PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Soundscape reproduction using headphones for web-based listening test

Anugrah Sabdono SUDARSONO[1]; Joko SARWONO[1]

[1] Kelompok Keahlian Fisika Bangunan Institut Teknologi Bandung, Indonesia

ABSTRACT

One of the methods to analyse a soundscape is by reproducing the acoustic environment. The acoustic environment reproduction usually conducted using multi-speaker systems in a laboratory. The experiment usually conducted individually resulting limited respondent of the experiment. A web-based experiment is one of the alternatives to do the soundscape study. By using the internet, the experiment can be done easier with more respondents.

The experiment using the internet usually conducted using an online questionnaire which can play a stereo sound or a video. In this study we are testing the validity of XY, Mid-Side, and Head Tracking decoders to reproduce acoustic environments using headphones. The XY and Mid-Side decoders were used to decode B-Format signals to stereo signal which can be played directly by the online questionnaire. The head tracking system is used to represent the experiment conducted using Youtube 360 of Facebook 360.

Analysis based on semantic score indicates that the listening test conducted using stereo decoders can gives the similar score with the listening test conducted using ambisonic reproduction system. Further analysis using semantic differential analysis shows that the experiment conducted using Mid-Side decoders can deliver the same soundscape dimensions with the in-situ experiment.

Keywords: Soundscape Reproduction, Headphones, Internet, Stereo Decoder

1. INTRODUCTION

Soundscape analysis is conducted using three methods: in-situ experiment, laboratory experiment and recalled in memory (1). The in situ experiment gives the best result in term of the validity of the experiment although it has the problem in the repeatability of the research (2).

The laboratory experiment has become another alternative for soundscape experiment. It usually conducted in three methods: by reproducing the acoustic environment (3), by simulating the acoustic environment (4,5), and by composing the acoustic environment (6,7). Those three methods implement ambisonic reproduction system to reproduce the sound in the lab.

Soundscape experiment using soundscape reproduction has been validated using semantic categorisation (8,9) and semantic differential analysis (3,10). The reproduction system verified by comparing the response in the laboratory condition to the original soundscape. Using semantic categorisation, Guastavino et al. compare stereo reproduction, two-dimensional ambisonic system and three-dimensional ambisonic system and suggest that the two dimensional ambisonic system is suitable for acoustic environment reproduction (8).

Another study also compares the soundscape dimension from ambisonic reproduction system. Davies et al. reproduce urban acoustic environment using three dimensional ambisonic (10) resulting the same dimension from Kang's experiment(11): Relaxation, Dynamic, Communication, and Space. Sudarsono et al. use the same method to compare the soundscape dimension in situ and in the laboratory experiment using two-dimensional ambisonic system resulting three soundscape dimension, which is consistent on both of the experiment: Relaxation, Dynamic, and Communication (12).

The laboratory experiment has an advantage in the repeatability of the experiment. However, since the sonic environment needs to be reproduced using the many speakers, a room with acoustic treatment is needed to build the system. The experiment must also be conducted individually in the room where the system is installed.

Another alternative of conducting soundscape research is by using headphones (13–16). The study

[1] anugrah@tf.itb.ac.id

using headphones has some advantages: the experiment can be conducted everywhere, the participation of the study can get higher than the laboratory experiment, and the system set up is simpler than the experiment using ambisonic reproduction system. Furthermore, the soundscape experiment using headphone open the possibility to do the experiment using the internet using online questionnaire (17).

There are two ways to reproduce the acoustic environment by the internet: stereo reproduction, and virtual reality. The stereo reproduction is a common method to reproduce sound in an online questionnaire and the virtual reality reproduction can be done using Youtube 360 or Facebook 360.

In this study, comparison between XY, Mid-Side, and Head Tracking is done in order to understand which decoders are suitable to do web-based listening test using headphone. The XY and Mid-Side decoders are used to decode B-Format signals to stereo signal which can be played directly by the online questionnaire. The head tracking system is used to represent the experiment conducted using Youtube 360 of Facebook 360.

