Psychophysiological Effect of Immersive Spatial Audio Experience Enhanced Using Sound Field Synthesis

Yasuhide Hyodo[1*], Chihiro Sugai[2], Junya Suzuki[3], Masafumi Takahashi[3], Masahiko Koizumi[2], Asako Tomura[2], Yuki Mitsufuji[3], and Yota Komoriya[1]
[1]*Tokyo Laboratory 25, R&D Center, Sony Group Corporation, Kanagawa, Japan*
[2]*Group 1, Corporate Technology Strategy Division, Sony Group Corporation, Tokyo, Japan*
[3]*Tokyo Laboratory 21, R&D Center, Sony Group Corporation, Tokyo, Japan*
**corresponding author: Yasuhide.Hyodo@sony.com*

2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII) | 978-1-6654-0019-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ACII52823.2021.9597435

***Abstract*—Recent advancements of spatial audio technologies to enhance human's emotional and immersive experiences are gathering attention. Many studies are clarifying the neural mechanisms of acoustic spatial perception; however, they are limited to the evaluation of mechanisms using basic sound stimuli. Therefore, it remains challenging to evaluate the experience of actual music contents and to verify the effects of higher-order neurophysiological responses including a sense of immersive and realistic experience. To investigate the effects of spatial audio experience, we verified the psychophysiological responses of immersive spatial audio experience using sound field synthesis (SFS) technology. Specifically, we evaluated alpha power as the central nervous system activity, heart rate/heart rate variability and skin conductance as the autonomic nervous system activity during an acoustic experience of an actual music content by comparing stereo and SFS conditions. As a result, statistically significant differences ($p < 0.05$) were detected in the changes in alpha wave power, high frequency wave power of heart rate variability (HF), and skin conductance level (SCL) among the conditions. The results of the SFS condition showed enhanced the changes in alpha power in the frontal and parietal regions, suggesting enhancement of emotional experience. The results of the SFS condition also suggested that close objects are grouped and perceived on the basis of the spatial proximity of sounds in the presence of multiple sound sources. It is demonstrating that the potential use of SFS technology can enhance emotional and immersive experiences by spatial acoustic expression.**

Index Terms —Immersive audio experience, perception of spatial sound, psychophysiological effect

I. Introduction

Music evokes strong emotions in listeners and has psychological and physiological beneficial effects on health and wellbeing. The relationship between music and emotions has been historically studied in the field of psychology, and music-induced emotions have been attributed to cognitive evaluation [1]. Recent advances in neuroscience have led to research into the principles underlying the elicitation of emotions induced by music. For example, several mechanisms have been proposed, including brainstem reflexes caused by basic acoustic characteristics such as timbre, intensity, and consonance/dissonance, and emotional contagion, in which listeners perceive the emotional features and expressions of music, and imitate these features and expressions as their own emotions [2][3][4]. As a result, it has been reported that listeners show various physiological responses involving various systems, including central and autonomic nervous systems, when they concentrate on listening to music [5]. On the basis of these neural mechanisms, many studies on the effects of musical experiences on immersive sensations and emotions have been conducted in the area of psychophysiology. Furthermore, in the research field of applied acoustics, there is increasing attention on psychophysiological effects of music to enhance the listener's immersive acoustic experience.

As a recent scientific approach from the perspective of the immersive acoustic environment, a soundscape is considered to be acoustically equivalent to a landscape and is defined as a human perception of the acoustic environment in context [6]. A soundscape can be perceived as the outcome of a single sound or as the result of a combination of sounds arising from various environments. Previous studies on soundscapes have shown that positively perceived sounds, such as the sounds of nature, are associated with a higher quality of life and improved psychological and physical health [7]. In a previous study [8], soundscapes were classified into three main areas on the basis of sound sources: *geophony*, which includes all environmental sounds produced by non-organic elements of nature such as waterfalls; *biophony*, which includes organic but nonhuman sources such as animal mating vocalizations; and *anthrophony*, which includes all environmental sounds produced by human sources such as the human voice and sounds associated with human activity. The soundscape concept has been systematized and psychological evidence of the effects of sound on the human nervous system is accumulating, but the mechanisms underlying the effects are not yet fully understood. In recent years, the advancements in psychophysiology and neuroscience have accelerated the objective evaluation of emotion and acoustic environments; however, most studies are focused on assessing only autonomic nervous system responses measured such as skin conductance and heart rate/heart rate variability, which are easy to measure, and limited to studies on the neural basis of emotion and cognitive process [9][10][11]. Therefore, to investigate the effect of an immersive acoustic environment on emotion, it is important to understand the physiological basis using not only the parameters associated with the autonomic nervous system such as skin conductance and heart rate/heart rate variability but also those associated with the central nervous system measured by a method such as electroencephalography (EEG).

As an element of acoustics that provides a sense of immersive and realistic experience, sound source localization is a field of research that needs attention. If it is possible to perceive the positions of sound sources above and below, left and right, and in front and behind a listener, it is expected to provide the listener a more immersive and realistic experience. From the viewpoint of the physiological basis of sound localization, spatial sound cues are already processed at the level of the brainstem. For example, it has been shown that neurons in the inferior colliculus (IC) are sensitive to interaural time and level differences (ITD and ILD, respectively)[12][13]. At the level of the cortex, neuroimaging studies have shown that nonprimary areas of the auditory cortex, particularly the temporal lobe plane and posterior superior temporal gyrus (STG), are sensitive to moving sound stimuli [14][15]. Furthermore, there is also a growing interest in decoding the spatial characteristics of an auditory scene from brain activity. The cortical representation of a continuously moving auditory stimulus has been investigated by scalp EEG [16]. These studies are very important to clarify the neural mechanisms of acoustic spatial perception; however, they are limited to the evaluation of mechanisms using basic sound stimuli. Therefore, it remains challenging to evaluate the experience of actual music contents and to verify the effects of higher-order neurophysiological responses including a sense of immersive and realistic experience.

