Applied Acoustics 150 (2019) 236–245

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Signal reconstruction of fast moving sound sources using compressive beamforming

Fanyu Meng [a,*], Yan Li [a], Bruno Masiero [b], Michael Vorländer [a]

[a] Institute of Technical Acoustics, RWTH Aachen University, Aachen 52074, Germany
[b] Faculty of Electrical and Computing Engineering, University of Campinas, Campinas-São Paulo 13083-852, Brazil

ARTICLE INFO

Article history:
Received 4 November 2018
Received in revised form 23 January 2019
Accepted 17 February 2019

ABSTRACT

Source signal is one of the main input parameters when auralizing moving sound sources in the Virtual Reality (VR) environments. This work utilizes compressive beamforming (CB) as a tool to reconstruct signals from fast moving sources. A pseudorandom microphone array is designed to meet the requirement of using CB and delay and sum beamforming (DSB), thus allowing for the signal reconstruction from the CB output and for the comparison between these two beamforming algorithms. Parameter studies through error analysis are conducted to evaluate how the reconstructed source signal is influenced by parameters, i.e. regularization parameter, window length, signal-to-noise ratio (SNR), basis mismatch and distance between the array and source trajectory. In general, CB outperforms DSB in signal reconstruction in terms of varying every parameter, except for the similar performance with SNR = 30 dB. We used the designed microphone array with both CB and DSB to reconstruct the signal of a known engine noise emitted by a loudspeaker installed on a moving car. The localization results delivered by CB are similar to DSB, which is in line with the simulation results. This behavior can result from potential coherence in the sensing matrix of CB due to similar time-domain transfer functions (TDTFs). However, CB still delivers lower reconstruction errors. Both simulation and measurement results indicate that CB is a viable option to reconstruct the signals of fast moving sound sources.

1. Introduction

Auralizing fast moving vehicles in urban environments requires the characterization of sound sources, including obtaining signals and spatial locations, which is one of the key questions in auralization [1]. Combining with follow-up synthesis and propagation models, the auralizations of moving vehicles, e.g. cars and aircraft, can be created [2,3], even with interactive real-time implementations in virtual reality [4–7]. Characterizing moving sources is more difficult compared to stationary sources, due to the time-dependent spatial location and frequency-shifted signal at the receiver's position caused by the non-stationary motion [8]. This paper investigates the characterization of fast moving sound sources, with the focus on reconstructing the signals based on the spatial locations of sources associated with vehicles such as intake, exhaust and tire noise.

The sound source signals can be obtained by forward and backward models according to the inherent principles [3]. The forward model requires physical or spectral information, or relies on the generation mechanism of sound sources, whereas the backward model utilizes either near-field or far-field recordings to extract sound source signals. Compared to the forward model, undertaking measurements in the backward model is more time consuming. Nevertheless, it saves the time to establish physical or empirical models. Moreover, signals obtained from recordings overcome the deficiency of low realism which is probably the main drawback of the forward model [6]. More problematically, theoretical models or empirical equations are not always achievable, such as in the case of aerodynamic noise caused by fast motion.

The backward model has been applied to generate the signals of moving sound sources. It was used to obtain aircraft noise signals from several recordings with microphones at discrete positions on the ground [9]. Peplow et al. [10] utilized the backward model to extract train noise signals by back propagating mono pass-by recordings of several distributed moving sources. However, for most of the cases, the locations of sound sources on moving vehicles are unknown, which is a key information for source characterization. Additionally, although the train passed by slowly, the recording of a particular target sound source was still contaminated by the presence of other sound sources. Bongini et al. [11] first applied a two-dimensional microphone array to localize the sound sources on a moving train. However, the impact of neighbor

* Corresponding author at: Microflown Technologies, Tivolilaan 205, Arnhem, The Netherlands.
E-mail address: meng@microflown.com (F. Meng).

https://doi.org/10.1016/j.apacoust.2019.02.012
0003-682X/

sources still exists because they also obtained the source signals by back propagating the individual pass-by recording from each microphone.

Delay and sum beamforming (DSB) was first applied in the backward model to extract the signal from a moving sound source [3]. Using beamforming benefits source characterization because along with the signal reconstruction, the localization was also obtained. The directional pattern of the beamformer overcomes the contamination of the desired source from other sources. However, DSB fails to yield high spatial resolution, which might result in reconstructed signals with noise from neighboring sources. In order to reconstruct signals more precisely, we need higher spatial-resolution beamforming methods, which can possibly be modified for moving sound sources.

Compressive beamforming (CB) is a method to achieve super resolution even with a small number of microphones, and it was indicated for the localization of moving sources [12,13]. However, CB has not been applied for the extraction of the source signal. Edelmann and Gaumond [14] mentioned the possibility to "listen to" the source by taking an inverse Fourier transform on the CB output, but it was not executed and yet the target was stationary sources. Therefore, we will explore CB for our goal of reconstructing non-stationary signals, thus extending the application of CB to source modeling for auralization.

The current research extends the application of CB to reconstruct the signal radiated by a fast moving sound source. The time-domain transfer function (TDTF) with incorporating Doppler effect [15] is adopted in CB as the sensing matrix. To start with, a framework for designing and optimizing pseudorandom microphone arrays for CB is proposed. Subsequently, errors in terms of regularization parameter, window length, SNR, basis mismatch and distance (between the source moving trajectory and array) using CB are analyzed. To conclude, the capability of using CB for the signal reconstruction of fast moving sources is performed on a known engine noise signal played by a moving loudspeaker attached to a car.

2. CB for moving sound sources

A scheme of how to use CB incorporating TDTF is proposed for moving sound sources. DSB for moving sources is also briefly introduced for later comparison with CB.

2.1. Moving sound source

2.1.1. Moving acoustic point source

The acoustic pressure field generated by a monopole point sound source moving along a straight line at constant speed v is described as [3,16]:

$$p(t) = \frac{1}{4\pi R(t)(1 - M\cos\theta(t))^2} s\left(t - \frac{R(t)}{c}\right). \tag{1}$$

where $p(t)$ is the sound pressure at the microphone in the sound field generated by the moving source, $s(t) \equiv \rho q'(t)$ is the source signal with ρ the density of the air and $q'(t)$ is the first derivative of the volume velocity $q(t)$ of the source, $R(t)$ is the distance between the source and the microphone, $M = v/c$ is the Mach number, and $\theta(t)$ is the angle between the moving direction of the source and source-microphone direction. $s(t)$ represents the strength and the characteristics of the source, and will be the signal to be reconstructed for auralization. An illustration of a point source moving rectilinearly at a constant speed is given in Fig. 1.

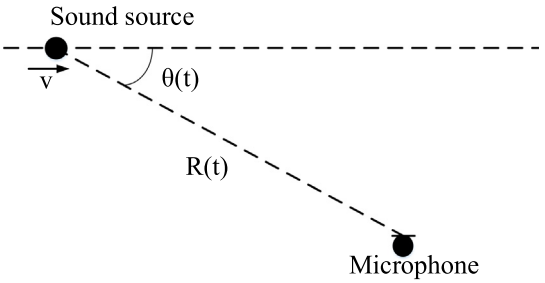

Fig. 1. Illustration of rectilinear motion of a point sound source moving at a constant speed.