2. METHODS

2.1 Acoustic Environment Recording

The A4 standard paper size will be used for all Conference papers with all margins set to 25 mm. This setting and other settings are automatically set if this template is used in Microsoft Word.

Please do not insert any page numbers. Do not include anything in the headers or footers except for the footnotes for e-mail addresses on the first page of the manuscript.

Four acoustic environment recordings were recorded in a stationary position using Soundfield ST-250 microphone and Roland R-44 digital recorder in Manchester city centre area. The entire acoustic environment is recorded in B-Format which consist of four outputs (W, X, Y, and Z). The location and description of the recording used in the experiment are shown in Table 1.

Table 1-Description of recording's locations

Locations	Descriptions	Dominant Sound Sources
Market Street	Outdoor shopping lane surrounded by two floors shopping building; selected because the space represents urban soundscape without traffic noise.	Hubbub and live music
St Ann Square	Pedestrianized square; away from traffic; mixed-use building around the space; two natural sound objects appear: water fountain and birds chirping; selected because this space has natural sound objects	Water fountain
Piccadilly Garden	Open public space; main bus and tram stops; selected because the space represents urban soundscape with traffic noise;	Live music, hubbub, and traffic noise
Food Market at Piccadilly Garden	Open public space; main bus and tram stops; busy food market; selected because in this space, the traffic noise is masked by the sound of hubbub.	Hubbub, live music, and traffic noise

2.2 Experiment

The experiment was conducted to analyse the soundscape reproduced by a different decoders, which could be implemented in web-based listening test. Three decoders were used in this experiment: Mid-Side decoder, XY decoder and head tracking decoder. The Mid-Side decoder used in this study is the Voxengo MSED VST Plug-in (18), the XY decoder used in this study is Soundfield SurroundZone2 VST Plug-in(19), and the head tracking decoder used Nx Ambisonic head tracker system from Waves (20). The sound level of reproduction is calibrated to the level 9.5 dB lower than the actual sound level.

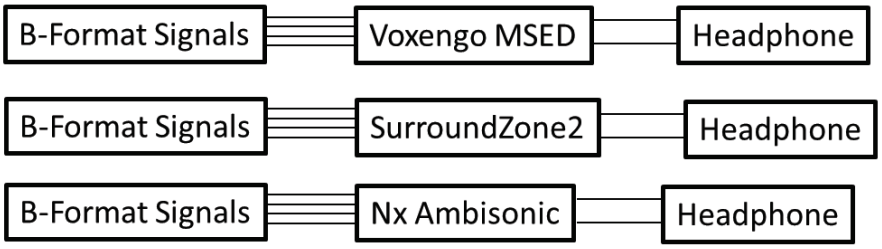
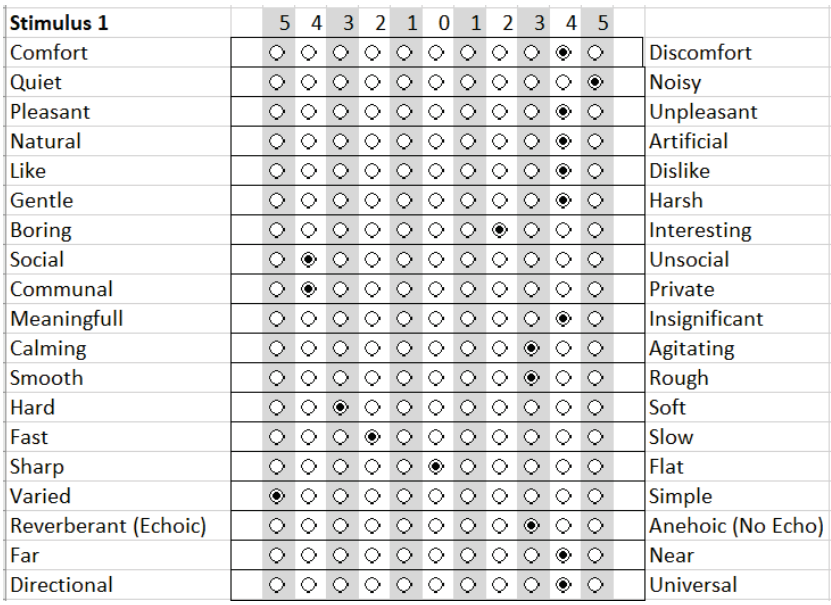


