Multiple source separation in the frequency domain using Negative Beamforming

P. Gómez, A. Álvarez, V. Nieto, V. Rodellar and R. Martínez

Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, Spain
pedro@pino.datsi.fi.upm.es

Abstract

The localization of acoustic sources in a room is essential in many applications, as security monitoring, video conferencing, automatic scene analysis [6], reverberation canceling [5], or robust Speech Recognition under multiple-party effect [7]. Through the present paper the design and operation of a negative beamformer for multiple source speech separation will be presented. The problems found for its proper operation when multiple sources are present on the same band will be pointed out and the solutions found will be commented and discussed showing the results of real experiments carried out on a recording scenario.

1. Introduction

The joint application of acoustic source localization (ASL) and scene video tracking is becoming an important support technology for hybrid audio-video platforms, as for example in video-conferencing or security systems in public places. For acoustic source localization positive array beamforming have been traditionally used with great success [4]. Nevertheless, positive array beamformers present certain inconveniences. These structures show non-neglectable side lobes which demand an important computational power to be reduced. The aperture of the main lobe is usually wide, and to narrow it either more sensors are needed or more power-demanding algorithms have to be used. On the other hand large number of sensors makes signal acquisition and conditioning interfaces expensive, rendering the system expensive and difficult to install and operate. Negative beamformers, on the other side do not require a large number of sensors, and their extension to beam-like detection is straightforward. Nevertheless negative beamformers need to be adapted to broad-band signal enhancement and tracking. In a preliminary paper [2] the authors have shown how to track and separate individual sinusoidal sources in different bands, and to extend the results to broad-band (speech) signals. Through this paper the separation of sources on the same band is addressed. Section 2 reviews the basics of negative beamforming to deal with broad-band signals. In Section 3 the problem of multiple source separation is treated. Section 4 shows some practical study cases.

2. Negative Beamforming

Negative Beamforming is based on the combination of signals recorded by a two-sensor array as the one shown in Figure 1.

![Figure 1. General aspect of a negative beamformer. The angle of arrival is $\phi$. The separation between the two microphones is $2D$. The angular steering factor is $\beta$. The delay interval is $T=\kappa \tau$. The sample time is $\tau$.](image-url)

Assuming that a plane sinusoidal wave given by (1) reaches the array with an arriving angle $\phi$ relative to the main array axis ($a-a'$), the system will show a transfer function with a module given by (4), in terms of the angular shifts given in (2) and (3):

$$x(t) = A \cos(2\pi \phi + \psi)$$

$$\alpha = \frac{2\pi D}{c} \sin \phi$$

$$\delta = 2k \frac{f}{f_s}$$

$$|H(\alpha, \delta)| = \left| \frac{1}{2} \left[ 1 - 2 \beta \cos \alpha \sin \frac{\delta}{2} - \sin \alpha \cos \frac{\delta}{2} \right] \right|$$

where $f$ is the frequency of the signal, $f_s$ is the sampling frequency, $c$ is the speed of sound, $k$ is the delay order and $\beta$ is the filter steering factor. This function shows a sharp notch at an angle given by:

$$\phi_c = \arcsin\left( \frac{c}{2\pi D} \arctg\left( \frac{1 - 2\beta}{\frac{\pi k f_s}{f}} \right) \right)$$

The bandwidth of this notch is theoretically zero, thus showing a much higher selectivity in frequency than positive beamformers. To show the behavior of this system both in angle and frequency (4) is plotted as a function of these two variables as in Figure 2. The transfer function shows two symmetrical lobes of quasi-circular shape for low frequencies, with increasing gain up to a point where a maximum is reached. For high frequencies a distortion appears as a clover leaf-like flattening of the circular pattern as shown in the picture.
Figure 2. Module of the transfer function plotted in polar coordinates as a function of the angle and frequency for \( f_s = 16,000 \) Hz, \( k = 1, \beta = 0.5 \). The notch for \( \phi = 0 \) is clearly seen.

As speech signals are broad-band the behavior of the filter do not show a null at the same angle for all frequencies when the value of the steering factor \( \beta \) is different from 0, 0.5 or 1. This can be inferred from (5) and observed in Figure 3.