2.1.2. Time-domain transfer function

Defining $t_e = t - \frac{R(t_e)}{c}$ as the emission time at the moving source and t as the reception time at the microphone, Eq. (1) can be written as

$$p(t) = \frac{1}{4\pi R(t_e)(1 - M\cos\theta(t))^2} s(t_e). \tag{2}$$

The TDTF is denoted by

$$H(t) = \frac{1}{4\pi R(t_e)(1 - M\cos\theta(t))^2}, \tag{3}$$

and it leads to $p(t) = H(t)s(t_e)$. Therefore, the transfer function is then expressed as

$$H(t_e) = \frac{1}{4\pi R(t_e)(1 - M\cos\theta(t_e))^2}, \tag{4}$$

and thus

$$p(t) = H(t_e)s(t_e). \tag{5}$$

In practice, noise should be introduced to any measurement model. Therefore, Eq. (5) with additive Gaussian noise is denoted by

$$p(t) = H(t_e)s(t_e) + n(t). \tag{6}$$

The time index t and t_e will be suppressed to simplify the notation. The problem above is extended to M_p microphones and N potential sources, which yields the following form

$$\mathbf{p} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{7}$$

where $\mathbf{p} = \left[p_1, \ldots, p_{M_p}\right]^T$, $\mathbf{s} = [s_1, \ldots, s_N]^T$ represents potential sources, $\mathbf{n} = \left[n_1, \ldots, n_{M_p}\right]^T$, and $H \in \mathbb{R}^{M_p \times N}$.

2.2. Compressive beamforming

A moving vehicle, in terms of noise, can be decomposed and represented by only a few main sources [2,17]. The beamforming algorithms are able to detect how many main sources there are and where they are located. The location can refer to the vehicle's component radiating sound, e.g. the engine and tire [2,17]. The presence of only a few sources enables exploiting the spatial sparsity of $\mathbf{s}$ in Eq. (7). The spatial sparsity here can be interpreted as the number of real sources is much smaller than the number of potential sources in Eq. (7). Apart from sparsity, if the columns of $\mathbf{H}$ are sufficiently incoherent, which indicates the correlation between the columns is sufficiently low, $\mathbf{s}$ can be solved by CB, specifically by minimizing the ℓ_0-norm [13,18], which counts the number of non-zero entries in the vector

$$\min_{\mathbf{s}\in\mathbb{R}^N} ||\mathbf{s}||_0 \text{ subject to } \mathbf{p} = \mathbf{H}\mathbf{s} + \mathbf{n}. \tag{8}$$

The ℓ_0-norm is a difficult non-convex problem and is thus normally replaced by the ℓ_1-norm

$$\min_{\mathbf{s}\in\mathbb{R}^N}||\mathbf{s}||_1 \text{ subject to } \mathbf{p}=\mathbf{H}\mathbf{s}+\mathbf{n}, \tag{9}$$

which can be recast as the unconstrained optimization

$$\min_{\mathbf{s}\in\mathbb{R}^N}||\mathbf{p}-\mathbf{H}\mathbf{s}||_2^2+\lambda||\mathbf{s}||_1, \tag{10}$$

where λ is the regularization parameter which balances the norm of the residual $||\mathbf{p}-\mathbf{H}\mathbf{s}||$and the sparsity of $\mathbf{s}$.

The stated solution is to solve the ℓ_1-norm optimization problem for a single time sample. For localization, single sample processing may find its application. However, if the values of one or some of the chosen time samples happen to equal or to be close to zero, the supposed spatial sparsity assumption would fail. Therefore, the problem is extended to multiple time samples. The cost function is reformulated as

$$\mathbf{P}=\mathcal{H}\mathbf{S}+\mathbf{N}. \tag{11}$$

where $\mathbf{P}=[\mathbf{p}(t_1),\ldots,\mathbf{p}(t_T)]\in\mathbb{R}^{M_p\times T},\mathbf{S}\in\mathbb{R}^{N\times T},\mathcal{H}\in\mathbb{R}^{M_p\times N\times T}$ which samples $\mathbf{S}$ temporally and spatially, $\mathbf{N}\in\mathbb{R}^{M_p\times T}$ and T is the number of time samples. Since sparsity is required in the spatial dimension but not necessarily in time [12], the ℓ_2-norm of all time samples of a particular focus point n is calculated, i.e. $s_n^{\ell_2}=||s_n(t_1),\ldots,s_n(t_T)||_2$. With the ℓ_1-norm of $\mathbf{s}^{(\ell_2)}=\left[s_1^{(\ell_2)},\ldots,s_N^{(\ell_2)}\right]$, the cost function becomes

$$\min_{\mathbf{s}^{\ell_2}\in\mathbb{R}^N}||\mathbf{P}-\mathcal{H}\mathbf{S}||_F^2+\lambda||\mathbf{s}^{\ell_2}||_1, \tag{12}$$

where $||\cdot||_F$ represents the Frobenius norm.

As the reception time t is calculated from the emission time t_e and the time-variant $R(t_e)$, the calculated time stamps in the reception time t are non-uniformly spaced. Therefore, $p(t)$ is interpolated in terms of t and delivered as $\tilde{p}(t)$, which substitutes $p(t)$ when using Eq. (12). The ℓ_1-norm optimization problem is solved in MATLAB using cvx toolbox [19].

After detecting the source position index n_s, the source signal reconstructed by CB is denoted as $\hat{s}_{n_s}(t), t=t_1,\ldots,t_T$.

2.3. Delay and sum beamforming

As DSB has been used for signal reconstruction [3], we will compare the results of CB and DSB in our case. In the following the modified DSB for moving sources for signal reconstruction is briefly introduced. More details about de-Dopplerization for the elimination of Doppler effect and the combination with DSB can be found in [3].

Due to the nonlinearity of $R(t)$ (or $R(t_e)$), the calculated reception time $t=t_e+\frac{R(t_e)}{c}$ would not coincide with the uniform time stamps by the microphone recording. Therefore, the recorded signal $p(t)$ is first interpolated in terms of the calculated reception time t, and delivered as $\tilde{p}(t)$. The de-Dopplerized signal $\hat{p}$ can be obtained [3]. Therefore, a stationary source case can be assumed as the Doppler effect has been eliminated. The sound field generated by a stationary point source $n(t)$ is expressed as [20]

$$p(t)=\frac{1}{4\pi R}s\left(t-\frac{R}{c}\right), \tag{13}$$

where R is the distance between the sound source and the microphone, $s(t)$ is the source signal and $n(t)$ is the noise part. If the de-Dopplerized signal is $\hat{p}(t)$ with assuming the array is moving with the source [3], the DSB equation [21] for a moving sound source is

$$\begin{aligned} y(t) &= \sum_{m=1}^{M_p} w_m\hat{p}_m(t+\tau_m) \\ &= \sum_{m=1}^{M_p} w_m\frac{1}{4\pi\widehat{R}_m^0}\hat{s}\left(t-\frac{\widehat{R}_m^0}{c}+\tau_m\right), \end{aligned} \tag{14}$$

where $\tau_m=\left(\widehat{R}_m^0-\widehat{R}^0\right)/c$, with $\widehat{R}_m^0$ representing the distance between the focus point of the microphone array and the "moving" array origin, $\widehat{R}^0$ is the distance between the focus point and the "moving" array origin, R_m^0 is the distance between the mth "moving" microphone and the source, and $\hat{s(t)}$ is the reconstructed source signal. If the focus point coincides with the source position, $\widehat{R}_m^0=R_m^0$ and $\widehat{R}^0=R^0$, Eq. (14) becomes

$$\begin{aligned} y(t) &= \frac{1}{4\pi}\left(\sum_{m=1}^{M_p}\frac{w_m}{R_m^0}\right)\hat{s}\left(t-\frac{R^0}{c}\right) \\ &= C\hat{s}\left(t-\frac{R^0}{c}\right), \end{aligned} \tag{15}$$

where $C=\frac{1}{4\pi}\sum_{m=1}^{M_p}\left(w_m/R_m^0\right)$ is a constant which depends on the weight and the positions of the potential sound source and microphones. The reconstructed source signal $\hat{s}(t)$ can be reconstructed by the time shift R^0/c and the division of the constant C on the beamforming output signal $y(t)$. Note that $\hat{s(t)}$ also contains noise recalling Eq. (6).

