Moving Sound Source Simulation Using Beamforming and Spectral Modelling for Auralization

Fanyu Meng [1], Frank Wefers[1,2], Michael Vorlaender[1]
[1] Institute of Technical Acoustics, RWTH Aachen University, 52074 Aachen, Deutschland, Email: fme@akustik.rwth-aachen.de
[2] now at HEAD Acoustics GmbH, 52134 Herzogenrath, Deutschland

Introduction

Noise emitted from trains is an important contribution to environmental noise, especially at high speed. To achieve better noise control and subsequent traffic planning in densely populated urban and rural areas, auralization is an efficient modern technique, which makes it possible to perceive simulated sound intuitively instead of describing the acoustic properties by abstract numerical quantities. Simply saying, auralization converts numerical acoustic data to audible sound files, basically through three procedures: sound source generation, sound propagation and reproduction [1].

The sound source signal can be recorded or synthesized. When auralization is carried out in real-time virtual reality environment, source parameters are changed due to different scenarios and listener movement. Therefore the listeners can move freely, leading to sound field changing by updating the parameters, which gives back more immerse perception. In this respect, simple recordings are not adequate since auralization is then confined to the scenarios where the recordings were made. Besides, for moving sources like cars or trains, the source signal in the recording is contaminated by the Doppler Effect and other outdoor propagation effects. Thus, parametric synthesizer is a proper way to reconstruct sound sources. The auralization is hence not limited to several scenarios by controlling the parameters, which can be obtained from recordings as well [2][3].

Overall, the sound source synthesis consists of forward and backward approaches. Forward method is based on a priori knowledge of the source, e.g. physical generation mechanism, spectral data to obtain the source signal, while backward method acquires the signal by inverting the propagation procedure (directivity, Doppler Effect, spherical spreading etc.) from recording.

For stable and small sound source, it can be placed in anechoic chamber and every condition can be well handled, so that it's feasible to obtain the source signal directly through near field recording. However, when it comes to multiple sources, especially moving ones, simple measurements are not applicable any longer. For example, aerodynamic noise from train can't be measured in static state. In this case, the inverse backward method can be utilized in moving sound source's synthesis for auralization.

Beamforming is a popular sound source localization method based on array technique [4]. In this research, beamforming method is not only applied for localization, but also extended to sound source synthesis. With the delay compensation of each microphone, an enhanced signal is obtained by summing the previous signals, which is the so-called delay and sum beamforming, or conventional beamforming (CB) [5]. Therefore, when the delay is well chosen with a corresponding incident angle, the sound source can be found accordingly. The beamforming output in frequency domain can not be directly considered as source spectrum, because it is the convolution of the spectrum and array's point spread function (PSF) [6]. However, the frequency information is not blurred, so spectral modelling is applicable in this case. Simply put, Spectral modelling synthesis (SMS) is to synthesize a piece of spectrum separately in deterministic and stochastic components [7]. For moving sound sources, a windowed part should be used instead of using the whole recording since the signal is changing due to the source movement. It is better to "sample" the detected moving plane discretely, and each part is analyzed when it passes the array origin [8]. At the same time, the window size should be determined properly according to the source moving speed. When the source moves at high speed and the window size is comparatively large, the source moving distance will be quite long which leads to lower resolution in the localization. Additionally, the Doppler Effect should be removed to achieve accurate CB results.

In this paper, de-Dopplerization technique is introduced, as well as with CB and SMS theories; in addition, a simulation with moving sound sources and microphone array is conducted to apply the previous theory for source synthesis; finally, localization and synthesized results are illustrated.

Theories

Conventional beamforming (CB) theory

When an array of microphones record simultaneously, the signals measured by each microphone can be denoted as $p_m(t)$. If we apply a time delay compensation $-\tau$ to a microphone compared to the array origin, the signal of this microphone is the same as that measured by the origin.

τ_m is expressed as

$$\tau_m = \frac{r_m - r_0}{c} \tag{1}$$

where r_m is the distance between source and the mth microphone, r_0 is the distance between source and array

origin. Then, the CB output is

$$y(t) = \sum_{m=1}^{M} w_m p_m(t + \tau_m) \tag{2}$$

where M is the number of microphones, w_m is the weight of the output signal from the mth microphone.

Moving sound source and de-Dopplerization

The sound field caused by a moving sound source is described by the following equation (constant speed) [9]:

$$p = \frac{1}{4\pi}\frac{q'[t-(R(t)/c)]}{R(1-Mcos(\theta(t)))^2} + \frac{q(t)}{4\pi}\frac{(cos(\theta(t))-M)v}{R(t)^2(1-Mcos(\theta(t)))^2} \tag{3}$$

where $q(t)$ is the source strength, $R(t)$ is the distance between source and receiver, c is the sound speed, v is the source moving speed, $M = v/c$ is the Mach number, $\theta(t)$ is the angle between source moving direction and source-receiver direction. Among which, $R(t)$ and $\theta(t)$ vary as the source moves, so they are the dependent variables of time.