Figure2 -Signal Decoding implemented in the experiment

The experiment was conducted using AKG K550 headphones in a quiet room. Fifteen participants (eight males and seven females, average age: 24.8, SD: 5.3) joined the experiment voluntarily. The entire participants were bachelor, master, and post-graduate students of Engineering Physics Department of Institut Teknologi Bandung. The experiment was conducted individually and last for about 45 minutes.

The entire participant listened to four recordings (decoded using three decoders) through the headphones and requested to fill the semantic scales made by Microsoft Excel as shown in Figure 3.



Stimulus 1	5	4	3	2	1	0	1	2	3	4	5	
Comfort	○	○	○	○	○	○	○	○	○	●	○	Discomfort
Quiet	○	○	○	○	○	○	○	○	○	○	●	Noisy
Pleasant	○	○	○	○	○	○	○	○	○	●	○	Unpleasant
Natural	○	○	○	○	○	○	○	○	○	●	○	Artificial
Like	○	○	○	○	○	○	○	○	○	●	○	Dislike
Gentle	○	○	○	○	○	○	○	○	○	●	○	Harsh
Boring	○	○	○	○	○	○	○	●	○	○	○	Interesting
Social	○	●	○	○	○	○	○	○	○	○	○	Unsocial
Communal	○	●	○	○	○	○	○	○	○	○	○	Private
Meaningfull	○	○	○	○	○	○	○	○	○	●	○	Insignificant
Calming	○	○	○	○	○	○	○	○	●	○	○	Agitating
Smooth	○	○	○	○	○	○	○	○	●	○	○	Rough
Hard	○	○	●	○	○	○	○	○	○	○	○	Soft
Fast	○	○	○	●	○	○	○	○	○	○	○	Slow
Sharp	○	○	○	○	○	●	○	○	○	○	○	Flat
Varied	●	○	○	○	○	○	○	○	○	○	○	Simple
Reverberant (Echoic)	○	○	○	○	○	○	○	○	●	○	○	Anehoic (No Echo)
Far	○	○	○	○	○	○	○	○	○	●	○	Near
Directional	○	○	○	○	○	○	○	○	○	●	○	Universal

Figure 3-The interface of questionnaire

The semantic scales are the scale used in the previous study about the validation of soundscape reproduction using ambisonic reproduction system (12). The scale was developed based on the study conducted by Kang et al. (21).

The calibration of the reproduction system was done using Dummy Head GRAS 45BB KEMAR Head and Torso to the level that represents -9.5 dB below the actual level. The calibration concept is based on the previous study about the effect of the sound level adjustment to the perception of the acoustic environment reproduction (12). This method has been validated in the previous study (22)

3. ANALYSIS

The effect of stereo decoder variation to the soundscape is analysed using two methods: score comparison and principal component analysis. The result of this study then compared to the previous experiment result using two-dimensional ambisonic system (12).

The first score is analysed using Analysis of Variance (ANOVA) between the score of semantic scales from four recordings reproduced using three stereo decoders. ANOVA shows that the difference between three stereo decoders (XY, Mid-Side, and Head Tracking) in general are not significant. Only the score of the scale of quiet from Piccadilly Garden (sig. 0.024), the scale of hard from Food Market (sig. 0.011) and the scale of directional from Food Market (sig. 0.032) indicate significant difference due to the stereo decoder variation.

Further analysis is conducted by comparing the average value of the perception according to the decoder variation with the result from the perception of the same acoustic environment reproduced by two-dimensional ambisonic reproduction. The mean comparison is shown in Figure 4. In general, the perception of the headphones reproduction and ambisonic reproduction is similar.