3. Design of pseudorandom microphone arrays

As mentioned in Section 2.2, the columns of the sensing matrix $\mathbf{H}$ are supposed to be incoherent to utilize CB. A random array is able to lower coherence in the sensing matrix [18], and the restricted isometry property (RIP) should be satisfied [22]. Gaumond et al. [23] proposed statistical restricted isometry property (StRIP) to help design sparse arrays. However, since DSB is also used as comparison to CB, its performance should also be taken into account. Gerstoft et al. introduced convex optimization to enhance the performance of beam patterns of 2D random arrays [24]. Good resolution and minimum maximum sidelobe level (MSL) were also used as criteria to design the planar random arrays [25–27]. A framework for the design and optimization of 2D pseudorandom microphone arrays which benefits both CB and DSB by considering RIP and beam patterns is proposed next.

3.1. Design concept

If the positions of microphones on an array aperture are randomized, it would probably lead to the microphones clumping in a small area, and thus reducing the spatial resolution if DSB is used [25]. This would also possibly increase of the coherence of the sensing matrix because of very similar $R(t)$ of the closely localized microphones. Therefore, restrictions are necessary to be introduced to the randomization in the design of random microphone arrays, which leads to pseudorandom microphone arrays.

According to Kook et al. [25], segmenting an array aperture into units, i.e. unit partition, can guarantee that the microphones are well distributed on the array aperture to avoid clumping. Afterwards, a baseline filter method is further introduced to ensure the scattered distribution of the microphones [26]. Here, baseline is defined as the distance between two arbitrary microphones in a microphone array, and baseline vector is the corresponding vector [25]. The baseline filter is able to keep the microphone

distribution scattered by controlling the appearance number of baseline vectors [26].

To employ CB, RIP should be considered and it is defined as

$$(1-\delta_p)\|\mathbf{S}\|_2^2 \leqslant \|\mathbf{HS}\|_2^2 \leqslant (1+\delta_p)\|\mathbf{S}\|_2^2, \tag{16}$$

for all p-sparse vectors. Here, δ_p is the isometry constant of matrix $\mathbf{H}$. As this work explores not only on the localization capability, but also the potential application for signal reconstruction, only the $p=1$ case is considered to test the RIP. Hence only one sound source is studied in the following.

The unit partition is able to deliver good beam patterns for DSB. In this sense, resolution and MSL are chosen as two of the design criteria. Therefore, the resolution, MSL and RIP are the criteria to optimize the design of pseudorandom microphone arrays. Fig. 2 exhibits the proposed framework. The details are elaborated in the following contents.

3.2. Array design and optimization

The arrays are designed on an aperture of 1.8 m × 1.8 m with 32 microphones. The aperture is discretized into 32 units, in each of which lays a microphone. Each unit has eight possible positions according to Zheng et al. [26]. The locations of all microphones in every unit are randomized. Irregular partition is adopted to ensure that the microphones are as scattered as possible [26], as can be seen in Fig. 4 that the units are either horizontal or vertical.

Only if the configuration meets the requirements of the baseline filter method [26], can it be saved as a candidate array. Following this rule, 1000 array configurations are generated. The RIP condition is then tested and we keep 50 arrays that meet the requirements. The resolutions for a 30° steering angle and MSLs for the remaining 50 arrays are shown in Fig. 3.

It is observed that the resolution does not differ significantly along the simulations. However, the MSLs can vary more than 10 dB between the selected array configurations. We select as the optimized pseudorandom array configuration with the smallest MSL (−19 dB MSL), which has 14° resolution at a steering angle of 30°. Fig. 4 shows the optimal array configuration, as well as the irregular unit partition.

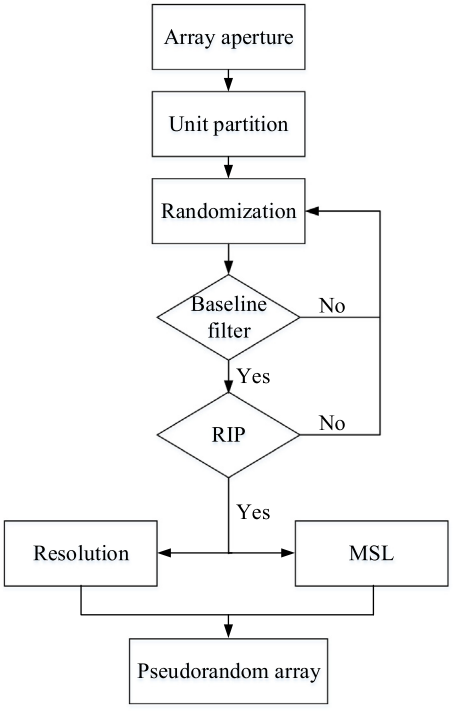

Fig. 2. The framework for designing and optimizing pseudorandom microphone arrays.

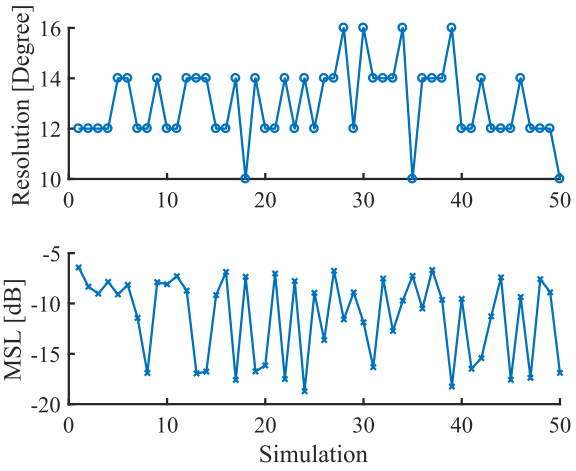

Fig. 3. The MSLs and resolutions of the 50 filtered arrays after baseline filtering and RIP test.

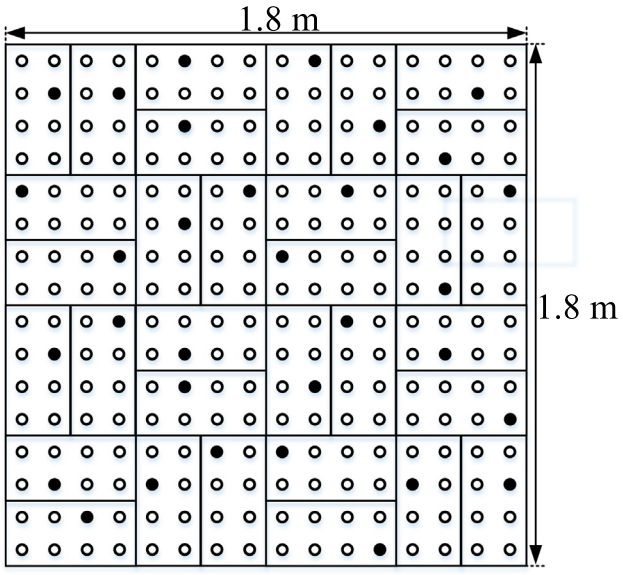

Fig. 4. The configuration of the optimized pseudorandom array with the scheme of irregular partition. The potential positions of the microphones are indicated by the "○", and the real positions of the microphones are represented by the "•.

4. Virtual scenario

A virtual scenario was created to test our algorithms. It is made of a moving plane with two sound sources which are "recorded" by the pseudorandom microphone array according to Section 2.1.1. The recorded data was analyzed using both DSB and CB to obtain the reconstructed source signals with the detected source locations.

4.1. Simulation setup

Two moving sound sources (S1 and S2) are simulated by Eqs. (2) and (7). Note that the equations are denoted in continuous time, but discrete time is required in digital processing. Thus the

calculated signals on the right-hand side of Eq. (2) are interpolated and resampled in terms of uniformly spaced time stamps to obtain $p(t)$ to simulate real recordings. S1 and S2 fixed in a plane which moves in the $-z$ direction at 20 m/s. The pseudorandom microphone array is placed 5 m away from the moving trajectory that is parallel with the array aperture. The moving plane is regarded as the reconstruction plane Ω, on which the beamforming calculations are conducted. Ω is meshed into grids and the distance between two grid points is 0.1 m. Each grid point is scanned as a potential sound source's position. S1 is on the origin of Ω, and S2 is 0.5 m located above S1 on the same vertical line. S1 and the origin of the array are both on the xz plane in the coordinate system. A sketch of the simulation can be found in Fig. 6. A 6 s recording of engine noise and a 6 s periodic signal are attached to S1 and S2, respectively. The periodic signal consists of a fundamental tone of 500 Hz and all its harmonics up to 8 kHz (with a random deviation on each harmonic of up to 50 Hz), plus a 200 Hz tone. The spectra of S1 and S2 are shown in Fig. 5.

