inter.noise
HAMBURG 2016

Compressive beamforming for moving sound source auralization

Fanyu MENG[1]; Bruno MASIERO[2]; Michael VORLAENDER[1]

[1] RWTH Aachen University, Germany

[2] University of Campinas, Brazil

ABSTRACT

Conventional beamforming (CBF) is currently widely used for sound source localization. However, the quality of the results obtained with CBF is compromised by the fact that the CBF's output does not represent the source only, but the convolution of the source and the array's point spread function. Furthermore, when it comes to source signal reconstruction, CBF delivers only spectral amplitude information, i.e., there is no phase information, which is essential in the process of auralization.

Compressive beamforming (CB) has been recently proposed to improve the resolution of array measurements. CB has the advantage that it not only may improve source localization ability, but also allows the reconstruction of the source signal. Different from CBF, CB performs well on very few data snapshots and even on single-snapshot data, a characteristic which is beneficial when dealing with moving sound sources. This paper utilizes compressive beamforming to first localize multiple moving sound sources, and then reconstruct the source signals. To conclude we compare results obtained with CBF and CB.

Keywords: auralization, moving sound sources, conventional beamforming, compressive beamforming

1. INTRODUCTION

Sound source signal generation is the key procedure in auralization. The sound source generation consists of forward and backward approaches. The forward method is based on the *a priori* knowledge of the sources, such as a physical model or estimated spectral data to obtain the source signal; while the backward method acquires the signal by inverting the propagation procedure (e.g., directivity, Doppler effect and spherical spreading) from the recording (1). For sound fields generated by multiple sources simultaneously, especially moving sources, sound source signals cannot be obtained by near field recordings in an anechoic chamber. In this case, the backward method can be utilized to synthesize moving sound sources for auralization.

For the backward method, beamforming is a popular sound source localization method based on the array technique (2). Conventional beamforming (CBF, also called delay-and-sum beamforming) output is the convolution of the source spectra and the array's point spread function (PSF) (3). Therefore, CBF is characterized by low resolution. Compressive beamforming (CB) increases the resolution by utilizing compressive sensing (CS), which reconstructs underlying sparse source signals by solving a convex minimization problem (4). Beside higher resolution than CBF, CB is also able to reconstruct the source phase, which is essential in auralization.

For moving sound sources, the recording signals cannot be directly utilized in CB or CBF because the signals include Doppler effect, which can be eliminated both in the time and frequency domains. The elimination procedure is called de-Dopplerization (5, 6). The de-Dopplerized signals can be regarded as stable sources at particular positions according to the selected analysis window.

This paper extends CB to moving sound source localization and signal reconstruction. In this regard, phases are restored as well as amplitudes, which makes the auralization more precise and realistic. Firstly, a virtual array is simulated to record moving sound sources. Then a de-Dopplerization method in the time domain is introduced to eliminate the Doppler effect. With the de-Dopplerized signals at each microphone, CB and CBF are applied to localize the moving sound sources. At last, the source signal reconstruction is conducted using CB.

[1] fme@akustik.rwth-aachen.de
[2] bsmasiero@gmail.com

2. DE-DOPPLERIZATION AND ANALYSIS WINDOW

2.1 De-Dopplerization Technique

The sound field generated by a moving sound source is denoted as (7):

$$p(t) = \frac{1}{4\pi} \frac{q'(t - R(t)/c)}{R(1 - M_a \cos(\theta(t)))^2} + \frac{q(t)}{4\pi} \frac{v(\cos\theta(t) - M_a)}{R(t)^2(1 - M_a \cos(\theta(t)))^2}, \quad (1)$$

where $q(t)$ is the source strength, $R(t)$ is the distance between source and receiver, c is the speed of sound, v is the source's moving speed, $M_a = v/c$ is the Mach number and $\theta(t)$ is the angle between source moving direction and source-receiver direction.

When the receiver is far away from the moving source, which is not running at a high speed (normally M_a < 0.2), the previous equation is rewritten by omitting the second term:

$$p(t) = \frac{1}{4\pi} \frac{q'(t - R(t)/c)}{R(1 - M_a \cos(\theta(t)))^2}. \quad (2)$$

To eliminate of the Doppler effect, the recordings need to be interpolated and re-sampled. The reception time is calculated by $t = t_e + R(t)/c$ by taking emission time as the reference time. Then the recorded signal is interpolated and re-sampled according to the equally spaced reception time to conduct the de-Dopplerization. After amplitude modification, the de-Dopplerized signal can be regarded as the signal received at a position, the distance of which to the assumed "stable" source is the same as the distance between the first position of the moving source and the receiver.