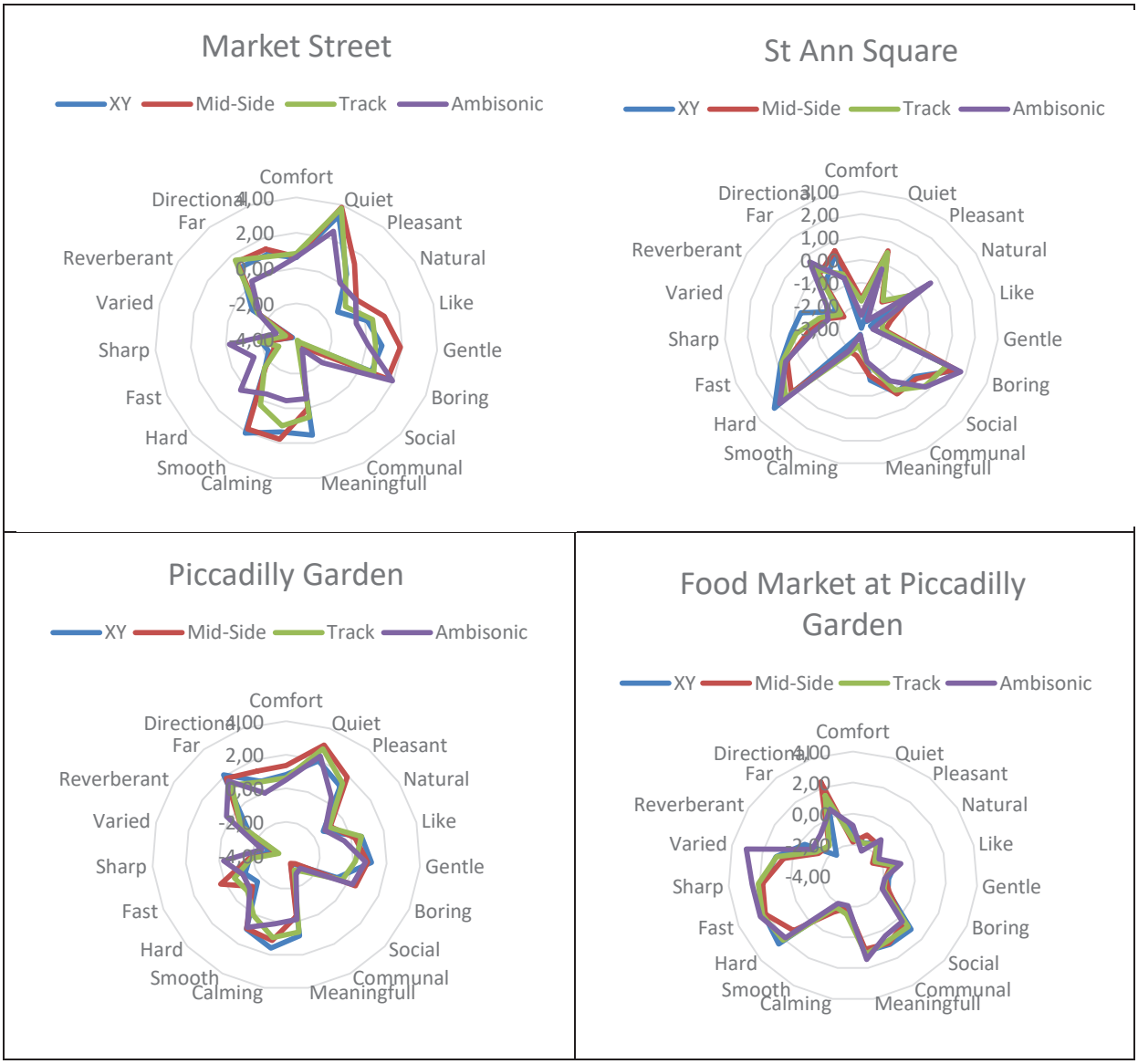


Figure 4-The average rating of soundscape dimension for different decoders

The semantic differential analysis is implemented to analyse the soundscape dimensions from the sonic environment recording. The Principal Component Analysis (PCA) from XY decoder experiment shows three reliable (Cronbach Alpha > 0.7) soundscape dimensions as shown in Table 2:

- The dimension related to the perception of dynamic and communication (33%). The dimension is represented by the semantic scale of Quiet, Gentle, Social, Communal, Calming, Smooth, Hard, Fast, Sharp, Varied, and Far.
- The dimension related to the perception of relaxation (21%). The dimension is represented by the semantic scale of Comfort, Pleasant, Natural, Like, Gentle, Calming.
- The dimension related to the perception of meaningful (10%). The dimension is represented by the semantic scale of Boring, and Meaningful.

The soundscape dimensions from the experiment using XY decoders shows a different result from the in-situ experiment and the laboratory experiment using ambisonic system (3). The in-situ experiment shows three reliable soundscape dimensions: Relaxation (24%), Dynamic (14%), and Communication (11%). The laboratory experiment using the ambisonic system also shows three

reliable soundscape dimensions: Relaxation (32%), Dynamic (18%), and Communication (12%).

The first dimension in the experiment using XY decoder is a combination of the dimension of dynamic and communication in the in-situ experiment. This dimension also represents the perception related to the sound object. The result indicates that the participant mainly relates the perception to the sound reproduced by the headphones and not the emotional aspect.

The experiment using the XY decoder shows that although the ANOVA shows the variation of different headphones is not affecting the perception much, the PCA shows different soundscape dimension. This result indicates that people react differently compared to the in-situ condition.

Table 2-Principal Component Analysis result of XY, Mid-Side, and Head Tracking Decoders

	Component			
	33%	21%	10%	7%
PCA XY Decoder (N= 84, KMO index 0.762, Bartlett's test of sphericity sig. 0.000)				
Comfort		.794		
Quiet	-.764			
Pleasant		.847		
Natural		.680		
Like		.869		
Gentle	-.610	.533		
Boring			-.875	
Social	.815			
Communal	.828			
Meaningful			.832	
Calming	-.573	.691		
Smooth	-.737			
Hard	.761			
Fast	.723			
Sharp	.755			
Varied	.758			
Reverberant				.752
Far	-.672			
Directional				.668
Cronbach's Alpha	0.93	0.91	0.75	0.28

	Component				
	34%	13%	12%	8%	7%
PCA Mid-Side Decoder (N= 84, KMO index 0.781, Bartlett's test of sphericity sig. 0.000)					
Comfort	.830				
Quiet	.736				
Pleasant	.847				
Natural				-.716	
Like	.869				
Gentle	.873				
Boring		-.610			
Social			.771		
Communal			.756		
Meaningfull				.632	
Calming	.797				
Smooth	.779				
Hard	-.757				
Fast		.525			
Sharp		.661			
Varied			.545		
Reverberant					.902
Far		-.817			
Directional					-.506
Cronbach's Alpha	0.94	0.73	0.83	0.22	0.29

	Component				
	23%	21%	15%	9%	8%
PCA Head Tracking Decoder (N= 84, KMO index 0.718, Bartlett's test of sphericity sig. 0.000)					
Comfort		.819			
Quiet	-.591	.552			
Pleasant		.829			
Natural		.703			
Like		.855			
Gentle			.761		
Boring	-.530				
Social	.651				
Communal	.649				
Meaningful					.756
Calming		.572			
Smooth			.831		
Hard	.604		-.566		
Fast	.702				
Sharp	.868				
Varied	.896				
Reverberant					.668
Far				-.701	
Directional				.847	
Cronbach's Alpha	0.88	0.88	0.86	0.57	0.38

The PCA from the experiment using Mid-Side decoder shows three reliable (Cronbach's Alpha>0.7) soundscape dimensions as shown in Table 2:

- The dimension related to the perception of relaxation (34%). The dimension is represented by semantic scale of Comfort, Quiet, Pleasant, Like, Gentle, Calming, Smooth, Hard.
- The dimension related to the perception of dynamic (13%). The dimension is represented by semantic scale of Boring, Fast, Sharp, and Far.
- The dimension related to the perception of communication (12%). The dimension is represented by the semantic scale of Social, Communal, and Varied.