The duration of S1 and S2, and the moving time of Ω is 6 s. The starting and stop positions of Ω are symmetric in terms of the origin O of the x axis.

S1 is the target source to be localized and reconstructed and S2 is regarded as an interference source. A wide frequency range of signal reconstruction can be studied by virtue of the broadband engine noise.

The time window along which a grid point travels in the viewing window [8] is named as steering window. In the following calculations, the signal-to-noise ratio (SNR) is 30 dB the length of the steering window is 256 samples and the sampling rate is 44100 Hz, unless stated otherwise.

4.2. Source detection

In Section 2.2, it was mentioned that the interpolated signal $\tilde{p}(t)$ is used instead of $p(t)$ during calculation. Recalling $t_e = t - \frac{R(t_e)}{c}$, in this paper $t_e = t - \frac{R(t_e)}{c}$, it can be seen that the interpolation can only be proceeded with the knowledge of $R(t_e)$, which depends on the locations of the source and receiver. It contradicts to our purpose of obtaining the source location.

Therefore, the calculation strategy is as follows. At t_1, the grid points on the vertical line L_1 as indicated by the solid dots in Fig. 7 form a group, and they are only processed when they pass by the viewing window, which is the spatial area between the two dashed lines [8]. The length of the viewing window is the product of the steering time window t_{win} and the source speed v. In each vertical line L_n, every grid point will be assumed as the source and thus there will be a set of interpolated received signals for all the 32 microphones. Subsequently, CB is applied on each point on L_n with the interpolated signals. The calculation continues point-wise on every point and vertical line, until all the points have been calculated. Fig. 7 exhibits the processing first at $[t_1, L_1]$, then at $[t_2, L_2]$. All the CB outputs are of the same length, t_{win}. Finally, a two-dimensional matrix with the root mean square (RMS) values of the amplitudes of CB output signals are derived. Large RMS values are detected as potential sound sources, with the corresponding CB outputs as the reconstructed signals.

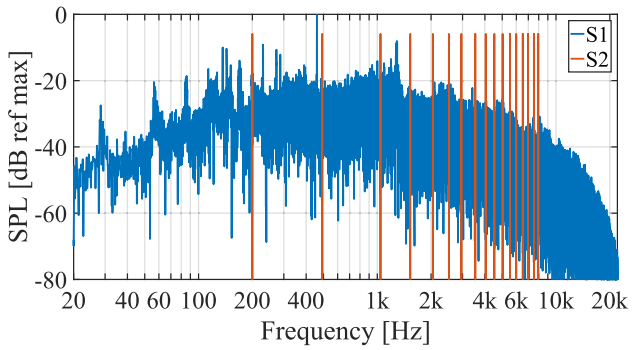

Fig. 5. Signal spectra of the two sound source signals S1 and S2.

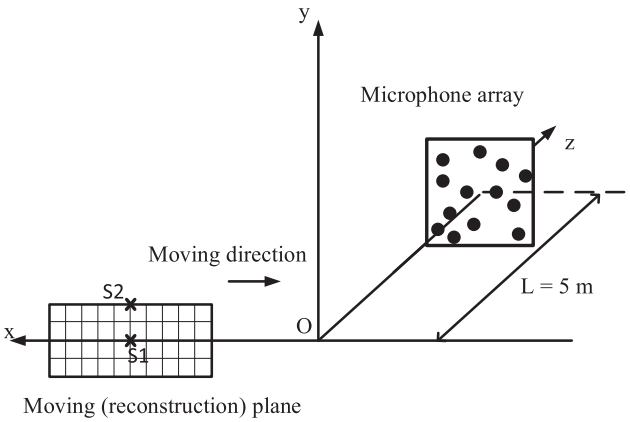

Fig. 6. The sketch of the simulation. The reconstruction plane Ω with sound sources $S1$ and $S2$ moves in the $-x$ direction along the x-axis. The microphone array is set 5 m away from the x-axis, with the array origin on the z-axis.

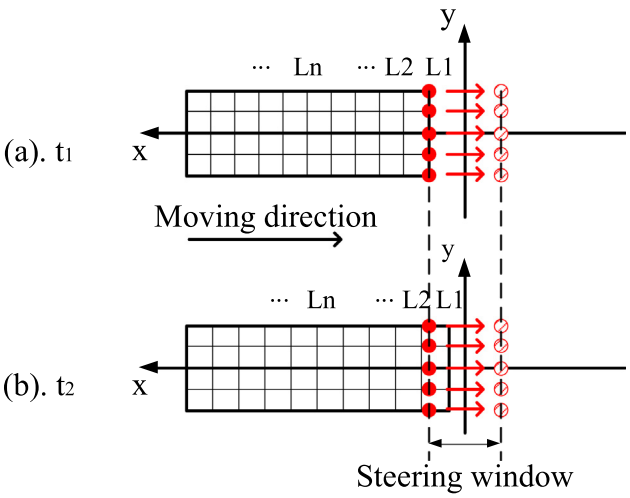

Fig. 7. The source detection procedure. The figure is the two-dimension view of Fig. 6 when looking in the $+z$ direction. The viewing window is the product of the time steering window t_{win} and the source speed v, and all grid points are only processed when they pass by the viewing window. The solid dots represent the grid points on a vertical line on the reconstruction plane. The arrows point to the positions of the grid points at the end of the spatial window. The grid points are piecewise processes from $[t_1, L_1]$ to $[t_n, Ln]$.

5. Error analysis

The proposed method using CB is evaluated by means of errors regarding localization and signal reconstruction. Regularization parameter, window length, SNR and mismatch are selected for the error analysis. Errors using DSB are also given to compare with CB.

5.1. Error description

The localization error e_{loc} is defined as

$$e_{loc} = \sqrt{(\hat{x} - x)^2 + (\hat{y} - y)^2}, \tag{17}$$

where $[\hat{x}, \hat{y}]$ and $[x, y]$ are the localized and original coordinates of the source. The z coordinate is omitted since Ω and the trajectory are both on the xy plane.

Jagla et al. [28] introduced a spectral method to evaluate the similarity between signals with accounting for human hearing by considering A-weighting and logarithmic sensitivity. They derived the difference between two signals by calculating the square of the subtraction between the absolute pressure values and then obtaining the logarithm of the square. However, log spectral distance which describes the difference between the logarithms of the two signals instead, has been applied to measured the perceptual distortion of speech processing [29]. It indicates that the logarithmic difference could be a criterion to quickly evaluate the perceptual difference. Therefore, we directly evaluate the absolute logarithmic error in the decibel scale as

$$e_S(f(k)) = \left| 20log_{10} \frac{|\widehat{S}(f(k))|}{|S(f(k))|} \right|, \quad k = 1, 2, \ldots, K \tag{18}$$

where S and $\widehat{S}$ are the Fourier transform of the reconstructed and original signals, K is the number of the frequency bins. With adding A-weighting to account for human hearing as in [28] and taking the average over the whole frequency range, e_{rec} is derived for the error estimation of the reconstructed signal:

$$e_{rec} = \frac{1}{K} \sum_{k=1}^{K} (e_S(f(k)) + w_A(k)). \tag{19}$$

e_{rec} is the measure used in this study to assess the perceptual difference between the reconstructed and original signals, or simply put, the reconstruction error. The small e_{rec} is, the more similar the reconstructed signal is to the original signal. Note that listening tests would be preferable to measure human perception more straightforwardly, however, the aforementioned measure is advantageous to quickly estimate errors in terms of various parameters. Listening tests could be included for future work to obtain more precise comparison from humans, and the results could also provide a correlation between subjective and objective measures.

