A Framework for Reporting Spatial Attributes of Sound Sources

Susanne Westphal, Michael Schoeffler and Jürgen Herre
International Audio Laboratories Erlangen, A joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Am Wolfsmantel 33, 91058, Erlangen, Germany

Abstract. Recently, a software package called webMUSHRA was published that allows to set up web-based auditory experiments without the need of programming expertise. This work proposes a framework that extends webMUSHRA by reporting methods for spatial attributes of sound sources. These spatial attributes are perceived width, height, depth, and location of sound sources, as well as apparent/auditory source width (ASW), and listener's envelopment (LEV). As newer multi-channel formats are not limited to the horizontal plane anymore, our framework explicitly supports reporting of spatial attributes in all three dimensions.
Reporting ASW and LEV has been evaluated by a listening test in which participants indicated their perceived ASW and LEV of different stimuli. The stimuli consisted of different source signals (cello, noise, and snare) simulated in different acoustical environments (anechoic chamber, studio room, small hall, and large hall). The results of the listening test were compared to the results of a previously conducted listening test which contained the same set of stimuli, but was carried out by other researchers using a different framework and procedure. The results analysis indicated that the proposed framework leads to comparable results.

1 Introduction and Related Work

Nowadays, either stereo or 5.1 surround sound systems are typically installed in living rooms. For commercial applications, e.g., sound systems in cinemas, there is a trend in developing new multi-channel formats towards adding additional loudspeakers in horizontal and especially elevated positions. A prominent example is the 22.2 multi-channel format that features 24 loudspeakers arranged in three horizontal layers [1]. The question might arise whether formats with more than two channels bring significant benefits that justify the development towards a larger number of channels. Answering this question might be dependent on the individual preferences of a listener. However, a study by a Silzle et al. [2] showed that "more channels" increases the overall sound quality. Furthermore, a study by Schoeffler et al. indicated that 5.1 surround sound systems significantly improve the overall listening experience compared to mono- and stereo systems [3].

Evaluating multi-channel formats with a large number of loudspeakers as well as elevated loudspeakers asks for new evaluation methods since many previous methods are not able to evaluate new benefits that are brought by these advanced multi-channel formats. For example, 5.1 surround sound systems consist of five channels in horizontal positions and one subwoofer. Therefore, when evaluating 5.1 surround systems, many evaluation methods have focused mainly on horizontal attributes, such as the width, which is defined as the perceived extent of a sound source in horizontal direction [4]. By developing multi-channel systems with elevated loudspeakers, the height, defined as the perceived extent of a sound source in vertical direction [4], becomes more important. In some cases, also the depth, which is the perceived extent of a sound source in radial direction [4], might be subject to evaluation. In localization tests, in which participants report spatial locations of stimuli, graphical user interfaces are often limited to the horizontal plane. However, especially in advanced multi-channel formats, listeners are fully surrounded by loudspeakers, which means evaluation methods must also support reporting spatial attributes in all three dimensions. To this end, we implemented a framework for reporting "three-dimensional" spatial attributes in listening tests. The framework has been integrated into webMUSHRA, an open-source listening test software, that enables to set up listening tests without the need of programming expertise [5]. The contributed framework enables to report width, height, depth, and location of stimuli. In addition, reporting the apparent/auditory source width (ASW) and listener envelopment (LEV) is also supported. The ASW is defined by Morimoto as the width of the sound image fused temporally and spatially with a direct sound's image [6]. LEV is defined by Norcross et al. as the listener's sense of being surrounded or enveloped by sound [7].

A subset of the framework, reporting ASW and LEV, was evaluated by means of a listening test. The results of the listening test are compared to the results of a previously conducted listening test published by Merimaa and Hess in 2004 [8]. The previously conducted listening test had a slightly different procedure, but the same reporting method and stimuli were used as in our listening test.

