inter.noise 2014
MELBOURNE AUSTRALIA
16-19 NOVEMBER

A three-stage method for sound field reproduction in rooms with reflection boundary:theory and experiments

Bo PENG[1]; Sifa ZHENG[2]; Xiangning LIAO[3]; Xiaomin LIAN[4]

State Key Laboratory Of Automotive Safety and Energy, Tsinghua University, China

ABSTRACT

This paper presents a three-stage method for sound field reproduction in ordinary rooms. Optimization of speaker positions is important to reduce the number of speakers and the calculation time should be acceptable so that the algorithm can be put into practise. In the first stage of the proposed algorithm, the least absolute shrinkage and selection operator(Lasso) is used to select the most important speakers' positions for all frequencies. Then *l*2-norm regularization is performed in the second stage to design the FIR filters. In the third stage, a fast convolution method based on Fast Fourier Transform(FFT) is carried out to reduce the time consuming of filtering algorithm. The performance of this three-stage method is investigated by experiments for different speaker numbers. The calculation time proves the efficiency of this method and the results of experiments show that compared with the widely used inverse filtering method, the proposed method can significantly reduce the speaker number without a serious side effects on the reproduction accuracy.

Keywords: Sound,Reproduction, Reflection I-INCE Classification of Subjects Number(s): 38.2

1. INTRODUCTION

Sound field reproduction(SFR) technique has become an important topic for decades. Ambience and localization cues of sound sources are precisely represented by reproduce the sound field in the listening space. Applications to acoustics include reproduction of primary acoustic signals, impressions regarding source locations and evaluation of sound quality.

After decades of research, there are mainly three kinds of method to reproduce sound. The first one is called ambisonics. The theoretical foundations of ambisonics were laid down in the 70s(1) and at the heart of this technique is the mode matching of orthogonal components of the desired sound field, such as cylindrical or spherical harmonics. This technique is primarily for circular or spheric loudspeaker arrangements while irregular arrangement of loudspeakers is studied in recent years.(2, 3)

Another important method is Wave Field Systhesis(WFS). It was firstly proposed by Berkhout in 1988.(4) This method is based on Helmholtz integral equation, which shows how a desired sound field in a closed source-free space can be reproduced by a continuous distribution of monople and dipole sources. The creators of WFS focused on linear loudspeaker setups and approximated the performance of planar source distributions.

The third method uses the least-squares(LS) criterion to design loudspeaker signals whose synthesis approximates the sound field. The desired field is sampled in space using an array of microphones and the loudspeaker amplitudes are designed in order to minimise the square error between the samples of the reproduced field and the desired one.(5) This method regards the reproduction as a numerical optimization problem and there is no strictly constraints on the position of loudspeakers.

Fazi's study shows that the three methods have an equivalent format background and give identical

[1] pb11@mails.tsinghua.edu.cn
[2] zsf@mail.tsinghua.edu.cn
[3] liaoxiangning@163.com
[4] lianxm@mail.tsinghua.edu.cn

reconstruction performance in the target frequency range when using the same number of transducers, regularly arranged over a sphere.(6) As for the former two methods, however powerful as theoretical tools, the mentioned approaches for sound field reproduction need to cope with limitations imposed by systems used in practise, such as non-ideal microphones, non-ideal loudspeakers, arrangement of loudspeakers and reflection of the walls. There is no such a requirement in LS-based method. As a result, there are already some experimental researches to verify the performance of LS-based method when reproduce a wideband mechanical noise(7, 8) while the experiments of the other two methods mainly conducted in the anechoic chamber with regular loudspeaker arrangement(spherical mostly). Recent study shows that LS-based method can outperform WFS method(9) that gives another reason to reproduce a practical sound field with LS based method.

