three dimensional sound reproduction in Vehicles based on data mining techniques

Maosheng Zhang, Ruimin Hu, Lin Jiang, Xiaochen Wang

National Engineering Research Center for Multimedia Software
School of Computer Science
Wuhan University, Bayi road, Wuhan, China

1 Introduction

Sound systems for vehicle have been well researched by scientists and engineers. Akitoshi Yamada developed a sound reproduction system for vehicle using only a pair of loudspeakers in 1982[1]. The system comprised a transfer function, a delay circuit, and a reverberation circuit. With the help of these components, a surrounding sound system was implemented. Honda Motor designed a sound reproducing apparatus for vehicle in 1990[2]. The apparatus takes advantage of a acoustic duct and a loudspeaker placing in the duct. In 2003, Takeshi reproduced a required sound image for the specified seat with a sound system consisting of two loudspeakers for Vehicles[3]. In addition, sound systems using more than two loudspeakers are developed to generate surrounding ambiance acoustic effects[4][5]. A sound entertainment system for determined positions in a vehicle is proposed by David in 2007. This system provides ultrasonic waves and cancels the unwanted noise[6]. FORD Motor Company invented a multichannel sound reproduction system for vehicles and applied for a patent in 2017[5]. The embodiments mentioned in this patent composed of several loudspeakers, including a low-frequency loudspeaker or sub-woofer, placing in pillars, door frames and vehicle roof. Obviously, the sound reproduction system for entertainment in vehicle are well researched. Lots of patents about sound reproduction in vehicle are applied by vehicle companies to recreate acoustic environment[6][7][8][9]. However, sound spatial perception is far from satisfactory and acoustic virtual reality has not been implemented in vehicle. Three-dimensional (3D) sound reproduction system provides immersive perception about sound sources and thus enhances the sensation of reality[10][11][12]. It is necessary to reproduce 3D sound to enjoy realistic acoustic environment and sound events in vehicle.

The most there popular 3-D sound reproduction algorithms include Wave Field Synthesis(WFS), Ambisonics and Amplitude panning. WFS is based on Huygens principle and it is able to reproduce the whole sound field and thus the real sound immersion was recreated[13]. However, WFS method is not practical since there are too many loudspeakers required in WFS system. Ambisonics system reproduces the sound pressure at listening point in the center of spherical loudspeaker array. While, it is impossible to configure a spherical loudspeaker array in vehicle. Amplitude panning is a widely used sound reproduction technique due to its computational efficiency. Vector base amplitude panning (VBAP) is a popular sound reproduction technique to render the sound direction and distance. And thus VBAP is considered as a promising technique to recreate sound events[14]. Unfortunately, VBAP system requires a spherical loudspeaker array, which is not satisfied in vehicle.

Though there are several methods to reproduce sound field in physics or mathematics method, the intrinsic goal of sound field reproduction is to reproduce sound pressure at every point in the listening field. Obviously, this is a typical data mining problem. In this paper, a regression algorithm is proposed to reproduce 3D sound in a specific location, the driving seat for example. A sound pressure data set is set up at the specific listening point based on physical sound theory. The sound pressure is related to the number of loudspeakers, the location of loudspeakers, frequency of sound, the distances between listening point and loudspeakers. Since both inputs and outputs are known, a supervised learning model is built to demonstrate the relationship between received sound pressure and the all factors. The mean square error (MSE) shows the proposed regression model predicts sound pressure accurately.

The rest of this paper is organized as follows: the theory foundation, including 3D sound filed reproduction method and regression model, is introduced in the next section. The proposed method to reproduce the 3D sound in vehicle is developed in section (3) experiment is conducted in section (4). And section (5) concludes this paper.

2 Fundamental Theory

2.1 3D sound reproduction

A realistic 3D sound reproduction system needs to reproduce sound pressure at the listening point or every point in a listening area. The Fourier transform of sound pressure generated by a sound source is shown in equation (1) [12][15][16].

$$p(r,\xi) = G\frac{e^{-ik|r-\xi|}}{|r-\xi|}s(\omega) \tag{1}$$

where the parameters are explained as follows:
ξ: $\xi = (\xi_x\ \xi_y\ \xi_z)$ is sound source location;
e: a constant irrational number which is the base of natural logarithm;
i: imaginary unit;
k: wave number, which is equal to $\frac{2\pi f}{c}$;
f: sound frequency;
c: sound velocity;
G: a constant number which represents the proportionality coefficient between sound pressure at a unit distance from a loudspeaker and the input to the loudspeaker;
$s(t)$: sound signals in time domain;
$s(\omega)$: sound signals in frequency domain, i.e. Fourier transformation of sound signal $s(t)$;
$r = (x,\ y,\ z)$: listening point.

