INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2024, VOL. 70, NO. 2, PP. 293-300
Manuscript received February 5, 2024; revised June, 2024. doi: 10.24425/ijet.2024.149544

Spatial sound and emotions: A literature survey on the relationship between spatially rendered audio and listeners' affective responses

Antonina Stefanowska, and Sławomir K. Zieliński

***Abstract*—With the development of the entertainment industry, the need for immersive and emotionally impactful sound design has emerged. Utilization of spatial sound is potentially the next step to improve the audio experiences for listeners in terms of their emotional engagement. Hence, the relationship between spatial audio characteristics and emotional responses of the listeners has been the main focus of several recent studies. This paper provides a systematic overview of the above reports, including the analysis of commonly utilized methodology and technology. The survey was undertaken using four literature repositories, namely, Google Scholar, Scopus, IEEE Xplore, and AES E-Library. The overviewed papers were selected according to the empirical validity and quality of the reported studies. According to the survey outcomes, there is growing evidence of a positive influence of the selected spatial audio characteristics on the listeners' affective responses. However, more data is required to build reliable, universal, and useful models explaining the above relationship. Furthermore, the two research trends on this topic were identified. Namely, the studies undertaken so far can be classified as either technology-oriented or technology-agnostic, depending on the research questions or experimental factors examined. Prospective future research directions regarding this topic are identified and discussed. They include better utilization of scene-based paradigms, affective computing techniques, and exploring the emotional effects of dynamic changes in spatial audio scenes.**

***Keywords*—spatial sound; emotions; affective responses**

I. Introduction

WITH the development of the entertainment industry, particularly Internet-based audio and audio-visual streaming services, the need for immersive and emotionally impactful sound design has emerged. Utilization of spatial sound is potentially the next step to improve the audio experiences for listeners in terms of their emotional engagement. While monophonic sound and its impact on listeners' emotions has been thoroughly examined during the last couple of decades [1]-[12], the influence of spatial sound on emotions is still unexplored, despite growing evidence that spatial audio with its inherent characteristics holds the potential for even greater immersivity and emotional impact on listeners [13]-[38].

This paper will focus on reviewing the existing studies examining the influence of spatial sound and its unique acoustic properties on listeners' emotional responses, with the purpose of identifying the existing empirical trends as well as formulating recommendations for future research. The research questions addressed in this paper are as follows:

- **RQ1**: *What are the methodologies used in the research investigating the relationships between spatial audio and emotions?*
- **RQ2**: *What have been the main findings so far?*
- **RQ3**: *What are the prospective future research directions?*

The content of the paper is as follows. The next section describes the way this literature survey was undertaken. The terminology used throughout the paper is introduced in Sec. III. The answers to the three above-mentioned research questions are given in Secs. IV, V, and VI, respectively. The conclusions along with the future research outlook are provided in the last section.

II. Literature Collection Procedure

The employed literature collection procedure was based on the PRISMA method [39], commonly utilized in systematic literature reviews. It comprised the following four stages:

1) *Identification*. The initial selection of the papers was identified using four popular literature repositories, namely, Google Scholar, Scopus, IEEE Xplore, and AES E-Library. A query employed during the search of the repositories is presented in Table I. All the records from the four search engines were collected on the 27th of June, 2023. As a result, 454 papers were identified. They were supplemented by six additional papers [16],[17],[26],[30],[31],[34], identified earlier by these authors during an informal exploratory literature review. Then, 57 duplicates were removed from the list, yielding a selection of 403 papers.
2) *Screening*. In order to exclude the studies outside the scope of the reviewed topic, the list of the papers identified in the previous stage was screened based on their metadata information, such as title, abstract, and keywords. This stage did not consider the contents of any of the articles. As a result, the repository of papers was limited to 37 articles.

The work was supported by the grants from Białystok University of Technology (WI/WI-IIT/3/2024, WZ/WI-IIT/5/2023) and funded with resources for research by the Ministry of Science and Higher Education in Poland.

A. Stefanowska and S. K. Zieliński are with Faculty of Computer Science, Białystok University of Technology, Poland (e-mail: antonina.stefanowska@sd.pb.edu.pl, s.zielinski@pb.edu.pl).