When far field and low Mach number are considered, the second term can be omitted because it is much smaller compared with the first one. Therefore, Equation 3 is simplified as

$$p = \frac{1}{4\pi}\frac{q'[t-(R(t)/c)]}{R(1-Mcos(\theta(t)))^2} \tag{4}$$

To get rid of Doppler Effect, the recordings need to be interpolated and re-sampled, as is illustrated in Figure 1. The reception time is calculated by $t = t_e + r/c$ with taking emission time as the reference time. Then the recorded signal is interpolated and re-sampled according to the reception time, whose interval between each sample is not equally spaced. The above procedure is called de-Dopplerization.

Hence, the CB output can be rewritten in the following form

$$y(t) = \sum_{m=1}^{M} w_m \bar{p}_m(t + \tau_m) \tag{5}$$

where $\bar{p}_m$ is the de-Dopplerized signal at mth microphone.

Spectral modeling synthesis (SMS)

This subsection introduces how to apply SMS with the CB output spectrum. Short-time Fourier Transform (STFT) is conducted on the given spectrum. The prominent peaks are detected in each magnitude spectrum, and then peak continuation is tracked along all the frames in time domain to pick out the deterministic part, or in other words, tones in the whole spectrum. Therefore, the deterministic component is synthesized by the sum of

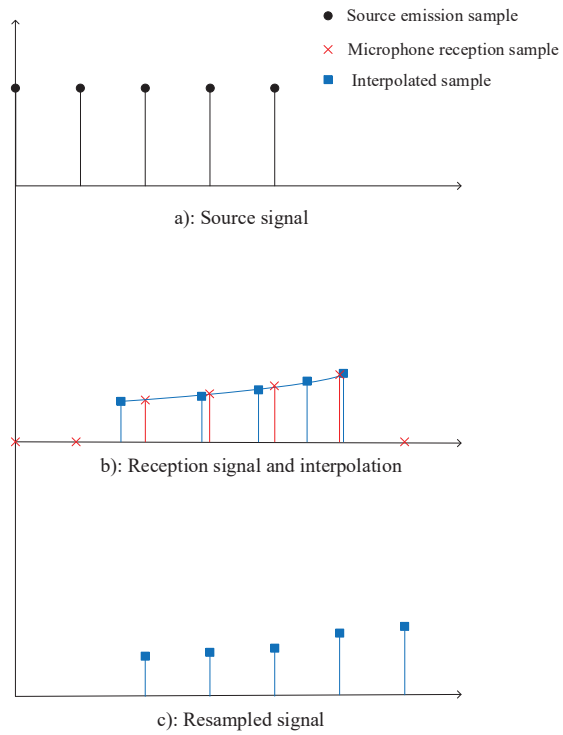

Figure 1: De-Dopplerization procedure. a): signal emitted from sound source; b): red cross represents the signal measured by microphone, blue square represents the interpolated samples according to the calculated reception time (interval not equally spaced) with regards to emission time; c): re-sampled interpolated signal with regards to equally-spaced-interval reception time

all the detected tones. Afterwards, the broadband component is modeled by the original spectrum subtracted by the synthesized tonal component. The procedure to execute SMS is described in Figure 2.

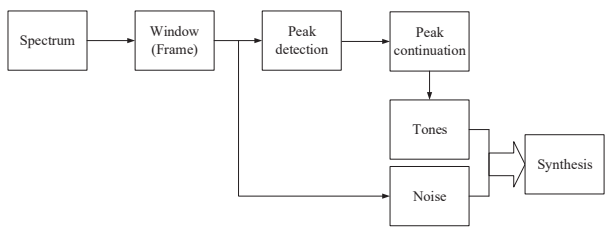

Figure 2: SMS diagram

Simulation and results

A plane (1.5 m x 5 m) is moving in the -y direction, carrying two point sources at the speed of 40 m/s. A microphone array is set 1.5 m away from the moving direction. The array origin is on z axis. Figure 3 shows the sketch of the simulation. The plane is then meshed into grids, with 5 cm spacing between each other. Each grid represents an assuming sound source, so that the array can steer its angle to "scan" the moving plane to look for sound source.