Alternatively, the sound pressure is also can be calculated in polar coordinates. Taking the listening point as an origin, sound pressure is shown in equation (2).

$$p(\omega) = G\frac{e^{-ik\sigma}}{\sigma}s(\omega) \tag{2}$$

where σ is the distance between sound source and listening point(the origin in polar coordinates).

It is concluded sound pressure is related to several factors after analyzing the sound pressure formula. In a multichannel reproduction system, the computational complexity grows with a exponential trend as the number of loudspeakers increases. However, whatever the 3D sound reproduction methods are, the ultimate goal of the reproduction is to recreate the sound pressure at listening point. As long as the volume of sound pressure data is large enough, data mining technology is a appropriate method to predict sound pressure.

2.2 Regression algorithm

Regression methods, popular techniques to predict system outputs, are well researched by scientists and engineers. Linear regression is one of the most well-known regression algorithm[17][18]. A linear equation is constituted by the product of a constant and a single variable with the power of either one or zero. A typical example of linear equation is shown in equation (3).

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + b = 0 \tag{3}$$

where $a_i, i = 1, 2, \dot{;} n$ is a coefficient and b is a constant.

if $a_2 = a_3 = \cdots = a_n = 0$, in other words, if there is only one variable, equation (3) is the simplest linear equation, which is written as shown in equation (4).

$$ax + b = 0 \tag{4}$$

where a, b are constants.

Linear regression is a linear approach to demonstrate the relationship between inputs $X = (x_1,\ x_2, \ldots,\ x_n)$ and output Y. In mathematics, the relationship is shown in equation (2.2)[19].

$$y = a_1x_1 + a_2x_2 + \cdots + a_nx_n + b \tag{5}$$

In data mining, given the known dataset X and Y, the linear regression is established as shown in equation (2.2).

$$y_i = a_1x_{i1} + a_2x_{i2} + \cdots + a_nx_{in} + b_i \tag{6}$$

In practice, matrix form is used to show the relationship, which is shown in equations (2.2) and (2.2).

$$y_i = x_i^T A + b_i \tag{7}$$

$$Y = X^T A + B \tag{8}$$

where $Y = (y_1, y_2, \ldots, y_n)$, $B = (b_1, b_2, \ldots, b_n)$, $A = (a_1, a_2, \ldots, a_n)$.

For given data X and Y, the parameters, i.e. coefficients vector A and constant vector B, are to be solved. As soon as the parameters are determined, equation (2.2) is able to be used as a predictor functions to predict Y according given data X.

Linear regression get widely practical uses in lots of areas such as engineering[20][21], biological[22], epidemiology[23], finance[21], economics[24], environmental science[25], computer vision[26]. Regression is also applied in acoustic science, especially in speaker recognition, emotion recognition, scenario recognition, audio event detection, audio information retrieval[27][28][29].

3 three dimensional sound reproduction algorithm in Vehicle

First of all, a sound pressure dataset is needed. In vehicle, several loudspeakers are placed around the vehicle. The sound system is a typical multichannel sound reproduction system. The sound pressure at listening point in a multichannel sound system is the superposition of all active loudspeakers, which is shown in equation (9).

$$p(\omega) = G\Big(\sum_{j=1}^{L} \frac{e^{-ik\sigma_j}}{\sigma_j}\Big)s_j(\omega) \tag{9}$$

where L is the number of loudspeakers, $s_j(\omega)$ is sound signal in frequency domain in the jth loudspeaker, $sigma_j$ is the distance between the ith loudspeaker and the origin.