5.2. Regularization parameter vs. window length

It is critical to select the regularization parameter as it determines the tradeoff between the fit of the solution to the original data versus the sparsity prior [12]. The selection of the regularization parameter still remains a difficult question, and trials through simulations were conducted to find out the optimal solution [12,30]. It was suggested that a low noise level could be employed for the selection of regularization parameter to guarantee capturing all nonzero elements [13]. It was also pointed out that the regularization parameters in the constrained and unconstrained forms are related [12]. Therefore, the regularization parameter β in the Dantzig Selector [31,32] is used as the search basis. $\beta = \epsilon_N \sigma$, where $\epsilon = \sqrt{2 \log N}$ and σ is the standard deviation of the noise. Simulations are conducted in the neighborhood of β to search for a good choice of the regularization parameter λ in the unconstrained form in Eq. (12).

The errors of varying λ from 0.5β to 2β are studied. Additionally, the length of a steering window determines the spatial and spectral resolution, which has been discussed for DSB [3]. Thus the regularization parameter and the window length are jointly investigated. The errors of localization and signal reconstruction are compared in Fig. 8(a) with SNR = 30 dB. Similar performance between DSB and CB can be observed except for some large variations for the windows with 32, 64 and 256 samples. For CB, most of the errors in terms of various λ achieve similar results. For the 64- and 256-sample window, no error is detected from the localization and signal reconstruction.

Now the SNR is decreased to 5 dB, a value more prone to be found in real measurement situations. The errors are shown in Fig. 8(b). e_{loc} and e_{rec} become larger with decreasing the SNR as expected, and the localization results of DSB and CB are still quite similar. Nevertheless, DSB and CB can be clearly distinguished in terms of signal reconstruction. For all the λ selected, the e_{rec} of CB are all below those of DSB, and each λ delivers a separate curve. The level difference could be quite large, e.g. around 6 dB for the case of 64-sample window with $\lambda = 0.5\beta$ and $\lambda = 1.75\beta$. The range of SNR is then extended to [15 dB, −5 dB] to have a better understanding of the influence of SNR (Fig. 9, $\lambda = 0.75\beta$). It can be seen that the reconstruction error increases as SNR decreases, and gradually CB outperforms DSB.

In Eq. (15), the reconstructed signal $\hat{s}(t)$ from DSB also contains noise, the incrementation of which would lead to increasing error in $\hat{s}(t)$. On the contrary, CB takes noise into account during the calculation as shown from Eq. (9) to Eq. (11). Thus CB outperforms DSB for signal reconstruction with the presence of strong noise, i.e. SNR = 5 dB in our case. It can be expected that both algorithms would perform similarly with the presence of slight noise (SNR = 30 dB), as in Fig. 8(a). However, the localization ability of CB shows no clear advantage over DSB in the current situation. The distance between Ω and the microphones is large compared to the microphone distances, which would result in potential coherence in the TDTF and cause errors. It could degrade the localization ability of CB leading to errors and similar performance with DSB, as well as the signal reconstruction errors. Another possible reason for the similarity between DSB and CB in localization could be that only two sound sources are considered. CB would outperform DSB with the presence of many sources according to literature. The literature has been mainly focused on stationary sources, based on which CB delivers better localization. Whereas in the current study, moving instead of stationary sources have been addressed, and CB is found not advantageous over DSB in terms of localization. However, localization will not be further discussed due to our aim of signal reconstruction, not localization.

For the signal reconstruction, no clear deviation of e_{rec} is observed when varying λ under SNR = 30 dB. Whereas, e_{rec} increases as λ increases with higher noise level, SNR = 5 dB. This can be explained by Eq. (12). When SNR is large, $||\mathbf{P} - \mathcal{H}\mathbf{S}||_F^2$ deviates little due to low noise level, which indicates that the reconstructed and original signals are quite similar. As noise increases (SNR decreases), $||\mathbf{P} - \mathcal{H}\mathbf{S}||_F^2$ can be largely deviated accordingly. Thus, the selection of lambda is critical, leading to e_{rec} varying more with changing λ in Fig. 11(b) than in Fig. 11(a). However, no obvious correlation between e_{rec} and the window length is found. The largest window length studies is with 512 samples, which provides a frequency resolution of 86 Hz under the sampling rate of 44.1 kHz. This leads to energy leakage over the whole frequency range. This leakage is not correlated to the frequency resolution, which results in the random change of e_{rec} with the window length considered in this work.

The lowest e_{rec} with the window length of 32 samples and $\lambda = 1.75\beta$ can be selected for signal reconstruction. However, the corresponding e_{loc} reaches over 0.5 m in this sense, which would lead to perceptional difference in auralization. Additionally, this window length is too limited to extract the characteristics of the source signal, e.g. low frequency information. It is thus necessary to select parameters according to both localization and signal reconstruction.

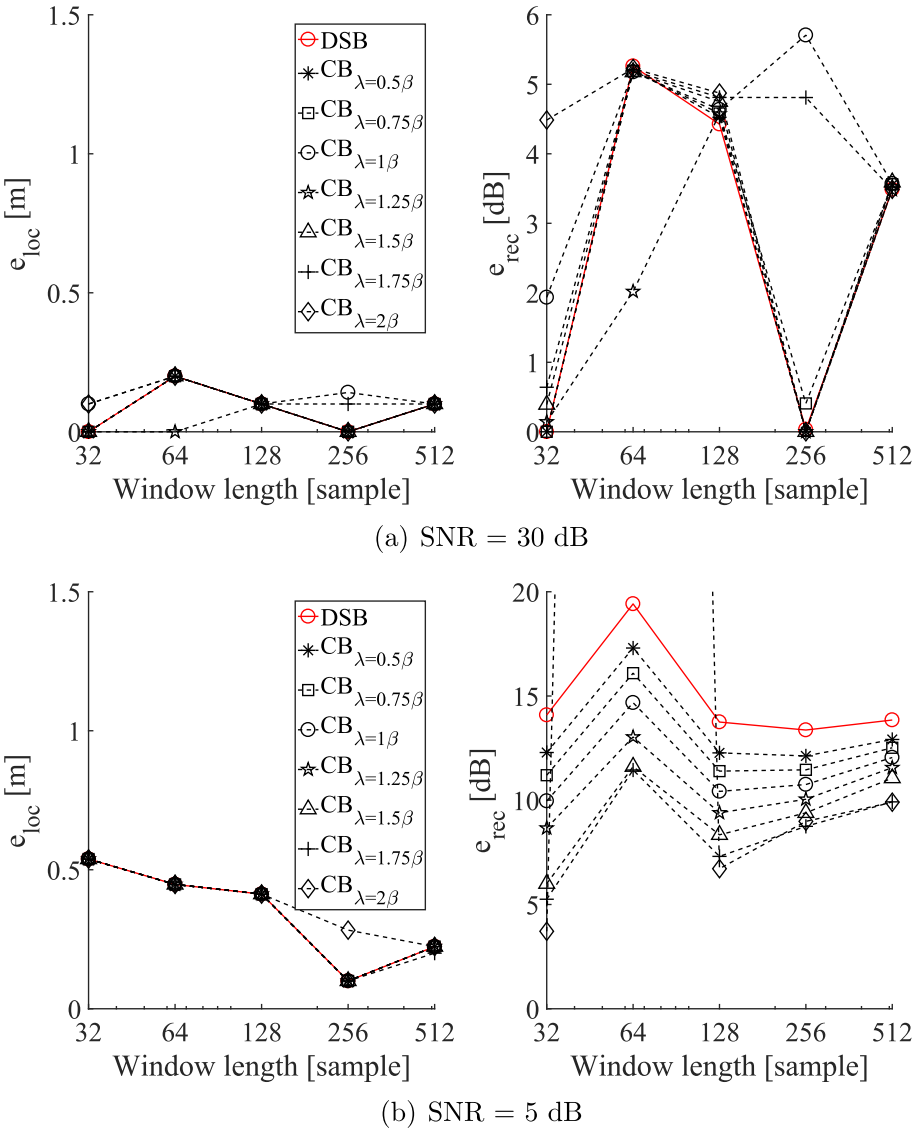

Fig. 8. The errors of localization and signal reconstruction versus window length for various regularization parameter λ with (a) SNR = 30 dB and (b) SNR = 5 dB using DSB and CB. The value of 64 samples with CB $\lambda = 2\beta$ is 174.5 dB, which is out of the range of the y-axis and not shown in the figure.

