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) \( \tau_{ij} \) of several sound sources from a two-channel mixture signal.

We consider a local angular spectrum \( \phi_{\text{local}}(\tau) \) describing sound direction in each time-frequency bin \((t, f)\),

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

\[
\phi(\tau) = \sum_{t, f} \phi_{\text{local}}(\tau),
\]

- 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 \( \phi_{\text{SNR}}(\tau) \) as the signal-to-noise ratio between the signal power along direction \( \tau \) vs. other directions.

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

SNR estimation by probabilistic modeling

An alternative approach is to jointly estimate the signal power \( v_s \) in direction \( \tau \) and the residual power \( v_r \) 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

\[
\Phi_{xx} = v_s \Phi_s + v_r \Phi_r
\]

where

\[
\Psi = \frac{1}{\sin(2\pi f d/c_{\text{f}})}
\]

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

We then compute the SNR by

\[
\phi_{\text{SNR}}(\tau) = \frac{v_s}{v_r}
\]

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

\[
\phi_{\text{SNR}}(\tau) = \frac{1 + 2 v_s/v_r + \sin(2\pi f d/c_{\text{f}})}{1 - \sin(2\pi f d/c_{\text{f}})}
\]

By inverting these equations, we define

\[
\phi_{\text{SNR}}(\tau) = \frac{v_s}{v_r} W_d \phi_{\text{SNR}}(\tau) + W_d - 1
\]

\[
\phi_{\text{SNR}}(\tau) = \frac{v_s}{v_r} W_d \phi_{\text{SNR}}(\tau) - \frac{1}{2}
\]

where

\[
W_d = \frac{1 - \sin(2\pi f d/c_{\text{f}})}{2}
\]

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

SNR estimation by weighted beamforming

Beamforming can be used to estimate the power along direction \( \tau \) and subtract it from the total power.

Denoting by \( d = [1, e^{-2\pi f d/c_{\text{f}}}]^T \) the steering vector associated with TDOA \( \tau \) and by \( \Phi_s \) the covariance matrix of the input signal, we get

\[
\phi_{\text{SLD}}(\tau) = \frac{d^T \Phi_s d}{2 \pi f \Phi_s d - d^T \Phi_s d}
\]

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

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)

\( \phi_{\text{MVDR}} \) outperforms all other algorithms both in terms of recall and precision for most microphone spacings and for all reverberation times.

Conclusion

The proposed angular spectrum \( \phi_{\text{SNR}}(\tau) \) 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