The purpose of a 3D sound reproduction system in vehicle is to reproduce the same sound pressure at listening point in vehicle using several loudspeakers as the sound pressure the sound source generated. Given the received signal s(t) in time domain, we need to reproduce s(t) in loudspeaker array. In an amplitude panning system, the sound signal in the jth loudspeaker is assigned as $s_j(\omega) = w_i s(\omega)$, where w_j is the jth weight. And then, equation (9) is equivalently transformed to equation (10).

$$p(\omega) = G\Big(\sum_{j=1}^{L} \frac{e^{-ik\sigma_j}}{\sigma_j}\Big) w_j s(\omega) \tag{10}$$

Apparently, it is necessary to solve all the weights $w = (w_1, w_2, \ldots, w_L)$ in equation (10). Unfortunately, it is hard to get a accurate solution to maintain sound pressure. Since sound pressure is depend on several factors and w_i is ranged from 0 to 1, regression model is an efficient method to estimate sound pressure as soon as a large amount of sound pressure data is available.

Sound pressure dataset is created with parameters sound signal frequency f, the ith loudspeaker location $L_i = (\theta_i, \delta_i, \sigma_i)$, listening point location O and $W = (w_1, w_2, \ldots, w_L)$ based on equation (**??**), where θ_i, δ_i is azimuth and elevator of the ith loudspeaker respectivly, w_i ranges from 0 to 1 with step-size h. Let every w_i traverses all the possible values and we calculate the corresponding sound pressure, a big sound pressure dataset is created. The data size of this dataset is shown in equation (3).

$$V = f_b \times L^(1/h) \tag{11}$$

where f_b is the number of frequencies.

It is obvious that the smaller the step-size h is, the larger the dataset is, and the more accurate the regression model is. Assuming there are five loudspeakers in a vehicle and step-size $h = 0.01$, there are $7.8886 * 10^69$ sound pressure values for each frequency. When we design program algorithm, the generating of each w_i is randly chosen by a rand function rather than a fixed step-size since the weights in reality is barely uniform. There is another constraint for weight vector where every w_i is non-negative and the sum of all elements in W is 1. The sound pressure dataset generating algorithm is shown in algorithm (1).

Since a sound pressure dataset is created, a stepwise linear regression model is going to be established. Taking parameters f, $L_i = (\theta_i, \delta_i, \sigma_i)$, O and $W = (w_1, w_2, \ldots, w_L)$ as predictors and sound pressure p as response, a linear regression model is proposed. In order to protect against over-fitting, we also design cross-validation technique by partitioning the dataset into 5 folds and estimating accuracy on each fold. When the training and cross validation are finished, the prediction function is determined and can be used to predict new data. Figure (1) shows the flowchart of the developed sound reproduction regression model.

Algorithm 1 Framework of generating sound pressure dataset.

Input:

1: sound signal frequency f;
2: the loudspeaker location $L_i = (\theta_i, \delta_i, \sigma_i), i = 1, 2, \ldots, L$;
3: listening point location O;

Output:

4: a dataset SPdata composed of sound pressure p, weight vector W and the above mentioned inputs;
5: loop variable i=0;
6: generating weight vector;
7: i=i+1;
8: estimating sound pressure p;
9: pushing sound pressure and parameters into dataset SPdata ;
10: if $i \leq v$ then goto state (5); else goto state (11);
11: **return** SPdata;

Getting loudspeaker array parameters Generating sound pressure dataset partitioning dataset Training cross validation prediction

4 program and experiment

4.1 implementation

We utilize Matlab 2017 to train the proposed model and analyze the model performance. The number of the loudspeaker array is 4. We assume the four loudspeakers are placed at left front(L_1), right front(L_2), right behind(L_3) and left behind(L_4). The location coordinates of all loudspeakers are shown in table (1).

Table 1: Loudspeakers location

loudspeaker	azimuth	elevator
L1	320°	0°
L2	80°	0°
L3	135°	0°
L4	225°	0°

The proposed model is programmed on a notebook manufactured in 2011. The hardware parameters are shown in table (2).

It takes 33.3 seconds to finish the regression model and vital parameters of the stepwise linear regression model is shown in table (3).

