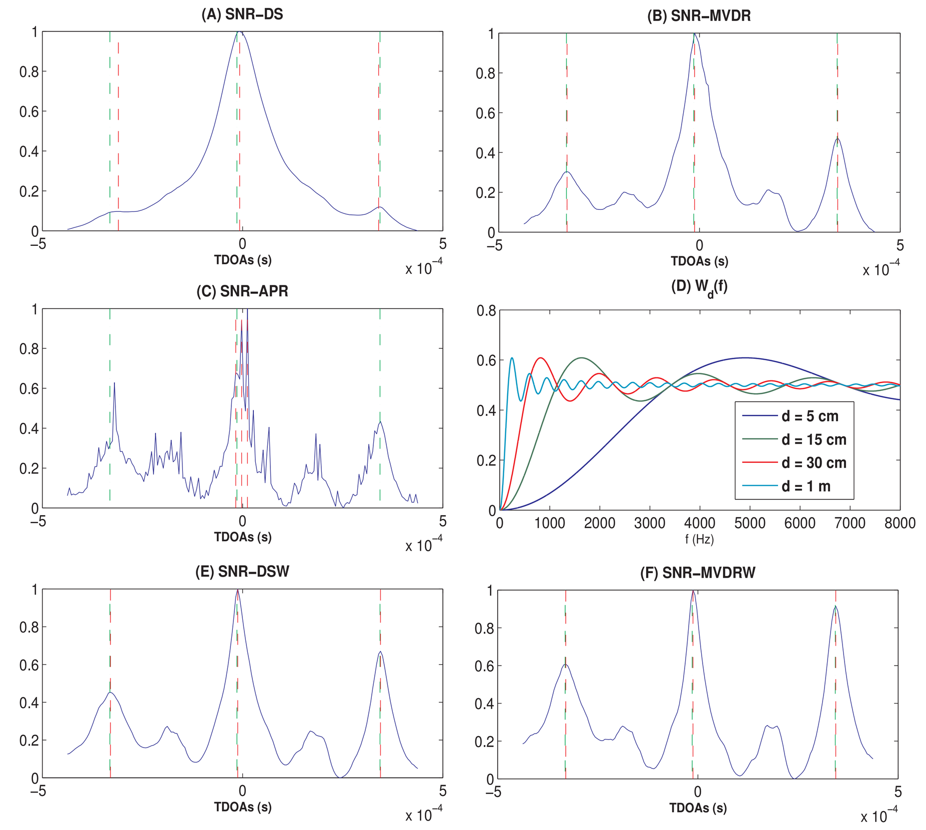

Figure 1: *Angular spectra and weighting curves.*

SNR estimation by probabilistic modeling

An alternative approach is to jointly estimate the signal power v_s in direction τ and the residual power v_b according to some probabilistic model.

Assuming that the predominant source and the residual are zero-mean Gaussian and that the residual is spatially diffuse, the mixture signal is zero-mean Gaussian with covariance matrix

$$\mathbf{\Phi}_{\mathbf{xx}} = v_s \mathbf{d}_\tau \mathbf{d}_\tau{}^H + v_b \mathbf{\Psi} \quad (4)$$

where

$$\mathbf{\Psi} = \begin{pmatrix} 1 & \mathrm{sinc}(2\pi f \frac{d}{c}) \\ \mathrm{sinc}(2\pi f \frac{d}{c}) & 1 \end{pmatrix}. \quad (5)$$

v_s and v_b can be estimated in closed form in the maximum likelihood (ML) sense using the algorithm in [3].

We then compute the SNR by

$$\phi_{\mathrm{APR}}(\tau) = \frac{v_s}{v_b} \quad (6)$$

This form of SNR increases the resolution of the peaks but results in secondary peaks around the true TDOAs.

SNR estimation by weighted beamforming

We combine both approaches to obtain smooth, high resolution spectra.

To this aim, we express the relationship between the above two forms of SNR when the input signal consists of a source of TDOA $\tau = 0$ and a diffuse noise.

By plugging (4) into (2) and (3), we obtain

$$\phi_{\mathrm{DS}}(\tau) = \frac{1 + 2\, v_s/v_b + \mathrm{sinc}(2\pi f \frac{d}{c})}{1 - \mathrm{sinc}(2\pi f \frac{d}{c})} \quad (7)$$

$$\phi_{\mathrm{MVDR}}(\tau) = \frac{1 + 2\, v_s/v_b}{1 - \mathrm{sinc}(2\pi f \frac{d}{c})} \quad (8)$$

By inverting these equations, we define

$$\phi_{\mathrm{DSW}}(\tau) = \frac{v_s}{v_b} = W_d\, \phi_{\mathrm{DS}}(\tau) + W_d - 1 \quad (9)$$

$$\phi_{\mathrm{MVDRW}}(\tau) = \frac{v_s}{v_b} = W_d\, \phi_{\mathrm{MVDR}}(\tau) - \frac{1}{2} \quad (10)$$

where

$$W_d = \frac{1 - \mathrm{sinc}(2\pi f \frac{d}{c})}{2} \quad (11)$$

is a frequency weighting curve depending on the distance d between microphones.