When LS-based method is used to do the SFR, the optimization of the loudspeaker positions should be taken into consideration for the sake of reducing loudspeakers. And when deal with time-domain sound siganls, FIR filters are necessary to get time-domain loudspeaker signals. The length of FIR filters is mainly decided by the frequency-domain resolution. To reproduce the sound field precisely requires a high resolution in the frequency-domain which leads to long FIR filters and computational diseconomy. The computational time should be acceptable when apply the SFR technique.

This paper present a three-stage method, which is a supplement of the widely used LS-based method. In the first stage, lasso is conducted to choose the most important loudspeaker positions which gives a solution to reduce the speaker number. And then LS-based method with l2-norm regularization is used to design the FIR filters with high frequency-domain resolution. In the third stage, a fast convolution method is carried out to make the convolution efficient. Experimental research also conducted to verify the proposed method.

2. SOUND FIELD REPRODUCTION

2.1 Process of LS-based Method

The process of LS-based method for SFR is almost the same. As shown in the following block diagram, $\mathbf{P}(f)=[P_1(f),P_2(f),\cdots,P_M(f)]$ are the measured sound pressures in the desired sound field. M is the number of microphones and f stands for frequency. $\mathbf{S}(f)=[S_1(f),S_2(f),\cdots,S_M(f)]$ are the sound pressures in the reproduction sound field measured with the same mircophone array. $\mathbf{G}(f)=[G_1(f),G_2(f),\cdots,G_L(f)]$ are the loudspeaker signals and L is the loudspeaker number. $\mathbf{H}(f)$ is defined as the acoustic transfer function(ATF) matrix of demension $L\times M$ which can be measured directly. The relationship between $\mathbf{S}(f)$ and $\mathbf{G}(f)$ can be de described as $\mathbf{S}(f)=\mathbf{G}(f)\mathbf{H}(f)$. This description laid the foundation of LS-based method. $\mathbf{R}(f)$ is an inverse matrix of demension $M\times L$ and it can be regarded as the inverse of $\mathbf{H}(f)$. $e(f)$ is defined as the sound reproduction error.

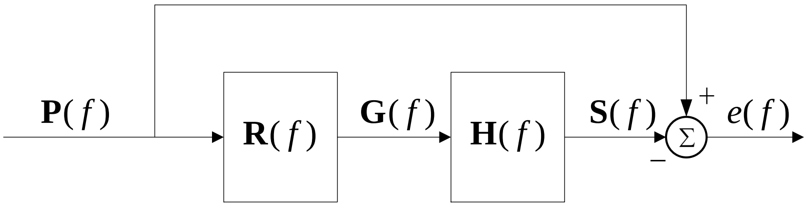


Figure 1 – Block diagram illustrateing the process of SFR with LS-based method

The process of LS-based method can be summarised as follows: first use the inverse matrix $\mathbf{R}(f)$ to multiply with the measured sound pressures $\mathbf{P}(f)$ to get the loudspeaker signals $\mathbf{G}(f)$. And then playback the signals to obtain the reproduced sound pressures $\mathbf{S}(f)$.

2.2 Error Discription

It is of great significance to decribe the error of SFR clearly and reasonably. The performance of SFR can be presented concisely as follows:

$$\min_{\mathbf{G}(f)} \left\|\mathbf{P}(f)-\mathbf{S}(f)\right\|_2^2 \tag{1}$$

Here three different standards are used to evaluate the reproduction accuracy according to equation (1). The first one called normalized reproduction error is defined by

$$e_1(f)=\frac{\left\|\mathbf{P}(f)-\mathbf{S}(f)\right\|_2}{\left\|\mathbf{P}(f)\right\|_2} \tag{2}$$

It describes, on the average, how accurately the sound field is reproduced over the microphone array. To quantify the reproduced sound environment with respect to timbre, the averaged magnitude spectrum error is defined by

$$e_2(f)=\frac{1}{M}\left\|20\log_{10}\frac{\left|\mathbf{P}(f)\right|}{2\times10^{-5}}-20\log_{10}\frac{\left|\mathbf{S}(f)\right|}{2\times10^{-5}}\right\|_2 \tag{3}$$