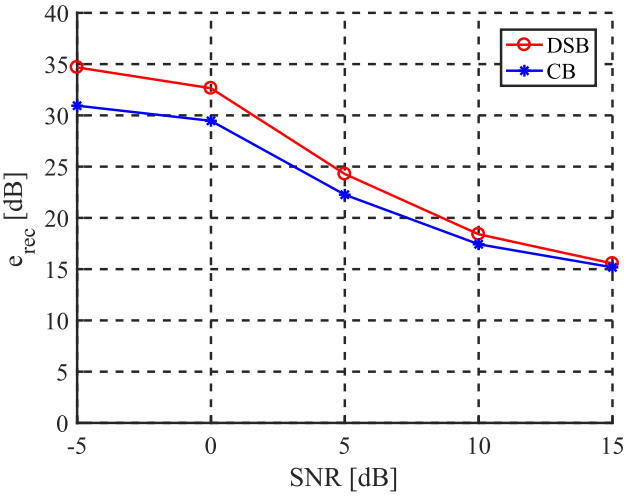

Fig. 9. The errors of localization and signal reconstruction versus SNR (Window length: 256 samples, $\lambda = 0.75\beta$).

5.3. Mismatch

Mismatch emerges when a sound source is between two grid points. In the DSB case, wrong delays would be introduced to the calculations and basis mismatch in the sensing matrix would occur in CB [13]. In this context, neither beamforming method is able to correctly localize the source. The sensitivity of compressive sensing to DFT basis mismatch was studied by Chi et al. [33]. For the application of sound source localization using CB, the basis mismatch was analyzed and several wrong localization results were presented [30].

S1 is placed from 0.01 m to 0.09 m away from the origin of Ω in the y direction with 0.02 m step to create mismatch $\Delta \in$ [0.01 m, 0.09 m]. SNR = 5 dB, $\lambda = 1.75\beta$ and the window length is 256 samples. The error results are exhibited in Fig. 10. The signal reconstruction using CB creates lower error than using DSB, the variation is around 5 dB. This is in line with the results from the previous section, that CB is more reliable than DSB under the given SNR if the parameters are selected properly. However, compared to the matched case, CB also yields larger e_{rec} due to mismatch compared to the matching cases. Note that e_{rec} does not vary much as Δ

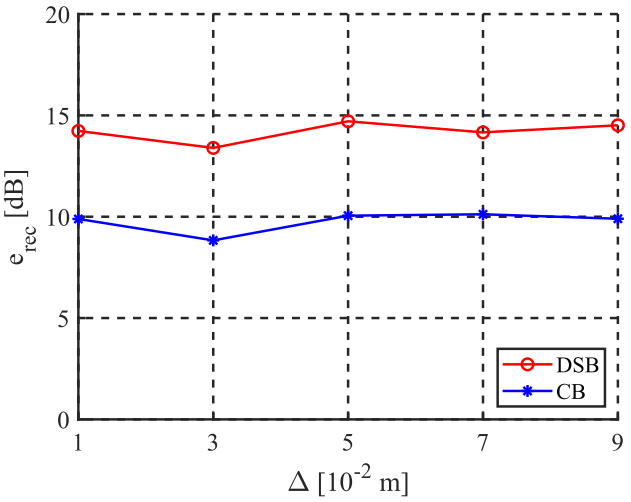

Fig. 10. The errors of localization and signal reconstruction versus mismatch Δ (Window length: 256 samples, SNR = 5 dB).

changes. It is because the mismatch studies is small compared to the distance L. This implies that in the current simulation setup, the reconstruction accuracy is robust in terms of mismatch.

5.4. Distance

The positions of microphones have been randomized and optimized to reduce the coherence of the sensing matrix. Recalling Eq. (3), large $R(t)$ would increase the similarity between TDTFs in the sensing matrix, which could reduce the coherence. Together with the microphone position, the distance L between the source trajectory and array plane should also be considered with respect to the coherence.

Fig. 11 shows the drop of e_{rec} with decreasing L until 2 m, and the curve of CB is below that of DSB. When L = 1 m, e_{rec} of CB rises and goes beyond DSB. This could be due to the regularization parameter λ. As L decreases the sensing matrix changes as well. λ was selected with L = 5 m, and it indicates that when L reaches 1 m λ is supposed to be reselected to balance the residual $||\mathbf{p} - \mathbf{Hs}||$ and the sparsity of $\mathbf{s}$.

However, in on-site measurements of pass-by vehicles, it is not always viable to place the microphone array close to the vehicle

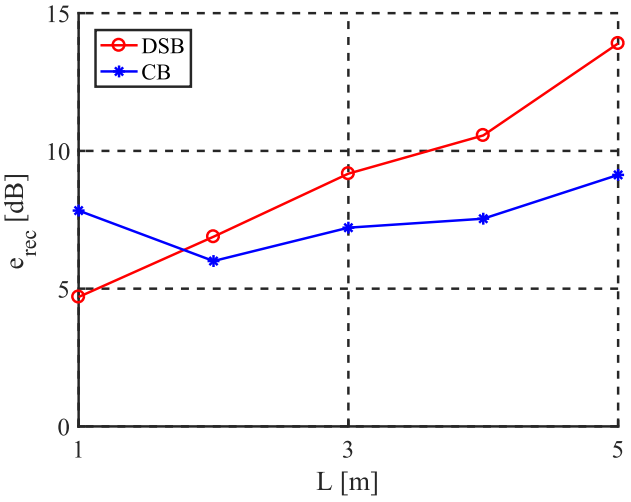

Fig. 11. The signal reconstruction error versus distance L (Window length: 256 samples, SNR = 5 dB).

trajectory. First of all, the turbulence between the car and air would introduce more noise to the microphones. Additionally, the risk exists that the array aperture would fall towards the car. It is thus necessary to keep the array distant from the trajectory. Note that in the safe distance range, CB outperforms DSB in signal reconstruction.

6. Application on a moving loudspeaker

6.1. Measurement setup

The measurements were performed on the Proving Ground of Institute for Automotive Engineering, RWTH Aachen University. The proving ground has a 400 m long test track with an acoustical part which was built referring to ISO 10844/94 [34]. The array was set 5 m away from the moving trajectory of the near-side surface of the car to keep a safe distance, on which the loudspeaker was installed. The car ran along the trajectory parallel to the array plane at different speeds. The near-side surface as the reconstruction plane was discretized into grids with 0.1 m spacing. During the measurements, the speeds were 20 km/h, 30 km/h, 50 km/h, 80 km/h and 100 km/h with two repetitions, respectively. The loudspeaker's position was measured and located at [1.2 m, 3.2 m] with reference to the front bottom point on the reconstruction plane. Fig. 12 shows the pass-by measurement setup.