The soundscape dimension from the experiment using Mid-Side Decoder shows the similar soundscape dimension to the in-situ and the laboratory experiment using ambisonic system: Relaxation, Dynamic, and Communication. Furthermore, the order of the dimension also consistent with the in situ and the validated laboratory experiment.

The dimension of Relaxation in this experiment shows the similar variance to the laboratory experiment using ambisonic system (38% in the validated laboratory experiment and 34% in our experiment. The dimension of dynamic variance shows 5% difference in the variance of the validated laboratory experiment. The dimension of communication shows the same variance to the dimension of communication in the validated laboratory reproduction experiment.

The Mid-Side experiment shows that the reproduction system can reproduce the similar soundscape dimension to the in-situ condition and the validated laboratory experiment. This result shows that the Mid-Side Reproduction system is adequate to be used in soundscape experiment.

The experiment using head tracking decoder shows three reliable (Cronbach's Alpha > 0.7) soundscape dimensions as shown in Table 2:

- The dimension related to the perception of dynamic and communication (23%). The dimension is represented by the semantic scale of Quiet, Boring, Social, Communal, Hard, Fast, Sharp, and Varied.
- The dimension related to the perception of relaxation (21%). The dimension is represented by the semantic scale of Comfort, Quiet, Pleasant, Natural, Like, and Calming.
- The dimension related to the perception of smoothness (10%). The dimension is represented by the semantic scale of Gentle, Smooth, and Hard.

The soundscape dimension from the experiment using head tracking decoder shows the similar result to the experiment using XY decoder. The result shows different perception to the in-situ experiment and the validated laboratory experiment. The result of this experiment is consistent with the experiment using XY decoders, which are dominated by the perception related to the rating of the sound objects.

Many study regarding soundscape dimension' study has determined several soundscape dimensions. Although there is some soundscape dimensions difference, the first dimension always represents the emotional aspect: relaxation, pleasantness, calmness, and emotional assessment. The sonic environment reproduction using headphones and Mid-Side decoder shows the same characteristic with the other soundscape dimension. Furthermore, the soundscape dimensions are consistent with the in-situ experiment conducted by Kang et al. (11) and Sudarsono et al. (3).

4. CONCLUSIONS

The study has compared three stereo decoders for soundscape reproduction using headphones: XY decoder, Mid-Side Decoder and Head Tracking Decoder. This study shows that the Mid-Side decoder is suitable for soundscape reproduction based on the semantic scales score comparison to the validated reproduction system. Furthermore, analysis using semantic differential analysis shows the same soundscape dimensions with the in-situ experiment and the validated soundscape reproduction: relaxation, dynamic, and communication. Hence, this result shows that the soundscape experiment using headphones using Mid-Side Decoder can be a good alternative in soundscape study especially the experiment conducted using the internet.