Regression model parameters in table (3) indicate the proposed model for 3D

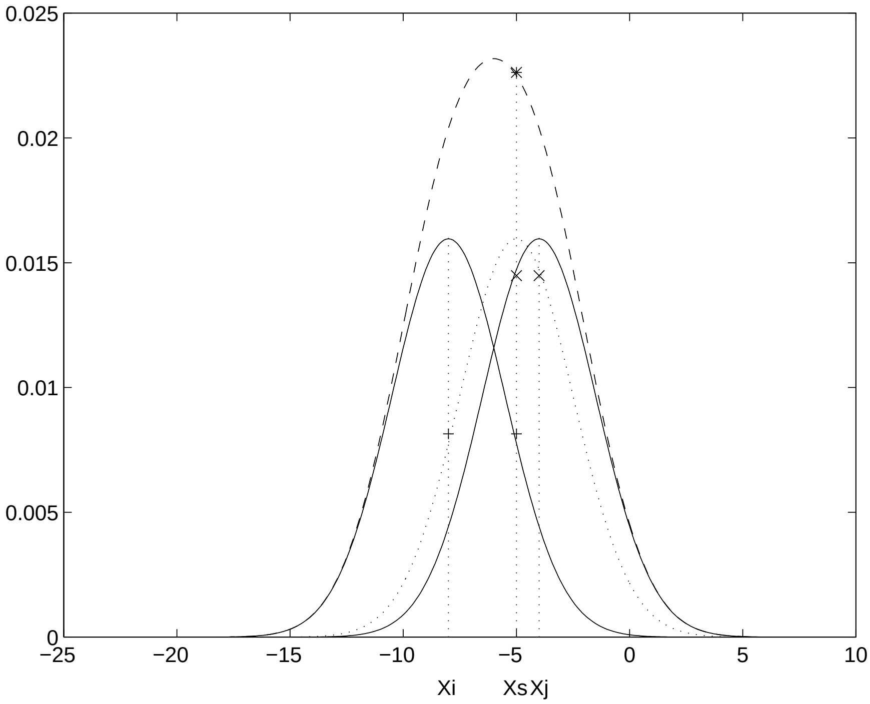


Fig. 1: The flowchart of the sound reproduction regression model in vehicle

Table 2: Notebook hardware parameters

Item	Parameter value
CPU	i7 3610QM, 2.30GH
RAM	4G
system	windows 10

sound reproduction in vehicle is accurate since both the MSE, RMSE are rather small. Due to the poor performance of the notebook, the prediction speed is slow and the training time is long.

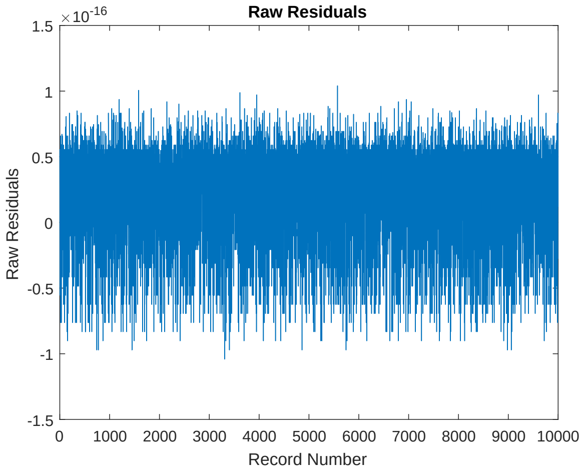
We also demonstrate the residuals in our model. The raw residuals, Pearson residuals, studentized residuals, standardized residuals is illustrated in figures (??)-(??) respectively. The tiny residuals shown int figures (2)-(3) concisely presents the accuracy on 3D sound reproduction.

4.2 experiment

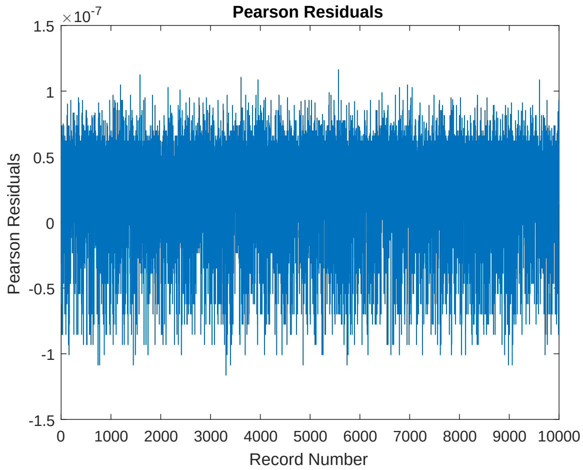
We conducted three experiments to evaluate the reproduction performance of the proposed technique. The result are compared with sound pressure in theory. The less residuals the proposed method are, the more accurate the reproduction method is. The loudspeaker configuration in the experiments is in shown in table