A set of photoelectric sensors were placed between the trajectory and the array. The sender was placed on the other side of the car's trajectory. The emitted infrared light could be subsequently received by the receiver sensor (can be seen in Fig. 12) with the absence of obstacles. A switch of receiving the light at the sensor would generate an impulse in the sensor's recording channel. Two impulses were excited due to the car approaching and leaving. The receiver sensor was connected to the same preamplifier with the microphones and thus synchronized. The pass-by time of the car is $[T_{Apr.}, T_{Lea.}]$. Note that $[T_{Apr.}, T_{Lea.}]$ is in terms of the emission time. Following what was shown in Fig. 7 in Section 4.2 and taking the near-side surface of the car as the reconstruction plane Ω, the loudspeaker can be localized and its signal emitted during passing by the steering window can be reconstructed. Additionally, with the knowledge of the car's length, the speeds were calculated.

The same engine noise signal as in the simulations was played during the pass-by measurements. A sweep signal was added and played before the engine signal. The impulse response of the sweep signal could indicate the delay in the recording channels in terms of the playback channel and thus synchronize the recording and playback, in order to extract the original signal from the playback channel according to the pass-by time $[T_{Apr.}, T_{Lea.}]$. The

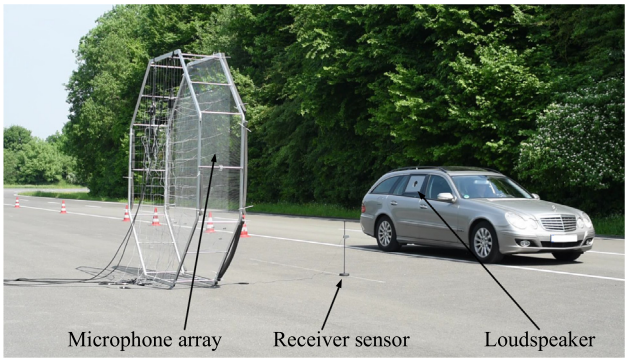

Fig. 12. The pass-by measurement of a car with a loudspeaker installed.

reference signal, from which the original signal was extracted, was recorded in an anechoic chamber.

The regularization parameter $\lambda = 1.75\beta$, and a steering window of 256 samples are applied.

6.2. Results

The localization results of the loudspeaker in terms of various speeds are shown in Fig. 13. $[x', y']$ are the coordinates with the front bottom of Ω as the origin of the local coordinate system. The dashed lines are the approximated x' and y' positions of the geometrical center of the loudspeaker surface. However, it is uncertain if the geometrical center matches the acoustic center. An interesting observation is that CB is slightly more accurate in localizing the loudspeaker, which was not implied in the simulations. However, it could result from the measurement uncertainties. The car could have not exactly followed the indicated line on the ground, especially at higher speeds. This trajectory offset could also explain the large variations in the y' direction in Fig. 13. Moreover, uncertainties also exist in the measured positions, e.g. positions of the microphones, photoelectric sensors and loudspeaker, which can introduce errors into the results.

Regarding the reconstruction results, all differences between CB and DSB are within 1 dB, with CB slightly outperforming DSB. Take the first 50 km/h run as an example, the e_{rec} of DSB and CB are 4.4 dB and 3.5 dB, respectively. Since the result values are even lower than the simulation results, they might be inaccurate since the aforementioned uncertainties could lead to incorrect distances and thus a wrong time interval for the extraction of the original signal, on the basis of which e_{rec} is computed. To inspect the possible e_{rec} variations caused by uncertain time intervals, the calculated time interval is shifted from −256 to 256 samples, leading to 512 different extracted original signals. Fig. 14 shows the e_{rec} in terms of the sample shift, demonstrating that the e_{rec} bias ranges from 0 dB to 10 dB (for CB is around 0–9 dB). Overall, the e_{rec} curve of CB is overall slightly below the DSB curve, which is in line with the simulations. Moreover, it also supports that CB is more robust under basis mismatch caused by the measurement uncertainties.

7. Discussion

In the simulations, the sample number of the steering window varied from 32 to 512 with the sampling rate of 44.1 kHz. A dip in the reconstruction error can be observed in Fig. 8(a) when the window length has 128 samples. Due to the computational cost of using CB, larger steering windows were not investigated. Small window size would cause loss of some low frequencies. Instead of using those short signal excerpts for auralization, the signal data are captured in parameters such as spectral and temporal envelopes. The actual signal synthesis will be done in combination with synthesis methods, such as spectral modeling synthesis (SMS) [35], to create audible sounds with arbitrary lengths to adapt to dynamic auralization scenes. Synthesizing pass-by sounds with combining parameters extracted from the reconstructed signals using DSB and SMS has been presented and could be perceived as realistic, although with small windows [36]. It indicates that the potential loss of low frequencies may not be audible. Another limitation of the proposed method is that the regularization parameter λ is determined by the particular simulations in Section 5.2. Obtaining λ through simulation trials has been used in previous CB work since it is difficult to discover a general solution [12]. It is thus not generic to other scenarios, for instance, the on-site measurements.

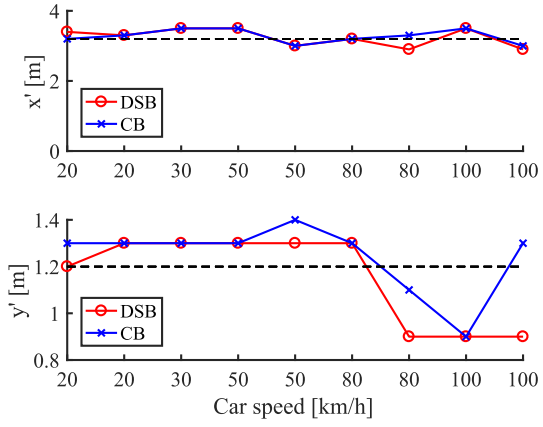

Fig. 13. Localization in x' and y' direction of the loudspeaker versus car speed with 0.1 m spaced grids. The dashed lines in the upper and bottom plots represent 3.2 m and 1.2 m, respectively. Here, $[x', y']$ are the coordinates with the front bottom of Ω as the origin of the local coordinate system.

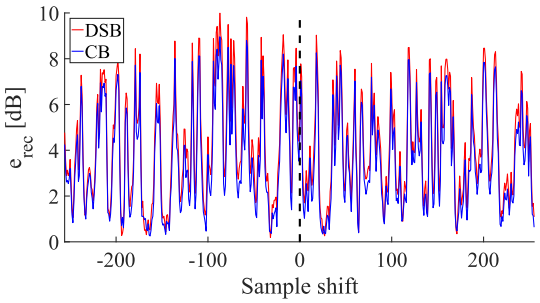

Fig. 14. The signal reconstruction error e_{rec} of DSB and CB in terms of sample shift of the time interval, which is used to extract the original signal.

In the simulation, it is straightforward to localize the time slot in the entire reference source signal to extract the original signal, which is recorded by microphones and post-processed to achieve the reconstructed signal. In the measurements, however, limited control of measurement conditions can result in biased results. As stated in the localization results of the measurements, the car might not follow the indication on the ground, leading to wrong extraction of the original signal from the reference source signal. Thus Fig. 14 was shown to look into how much it would deviate. Another concern that needs to be mentioned is that the reference source signal was measured in an anechoic chamber, which introduced uncertainties due to resetting the measurement setups. On-site measuring the reference signal is not preferable since it is difficult to capture the accurate signal played by the loudspeaker in an outdoor measurement environment. Moreover, the directivity pattern of the loudspeaker should have been taken into account in the measured data. It is neglected in this work due to the slight change in the directivity pattern, resulting from the small loudspeaker-microphone angle variations during the short pass-by distance within the steering window.