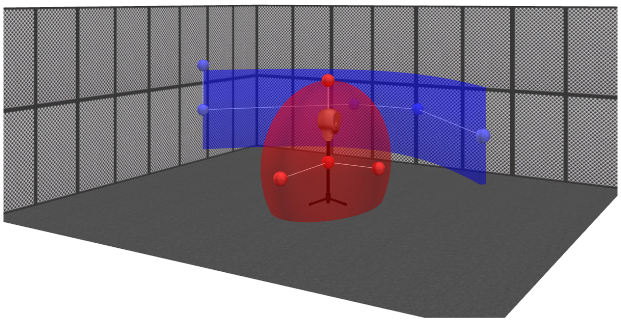

Figure 1 The framework's 3D view showing a scene with a blue extruded curve (for reporting width, height, and depth) and a red ellipsoid (for reporting LEV).

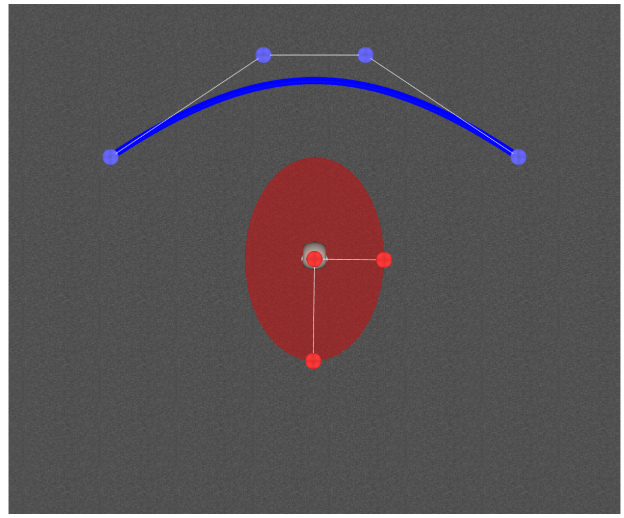

Figure 2 A 2D top view of a scene with a blue curve (for reporting ASW) and a red ellipse (for reporting LEV).

2 Proposed Reporting Methods

2.1 Scene

A scene is a virtual representation of a listening room in which a participant reports the perception of spatial attributes of sound sources. The proposed framework allows to visualize a scene by so-called "3D views" or "2D views". When a 3D view is used, the scene is captured from a virtual camera at an arbitrary position (see Figure 1). 2D views show a scene by an orthographic camera that is placed at a fixed position (top, side, and front view). A scene with a 2D top view is depicted in Figure 2. The differences between 3D and 2D views for localization listening tests have been evaluated by a previous study [9]. The results of the study show that the type of view has no significant effect on the reported location of a sound source. However, participants of the study indicated that 2D views are more convenient than 3D views.

2.2 Width, Height, Depth

A blue extruded curve represents the width, height, and depth of a sound source (see Figure 1). The appearance of the blue extruded curve can be changed by moving the blue control points. The adjustment of width and curvature is based on a cubic Bézier curve and can be controlled by four control points. Height and depth of the blue extruded curve are adjusted by two additional control points.

2.3 Localization

The framework's method for reporting locations of sound sources has already been published and evaluated [9]. A colored sphere is used to indicate the location of a sound source. Furthermore, if required, the size of the sphere can be changed to report the perceived broadness of a sound source.

2.4 Apparent/Auditory Source Width (ASW) and Listener Envelopment (LEV)

Since the ASW is similarly defined as the width of a sound source, the ASW is reported with the extruded blue curve without the control points for the height and depth (see Figure 2). As mentioned before, the width of the blue extruded curve is based on a cubic Bézier curve. The reason for choosing a cubic Bézier curve is that it can also be shaped as an arc, which has already been used by other researchers to represent the ASW of a sound source [8].

The LEV of a sound source is represented by a red ellipsoid. The form and position of the red ellipsoid can be changed by three control points, two control points for the height and width of the ellipsoid, and one control point for the position. In Figures 1 and 2, the red ellipsoid representing the LEV is shown.