It represents the reproduction error in terms of power density function of the desired sound pressures and reproduced sound pressures. In other words, it does not take into account the spatial distribution of the phase but only the spatial distribution of sound pressure amplitude. The third standard, averaged sound pessure level error, takes the time dimension into account and the definition is given by

$$e_3(t)=\frac{1}{M}\left\|SPL[\mathbf{p}(t)]-SPL[\mathbf{s}(t)]\right\|_2 \tag{4}$$

In equation (4), $SPL[\cdot]$ denotes to get sound pressure level versus time. $\mathbf{p}(t)$ and $\mathbf{s}(t)$ are time-domain signals of desired sound field and reproduced sound field respectively. In other words they are the time-domain expression of $\mathbf{P}(f)$ and $\mathbf{S}(f)$. This standard gives a comprehensive evaluation of the reproduction.

3. THE THREE-STAGE METHOD

The proposed three-stage method is an extension of LS-based method and it will be introduced in detail in this section.

3.1 First stage Position Slection

As shown in equation (1), sound field reproduction can be regarded as an optimization problem. Regularization is always needed to improve the robustness. When the regularization is performed with l1-norm, it enables parsimonious variable selection. Another name of this algorithm is called the least absolute shrinkage and selection operator(Lasso) method and it has already been used for SFR in references (10, 11). Here lasso method is used for choosing loudspeaker positions only. After discretizing the frequency with a high resolution Δf , the lasso method is given by

$$\mathbf{G}_{Lasso}(f_n):=\min_{\mathbf{G}_0(f_n)}\left[\frac{1}{2}\left\|\mathbf{G}_0(f_n)\mathbf{H}_0(f_n)-\mathbf{P}(f_n)\right\|_2^2+\alpha(f_n)\left\|\mathbf{G}_0(f_n)\right\|_1\right] \tag{5}$$

Where $\mathbf{G}_0(f_n)$ stands for the candidate loudspeakers with L_0 (L_0>L) length and $\mathbf{H}_0(f_n)$ is the corresponding acoustic transfer function matrix. The lasso solution can be obtained by an iteration method which is introduced in detail in (10). Some entries of $\mathbf{G}_{Lasso}(f_n)$ are forced to zero and the sparsity parameter $\alpha(f_n)$ controls the number of nonzero ones in the sense that larger values of $\alpha(f_n)$ yield fewer nonzero entries. By making use of this property, the loudspeaker positions can be chosed by

$$\mathbf{F}=\text{find}\left[\sum_{n=1}^{N}\mathbf{G}_{Lasso}(f_n)>\delta\right] \tag{6}$$

In (6), $f_n=f_{start}+(n-1)\Delta f$ is the discretized frequency and f_{stop} is got when $n=N$. δ is called loudspeaker number weight parameter. The notation $\text{find}\left[\mathbf{A}>\delta\right]$ returns the linear indices of the entries of vector $\mathbf{A}$ that are greater than δ. The returned indices stand for the chosen loudspeaker positions and the length of the indices vector $\mathbf{F}$ equals to the new loudspeaker number.

With the chosen loudspeakers, FIR filters can be design in the next subsection.

3.2 Second Stage FIR Filters Design

It is necessary to use FIR filters to deal with time domain sound pressure signals. Lasso is a powerful method to select the most significant components in the target vector but single-stage lasso selects a different set of active speakers for each discrete frequency so that it is difficult to transfrom the results to time-domian FIR filters. Here *l*2-norm regularization is used to accomplish this mission. Assume the number of chosen loudspeakers is *L*, the SFR problem can be stated as:

$$\mathbf{G}_L(f_n) := \min_{\mathbf{G}(f_n)} \left[\left\| \mathbf{G}(f_n)\mathbf{H}(f_n) - \mathbf{P}(f_n) \right\|_2^2 + \lambda(f_n) \left\| \mathbf{G}(f_n) \right\|_2^2 \right] \tag{7}$$