Table 3: Parameters in subjective experiments

Parameter	value
MSE	7.9968e-19
RMSE	8.9425e-10
R-squared	1
DFE	9996
SSE	7.9936e-15
SST	3.9626
SSR	3.9626
MAE	0
Prediction speed	130000 obs/sec
Training time	33.306 sec



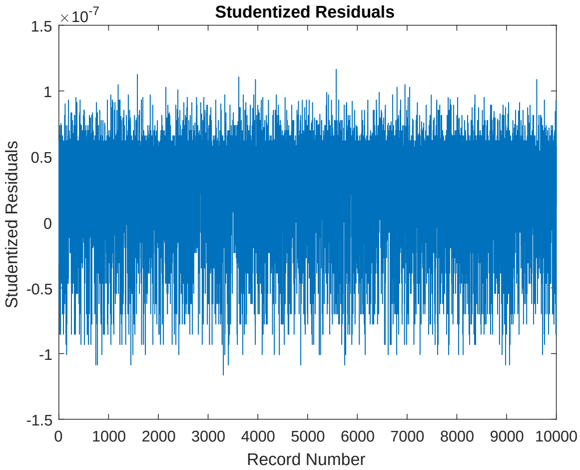
(a) Raw residuals



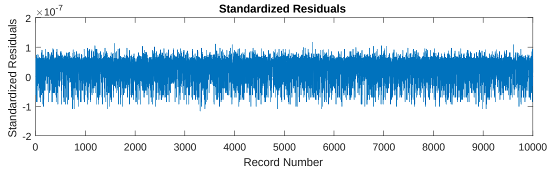
(b) Pearson residuals

Fig. 2: Raw residuals and c

(1). In order to verify the over fitting and under fitting performance, different data volume was under test. In experiment 1, we tested 100 sound pressure while in the other two experiments 1000 and 10000 computations are conducted. We randomly generated different weights in each computation. To show detailed comparison, raw residuals arerepresented graphically. Figures (4)-(6) show the residuals in experiments. The negligible residuals exhibite the proposed regression model accurately reproduce the sound pressure.



(a) Studentized residuals



(b) Standardize residuals

Fig. 3: Studentized residuals and Standardize residuals

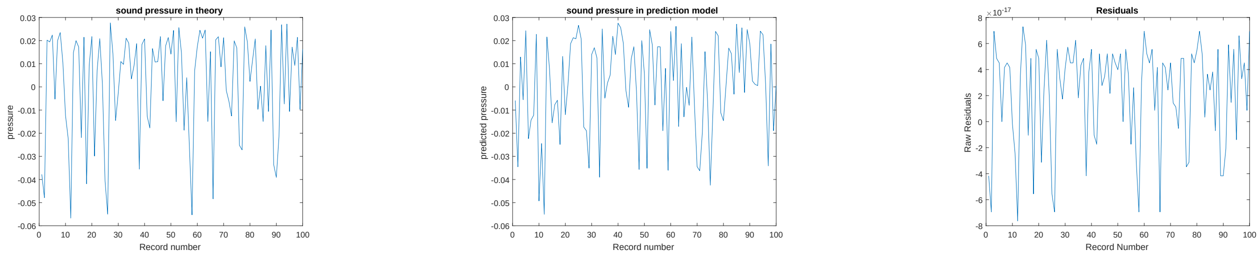


Fig. 4: prediction residuals in experiment (100 different weights)

5 conclusion

This study shows a 3D sound reproduction system for vehicle. A stepwise linear regression is proposed to reproduce the sound pressure. The residuals including craw residuals, Pearson Residuals, Studentized residuals and Standardize residuals are rather small. The experiments show whatever the data volume is, the proposed reproduction system accurately simulates the theoretical reproduction system.