2.2 Analysis Window Selection

For moving sound sources, the received signals change as the distance between the source and the receiver changes. Therefore, it is necessary to choose an analysis window to process de-Dopplerization and prepare for further study. In this research, the analysis window starts from the assumed point source when it is in "front" of the microphone (or the array origin when an array is applied). Here, "front" means that the direction between the point source and the array origin is perpendicular to the source moving direction.

A moving object can be normally simplified to a moving plane for source localization analysis. The plane is meshed into grids for the sound source localiztion. For example, A and B are two grids on the moving plane. When A is the assumed sound source, the analysis window starts at t_A and ends at $t_A + t_{win}$. The de-Dopplerization of A takes $[t_A, t_A + t_{win}]$ as the emission time interval. The de-Dopplerized signals are then applied in the following CBF and CB. The selection rule for the analysis window remains the same for point B. There is an overlap between the time windows of two adjacent points to ensure that no potential sources are omitted. The diagram is shown in Figure 1.

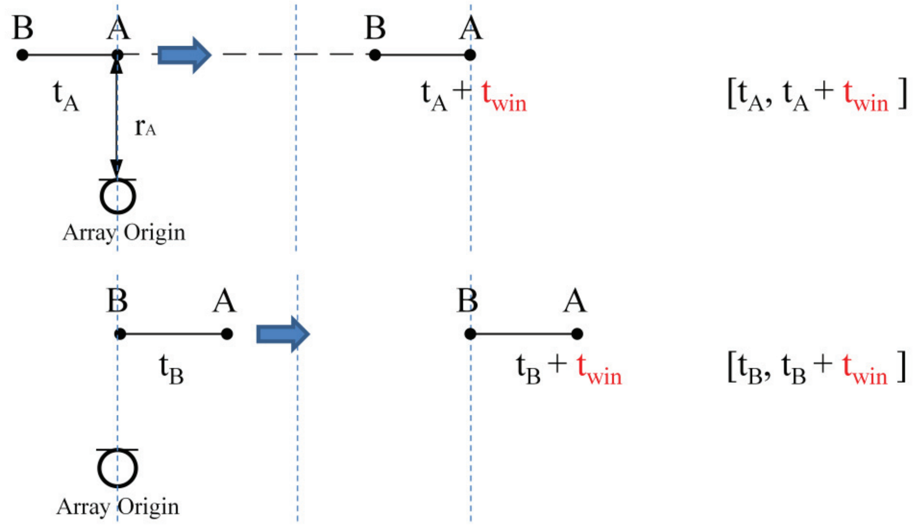


Figure 1 - Analysis widow selection for two adjacent points

3. BEAMFORMING THEORIES

When an array of microphones is applied to record a sound source, the signal $s(t)$ received at the mth microphone is

$$p_m(t) = s(t - \Delta_m), \tag{3}$$

where Δ_m is the propagation time from the source to the mth microphone. In the frequency domain,

$$P_m(\omega) = S(\omega)e^{-j\omega\Delta_m}, \tag{4}$$

where $P(\omega)$ and $S(\omega)$ are the Fourier transforms of $p(t)$ and $s(t)$ respectively. To simplify the notation, ω is suppressed. For spherical wave, the incident unit vectors from the nth source to the array origin and the mth microphone are a_{n0} and a_{nm} respectively, and r_{nm} is the distance between the nth source and the mth microphone. Then we get $\Delta_{nm} = (\|a_{nm}\| - \|a_{n0}\|)/c$. Thus, $\omega\Delta_{nm} = \frac{\omega}{c}(\|a_{nm}\| - \|a_{n0}\|) = k(\|a_{nm}\| - \|a_{n0}\|)$, in which k is the magnitude of the wave number. If we define $\boldsymbol{S} = (S_1, S_2, \ldots, S_N)'$ with N representing the reconstruct point source number, $\boldsymbol{P} = (P_1, P_2, \ldots, P_M)'$ with M representing the microphone number, the manifold vector is