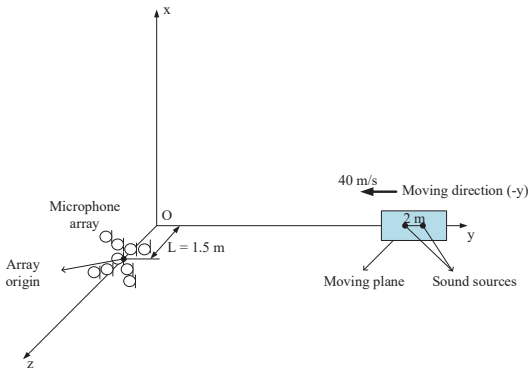

Figure 3: Moving sound source measured by microphone array

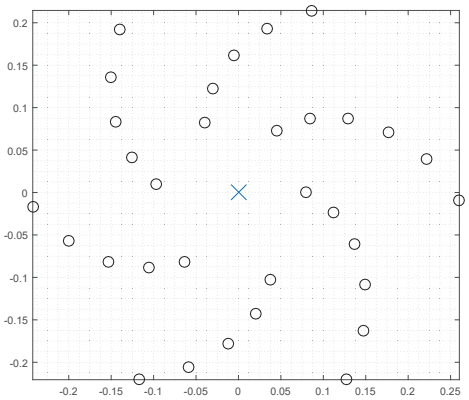

Figure 4: Spiral microphone array with 32 microphones (o: microphone, ×: array origin)

A spiral microphone array is applied (Figure 4). Its basic parameters can be found in Table 1. First of all, this is a two-dimensional array, which is capable to steer both horizontally and vertically; Besides, it has high Maximum Sidelobe Level (MSL) and avoids grating lobes due to its irregular arrangementarrangement of microphone positions [10].

Table 1: Basic parameters of the spiral microphone array (Resolution is obtained at 3 kHz, with array steering to 30°)

Mic number	Spacing	Diameter	Resolution
32	0.04-0.06m	0.50m	0.64m

The window size is chosen according to Equation 6. There are 1024 samples in each window in the following analysis due to 40 m/s moving speed of the sound source, considering the maximum steering angle as 30° at the same time.

$$t_{win} = \frac{2Ltan(\theta)}{v} \tag{6}$$

Now taking two tones (2 kHz, 4 kHz) as the sources, with 10 dB SNR additive white Gaussian noise (AWGN) at each microphone, the localization results are shown in Figure 5. It can be seen that the center of the red dots represent the sound source location. Focus the array steering direction at the two source locations, the spectra are obtained from the beamforming output as in Figure 6. Two peaks, representing the two tones, are obviously illustrated in the two figures.

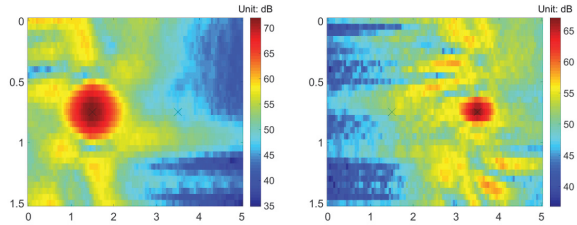

Figure 5: Localization results of moving tones (2 kHz and 4 kHz)

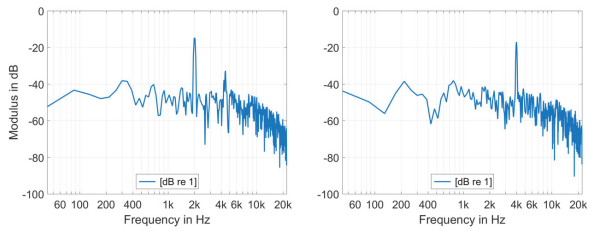

Figure 6: The beamforming output spectra with array steering to the 2 kHz and 4 kHz tonal sources

In SMS, a lot of parameters are involved to perceive a better result, e.g. window size, type, maximum peak amplitude (MPA), initial peaks in the first frame, peak deviation during continuation detection etc. Since the frequency resolution is limited when the window size is small, interpolation is conducted in the spectra before using SMS to increase the resolution to 1 Hz. The peak detection results are shown in Figure 7 with varying MPA. As can be seen, when MPA = 0.009 Pa, a clear 4 kHz tone can be tracked along all the frames; however, for the other cases, either incorrect or wrong trajectories are found. Since in this example the source information has already been given, the result can be verified directly; when the source is unknown, there is no a priori knowledge. Therefore, the parameters need to be determined deliberately with proper simulation and verification.