This is a classic optimization problem, $\lambda(f_n)$ is called the regularization parameter. The solution is optimized when

$$\mathbf{G}_L(f_n) = \mathbf{P}(f_n)\mathbf{H}^H(f_n)\left[\mathbf{H}(f_n)\mathbf{H}^H(f_n) + \lambda(f_n)\mathbf{I}\right]^{-1} \tag{8}$$

$(\cdot)^H$ denotes the Hermitian transpose and $\mathbf{I}$ is an identity matrix of dimension $L \times L$. There are several methods to decide $\lambda(f_n)$.(12) Here L-curve criterion is used.(13) From equation (8) it is easy to define the discretized inverse matrix $\mathbf{R}(f_n)$ by

$$\mathbf{R}(f_n) = \mathbf{H}^H(f_n)\left[\mathbf{H}(f_n)\mathbf{H}^H(f_n) + \lambda(f_n)\mathbf{I}\right]^{-1} \tag{9}$$

Then FIR filters can be designed with the frequency sampling method(14) and the FIR filters can be expressed in matrix form as

$$\mathbf{r}(\tau_i) = \begin{bmatrix} r_{11}(\tau_i) & r_{12}(\tau_i) & \cdots & r_{1L}(\tau_i) \\ r_{21}(\tau_i) & r_{21}(\tau_i) & \cdots & r_{2L}(\tau_i) \\ \vdots & \vdots & \ddots & \vdots \\ r_{M1}(\tau_i) & r_{M1}(\tau_i) & \cdots & r_{ML}(\tau_i) \end{bmatrix} \tag{10}$$

where τ_i is the discretized time and *I* is the time length. Every entry of $\mathbf{r}(\tau_i)$ is a FIR filter. To achieve a high accuracy of reproduction the time length of the filters should be long enough to cooperate with the high frequency resolution Δf .

3.3 Third Stage Fast Convolution

The process of filtering is illustrated in figure 2. $p_1(t_k), p_2(t_k), \cdots, p_M(t_k)$ are the discretized sound pressures in the desired field and $g_1(t_k), g_2(t_k), \cdots, g_L(t_k)$ are the discretized signals of loudspeakers. It is a relatively complicated filter network.

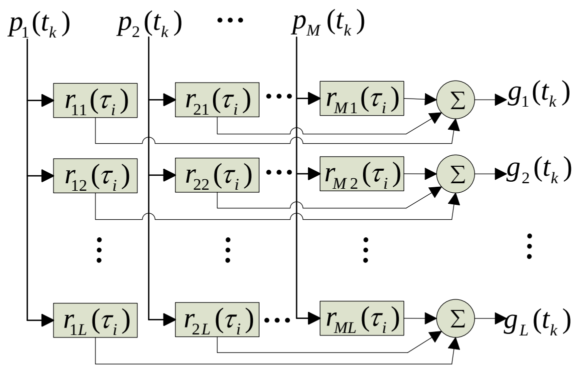


Figure 2 – Block diagram illustrateing the process of filtering

The filtering is accomplished by convolution. Because of the high frequency resolution, the calculation is time consuming. Define $b_m(k)$ to decribe the convolution result of every FIR fliter and in order to focus on mathematics itself, $b_m(t_k)$ is defined by

$$b_m(k) = r_{lm}(i) * p_m(k) \tag{12}$$

In definition (12), *i*,*k* are both nonnegative integers coresponding to τ_i and t_k respectively. It is a linear convolution operation and only the valid party of the result, which means the convolution result gotten when all the data in the filters are used, is meaningful. So the lengths of $r_{lm}(i), p_m(k), b_m(k)$ are *I*,*K*(*K*>*I*) and *K*-*I*+1. Then the fast convolution method is given by

$$b_m(k) = \mathrm{ifft}\left[\mathrm{fft}\left[p_m(k), K\right] \circ \mathrm{fft}\left[r_{lm}(i), K\right]\right] \tag{13}$$

where $\circ$ is the Hadamard product. $\mathrm{fft}\left[r_{lm}(i), K\right]$ means the FFT operation of length *K*. Zero padding should be performed if the vector length is less than *K*. Notation $\mathrm{ifft}\left[\cdot\right]$ declares inverse FFT operation. The length of $b_m(k)$ in (13) is *K* and the last *K*-*I*+1 entries are exactly what we want.