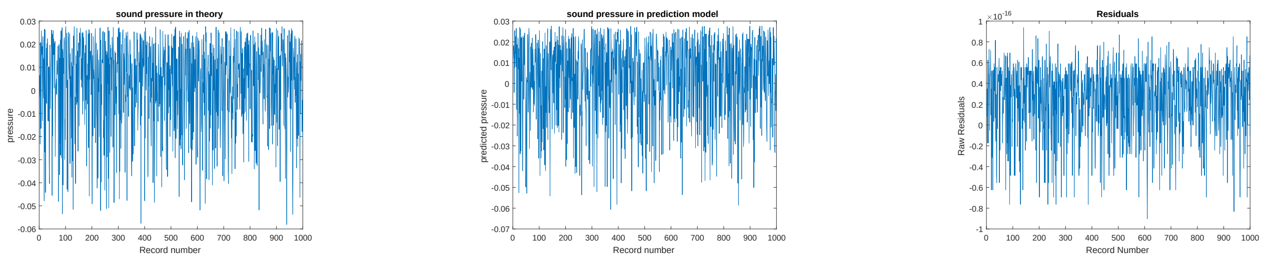


Fig. 5: prediction residuals in experiment(1000 different weights)

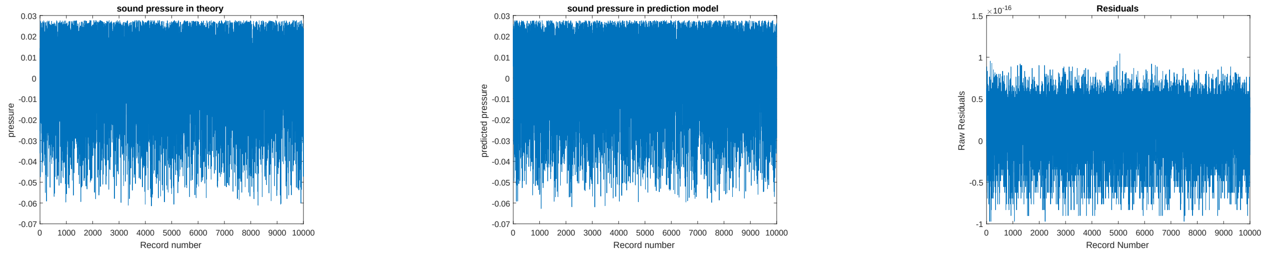


Fig. 6: prediction residuals in experiment(10000 different weights)

6 ACKNOWLEDGMENT

This research was supported by National Nature Science Foundation of China (No.61761044, 61762005, 61701194, 61471271, 61671335), National Nature Science Foundation of China (No. U1736206), Guangxi Colleges and Universityies Key Laboratory of Complex System Optimization and Big Data Processing(No.2016CSOBDP0005).

References

[1] A. Yamada, "Sound reproduction system for motor vehicle," *Journal of the Acoustical Society of America*, vol. 72, no. 3, pp. 1101–1101, 1982.

[2] K. Terai, S. Saiki, K. Murata, K. Satoh, Y. Kumura, Y. Nakama, M. Ogawa, and S. Obata, "Sound reproducing apparatus for use in vehicle," *Journal of the Acoustical Society of America*, vol. 88, no. 5, pp. 2518–2518, 1990.

[3] T. Enya, Y. Sato, and I. Aichi, "Sound output apparatus for an automotive vehicle," 2003.

[4] D. L. Clark and J. W. Steuber, "Vehicle audio system," *Journal of the Acoustical Society of America*, vol. 104, no. 6, p. 3155, 1998.

[5] M. C. A. S. N. M. Orellana, Fernando Mar (Benito Juarez, "Loudspeaker arrangement in a vehicle," Patent 9 725 047, August, 2017. [Online]. Available: http://www.freepatentsonline.com/9725047.html

[6] D. S., B. E., D. VallWendell, and C. Johnson, "Audio reception control arrangement and method for a vehicle," *Journal of the Acoustical Society of America*, vol. 155, no. 6, pp. 3151–3156, 2007.