Figure 3. Module of the transfer function for a steering factor of \( \beta = 0.75 \) plotted in magnitude as a function of the arrival angle given in radians (x axis). Low frequencies are found at the front side; high frequencies at the rear side.

Figure 4. Sub-band signal separation in 16 channels using an equally spaced band-pass filter bank for a sample frequency of \( f_s = 44,100 \) Hz (x axis in Hz).

To compensate the notch behavior for wide-band signals sub-band separation may be implemented by means of a filter bank, as the one shown in Figure 4. Different implementations of the basic negative beamformer with different steering coefficients may be run concurrently in signal source separation resulting a structure as the one shown in Figure 5.

Figure 5. Structure of the Broad-Band Signal Enhancing Beamformer (BSEB) for a two-channel structure.

The anti-alias filters (AAF) limit the frequency span to be treated. Two filter banks (BPF) separate the broad-band signals into a set of channels (16 in this case). Each two corresponding signals from the filterbanks are treated by a negative beamformer (NBF). The best steering factors are estimated by a specific algorithm as commented in the next section. Resulting signals are finally combined to produce the desired output (enhanced or suppressed).

3. Multiple source detection

To separate two sound sources coming from different planes of arrival the steering factors associated to the corresponding angles of arrival must be evaluated from estimations of the energy of the signal recorded at the negative beamformer output. The following cases will have to be considered:

- Two (or more) sinusoidal sources are present in different bands (no spectral overlap).
- Two (or more) sinusoidal sources are present within the same band (spectral overlap).

The first case (two or more sources present in different bands) may be reduced to the case of a single source in a given band searching for the steering factor rendering the output power of each negative beamformer in Figure 5 to a minimum accordingly with expression (4). Figure 6 describes the behavior of the amplitude of the output signal as predicted from such expression, showing a "V" shaped pattern, where the bottom of the groove points to the value of the optimum steering factor to cancel the detected sinusoid.

Figure 6. Amplitude of the negative beamformer output (y axis) for a 1,000 Hz sinusoid as a function of the steering factor (x axis). The apex of the v-grooved shape points out the best value of the steering factor (\( \beta = 0.2 \)).

The second case (two or more sources present in the same band) produce a different behavior in the amplitude of the
output power, resulting in a "U" shaped pattern, where the
bottom of the groove do not point to the value of the canceling
steering factor. This situation is described in Figure 7.

This second case is rather frequent when two or more broad-
band sources are active on the same scenario (for example
with two speakers active at a time, or with a speaker in the
presence of one or several broad-band noise sources). A
possible strategy to treat this problem would be to raster for
channels where a single source would be active at a time,
annotating the source angle, correlating results among other
bands with single-source activity. Although his technique has
been successfully checked [2] other more powerful techniques
should be devised to cover cases where single-source active
bands are difficult to detect. This case could be treated taking
estimates of the output power for different values of the
steering factor. As an example, for a two source case with
similar amplitudes and angular shifts \( \alpha_1 \) and \( \alpha_2 \) from (2)
corresponding to arrival angles \( \phi_1 \) and \( \phi_2 \) it may be shown that
the following relations hold:

\[
A^2 = \sin \alpha_1 + \sin \alpha_2 = \frac{P_m(0.5)}{2 \cos^2 \frac{\delta}{2}}
\]

\[
B^2 = \cos \alpha_1 + \sin \alpha_2 = \frac{P_m(1)+P_m(0)-P_m(0.5)}{2 \sin^2 \frac{\delta}{2}}
\]

where \( P_m(0) \), \( P_m(0.5) \) and \( P_m(1) \) are the values of the output
power for the respective values of the steering factor. This
system may be solved for the angular shifts as:

\[
\alpha_1 = \frac{2 \arctg \left[ \frac{A^2}{B^2} \right] + \arccos \left[ \frac{A^2 + B^2}{2} - 1 \right]}{2}
\]

\[
\alpha_2 = \frac{2 \arctg \left[ \frac{A^2}{B^2} \right] - \arccos \left[ \frac{A^2 + B^2}{2} - 1 \right]}{2}
\]

and from these expressions, the directions of arrival may be
easily determined as:

\[
\phi_1 = \arcsin \left[ \frac{\alpha_1 c}{2\pi D} \right]
\]