To bridge the gap between applied acoustics and neuroscience/psychophysiology in examining the effects of spatial acoustic experiences, in this paper, we investigate the effects of actual music contents on neurophysiological responses related to emotion elicitation and the spatial perception of sound movement during audio experience. Specifically, to verify the effects of enhancement of emotional and immersive experience by sound spatial perception, we create spatial acoustic contents as experimental stimuli in which the sound objects move in three dimensions using *sound fields synthesis* (SFS) technology [17]. SFS technology synthesizes physical wavefronts of multiple virtual sound sources simultaneously by changing delays and gains of driving signals for individual loudspeakers, and the synthesized wavefronts give the listener a spatial acoustic experience. We evaluated differences in psychophysiological responses between SFS and stereo sound sources conditions, and verified the effects of spatial audio experience on enhancing the immersive experience and emotion elicitation in the music experience through human spatial perception.

II. METHOD

In this study, we used SFS technology [17] to investigate the effect of emotion and spatial perception of sound object movement. On the basis of the literature on immersive acoustic environments such as soundscapes, we hypothesized that SFS technology could enhance a listener's emotional experience. Specifically, we hypothesized that the content that induces arousal will further arouse the listener and the content that induces relaxation will further relax the listener through the SFS technology. To verify these effects, we created new spatial acoustic contents to induce emotions and conducted psychophysiological experiments with measurements by EEG, electrocardiography (ECG), and electrodermal activity (EDA), as non-invasive methods of measuring physiological signals of nervous activity.

A. Content Creation

In creating the acoustic contents, we prepared two types of content: one for making the participants get accustomed to the two listening conditions, stereo and SFS (i.e., habituation), and the other for the actual measurements. These contents were created by a sound artist, evala.

1) Content for habituation: In this experiment, the participants needed to get used to wearing sensors, the experimental environment and the loudness of the loudspeakers in order to eliminate unnecessary stress. To achieve this, we prepared a two-minute melody similar to that of the signal sound was repeated at a constant rhythm. This content was prepared for both the SFS and stereo conditions.

2) Content for measurement (a target sound stimulus): To observe if the spatial sound of SFS technology contributes to the participants' emotional experience, we created a nine-minute spatial sound content. The content was created using simple sound sources and melodies so that the participants could actually listen to the spatial sound jumps and movements created by the SFS technology. The content was composed of two parts: the first half was for investigating the relaxation effect of SFS technology, and the latter half for investigating the concentration and arousal effects of SFS technology. The music was designed as a signal to change scenes after around 30 s considering the biological reaction time. Table I. shows the scene details in the target stimulus. The SFS reflected natural sounds such as rain and animal sounds, as well as melodies from electronic sounds and human voices. With time, these sound sources jump out spatially and move three-dimensionally around the participant, creating a sound content that gives a sense of space, as if the participant was entering the scenery of the music.

TABLE I. SCENE DETAILS IN TARGET SOUND STIMULUS

	Scene	Start time	End time	Duration	
1	Opening	00:00 (0 s)	00:10 (10 s)	10 s	Relaxation part 293 s
2	Sound of water droplets and muddy streams	00:10 (10 s)	01:25 (85 s)	75 s	
3	Sound of a wind	01:25 (85 s)	01:49 (109 s)	24 s	
4	Sound of rain	01:49 (109 s)	02:23 (143 s)	46 s	
5	Insect and animal sounds.	02:23 (143 s)	03:53 (233 s)	90 s	
6	Sound of wavelets	03:53 (233 s)	04:53 (293 s)	60 s	Concentration part 232 s
7	Melody A	04:53 (293 s)	05:22 (322 s)	29 s	
8	Chime of a bell	05:22 (322 s)	05:50 (350 s)	28 s	
9	Bird's chirping	05:50 (350 s)	06:10 (370 s)	20 s	
10	Melody B	06:10 (370 s)	06:37 (397 s)	27 s	
11	Melody C	06:37 (397 s)	08:33 (513 s)	116 s	
12	Woman's breath	08:33 (513 s)	08:45 (525 s)	13 s	

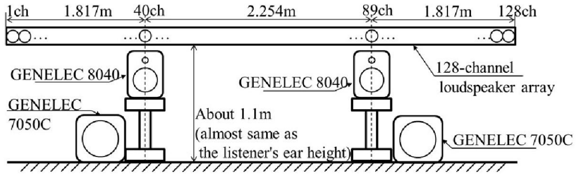
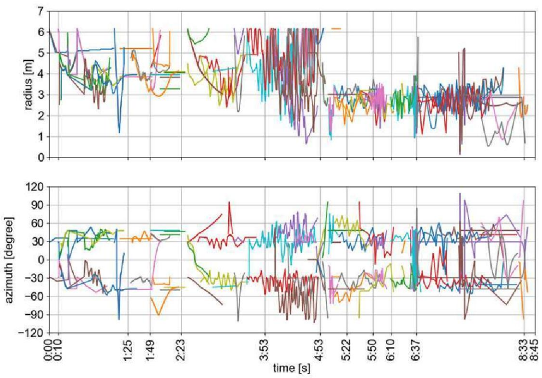
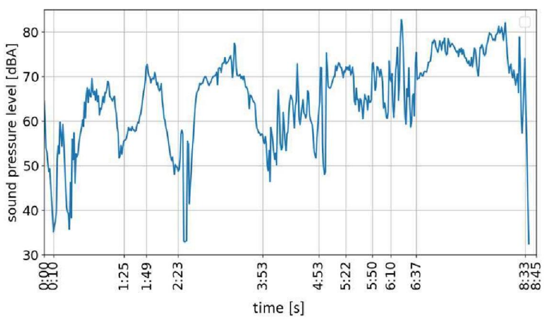


Fig. 1. Placement of loudspeakers.