8. Concluding remarks

CB is applied to the signal reconstruction of a fast moving engine signal. A pseudorandom microphone array is designed to fulfill the requirement of CB, as well as assuring optimized localization performance of DSB in order to compare the two beamforming algorithms. The parametric studies indicate that CB outperforms DSB in terms of signal reconstruction under noisy situations, basis mismatch and large distance L between the array and source moving trajectory, while under ideal noise condition, e.g. SNR = 30 dB,

the performance of both algorithms are quite similar. Additionally, the performance of CB varies in terms of the window length and distance L. The reconstruction error increases with increasing L. In the measurement application, the signal reconstruction performance of CB and DSB are similar, just like the parameter analysis for SNR = 30 dB, but still delivers lower reconstruction errors. For localization, CB and DSB are quite similar in this study, both in simulations and measurements. Potential coherence in the sensing matrix due to large distance L and small number of sources could explain the localization similarity of the two algorithms. Nevertheless, for the purpose of this paper, signal reconstruction, CB has been demonstrated to be advantageous by means of simulations with various parameters and measurements, and thus be able to reconstruct the signals from fast moving sources.

The measure e_{rec} proposed in this study is advantageous to quickly evaluate the reconstructed signals, especially taking into account the high computational demand of CB. However, although e_{rec} accounts for human perception, listening tests are still desired to acquire more straightforward perceptual difference between the reconstructed and original signals. Furthermore, source signals with arbitrary lengths can be created combining parameters extracted from the reconstructed signal, with proper signal synthesis approach, e.g. spectral modeling synthesis (SMS). Pass-by vehicles in real scenarios will be measured by the designed array and post-processed by the proposed method. Thus, with proper propagation models and reproduction techniques, dynamic pass-by auralizations can be created to compare with real-scenarios to help validate and improve the method we proposed. The proposed work would be applicable with further array design and algorithm optimization to increase the reconstruction accuracy for real complex urban scenarios to obtain various source signals from moving vehicles.

Acknowledgments

The authors would like to thank Prof. Armin Kohlrausch from Eindhoven University of Technology for the discussion about the error evaluation of signal reconstruction, and the staff from the workshop of the Institute of Technical Acoustics, RWTH Aachen University for constructing the microphone array. Gratitude also goes to the Institute for Automotive Engineering, RWTH Aachen University for providing the Proving Ground.

References

[1] Vorländer M. Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. Berlin Heidelberg, Berlin, Heidelberg: Springer; 2007.
[2] Pieren R, Heutschi K, Wunderli JM, Snellen M, Simons DG. Auralization of railway noise: emission synthesis of rolling and impact noise. Appl Acoust 2017;127:34–45.
[3] Meng F, Behler G, Vorländer M. A synthesis model for a moving sound source based on beamforming. Acta Acustica United Acustica 2018;104(2):351–62.
[4] Sahai AK, Snellen M, Simons DG. Objective quantification of perceived differences between measured and synthesized aircraft sounds. Aerospace Sci Technol 2018;72:25–35.
[5] Jiang L, Masullo M, Maffei L, Meng F, Vorländer M. How do shared-street design and traffic restriction improve urban soundscape and human experience? —an online survey with virtual reality. Build Environ 2018;143:318–28.
[6] Jiang L, Masullo M, Maffei L, Meng F, Vorländer M. A demonstrator tool of web-based virtual reality for participatory evaluation of urban sound environment. Landscape Urban Planning 2017.
[7] Wefers F, Vorländer M. Flexible data structures for dynamic virtual auditory scenes. Virtual Reality 2018;22(4):281–95.
[8] Barsikow B, King W, Pfizenmaier E. Wheel/rail noise generated by a high-speed train investigated with a line array of microphones. J Sound Vib 1987;118 (1):99–122.
[9] Rietdijk F, Heutschi K, Zellmann C. Determining an empirical emission model for the auralization of jet aircraft. In: Proceedings of the 10th European Congress and Exposition on Noise Control Engineering (31 May–3 June 2015).
[10] Peplow A, Forssén J, Lundén P, Nilsson ME. Exterior auralization of traffic noise within the listen project. In: Proceedings of the European Conference on Acoustics (Forum Acusticum 2011) (23 June–1 July 2011).
[11] Molla S, Bongini E, Gautier PE, Habault D, Mattei PO, Poissen F. Vehicle passby noise prediction and audio synthesis. In: Proceedings of the 19th International Congress on Acoustics (2–7 September 2007).
[12] Malioutov D, Cetin M, Willsky AS. A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans Signal Processing 2005;53(8):3010–22.
[13] Xenaki A, Gerstoft P, Mosegaard K. Compressive beamforming. J Acoust Soc Am 2014;136(1):260–71.
[14] Edelmann GF, Gaumond CF. Beamforming using compressive sensing. J Acoust Soc Am 2011;130(4):EL232–7.
[15] Sijtsma P, Oerlemans S, Holthusen H. Location of rotating sources by phased array measurements. In: Proceedings of 7th AIAA/CEAS Aeroacoustics Conference and Exhibit.
[16] Morse PM, Ingard KU. Theoretical acoustics. Princeton: Princeton University Press; 1968. 1986.
[17] Kephalopoulos S, Paviotti M, Ledee FA. "noise assessment methods in europe (cnossos-eu)," Common noise assessment methods in Europe (CNOSSOS-EU) (2012).
[18] Candè EJ, Wakin MB. An introduction to compressive sampling. Signal Process Mag, IEEE 2008;25(2):21–30.
[19] Grant M, Boyd S. Graph implementations for nonsmooth convex programs. Recent Adv Learn Control 2008;95–110.
[20] Kuttruff H. Acoustics: an introduction. London and New York: Taylor & Francis; 2007.
[21] Johnson DH, Dudgeon DE. Array signal processing: concepts and techniques. Simon & Schuster; 1992.
[22] Candes EJ. The restricted isometry property and its implications for compressed sensing. Comptes Rendus Mathematique 2008;346(9–10):589–92.
[23] Gaumond CF, Edelmann GF. Sparse array design using statistical restricted isometry property. J Acoust Soc Am 2013;134(2). EL191–7.
[24] Gerstoft P, Hodgkiss WS. Improving beampatterns of two-dimensional random arrays using convex optimization. J Acoust Soc Am 2011;129(4). EL135–40.
[25] Kook H, Davies P, Bolton JS. Statistical properties of random sparse arrays. J Sound Vib 2002;255(5):819–48.
[26] Zheng S, Xu F, Lian X, Luo Y, Yang D, Li K. Generation method for a two-dimensional random array for locating noise sources on moving vehicles. Noise Control Eng J 2008;56(2):130–40.
[27] Bai MR, Chen YS, Lo YY. A two-stage noise source identification technique using a far- field random array. In: Proceedings of the 45th International Congress and Exposition on Noise Control Engineering (INTER-NOISE 2016) (21–24 August 2016).
[28] Jagla J, Maillard J, Martin N. Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-and-add method. J Acoust Soc Am 2012;132(5):3098–108.
[29] Rabiner LR, Juang BH. Fundamentals of speech recognition. New Jersey: Prentice-Hall International (UK), Englewood Cliffs; 1993.
[30] Gerstoft P, Xenaki A, Mecklenbräuker CF. Multiple and single snapshot compressive beamforming. J Acoust Soc Am 2015;138(4):2003–14.
[31] Candes E, Tao T. The dantzig selector: Statistical estimation when p is much larger than n. Ann Stat 2007:2313–51.
[32] Gurbuz AC, McClellan JH, Cevher V. A compressive beamforming method. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2008. IEEE; 2008. p. 2617–20.
[33] Chi Y, Scharf LL, Pezeshki A, Calderbank AR. Sensitivity to basis mismatch in compressed sensing. IEEE Trans Signal Process 2011;59(5):2182–95.
[34] ISO 10844-1994, Acoustics. specification of test tracks for the purpose of measuring noise emitted by road vehicles, (15 March 1995).
[35] Pieren R, Bütler T, Heutschi K. Auralization of accelerating passenger cars using spectral modeling synthesis. Appl Sci 2015;6(1):5.
[36] Meng F, Georgiou F, Stienen J, Vorländer M, Hornikx M. A concept of auralizing urban pass-by vehicles by measurement- and simulation-based synthesis of sound sources and impulse responses. In: Proceedings of the 11th European Congress and Exposition on Noise Control Engineering (2018).