3) *Eligibility*. The list of articles that passed the screening stage was analysed by both authors independently, based on their contents and the specified eligibility criteria. Articles that did not meet the conditions (in our case eleven papers) were excluded from the review. The eligibility criteria were as follows:
    - The publication must be a journal article or a conference paper.
    - The article must report an empirical study performed on human subjects.
    - Spatial audio and/or its properties must constitute at least one of the experimental factors.
    - The study must include some form of subjects' emotional state evaluation (based on questionnaires and/or physiological data).
    - The paper must include a description of the methodology and the equipment used in the study.
4) *Included*. The final list of 26 articles that passed through all the previous stages was obtained and subjected to further analysis. For clarity, a flow diagram of the literature collection procedure based on the PRISMA method is presented in Fig. 1.

TABLE I
A QUERY APPLIED DURING THE PAPERS IDENTIFICATION

(“spatial sound” OR "spatial audio" OR "3d sound" OR "3d audio" OR "immersive sound" OR "immersive audio") AND ("emotion" OR "emotional" OR "affective")

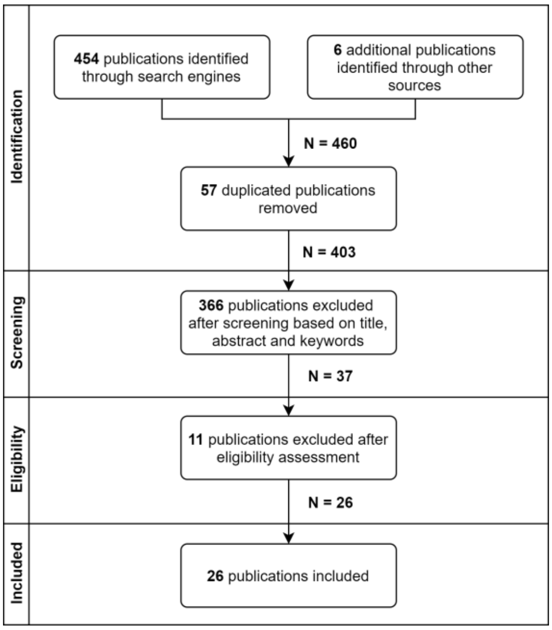
Identification
454 publications identified through search engines
6 additional publications identified through other sources
N = 460
57 duplicated publications removed
N = 403
Screening
366 publications excluded after screening based on title, abstract and keywords
N = 37
Eligibility
11 publications excluded after eligibility assessment
N = 26
Included
26 publications included

Fig. 1. Flow diagram of the literature collection procedure based on the PRISMA [39] method

III. Terminology Used in the Paper

We use the following terms to describe the methods of sound reproduction:

- *mono* – A condition whereby sound is delivered to a listener using either a single loudspeaker or via headphones reproducing the same signal.
- *stereo* – A condition under which two-channel sound is delivered to a listener employing two loudspeakers arranged in the standard configuration (at the angles of ±30° degrees [40]) or using headphones.
- *spatial sound* (a.k.a. 3D *audio*) – A condition whereby sound is delivered to a listener in a way that allows him/her to perceive audio sources as arriving from various directions horizontally and/or vertically. Technologically, this is accomplished using either an array of loudspeakers or a pair of headphones employing binaural technology. For a comprehensive overview of the technologies employed to reproduce spatial audio, see [41]-[43].

The following terms are used in this paper to describe listeners' emotional reactions to spatial sound:

- *affective response* – A change in a person's emotional state in reaction to a stimulus;
- *valence* – A characteristic of an emotional state describing how positive the experienced feeling is [44];
- *arousal* – A characteristic of an emotional state describing how intense the experienced feeling is [44];
- *dominance* – A characteristic of an emotional state describing how conscious the experienced feeling is [45];
- *discrete emotional space* – A model in which emotional states are defined by a finite set of emotional categories such as six universal emotions proposed by Ekman [46] or those described by Plutchik's wheel of emotions [47];
- *continuous emotional space* – A model under which emotional states are defined by points in a continuous multidimensional space. The most popular model was proposed by Russell [44]. It consists of two dimensions: valence and arousal.