(a) Sound object movement in target stimulus. (Top) distance between sound objects and position of listener. (Bottom) azimuth of sound objects from position of listener.



(b) Equivalent continuous A-weighted sound pressure level for every second.

Fig. 2. Parameters of target sound stimulus.

B. Acoustic Equipment and Conditions

As the playback equipment, we used a 128-channel linear array of loudspeakers, which is similar to that used at Acoustic Vessel Odyssey [17]. In this examination, these loudspeakers covered the frequency range from 250 Hz to 24 kHz. To compensate for low frequencies, we also used two GENELEC8040 woofers that covered the range from 80 to 250 Hz, and two GENELEC7050C subwoofers covered frequencies below 80 Hz. Fig. 1 shows the placement of these loudspeakers in detail. The height of the linear array was set to about 1.1m which was the same as that of the listener's ear. The woofers were placed under the 40th and 89th channels of the linear array. The subwoofers were placed nearby the woofers. The listener sat in front of the speaker, and the distance between the listener and the speaker was 2.5 m.

C. Content parameters

Fig. 2 shows parameters of the target sound stimulus as described in Table I. Fig. 2(a) shows the sound object movement in the content for measurement. Each color in the graph indicates a sound object. The top panel in Fig. 2(a) shows the distance (radius) between the sound objects and the position of the listener. The bottom panel in Fig. 2(a) shows the angle (azimuth) of the sound objects from the position of the listener. The angle is positive in the left-hand direction of the listener and in Fig. 2(a) in the front. Fig. 2(b) shows the equivalent continuous A-weighted sound pressure levels, which are used to account for the relative loudness perceived by the human ear per second.

For comparison, we created the stereo contents from the SFS contents as explained in this section. In a lower frequency range, we used the same woofers and subwoofers and these loudspeakers produced the same sound. In this 128-channel linear loudspeaker array, we used only the 40th and 89th channels. The sound signals of these two loudspeakers were generated by adjusting the left–right balance of each sound objects according to the positions of the sound objects based on the tangent law. The gain of each sound objects was adjusted according to the distance of the sound object from position of listener and the entire gain was adjusted so that the emitted sound intensity was the same as that of the SFS content.

D. Experimental Procedure

This study was preapproved by the Sony Bioethics Committee. There were 16 healthy participants (9 males and 7 females), aged 25–55 years.

We conducted the experiment indoor in a noise-free environment. The participants were seated at 2.5 m from the loudspeakers. During data collection, the participants were instructed to remain in the resting state and to close their eyes during the task. The EEG data were recorded using a 20-channel CGX Quick-20 acquisition system (CGX, United States) at a sampling rate of 500 Hz. To monitor autonomic nervous system activity, ECG and EDA were recorded using a Shimmer3 ECG/GSR system (Shimmer, Ireland) at a sampling frequency of 256 Hz（Fig. 3）.

The participants performed the tasks shown in Fig. 4 under two different conditions: stereo and SFS. Specifically, in Task A, the participants listened to a 150 s sound to familiarize themselves with the condition with eyes closed. Second, each participant performed a baseline task in the rest state for 120 s with eyes closed. Finally, in Task B, the participants listened to the target stimulus (a 530 s evaluation sound) with eyes closed. To remove the order effect, the order of the two conditions was randomized.

The participants were instructed not to make any unnecessary movements in order to minimize the effect of noise caused by movement during the EEG measurement. In addition, the participants were instructed not to fall asleep during the experiment, and this was confirmed by post-intervention interviews.

The effects of relaxed/concentrated music listening were assessed using the visual analog scale (VAS) ratings of arousal. VAS ratings were taken after the completion of Task B to evaluate the participants' subjective experience in addition to physiological responses. The VAS ratings were normalized from 0 to 100, and the participants were instructed to answer closer to 0 if they felt relaxed, closer to 100 if they felt excited, and closer to 50 if they felt neutral.

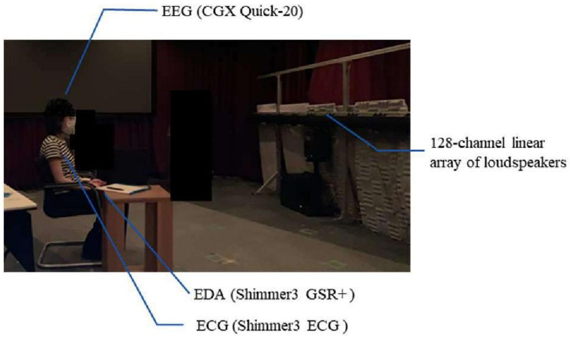


Fig. 3. Experimental setup of equipment.

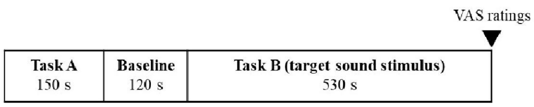


Fig. 4. Procedure of experiment.