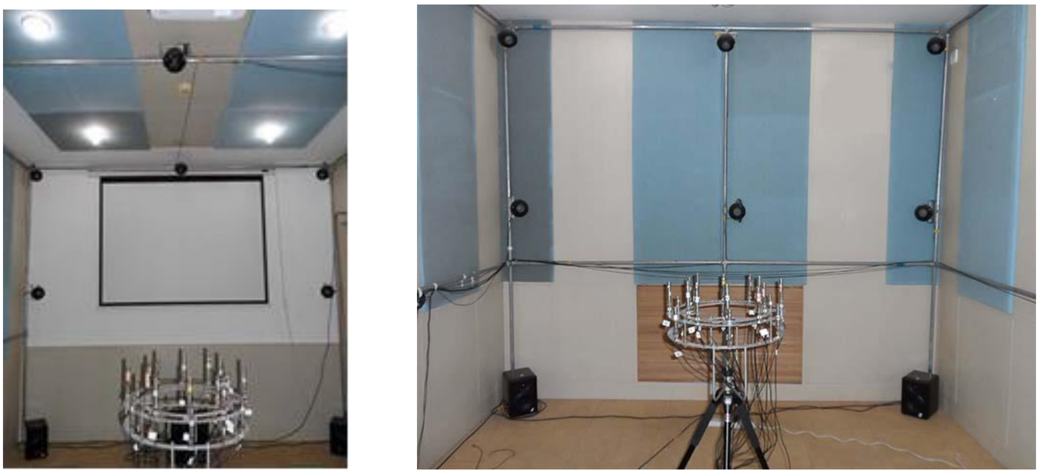
When *K* is too large compared with *I*, $p_m(k)$ can be divided into several parts. Then use the algorithm above to calculate the result of each part and the total calculation time can be reduced.

4. EXPERIMENTS

In this section a sound field measured in a running vehicle is reproduced using the proposed three-stage method in the laboratory.

4.1 Configurations

Although there are no strictly requirements on the shapes of loudspeaker array and microphone array, it is still necessary to arrange the loudspeakers and microphones reasonably. Figure 3 shows the arrangement used in the experiments. The basical idea of the design is to make the laboratory more convenient. The size of the laboratory is 4.2m×3.1m×2.6m. There are two kinds of loudspeakers(20 loudspeakers in total) are used to cover a wide range in the frequency domain. The components below 100Hz are reproduced only by 4 loudspeakers while the components above 100Hz are reproduced by all the loudspeakers(Including the former mentioned 4). So there are 4 loudspeakers playback the signal in the whole frequency range and 16 loudspeakers only in the high frequency party. The loudspeaker selection is carried out in the high frequency party. The total number of microphones is 18 and it is arranged as a cylindrical envelope. The diameter of the cylindrical is 0.4m.