ACKNOWLEDGEMENTS

This research is funded by Institut Teknologi Bandung

REFERENCES

1. Aletta F, Kang J, Axelsson Ö. Landscape and Urban Planning Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landsc Urban Plan [Internet]. 2016;149:65–74. Available from: http://dx.doi.org/10.1016/j.landurbplan.2016.02.001
2. Adams, Bruce, Davies, Cain, Jennings, Carlyle, et al. Soundwalking s Methodology for Understanding Soundscapes. In: Institute of Acoustics Spring Conference 2008 [Internet]. Reading: the Institute of Acoustics; 2008. Available from: http://usir.salford.ac.uk/2461/
3. Sudarsono AS, Lam YW, Davies WJ. The effect of sound level on perception of reproduced soundscapes. Appl Acoust [Internet]. 2016;110:53–60. Available from: http://dx.doi.org/10.1016/j.apacoust.2016.03.011
4. Hong JY, Jeon JY. Designing sound and visual components for enhancement of urban soundscapes. J Acoust Soc Am [Internet]. 2013 Sep [cited 2013 Dec 9];134(3):2026–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23967935
5. Sudarsono AS, Lam YW, Davies WJ. The Validation of Soundscape Environment Simulator to Determine the Relationship between Sound Objects and the Perception of Soundscape. Acta Acust united with Acust. 2017;
6. Sudarsono AS, Lam YW, Davies WJ. Soundscape Simulator as A Tool to Predict the Perception of Acoustic Environment. Inter Noise 2017. :1073–8.
7. Sudarsono AS, Lam YW, Davies WJ. Soundscape Perception Analysis Using Soundscape Simulator. In: Inter Noise 2016. Hamburg; 2016. p. 6868–75.
8. Guastavino C, Katz BFG, Polack J, Levitin DJ. Ecological Validity of Soundscape Reproduction. Acta Acoust United with Acoust [Internet]. 2005;91(September 2004):333–41. Available from:

http://www.ingentaconnect.com/content/dav/aaua/2005/00000091/00000002/art00015
9. Guastavino C, Katz BFG. Perceptual evaluation of multi-dimensional spatial audio reproduction. J Acoust Soc Am [Internet]. 2004 [cited 2013 Nov 12];116(2):1105. Available from: http://link.aip.org/link/JASMAN/v116/i2/p1105/s1&Agg=doi
10. Davies WJ, Bruce NS, Murphy JE. Soundscape Reproduction and Synthesis. Acta Acust united with Acust [Internet]. 2014 Mar 1 [cited 2014 Mar 4];100(2):285–92. Available from: http://openurl.ingenta.com/content/xref?genre=article&issn=1610-1928&volume=100&issue=2&spage=285
11. Kang. Urban Sound Environment. 1st ed. Oxon: Taylor and Francis; 2007. 88-89 p.
12. Sudarsono AS, Lam YW, Davies WJ. The effect of sound level on perception of reproduced soundscapes. Appl Acoust [Internet]. 2016 Sep;110:53–60. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0003682X16300482
13. Axelsson Ö, Nilsson ME, Berglund B. A principal components model of soundscape perception. J Acoust Soc Am [Internet]. 2010 Nov [cited 2013 Nov 11];128(5):2836–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21110579
14. Cain R, Jennings P, Poxon J. The development and application of the emotional dimensions of a soundscape. Appl Acoust [Internet]. 2013 Feb [cited 2013 Nov 12];74(2):232–9. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0003682X1100301X
15. Hall DA, Irwin A, Edmondson-Jones M, Phillips S, Poxon JEW. An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes. Appl Acoust [Internet]. 2013 Feb [cited 2013 Nov 12];74(2):248–54. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0003682X11000740
16. Parizet E, Hamzaoui N, Jacquemoud J. Noise assessment in a high-speed train. 2002;63:1109–24.
17. Science E. The Development of a Web-Based Urban Soundscape Evaluation System The Development of a Web-Based Urban Soundscape Evaluation System. 2018;0–11.
18. Vanee A. Voxengo MSED [Internet]. Available from: http://www.voxengo.com/product/msed/
19. Soundfield SurroundZone2 [Internet]. Available from: http://www.soundfield.com/products/surroundzone2
20. Nx – Virtual Mix Room over Headphones [Internet]. Available from: https://www.waves.com/plugins/nx#introducing-nx-virtual-mix-room
21. Kang J, Zhang M. Semantic differential analysis of the soundscape in urban open public spaces. Build Environ [Internet]. 2010 Jan [cited 2013 Nov 6];45(1):150–7. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0360132309001309
22. Sudarsono AS, Sarwono J. SOUND LEVEL CALIBRATION ON SOUNDSCAPE REPRODUCTION USING HEADPHONE. :1–7.