E. *Preprocessing Steps*

1) EEG data preprocessing: All EEG data were processed offline in MATLAB 2020b using custom written scripts and the EEGLAB toolbox [18]. The data were filtered using an 8th-order IIR filter between 0.02 and 30 Hz and downsampled to 128 Hz. Data from unsatisfactory channels were interpolated from those from the surrounding channels. In this study, we focused on the analysis of fluctuations in the alpha frequency range of 8–13 Hz, because alpha power, or the amplitude of brainwaves in the 8–13 Hz frequency band, has consistently been shown to be inversely correlated with mental activity [19] and directly correlated with relaxation; thus it has the potential as an indicator of spatial perception of sound localization [16]. For the visualization of brain activation patterns by EEG, we used MNE python [20].

2) EDA data preprocessing: EDA data were processed offline in MATLAB using custom written scripts and the Ledalab toolbox[21]. The data were downsampled to 16 Hz and then decomposed into Skin Conductance Level (SCL) and Skin Conductance Responses (SCR) by Continuous Decomposition Analysis of Ledalab. Both of SCL and SCR are activated only by sympathetic nervous system [22]: therefore, we focused on the evaluation SCL/SCR as arousal indicator from EDA signal.

3) ECG data preprocessing: ECG data were processed using ConcensysPRO software (Shimmer, Ireland) to extract interbeat interval (IBI) time series. Then, we used custom written scripts to extract HRV indices. In particular, we focused on mean heart rate and high-frequency (HF) activity (0.15 to 0.40 Hz) in this study because both indices are robust indicators of ultrashort-term HRV analysis and are suitable to confirm changes of physiological phenomena in time series [23]. We expected HFs as a relaxation index activated by the parasympathetic nervous system. In this study, we used a 90-sec sliding window with 89-sec overlap to extract HRV features as time series data to examine the effect on time series data of each scene in the target stimulus.

III. Results

A. *Subjective Evaluation by VAS ratings*

First, the VAS ratings were analyzed to check the differences between the stereo and SFS conditions, and to evaluate the subjective effects of the experimental experience on the participants. A plot of the mean values with 95% confidence intervals of the VAS ratings of arousal in 16 participants is shown in Fig. 5. The results shown in Fig. 5 suggested that the difference in VAS ratings between the two conditions, stereo and SFS, was small. The mean score was below 50 under both conditions, suggesting that the subjective experience was relaxing under both conditions. One-way analysis of variance (ANOVA) was conducted and there was no statistically significant difference in the VAS ratings of arousal between the SFS and stereo conditions, $F(1,30) = 0.017$, $p = 0.896$, $\eta^2 = 0.001$.

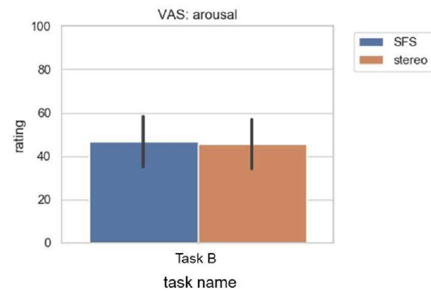


Fig. 5. Subjective evaluation from VAS ratings of arousal.

B. *Changes in Autonomic Nervous System Activity*

In addition to the VAS ratings, autonomic nervous system activity was assessed to confirm the effect of immersive spatial audio experience on physiological arousal. Specifically, to compare physiological responses to each scene, we focused on time series data analysis of (1) mean heart rate as a robust index of arousal by cardiovascular system, (2) HF as an index of relaxation by parasympathetic nervous system activity, and (3) SCL and (4) SCR as indices of tension caused by sympathetic nervous system activity. To exclude the effect of individual differences in the absolute values of physiological responses, normalization with baseline subtraction was introduced to consider *the law of initial values* (LIV) in biological science [24]. In Fig. 6, the lines indicate the mean values of the population data for each index at each time and are shown with 95% confidence intervals. As shown in these four indices, under both stereo and SFS conditions, the mean heart rate tended to decrease slightly, HF tended to increase slightly, SCL tended to decrease, and SCR tended to remain unchanged, suggesting that the participant was relaxed throughout the musical experience. The results are consistent with the VAS ratings and results in the literature [23]. The output differences in the autonomic nervous system responses between the SFS and stereo conditions were examined in the level of time series using the Mann–Whitney U test. However, no significant difference in autonomic nervous system activity was detected between the stereo and SFS conditions at the level of time series data.

Furthermore, we conducted 2-way repeated measures ANOVAs (Sound Type (SFS vs stereo) × Scene (#1 to #12 as shown in Table 1.)) to investigate statistical test beyond the level of the time series. We found that there was no significant

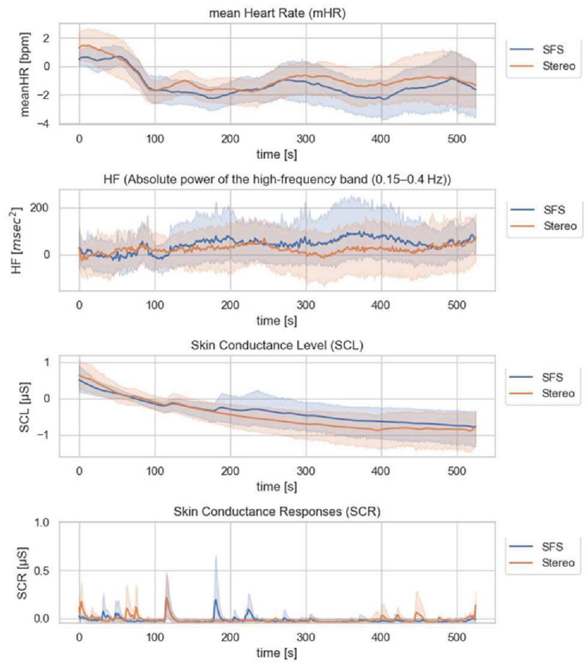


Fig. 6. Changes in autonomic nervous system responses (mean heart rate, HF, SCL, and SCR) during music experience (task B, the music listening task for the target sound stimulus) that were scaled by subtracting the mean of the baseline task prior to task B.