$$\boldsymbol{A} = \begin{bmatrix} e^{-jk(\|\boldsymbol{a}_{11}\| - \|\boldsymbol{a}_{10}\|)}/r_{11} & e^{-jk(\|\boldsymbol{a}_{21}\| - \|\boldsymbol{a}_{20}\|)}/r_{21} & & e^{-jk(\|\boldsymbol{a}_{N1}\| - \|\boldsymbol{a}_{N0}\|)}/r_{N1} \\ e^{-jk(\|\boldsymbol{a}_{12}\| - \|\boldsymbol{a}_{10}\|)}/r_{12} & e^{-jk(\|\boldsymbol{a}_{22}\| - \|\boldsymbol{a}_{20}\|)}/r_{22} & \cdots & e^{-jk(\|\boldsymbol{a}_{N2}\| - \|\boldsymbol{a}_{N0}\|)}/r_{N2} \\ \vdots & \vdots & & \vdots \\ e^{-jk(\|\boldsymbol{a}_{1M}\| - \|\boldsymbol{a}_{10}\|)}/r_{1M} & e^{-jk(\|\boldsymbol{a}_{2M}\| - \|\boldsymbol{a}_{20}\|)}/r_{2M} & & e^{-jk(\|\boldsymbol{a}_{MN}\| - \|\boldsymbol{a}_{N0}\|)}/r_{NM} \end{bmatrix}. \tag{5}$$

Then,

$$\boldsymbol{P} = \boldsymbol{AS}. \tag{6}$$

3.1 Conventional Beamforming

Beamforming is a general way to localize sound sources based on temporal and spatial filtering using microphone array (8). CBF reinforces the signal by delaying the received signal at each microphone and adding all of them. The output of CBF is denoted as

$$y(t) = \sum_{m=1}^{M} w_m \tilde{p}_m(t + \tau_m), \tag{7}$$

where w_m is the weight of the mth microphone, τ_m is the time delay between the mth microphone and the array origin, $\tilde{p}_m(t)$ is the de-Dopplerized signal of the mth microphone. As discussed in (2), the signal obtained with CBF is the spatial convolution between the impinging sound field and array's PSF, which results in a "blurred" signal.

3.2 Compressive Beamforming

With additive noise $\boldsymbol{n}$, Equation (6) is denoted as

$$\widetilde{\boldsymbol{P}} = \boldsymbol{AS} + \boldsymbol{n}. \tag{8}$$

When vector $\boldsymbol{S}$ is sparse, the compressive samples in $\boldsymbol{S}$ can be reconstructed by solving the l_1 minimization problem (4)

$$\min \|\boldsymbol{S}\|_1 \text{ subject to } \left\|\boldsymbol{AS} - \widetilde{\boldsymbol{P}}\right\|_2 \leq \epsilon. \tag{9}$$

The problem above is solved with SPGL1, which is designed for the basis pursuit denoising (BPDN) (9). We set $\epsilon = 0.01\left\|\widetilde{\boldsymbol{P}}\right\|_F$, in which $\|\cdot\|_F$ represents the Frobenius norm (10).

4. SOURCE RECONSTRUCTION

4.1 Simulation

A plane (1.5 m x 5 m) is moving in the x-direction, carrying two point sources at the speed of 40 m/s. These two sources are placed on the plane with 0.5 m spacing. A microphone array is set at 1.5 m away from the moving direction. The array origin is on the z-axis. The plane is meshed into 11x21 grids with 15 cm vertical spacing and 25 cm horizontal spacing.

CB will only show good performance if the columns of the sensing matrix are sufficiently uncorrelated, i.e., if the array presents a random geometry (4). Therefore, a random microphone array with 32 microphones is applied in this simulation. Figure 2 illustrates the array geometry.

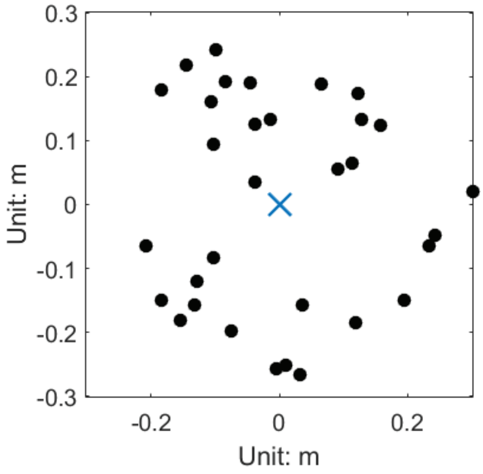


Figure 2 – Random microphone array

4.2 Localization

Figure 3 shows the localization results at 2 kHz. The reconstructed amplitudes using the two methods at the peak position are equalized in both vertical and horizontal directions. In the vertical localization result, the peak at the source position using CB is narrower than using CBF. The sidelobe levels of CB are also lower than those of CBF. In the horizontal direction, two clearer peaks can be seen using CB than using CBF, which means that CB has better resolution than CBF.