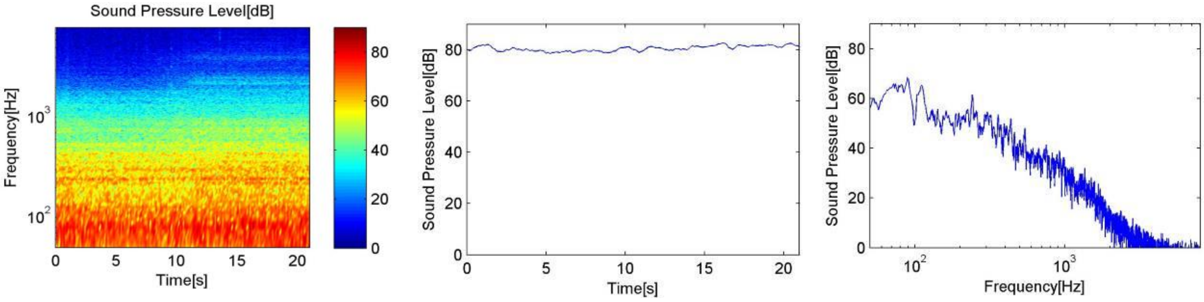


(a) backview

(b) frontview

Figure 3 – Arrangement of loudspeakers and microphones

The desired sound field is measured in a running vehicle with the same microphone array shown in figure 3. Sound pressure signal measureb by microphone #1 is given in figure 4 with different ways. The subplot (a) shows that the main components are in the low frequency part and the total time length of this signal is about 20 seconds. Subplot (b) indicates the sound pressure of this signal is stable as time changes. Subplot (c) is gotten using the data between 3s~4s and it gives an insight view of the magnitude of components in frequency domain.



(a) waterfall map (b) sound pressure level vs time (c) spectrum(3s~4s)

Figure 4 – Typical vehicle noise(measured by micphone #1)

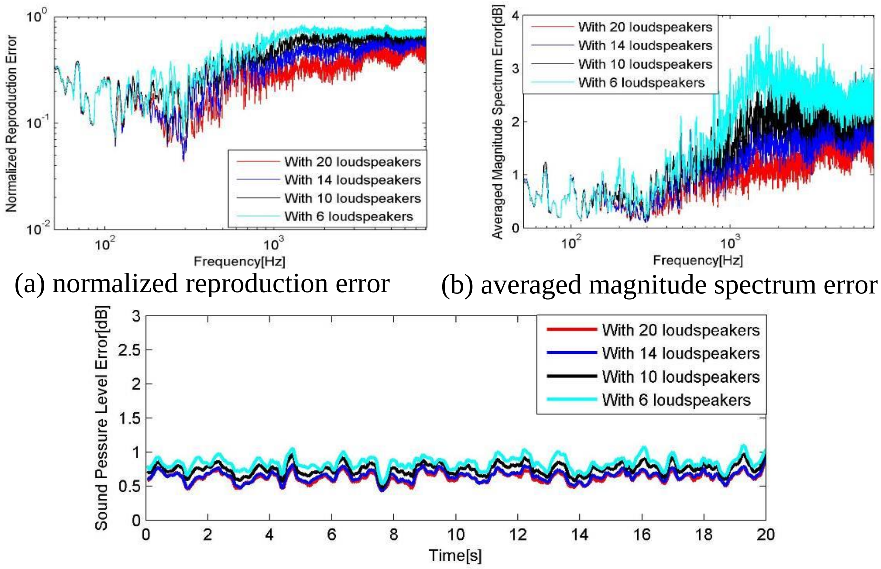
4.2 Computational Time

The frequency resolution Δf is 1Hz and the sampling frequency is 44100Hz so that the length of FIR filters I is 44100. The time length of the sound pressures is about 20s which means K is a very big value. Divide the sound pressures into short parts and each part has the length of 48200. The Computational time of $b_m(k)$ using two different methods is shown in the following table 1. The two methods are the direct method(calculate the convolution as definition) and the proposed fast method. The tests are performed using a laptop computer with core i5. The result shows that the fast method is about 191 times faster than direct method. If all the loudspeakers are used for reproduction, the convolution time of the fast method is about 206s while the direct method takes about 10 hours.