interaction effect of sound type and scene for all the indices. For mean heart rate, it was found that there was no significant main effect of sound type or no significant effect of scene. For HF, it was found that there was a significant main effect of sound type under the significance level at 0.05, $F(1{,}180) = 5.078, p = 0.025, \eta^2 = 0.009$, but there was no significant main effect of scene. For SCL, we confirmed that there was a significant main effect of sound type, $F(1{,}180) = 6.795, p = 0.010, \eta^2 = 0.009$, and there is a significant effect of scene, $F(11{,}180) = 6.112, p < 0.001, \eta^2 = 0.272$. For SCR, there was no significant main effect of sound type and that of scene. Since the main effects of sound type for HF and SCL were found, *post hoc* tests were conducted for them. The *post hoc* tests revealed that HF was greater in the SFS condition than in the stereo condition ($p = 0.023, d = 0.166$), and that SCL was greater in the SFS condition than in the stereo condition ($p = 0.010, d = 0.188$). Although no interaction effect was found, the simple main effect analysis of sound type was conducted to investigate difference according to each scene; however, no significant differences were detected.

C. Changes in EEG power

Finally, changes in EEG power during a music experience were evaluated to investigate the activity of the central nervous system. EEG power is physiologically more robust than autonomic nervous system activity; however, from the VAS ratings and autonomic nervous system activity, it was assumed that the difference between the two conditions was small. Therefore, to enhance changes in EEG power during a music experience, we used log-transformation for the scaling of the magnitude of EEG alpha power on the basis of a previous study [25]. Furthermore, normalization with baseline subtraction was introduced also as in the previous section. Fig. 7 shows the change in log-transformed alpha power in the music listening task with the target sound stimulus (hereafter referred to as task B) relative to the previous baseline task shown in Fig. 4. In addition, the output differences in log-transformed alpha power between the SFS and stereo conditions were examined using the Mann–Whitney U test in Fig. 8. As shown in Fig. 8, focusing on the details of the scene, statistically significant differences in the change in alpha power were detected between the two conditions. Especially, Melody A (scene #7) and Melody C (scene #11) contained scenes in which significant differences ($p < 0.01$) in the changes in alpha power in the frontal lobe were detected under the stereo and SFS conditions.

Furthermore, as well as for the physiological indices of autonomic nervous system activity, 2-way repeated measures ANOVAs (Sound Type × Scene) were also conducted for changes in alpha power in each scalp region of the EEG channels. For the frontal region (Fp1, Fp2, F3, F4, F7, F8, and Fz), we found that there was a significant main effect of sound type under the significance level at 0.05, $F(1{,}180) = 6.236, p = 0.013, \eta^2 = 0.012$, but there was no significant main effect of scene. There was no significant interaction effect of sound type and scene for the frontal region. For the parietal region (C3, C4, P3, P4, and Pz), we found that there was a significant main effect of sound type under the significance level at 0.05, $F(1{,}180) = 5.066, p = 0.026, \eta^2 = 0.007$, but there was no significant main effect of scene. There was no significant interaction effect of sound type and scene for the parietal region. For the temporal region (T3, T4, T5, and T6) and the occipital region (O1 and O2), we found that there were no significant main effects of sound type, no significant effects of scene, or no significant interaction effects. The *post hoc* tests revealed that change of frontal alpha power was greater for the SFS condition than for the stereo condition ($p = 0.012, d = 0.182$) and revealed that change in parietal alpha power was greater in the SFS condition than in the stereo condition ($p = 0.025, d = 0.163$). Although no interaction effect was found, the simple main effect analysis of sound type was conducted to investigate difference according to each scene. As a result, significant differences were found in changes in alpha power in the frontal region for Melody A ($p = 0.025$).

IV. Discussion

In this section, the results are discussed below from two perspectives: 1) emotional responses and 2) spatial perception of sound object movement.

A. Emotional responses

In the previous section, the results suggest that no significant differences in subjective evaluation were found, whereas significant differences were detected in HF from ECG, SCL from EDA, and alpha power from EEG. In the changes in HF and SCL, the results suggested that both parasympathetic activity and parasympathetic activity were more activated in the SFS condition than in the stereo condition. The results of the SCL waveform check through Task B showed that the waveform decayed in both Stereo and SFS conditions, and the HF also tended to increase from the baseline, indicating that the participants were basically in a relaxed state in both conditions in terms of autonomic nervous system activity. In the changes in alpha waves, we found that the alpha waves in the frontal and parietal regions were greater in the SFS condition than in the stereo condition.