However, the localization results on moving sound sources are not as good as the results on stable sources, such as the results of Xenaki (6). There are interpolations when simulating the measured signals on moving sources and de-Dopplerization. This can cause inaccuracies, whereas for stable sources no interpolation is included.

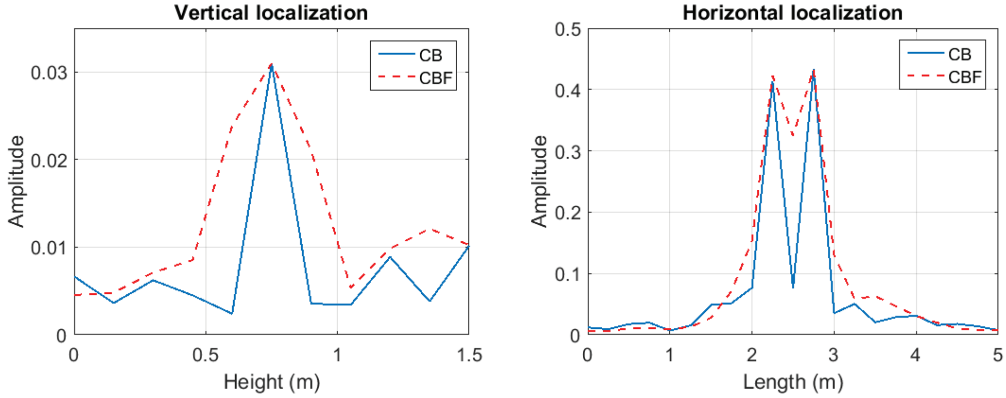
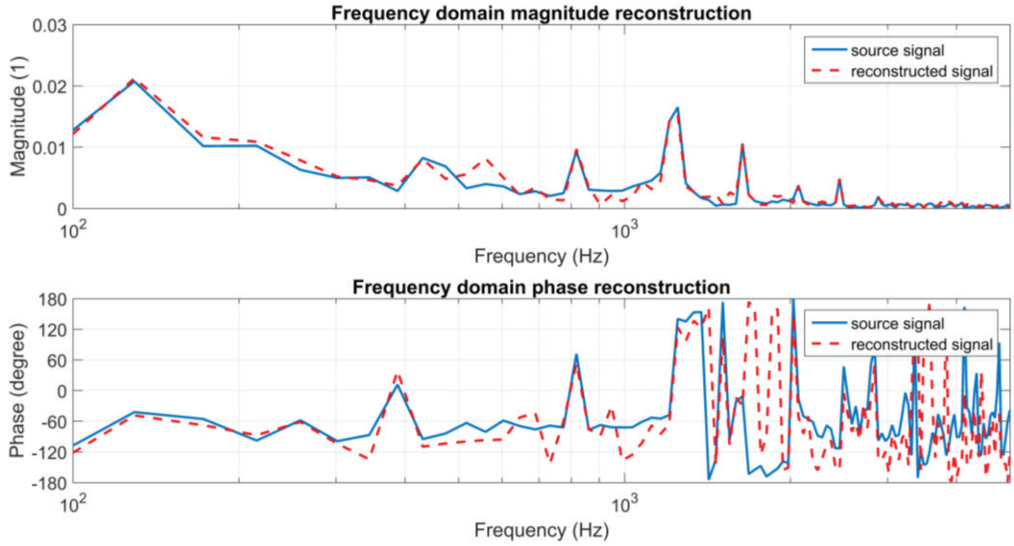