[7] J. G. G Simon, "Method and apparatus for control of personal digital media devices using a vehicle audio system," Patent 10/870,424, August, 2005. [Online]. Available: https://academic.microsoft.com/#/detail/1600054936

[8] M. Vu, B. Boblett, N. Penke, K. Hsieh, and J. Nuxoll, "Vehicle audio system interface," Patent 13/671 660, August, 2014. [Online]. Available: http://www.freepatentsonline.com/y2014/0096003.html

[9] J. S. Gibson, "Vehicle human machine interface with auto-customization," Patent 14/672 698, August, 2015. [Online]. Available: http://www.freepatentsonline.com/y2015/0277735.html

[10] A. Gupta and T. D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, NO. 5, JULY 2011*, vol. 19, no. 5, pp. 1149–1160, July 2011.

[11] D. Comminiello, S. Cecchi, M. Scarpiniti, M. Gasparini, L. Romoli, F. Piazza, and A. Uncini, "Intelligent acoustic interfaces with multisensor acquisition for immersive reproduction," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 591–598, AUGUST 2015.

[12] M. Zhang, R. Hu, S. Chen, X. Wang, D. Li, and L. Jiang, "Spatial perception reproduction of sound events based on sound property coincidences," in *International conference on Multimedia and Expro.* IEEE, 2015.

[13] G. Firtha, P. Fiala, F. Schultz, and S. Spors, "Improved referencing schemes for 2.5d wave field synthesis driving functions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 5, pp. 1117–1127, MAY 2017.

[14] V. Pulkki, "Spatial sound generation and perception by amplitude panning techniques," Ph.D. dissertation, Helsinki University of Technology, 2001.

[15] A. Ando, "Conversion of multichannel sound signal maintaining physical properties of sound in reproduced sound field," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1467–1475, 2011.

[16] S. Wang, R. Hu, B. Peng, Y. Yang, and H. Wang, "Sound intensity and particle velocity based three-dimensional panning methods by five loudspeakers," in *IEEE International Conference on Multimedia and Expo (ICME 2013)*, 2013.

[17] J. M. Hahne, F. B. Jiang, H. R. Farina, F. C. M.-R. Muller, and L. C. Parra, "Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 2, pp. 269–27, 2014.

[18] P. C. Austin and E. W. Steyerberg, "The number of subjects per variable required in linear regression analyses," *Journal of Clinical Epidemiology*, vol. 68, no. 6, pp. 627–636, 2015.

[19] C. Lewis-Beck and M. S. Lewis-Beck, *Applied Regression: An Introduction.* SAGE Publications, 2015.

[20] J. Rice and van Zwet, "A simple and effective method for predicting travel times on freeways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 3, pp. 200–207, 2004.

[21] V. O. Ongore and G. B. Kusa, "Determinants of financial performance of commercial banks in kenya," *International Journal of Economics and Financial Issues*, vol. 3, no. 1, pp. 237–252, 2013.

[22] V. M. N. C. S. Vieira, J. Creed, R. A. Scrosati, A. Santos, G. Dutschke, F. Leito, A. H. Engelen, O. R. Huanel, M. L. Guillemin, and M. Mateus, "On the choice of linear regression algorithms for biological and ecological applications," *Vieira2016On*, vol. 11, no. 12, pp. 925–929, 2016.

[23] P. D. Sampson, M. Richards, A. A. Szpiro, S. Bergen, L. Sheppard, T. V. Larson, and J. D. Kaufman, "A regionalized national universal kriging model using partial least squares regression for estimating annual pm2.5 concentrations in epidemiology," *Atmospheric Environment*, vol. 75, pp. 383–392, 2013.

[24] E. Smith, *Modern Labor Economics (10th international ed.).* London: Addison-Wesley, 2008.

[25] W. W. Piegorsch and A. J. Bailer, *Analyzing environmental data.* John Wiley & Sons, 2005.

[26] X. Chai, X. Chai, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1716–1725, 2007.

[27] A. Khasanova, J. Cole, and M. Hasegawa-Johnson, "Detecting articulatory compensation in acoustic data through linear regression modeling," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pp. 925–929, 1 2014.

[28] Y. Zhao, J. Li, J. Xue, and Y. Gong, "Investigating online low-footprint speaker adaptation using generalized linear regression and click-through data," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 4310–4314.

[29] Y. A. Chen, J. C. Wang, Y. H. Yang, and H. Chen, "Linear regression-based adaptation of music emotion recognition models for personalization," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 2149–2153.