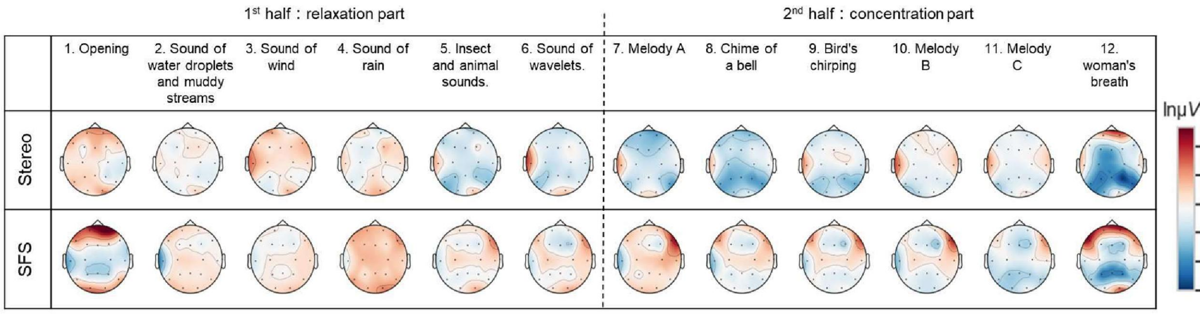


Fig. 7. Changes in alpha power during music experience that were scaled by subtracting the mean of the baseline task prior to task B.

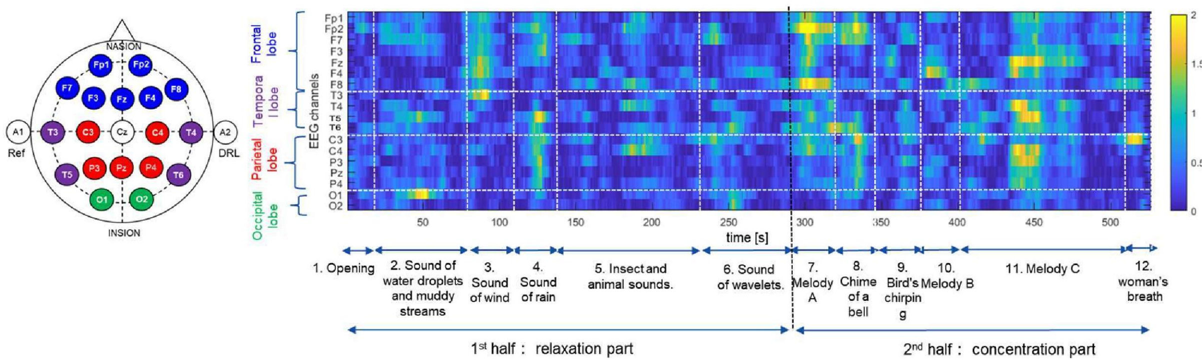


Fig. 8. Time series data of p-values for changes in alpha power on each EEG channel (Mann–Whitney U test).

It is known that the frontal lobe plays a role in regulating emotions, and that alpha power increases in a relaxed state [26], which is consistent with our results for EEG analysis. In particular, in Melody A, there was a significant difference in frontal alpha power at the level of time series, and participants were able to easily perceive the difference between the two conditions as a scene close to the boundary between the first and second half parts, suggesting that the relaxation effect was high.

In order to further investigate the effect of SFS on emotional responses to each scene, we conducted a detailed analysis of the brain activated area according to the results in Figs. 7 and 8. Specifically, we compared the results with those of previous studies in the context of *mindfulness meditation*, in which relaxation effects on the resting state are being studied [27]. Mindfulness meditation is a general term for types of meditations that originated in primitive Buddhism. There are two types of meditation technique: focused attention meditation (FA), which focuses attention deliberately and continuously on a specific object, and open monitoring meditation (OM), which focuses attention on an experience occurring in the present moment and accepts it as it is without judgment. Results of previous studies suggest that experienced meditators have increased alpha power in the prefrontal and parietal regions [28][29][30]. On the other hand, there is no strong consensus on the existence of parieto-occipital alpha power, with some studies suggesting an increase in occipital alpha power and others suggesting a decrease in alpha in FA and OM meditation.

From our results shown in Figs. 7 and 8, we confirmed that the alpha powers in the frontal and parietal regions increased more under the SFS condition overall, especially from scene #4 to scene #12 shown in Table I. On the other hand, in the occipital region, the trend of the changes in alpha power was not consistent between the SFS and stereo conditions. Although further investigation is required, our results suggest that SFS enhances the effects of meditation observed in previous studies, especially when focusing on changes in the frontal and parietal regions during meditation [28][29][30][31], and that SFS may have the potential to produce effects that allow people to relax and still be able to concentrate.

B. *Spatial perception of sound object movement*

In this subsection, we discuss the results shown in Figs. 7 and 8 in detail, taking into account the content parameters shown in Fig. 2. The sound location is spatially perceived using ITD and ILD as cues. In addition, previous studies have shown that the localization of sound sources depends on orientation and sound pressure level, which may affect the spatial perception of sound object movement [32]. These previous findings should be considered when interpreting the spatial perception effects of sound source motion in the SFS condition. We qualitatively evaluated the results shown in Figs. 7 and 8 and the content parameters shown in Figs. 2 and 3. In the neuroscience field, the parietal lobe contributes to human spatial perception. Furthermore, a recent study of sound source localization shows the response of the parietal lobe contributes significantly to the alpha power decoder [16].