IV. Methodology Overview

This section overviews the methodology applied in the studies investigating the relationship between spatial audio and emotions. It addresses the first research question posed at the outset of this review (RQ1).

A. Research Questions and Experimental Factors

In this paper, we categorize the existing studies as either technology-oriented or technology-agnostic, depending on the research questions or experimental factors examined. The former category refers to the experiments where the impact of spatial audio technology on human emotions is investigated (e.g., the equipment used to render spatial audio). The latter category represents the experiments whereby the researchers endeavor to establish the link between changes in ‘spatial audio scenes’ [48]-[50] and affective responses, regardless of the technology used.

1) Technology-oriented Studies

Out of 26 reviewed papers, 15 articles belong to the technology-oriented category [13],[15],[20]-[22],[25],[27]-[29],[32],[33],[35]-[38]. The research questions posed in these studies reflect the researchers' quest to explore how the changes in the audio reproduction systems affect the emotions felt by the listeners. The examples of the examined experimental factors are as follows: mono vs. spatial sound reproduced over headphones [13],[35]-[38], mono vs. stereo vs. surround sound [27], or stereo vs. spatial sound [15],[21],[22], [25],[32],[33],[38]. Moreover, one study aimed to compare the emotional effects evoked by the changes in the number and configurations of the loudspeaker placements, including a state-of-the-art 22.2 loudspeaker array [28]. The study of Ooishi *et al.* [29] is the only one that directly compared a playback system utilizing headphones with an array of 96 loudspeakers. In this case, spatial sound was reproduced either binaurally over the headphones or through a system of the loudspeakers employing a technique of wave field synthesis. The results of their empirical work suggest that spatial sound reproduced with a system of loudspeakers evokes more intense and, what might be considered as an unexpected outcome, more negative emotions than those aroused by headphones. More findings from the reviewed studies are provided in Sec. V.

2) *Technology-agnostic Studies*

The remaining eleven papers reviewed in this study belong to the technology-agnostic category [14],[16]-[19],[23],[24],[26], [30],[31],[34]. In contrast to the technology-oriented studies, in this category, a single reproduction system is normally used as a means of exploring how different spatial audio scenes change listeners' affective responses. In other words, the researchers' goal is to analyze the influence of the selected characteristics of acoustic scenes regardless of the device or technology used for their reproduction. For example, the researchers explore how the angular position of a single sound source affects the intensity of the perceived emotions [16],[31], they investigate the difference between the front and back-positioned audio sources [17],[18], or between front and side-positioned sound sources [23],[24]. Moreover, some researchers explore the influence of the spatial properties of the acoustic environments [14],[30], including the link between a type of concert halls and felt emotions [30].

As far as the technology-agnostic studies are concerned, the researchers typically treat spatial audio as a cause and evoked emotions as an effect, with the exception of the work by Pinheiro *et al.* [31], who demonstrated the opposite 'reaction'. Namely, they showed that emotions experienced by listeners may affect their perception of spatial audio, influencing their localization capabilities.

We argue that the technology-agnostic studies could be further categorized, in more detail, using a spatial audio scene description taxonomy. The proposed way of the additional categorization is based on the level of description of spatial acoustic scenes. Drawing inspiration from Rumsey's spatial audio scene description paradigm [46]-[48], we consider the three following categories that are illustrated in Fig. 2:

- *low-level description* – based around a single isolated sound source,
- *mid-level description* – considering an ensemble of sound sources,
- *high-level description* – pertinent to an entire acoustic environment with its unique reverberations, ambiences, etc.

Out of eleven studies belonging to the technology-agnostic group, seven investigated spatial scenes at the low level [16]-[18],[23],[24],[31],[34] and two at the high level [14],[30]. The work by Tajadura-Jiménez *et al.* [34] explored the emotional influence of spatial scenes both at the low and high levels, while the study undertaken by Gong *et al.* [19] could not be classified using this taxonomy as they investigated the influence of 'sound maps' in a computer game on the emotions of the participants. Surprisingly, no studies investigating the relationship between a mid-level scene description and emotions have been identified, indicating a potential research niche.