Experimental evaluation

We evaluated the proposed methods on 4446 mixture signals with

- 2 to 6 sources
- reverberation times from 50 ms to 750 ms
- microphone spacings from 5 cm to 1 m
- distances between the sources and the microphones from 20 cm to 2 m
- several source DOAs
- three source types (male speech, female speech and music)

ϕ_{MVDRW} outperforms all other algorithms both in terms of recall and precision for most microphone spacings and for all reverberation times.

	SNR					GCC		
	DS	MVDR	APR	DSW	MVDRW	PHAT	MUSIC	cSCT
$\mathcal{R}$	0.41	0.58	0.62	0.63	0.66	0.56	0.55	0.61
$\mathcal{P}$	0.48	0.61	0.37	0.67	0.69	0.65	0.33	0.64
$\mathcal{F}$	0.43	0.59	0.46	0.65	0.67	0.58	0.41	0.62

Table 1: *Average recall, precision and F-measure over all configurations (tolerance* $0.05 \times d/c$*).*

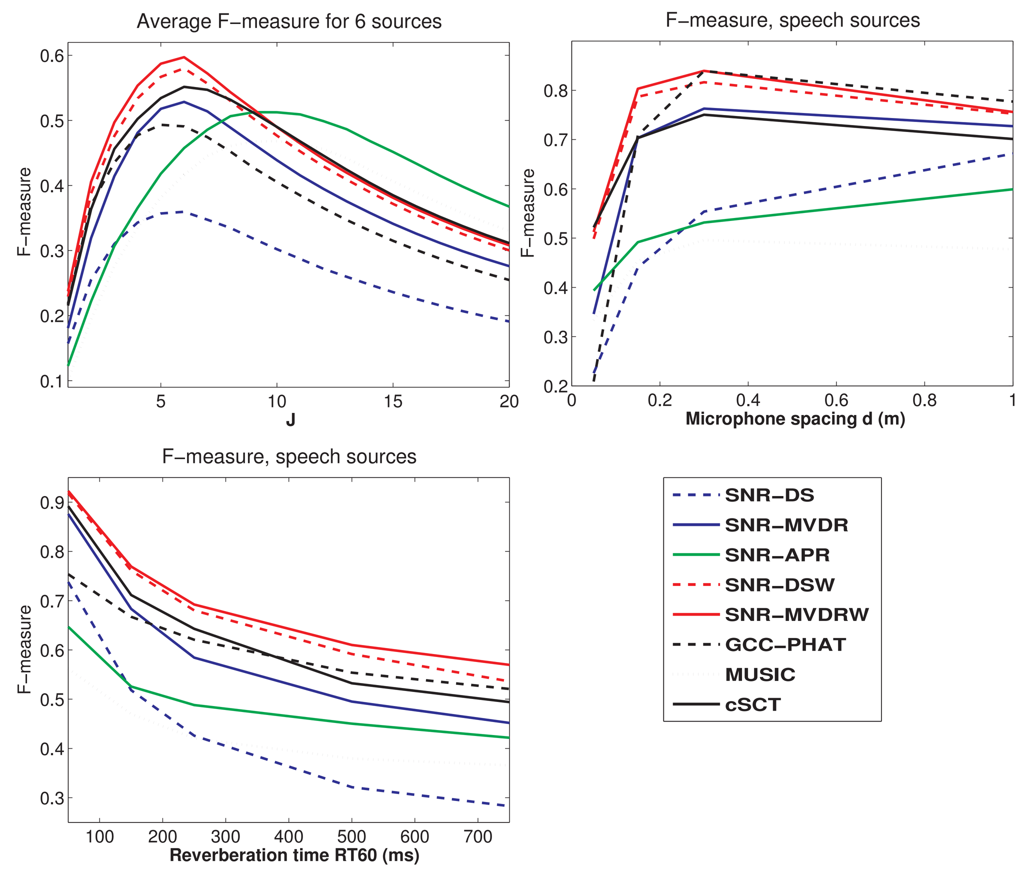

Figure 2: *Average F-measure as a function of the number of selected peaks* J*, the microphone spacing* d*, or the reverberation time* RT_{60}*.*

Conclusion

The proposed angular spectrum ϕ_{MVDRW} is robust to reverberation thanks to MVDR beamforming and to small microphone spacing thanks to frequency weighting.

Our latest work indicates no improvement compared to GCC-PHAT however when replacing the sum over time in (1) by the maximum [2].

References

[1] S. Arberet, R. Gribonval, and F. Bimbot. A robust method to count and locate audio sources in a multichannel underdetermined mixture. *IEEE TSP*, 2010.

[2] C. Blandin, A. Ozerov, and E. Vincent. Multi-source TDOA estimation in reverberant audio using angular spectra and clustering. *Technical report*, INRIA, 2011.

[3] N. Q. K. Duong, E. Vincent, and R. Gribonval. Spatial covariance models for under-determined reverberant audio source separation. In *Proc. WASPAA*, 2009.

[4] F. Nesta, P. Svaizer, and M. Omologo. Cumulative state coherence transform for a robust two-channel multiple source localization. In *Proc. ICA*, 2009.

ICASSP 2011, IEEE International Conference on Acoustics, Speech, and Signal Processing - Prague, Czech Republic - May 22–27, 2011