In our results, only Melody C (scene #11) contained scenes in which significant differences ($p < 0.01$) in the changes in alpha power in the parietal lobe were detected under the stereo and SFS conditions. Furthermore, during Melody C, there were mixtures of segments where significant

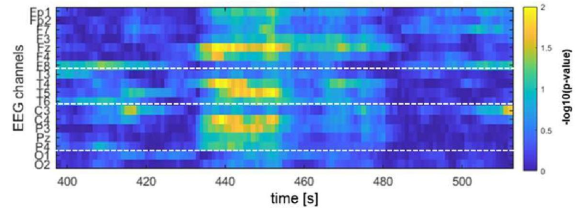
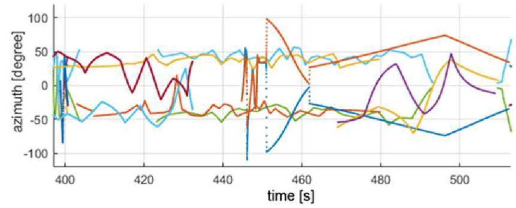
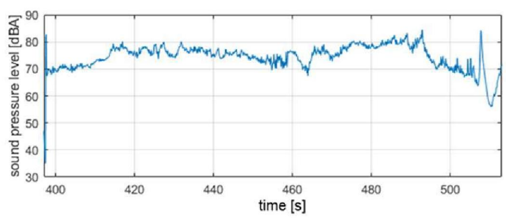


Fig. 9. Time series data of p-values for changes in alpha power on each EEG channel (Mann–Whitney U test) during Melody C in Task B.



(a) Azimuthal movement of sound objects.



(b) Equivalent continuous A-weighted sound pressure level during Melody C.

Fig. 10. Content parameters during Melody C in Task B.

differences were detected and segments where no significant differences were detected. Therefore, we focused on Melody C and compared the time series data with significant differences with the details of the content parameters. Fig. 9 shows the time series data of p-values of the changes in alpha power, which is excerpted from Fig. 8. Fig. 10 shows content parameters during Melody C, which is excerpted from Fig. 2. Specifically, Fig. 10(a) shows the sound object movement information of the azimuth as a content parameter, where each color shows a sound object, and Fig. 10(b) shows equivalent continuous A-weighted sound pressure level during Melody C. As shown in Fig. 9, significant differences between the stereo and SFS conditions were detected between 435 and 455 seconds in the change in alpha power in the parietal lobe. In addition, we compared the azimuthal movements of the sound sources at intervals where significant differences were detected and at the other intervals where no significant differences were not detected. We confirmed that the p-values were significantly lower at the intervals where multiple sound objects were moving in a spatially and temporally dispersed manner. Since the number of sound sources remained the same at both intervals, spatial proximity between multiple sound sources is suggested to be related to people's spatial perception. In addition, when sound objects that are not spatially close to each other move in a distributed motion, it is suggested that the sound objects to which attention was paid may have differed among the participants of the experiment. Our results provide findings similar to those in the literature [33], that is, close objects are grouped and perceived on the basis of the spatial proximity of sounds in the presence of multiple sound sources. On the other hand, these findings are still limited to focusing on sound source movement as one of the acoustic parameters. It remains to be interpreted from more multidimensional and principles-based perspectives, considering other content elements that contribute to emotions and contextual information in music contents. This study is only a preliminary to examine the physiological and psychological effects of mechanisms and real-world content experiences. Further evaluation will be carried in our future work.

V. Conclusion

In this study, we verified the psychophysiological responses of immersive spatial audio experience using SFS technology for examining the effect of spatial acoustic experiences to bridge the gap between applied acoustics and neuroscience/psychophysiology. For psychophysiological verification, we analyzed alpha power from EEG, heart rate/heart rate variability from ECG, and skin conductance level/skin conductance responses from EDA analysis during an audio experience using SFS technology. As experimental stimuli, we created spatial audio contents using SFS technology. For comparison, we created stereo contents from SFS contents. As a result of psychophysiological verification, in the subjective evaluation and autonomic nervous system activity indices, the results of statistical analysis suggest that the participants were in the relaxed state during the music experience under both stereo and SFS conditions. We found significant main effects of sound type (stereo vs SFS) in changes in HF, SCL, and alpha power in the frontal and parietal regions ($p < 0.05$). At the level of time series data, in changes in alpha power, significant differences between conditions were detected ($p < 0.01$). The results under the SFS condition showed enhanced changes in alpha power in the frontal and parietal regions, suggesting a trend similar to previous studies on meditation. Furthermore, the results under the SFS condition also suggested that close objects are grouped and perceived on the basis of the spatial proximity of sounds in the presence of multiple sound sources, demonstrating the effect of spatial perception of sound object movement.

The results suggest the potential use of SFS technology to enhance emotional and immersive experiences by spatial acoustic expression. However, there are some limitations in generalizing the results due to limited variations of music stimuli. This study is still limited to the analysis of the alpha power obtained by EEG carried out to understand the neural mechanism in detail. From the viewpoint of content parameters, they remain to be analyzed on the basis of a more multidimensional and principles-based investigation, considering other content elements and contextual information in music contents. In the future, we will explore other physiological analysis methods to further clarify the neural mechanism and content evaluation methods to identify which neural mechanism may be applied to enhance the listener's acoustic immersive experience.

VI. Acknowledgements

We wish to acknowledge that the SFS acoustic stimuli were created by a sound artist, evala (https://evala.jp/About). We wish to thank Naoya Sazuka, Mao Katsuhara, and Koki Katsumata for advice on experimental design, and Taiki Endo for support on experimental operation.

REFERENCES

[1] L. B. Meyer, Emotion and meaning in music. University of Chicago Press., 2008.