Similarly, the localization and peak detection results of moving tone and noise sources are given in Figure 8, with the same AWGN (SNR=10 dB). Figure 9 shows the spectra of the beamforming output with array steering at the two source positions. The track of tone signal is the same as the example above, while the noise part is synthesized in a different way. Broadband white noise is firstly generated, and then filtered by the noise spectrum which is shown in the second subfigure. For the amplitude of synthesized signal, the beamforming output should be compensated by the propagation distance to reach the original level of the source. With the amplitude and fre-

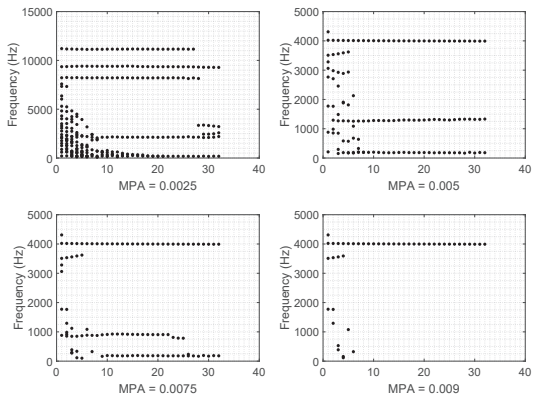

Figure 7: Peak detection with varying MPA in applying SMS to two moving tones

quency/amplitude and spectrum, the tonal and broadband signal can be synthesized separately.

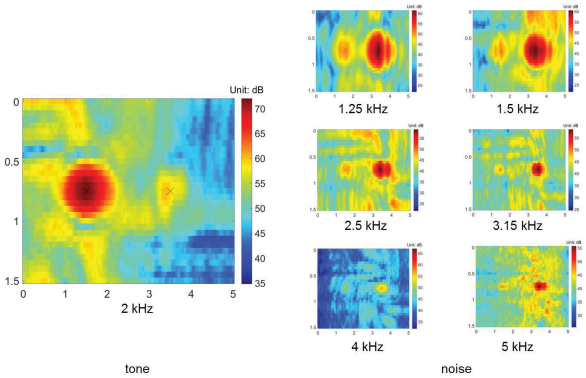

Figure 8: Localization results of moving tone (2 kHz) and noise

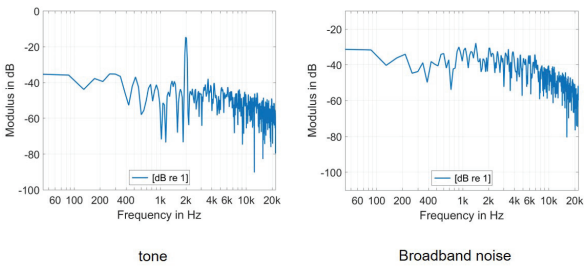

Figure 9: The beamforming output spectra with array steering to the 2 kHz tone and broadband noise sources

Conclusions and Outlook

In this paper, the reconstruction procedure of moving sound sources using microphone array is introduced. As an array processing method, delay and sum beamforming is applied for not only localization but also sound source synthesis on the basis of beamforming output spectrum. In addition, de-Dopplerization technique and spectral modeling synthesis are joint together with beamforming to achieve the goal. A simulation with different tonal and broadband moving sources is run and the results show that this synthesis procedure is plausible and promising. In the future study, a more detailed parameter selection method should be developed in SMS. What are also important are verification approaches, e.g. measurement and listening test. Besides, when it comes to high-speed moving sources, the window size should be narrowed, and only using interpolation may increase uncertainty in the spectral synthesis. Hence how to overcome this problem is a topic to be studied as well. Over all, a synthesis model can be generalized for multiple moving sound sources with the help of beamforming and SMS.

References

[1] Michael Vorlaender. *Auralization*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[2] Reto Pieren, Kurt Heutschi, Menno Müller, Madeleine Manyoky, and Kurt Eggenschwiler. Auralization of wind turbine noise: Emission synthesis. *Acta Acustica united with Acustica*, 100(1):25–33, 2014.

[3] Nadine E. Miner and Thomas P. Caudell. Method of sound synthesis, 2004.

[4] Harry L. van Trees. *Optimum Array Processing*. John Wiley & Sons, Inc, New York, USA, 2002.

[5] Don H. Johnson and Dan E. Dudgeon. *Array signal processing: concepts and techniques*. Simon & Schuster, 1992.

[6] Thomas F. Brooks and William M. Humphreys. A deconvolution approach for the mapping of acoustic sources (damas) determined from phased microphone arrays. *Journal of Sound and Vibration*, 294(4):856–879, 2006.

[7] Xavier Serra and Julius Smith. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, pages 12–24, 1990.

[8] Fanyu Meng, Frank Wefers, and Michael Vorlaender. Acquisition of exterior multiple sound sources for train auralization based on beamforming. In *EuroNoise 2015 : Maastricht, May 31 - June 3, 2015 / 10th European Congress and Exposition on Noise Control Engineering.*, pages 1703–1708. May 2015.

[9] Philip M. Morse and K. Uno Ingard. *Theoretical acoustics*. Princeton University Press, Princeton, 1986(1968).

[10] J. Hald J.J. Christensen. Technical review: Beamforming, 2004.