Table 1 – Computational time of $b_m(k)$

Method	Time[s]
Direct	0.54677
Fast	0.0028586

4.3 Results and Analysis



(a) normalized reproduction error (b) averaged magnitude spectrum error

(c) averaged sound pessure level error

Figure 5 – Reproduction error with different number of loudspeakers

The proposed three-stage method can reduce the number of loudspeakers and the loudsepaker number weight parameter δ can control the number of chosen loudspeakers in the sense that larger values of δ yield fewer loudspeakers. In the experiments the reproduction is performed using 14, 10 and 6 loudspeakers individually. The prorduction error is evaluated with the three standards introduced in subsection 2.2. The results are compared with the one all the loudspeakers are used and the comparisons are shown in figure 5. The three standards all tell that the more loudspeakers are used the better the production is. In subplot (b) and (c), the magnitudes of error in low frequencies are almost the same. Because the main components of the soundpressures are in the low frequency part, in other words the magnitudes of high frequency components are small, so that the magnitudes of error in high frequencies are relatively big in subplot (a) and (b). Subplot (c) declare that even if the number of loudspeakers is reduced by 6, sound pessure level error is almost the same. So it is reasonable to reduce 6 loudspeakers at least without a big influence on the reproduction accuracy.

5. Conclusions

It is shown that the proposed three-stage sound field reproduction have two main advantages: first, it provides a solution to choose the loudspeaker positions. Second, this method can implement the sound field reproduction efficiently with a high frequency precise which is decided by the frequency resolution.

REFERENCES

1. Gerzon MA. Periphony: With-height sound reproduction. J Audio Eng Soc. 1972 ;21(1):2-10.
2. Trevino J, Okamoto T, Iwaya Y, Suzuki Y. High order Ambisonic decoding method for irregular loudspeaker arrays. Proceedings of 20th International Congress on Acoustics; 23–27 August 2010; Sydney, Australia 2010.
3. Zotter F, Pomberger H, Noisternig M. Ambisonic decoding with and without mode-matching: A case study using the hemisphere. Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics; 6-7 May 2010; Paris, France 2010.
4. Berkhout AJ. A holographic approach to acoustic control. J Audio Eng Soc. 1988;36(12):977-95.
5. González A, López JJ. Fast transversal filters for deconvolution in multichannel sound reproduction. Speech and Audio Processing, IEEE Transactions on. 2001;9(4):429-40.
6. Fazi FM, Nelson PA. A theoretical study of sound field reconstruction techniques. 19th International Congress on Acoustics; 2-7 September 2007; Madrid,Spain 2007.
7. Gauthier P, Camier C, Gauthier O, Pasco Y, Berry A. Aircraft sound environment reproduction: Sound field reproduction inside a cabin mock-up using microphone and actuator arrays. Proceedings of Meetings on Acoustics; 2-7 June 2013;Montreal, Canada 2013.
8. Li S, Zheng S, Peng B, Lian X. Experimental research on 3D sound reproduction with reflection boundary. Proc INTER-NOISE 2012; 19-22 August, 2012; New York, USA 2012.
9. Kolundzija M, Faller C, Vetterli M. Reproducing sound fields using MIMO acoustic channel inversion. J Audio Eng Soc. 2011 2011-01-01;59(10):721-34.
10. Lilis GN, Angelosante D, Giannakis GB. Sound field reproduction using the lasso. Audio, Speech, and Language Processing, IEEE Transactions on. 2010 2010-01-01;18(8):1902-12.
11. Radmanesh N, Burnett IS. Wideband sound reproduction in a 2D multi-zone system using a combined two-stage Lasso-LS algorithm. 2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop; 17-22 June 2012; Hoboken, USA 2012.
12. Choi HG, Thite AN, Thompson DJ. Comparison of methods for parameter selection in Tikhonov regularization with application to inverse force determination. J Sound Vib. 2007;304(3):894-917.
13. Calvetti D, Morigi S, Reichel L, Sgallari F. Tikhonov regularization and the L-curve for large discrete ill-posed problems. J Comput Appl Math. 2000;123(1):423-46.
14. Orfanidis SJ. Introduction to signal processing. Prentice-Hall, Inc.; 1995.