[2] P. N. Juslin, and D. Vastfjall, "Emotional responses to music: The need to consider underlying mechanisms," Behavioral and brain sciences, 2008, 31(5), 559.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] S. Koelsch, "Toward a neural basis of music perception–a review and updated model." Frontiers in psychology 2 (2011): 110.

[5] A. Sloboda, "Music structure and emotional response: Some empirical findings." Psychology of music 19.2 (1991): 110-120.

[6] M. Erfanian A.J. Mitchell, J. Kang, and F. Aletta, "The psychophysiological implications of soundscape: A systematic review of empirical literature and a research agenda." International journal of environmental research and public health 16.19 (2019): 3533.

[7] Fiebig, A., Jordan, P., & Moshona, C. C. (2020). Assessments of Acoustic Environments by Emotions–The Application of Emotion Theory in Soundscape. Frontiers in Psychology, 11, 3261.

[8] Pijanowski, Bryan C., et al. "Soundscape ecology: the science of sound in the landscape." BioScience 61.3 (2011): 203-216.

[9] Gomez, Patrick, and Brigitta Danuser. "Affective and physiological responses to environmental noises and music." International Journal of psychophysiology 53.2 (2004): 91-103.

[10] Hume, Ken and Mujthaba Ahtamad. "Physiological responses to and subjective estimates of soundscape elements." Applied Acoustics 74.2 (2013): 275-281.

[11] Medvedev, Oleg, Daniel Shepherd, and Michael J. Hautus. "The restorative potential of soundscapes: A physiological investigation." Applied Acoustics 96 (2015): 20-26.

[12] Middlebrooks, John C., and David M. Green. "Sound localization by human listeners." Annual review of psychology 42.1 (1991): 135-159.

[13] Stecker, G. C. and F. J. Gallun. "Binaural hearing, sound localization, and spatial hearing." Translational perspectives in auditory neuroscience: Normal aspects of hearing 383 (2012): 433.

[14] Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H. J., & Scheich, H. (1999). A movement-sensitive area in auditory cortex. Nature, 400(6746), 724-726.

[15] Bednar, Adam, Francis M. Boland, and Edmund C. Lalor. "Different spatio - temporal electroencephalography features drive the successful decoding of binaural and monaural cues for sound localization." European Journal of Neuroscience 45.5 (2017): 679-689.

[16] Bednar, Adam and Edmund C. Lalor. "Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG." NeuroImage 181 (2018): 683-691.

[17] Mitsufuji, Yuki, Asako Tomura, and Kazunobu Ohkuri. "Creating a Highly-Realistic" Acoustic Vessel Odyssey" Using Sound field Synthesis with 576 Loudspeakers." Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science. Audio Engineering Society, 2018.

[18] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," J. of neuroscience methods, vol. 134, no. 1, pp. 9–21, 2004.

[19] Baumgartner, Thomas, Michaela Esslen, and Lutz Jäncke. "From emotion perception to emotion experience: Emotions evoked by pictures and classical music." International journal of psychophysiology 60.1 (2006): 34-43.

[20] MNE Python: https://mne.tools/stable/index.html

[21] Benedek, M. & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. Journal of Neuroscience Methods, 190, 80-91.

[22] Boucsein, Wolfram. Electrodermal activity. Springer Science & Business Media, 2012.

[23] Castaldo, Rossana, et al. "Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life." BMC medical informatics and decision making 19.1 (2019): 1-13.

[24] Wilder, Joseph. "Basimetric approach (law of initial value) to biological rhythms." Annals of the New York Academy of Sciences 98.4 (1962): 1211-1220.

[25] Smulders, F. T., Ten Oever, S., Donkers, F. C., Quaedflieg, C. W., & Van de Ven, V. (2018). Single - trial log transformation is optimal in frequency analysis of resting EEG alpha. European Journal of Neuroscience, 48(7), 2585-2598.

[26] Cacioppo, John T., Louis G. Tassinary, and Gary Berntson, eds. Handbook of psychophysiology. Cambridge university press, 2007.

[27] Lee, D. J., Kulubya, E., Goldin, P., Goodarzi, A., & Girgis, F. (2018). Review of the neural oscillations underlying meditation. Frontiers in neuroscience, 12, 178.

[28] Takahashi, T., Murata, T., Hamada, T., Omori, M., Kosaka, H., Kikuchi, M., ... & Wada, Y. (2005). Changes in EEG and autonomic nervous system activity during meditation and their association with personality traits. International Journal of Psychophysiology, 55(2), 199-207.

[29] Henz, Diana, and Wolfgang I. Schöllhorn. "Temporal courses in EEG theta and alpha activity in the dynamic Health Qigong techniques Wu Qin Xi and Liu Zi Jue." Frontiers in psychology 8 (2018): 2291.

[30] Travis, Frederick. "Autonomic and EEG patterns distinguish transcending from other experiences during Transcendental Meditation practice." International Journal of psychophysiology 42.1 (2001): 1-9.

[31] Dentico, D., Ferrarelli, F., Riedner, B. A., Smith, R., Zennig, C., Lutz, A. (2016). Short meditation trainings enhance non-REM sleep low-frequency oscillations. PLoS ONE 11:e0148961. doi: 10.1371/journal.pone.0148961

[32] van der Heijden, K., Rauschecker, J. P., Formisano, E., Valente, G., & de Gelder, B. (2018). Active sound localization sharpens spatial tuning in human primary auditory cortex. Journal of Neuroscience, 38(40), 8574-8587.

[33] Bregman, Albert S. Auditory scene analysis: The perceptual organization of sound. MIT Press, 1994.