3 Evaluation of Reporting ASW and LEV

3.1 Method

3.1.1 Participants

Seven participants (two females, five males) volunteered to participate in the listening test. The participants had an average age of 25.9 years (*SD* $= 4.1$) and four identified themselves as expert listeners in either timbre or spatial audio*. Four participants indicated that they have never taken part in a listening test about spatial attributes before.

3.1.2 Stimuli

In the listening test, the same stimuli were used as in [8] to compare the results with each other. In total, the participants listened to twelve stimuli which consisted of three sound source signals convolved with four binaural room impulse responses (BRIRs) corresponding to an anechoic chamber, a studio room, a small hall, and a large hall. The sound source signals were pink noise as well as a cello-, and a snare recording. More information about the sound source signals and the rooms of the BRIRs can be found in [8]. For loudness calibration, the pink noise in the anechoic chamber was set to 72 dBA SPL.

3.1.3 Materials and Apparatus

As graphical user interface, we used the webMUSHRA listening test software, extended by the proposed reporting methods for ASW and LEV. The room of the virtual scene had measurements of (W x D) 600 $\times$ 500 cm. Participants listened to the stimuli using electrostatic headphones (Stax SR-507) driven by a Stax SRM 600 amplifier.

3.1.4 Procedure

In the beginning of the listening test, the definitions of ASW and LEV were presented to the participants. ASW was explained as the perceived extent/width of a sound source in the horizontal direction and LEV as the sensation of being surrounded or enveloped by sound. Next, participants rated the ASW and LEV of two stimuli for training reasons. The first stimulus (snare recording in the anechoic chamber) was expected to have a rather less wide ASW and a rather less pronounced LEV. The second stimulus (pink noise in the small hall) was expected to have a rather wider ASW and a rather more pronounced LEV. Participants were provided with the information about the expected ASW and LEV of the stimuli. Moreover, the participants were also provided with detailed information on how to control the blue extruded curve and the red ellipsoid. In the training as well as in the actual listening test, the scene was visualized by a 2D top view. After the participants were familiarized with the graphical user interface, as well as the definitions of ASW and LEV, they rated the twelve stimuli in random order. Finally, the participants were asked about their age, gender, and whether they are expert listeners in judging timbre and in spatial audio. Furthermore, they were asked whether they have already taken part in a listening test about spatial audio, and whether they have a professional background in audio.

3.2 Results and Discussion

On average the participants needed 11.0 minutes (*SD* = 5.0) to complete the listening test. The boxplots depicted in Figures 3 and 4 show the responses to the perceived ASW and LEV. According to the means of ASW and LEV, the small hall turned out to have the widest ASW ($M = 107.0°, SD = 45.9$) and the most pronounced LEV ($M = 4.9\,m^2, SD = 4.6$) followed by the large hall (ASW: $M = 106.3°, SD = 52.3$; LEV: $M = 4.6\,m^2, SD = 4.4$). The studio room had a more wide ASW ($M = 97.8°, SD = 53.5$) and more pronounced LEV ($M = 2.0\,m^2, SD = 2.5$) than the anechoic chamber (ASW: $M = 84.6°, SD = 58.9$; LEV: $M = 1.2\,m^2, SD = 1.2$). According to results of a Kruskal-Wallis rank sum test, the differences in ASW between the rooms are not significant† ($H(3)$ = 2.186, $p =$ .535). However, the differences in LEV turned out to be significant ($H(3)$ = 26.536, $p <$ .001). As post-hoc analysis a pairwise comparisons using Nemenyi test was carried out to analyze the individual differences in LEV between the rooms. The Nemenyi test revealed that the anechoic chamber was significantly different from the small hall ($p <$ .001)

* M = mean, SD = standard deviation.