An important point of discussion is how to determine if the
case under study is one of a single source or a multiple source
on a given band. This can be solved measuring the "V" or "U"
shape of the groove bottom using the following criterion:

\[
c_j = \frac{P_m(\beta_{\text{min}})}{P_m(0)+P_m(1)} > \vartheta_j
\]

where:

\[
\beta_{\text{min}} = \arg \left[ \min \{P_m(\beta)\} \right]
\]

is the value of the steering factor at the minimum value in the
output power is obtained, and \( \vartheta_j \) is an empirically established
threshold [2].

4. Results and Discussion

The technique described in the present paper has been
checked using real signals recorded on a sound proof chamber
as the one depicted in Figure 8.

\[
\phi_2 = \arcsin \left[ \frac{\alpha_2 c}{2\pi D} \right]
\]

(11)

Figure 7. Amplitude of the negative beamformer output (y
axis) for two 1,000 Hz sinusoidal sources at symmetric
angular positions (\( \phi_1=22^\circ30' \) and \( \phi_2=-22^\circ30' \)) with
respect to the microphone array as a function of the steering factor
(\( 0^\circ \leq \beta \leq 1^\circ \) x axis).

This situation is described in Figure 7.

Figure 8. Schematic diagram of the recording scenario.

Figure 9. Power spectrum of the output signal of the
beamformer after a 1024-point Fast Fourier Transform has
been carried out on the different Negative Beamformer
outputs ranging the steering factor in steps of value
0.01 each (right horizontal axis). Only the first 32
frequency bins have been shown out of the 512 produced by the FFT
(left horizontal axis).
Two sound sources $s_1$ and $s_2$ are placed at certain distances relative to the array main axis ($y$ axis), resulting in angles of arrival given by $\varphi_1$ and $\varphi_2$ to microphones $m_1$ and $m_2$. The first experiment used two sinusoidal sources of 1000 Hz and 2000 Hz for $s_1$ and $s_2$. The resulting signals in $m_1$ and $m_2$, separated 4 cm from each other ($D=2cm$) were treated by a negative beamformer with a sweeping steering coefficient, and the FFT of the resulting output was evaluated for each value of $\beta$ giving as a result the power spectrum presented in Figure 9. It may be seen that two main spectral lines spread around bins 13 and 24 corresponding to $f_1=1,120$ Hz and $f_2=2,067$ Hz showing clear "V" grooved shapes. The steering factors derived from these values are $\beta=0.2$ and $\beta=0.76$. In a second experiment two sinusoidal sources of 1000 Hz each were used for $s_1$ and $s_2$ at angles $\varphi_1=15^\circ$ and $\varphi_2=-45^\circ$. The resulting signals in $m_1$ and $m_2$ were treated by a negative beamformer, and $P_m$ evaluated for $\beta=0$, 0.5 and 1. The application of expressions (6)-(11) rendered as estimates for the arrival angles $\varphi_1=14^\circ49'$ and $\varphi_2=-44^\circ46'$.

In the last experiment two isolated words (/abajo/ and /double/) in Figure 10 are delivered respectively by sources $s_1$ and $s_2$ in Figure 8 placed at angular positions given by $\varphi_1=22^\circ30'$ and $\varphi_2=-22^\circ30'$ and 2m apart from the array center. The corresponding power spectra may be seen in Figure 10. The separation process consisted in detecting the arrival angles for the two sources, obtaining two traces where one of the sources have been removed using negative beamforming (cancellation). These intermediate signals are used later to reconstruct (selective enhancement) the target frame (/abajo/) as depicted in Figure 11, showing the possibilities of the technique exposed. This methodology be further extended to more than two sources using generalized de-coupling techniques [1][3].

5. Acknowledgments

This research is being funded by grant No. TIC99-0960 from the National Program for the Information and Communication Technologies of Spain, grant No. 07T-0001-2000 from Madrid Regional Research Program, and a Contract between the Polytechnic University of Madrid and the Centre Suisse d'Electronique et de Microtechnique.

6. References