Figure 3 – Localization comparison using CB and CBF

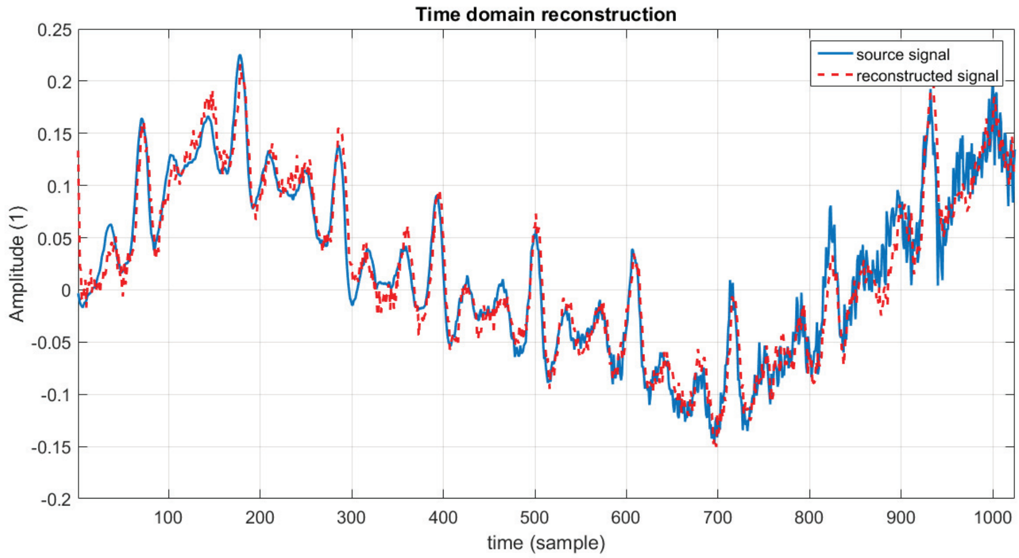
4.3 Source signal reconstruction

The first source is replaced by a piece of Jazz music, and the second source changes to white noise. The positions of the two sources remain the same. The reconstruction is conducted using CB in the frequency domain, and the time domain signal is reconstructed based on the magnitude and phase information.

Figure 4 shows the reconstruction of the source with Jazz signal during the time window when it passes the array origin. There are several tones in the Jazz music (Figure 4(a)), so the focus is only on the tonal information reconstruction. Between 400 Hz and 500 Hz, there is distortion in the magnitude and phase reconstruction. The reconstructed magnitudes and phases at the other tonal frequencies are mostly identical as the source signal. Figure 4(b) shows the time domain reconstructed signal based on the results in Figure 4(a).



(a) Frequency domain reconstruction



(b) Time domain reconstruction

Figure 4 – Source signal reconstruction in the frequency and time domains

5. CONCLUSIONS

This paper introduced the reconstruction procedure of moving source signals based on CB. A time domain de-Dopplerization technique was applied to eliminate the Doppler effect in the recording signals. CB and CBF then utilized the de-Dopplerized signals to localize the source positions using a random microphone array. CB and CBF were applied to localize the moving sources, and CB turned out to obtain better localization ability. With the source position information, the array steered to the source and its magnitude and phase information is reconstructed using CB.

Some parameters could influence the reconstruction results, such as the analysis window selection, window size, microphone array (array size, type, microphone number etc.) and source speed. Therefore, more simulations varying those parameters need to be conducted for the future study to improve the reconstruction model and make it more robust. In addition, on-site measurements are necessary to validate this reconstruction algorithm and the simulation results.

REFERENCES

1. Meng F, Wefers F, Vorlaender M. Moving Sound Source Simulation Using Beamforming and Spectral Modelling for Auralization. DAGA, 42. JAHRESTAGUNG FÜR AKUSTIK, Aachen, 14.-17. March

2016, In CD-ROM.
2. Van Trees H. L., Harry L. Detection, estimation, and modulation theory, optimum array processing. John Wiley & Sons, New York, USA, 2002.
3. Brooks T. F., Humphreys W. M. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. Journal of Sound and Vibration, 294 (4), 2006, pp 856–879.
4. Xenaki A, Peter G, Klaus M. Compressive beamforming. The Journal of the Acoustical Society of America 136.1 (2014): 260-271.
5. Camargo H. E. A Frequency Domain Beamforming Method to Locate Moving Sound Sources. Diss. Virginia Polytechnic Institute and State University, 2010.
6. Kook H., et al. An efficient procedure for visualizing the sound field radiated by vehicles during standardized passby tests. Journal of Sound and vibration 233.1 (2000): 137-156.
7. Morse P. M., Ingard, K. U. Theoretical acoustics. Princeton: Princeton University Press, 1968.
8. Johnson D. H., Dudgeon D. E. Array signal processing: concepts and techniques. Simon & Schuster, 1992.
9. Van Den Berg E, Michael P. F. Probing the Pareto frontier for basis pursuit solutions. SIAM Journal on Scientific Computing 31.2 (2008): 890-912.
10. Ribeiro F. P., Vítor H. N. Fast near-field acoustic imaging with separable arrays. Statistical Signal Processing Workshop (SSP), IEEE (2011): 429-432.