† The significance level was set to $\alpha = 0.05$

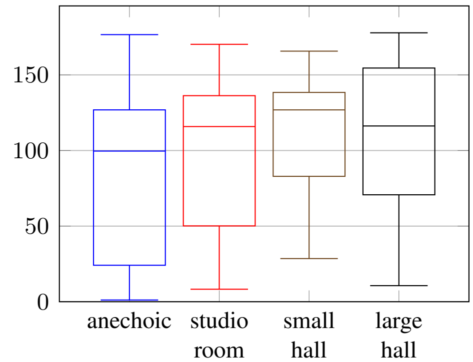

Figure 3 Boxplot of the opening angle of the ASW curve (in degree) relative to the participant.

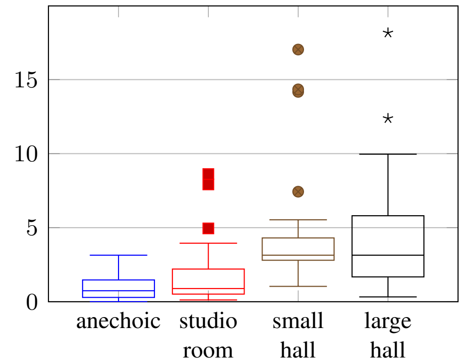

Figure 4 Boxplot of the area (in m^2) of the LEV.

and from the large hall (p = .002). Furthermore, the studio room was significantly different from the small hall (p = .013) and the large hall (p = .028). However, no significant differences were found between the studio room and the anechoic chamber (p = .875). Moreover, the large hall and the small hall did not differ significantly (p = .994).

Although not all results turned out to be statistically significant, the findings are in accordance with the results obtained by Merimaa and Hess' listening test. However, in their test, the differences between the rooms were more significant. The main reason for that might be the differences in procedure. Additionally, more time was spent to familiarize the participants with the definition of ASW and LEV as they had two sessions and multiple trials. Since our results are not in contrast to Merimaa and Hess' results, our framework's method for reporting ASW and LEV is expected to lead to valid results.

4 Conclusion

In this work, a framework for listening tests was implemented that features reporting methods for width, height, depth, location, ASW, and LEV of sound sources. The framework was evaluated by a listening test in order to test the validity of the reporting method for ASW and LEV. Future research of the framework should include evaluating the validity of reporting the spatial attributes that were not considered in the listening test. Moreover, the framework suitability for evaluating multi-channel formats with a large number of channels has to be tested.

References

1 K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando. A 22.2 Multichannel Sound System for Ultrahigh-Definition TV (UHDTV). *Society of Motion Picture & Television Engineers*, 117(3):40–49, 2008.

2 A. Silzle, S. George, E.A.P. Habets, and T. Bachmann. Investigation on the Quality of 3D Sound Reproduction. *International Conference on Spatial Audio*, 2011.

3 M. Schoeffler, S. Conrad, and J. Herre. The Influence of the Single/Multi-Channel-System on the Overall Listening Experience. In *Proc. of the AES 55th Conference on Spatial Audio*, Helsinki, Finland, 2014.

4 A. Lindau. Spatial Audio Quality Inventory (SAQI). Test Manual 1.2, 2014.

5 M. Schoeffler, F. Stöter, B. Edler, and J. Herre. Towards the Next Generation of Web-based Experiments: A Case Study Assessing Basic Audio Quality Following the ITU-R Recommendation BS.1534 (MUSHRA). In *1st Web Audio Conference*, Paris, France, 2015.

6 M. Morimoto. The Role of Rear Loudspeakers in Spatial Impression. In *Audio Engineering Society Convention 103*, New York, United States, 1997.

7 S. G. Norcross, G. A. Soulodre, and M. C. Lavoie. Temporal Aspects of Listener Envelopment in Multichannel Surround Systems. In *Audio Engineering Society Convention 114*, Amsterdam, Netherlands, 2003.

8 W. Hess and J. Merimaa. Training of Listeners for Evaluation of Spatial Attributes of Sound. In *Audio Engineering Society Convention 117*, San Francisco, United States, 2004.

9 M. Schoeffler, S. Westphal, A. Adami, H. Bayerlein, and J. Herre. Comparison of a 2D- and 3D-Based Graphical User Interface for Localization Listening Tests. In *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany, 2014.