The technology-agnostic studies could also be characterized based on whether the sound sources within acoustic scenes are static (staying in the same positions) or dynamic (moving within the scene's bounds). So far, the majority of the technology-agnostic studies have focused on the static sound sources [14], [16],[17],[23],[24],[30],[31],[34]. A couple of relatively new studies introduced the movements of sound sources. For example, in the study of Cuadrado *et al.* [15], selected sound effects in the spatial adaptation of the story moved around the acoustic scene. The studies of Warp *et al.* [36],[37] constitute two other examples of experiments with dynamic scenes. In their experiments, the sound sources moved based on the listener's position in the virtual environment. Moreover, the work of Filippi *et al.* [18] and Li *et al.* [26] also involved dynamic scenes, with the source switching positions between front, back, left, and right directions in at least one of the test conditions. However, up to now, there is no study comprehensively examining the 'dynamism' of the sound source as the experimental factor.

In summary, the reviewed studies can be classified as either technology-oriented or technology-agnostic, depending on the research questions or experimental factors examined. The latter category could be further subdivided according to the level of the description of spatial audio scenes. None of the reviewed studies have investigated the influence of mid-level spatial audio scene characteristics on emotions, indicating a prospective research direction. Furthermore, a link between dynamically positioned sound sources and evoked emotions requires a more comprehensive investigation.

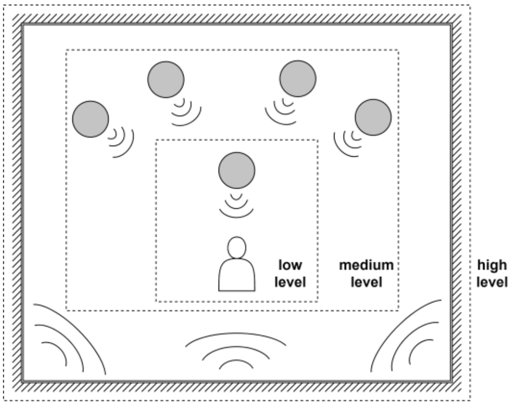

Fig. 2. Illustration of three acoustic scene description levels: low, medium and high. Circles represent foreground audio sources.

B. *Stimuli*

Twelve out of 26 reviewed studies utilized exclusively musical stimuli to elicit and compare the emotional responses of the listeners [13],[18],[20],[25],[27],[29],[30],[33],[35]-[38]. The usage of speech stimuli [14],[15],[22],[31] and natural or urban background sounds can also be encountered in the literature [23],[24],[26],[28]. Other studies make use of multiple sounds of different origins [16],[21],[34], sometimes including artificially synthesized signals [21],[34]. While not a common practice, the study of Gong *et al.* [19] utilized an audio-only game as the emotion-eliciting environment, which consisted of different sound effects serving as the only possible feedback to player's movements and actions.

Eight out of 26 scrutinized studies used a very limited collections of sound stimuli, which often consisted of just a single audio track or a video recording modified to different test conditions [13],[15],[22],[30],[33],[35]-[37]. In turn, those types of stimuli tended to be longer and more complex, e.g., song excerpts, narrated stories, video recordings, etc. This category of stimuli is often utilized in studies on immersion and emotional impact of virtual reality (VR) experiences [13],[14],[33],[36],[37].

C. Participants

In their study investigating listeners' preferences of spatial sound reproduction formats incorporating affective responses, Moiragias and Mourjopoulos [27] observed an interesting phenomenon of the opposite arousal scores between the individuals. This observation suggests that affective responses to spatial sound could be specific to certain groups of participants. Hence, to capture this effect, the number of participants should be relatively large, and the experimental data ought to be analyzed not only collectively but also separately for listeners exhibiting different emotional characteristics. However, researchers studying the link between spatial sound and emotions tend to collect data from a relatively small number of subjects. In 17 out of 26 reviewed publications, the number of participants did not exceed 40 [13],[14],[17],[18],[21],[23],[24],[26]-[32],[34],[36],[37]. Five other papers reported the number of subjects in the range of 40−140 [19],[20],[22],[33],[35]. In only four studies the researchers managed to collect data from over 200 participants [15],[16],[25],[38], two of which employed remote listening tests via online platforms [16],[38]. Despite significant differences in the number of participants, various studies examining similar experimental factors seem to reach mostly matching conclusions, as summarized in Sec. V.

None of the reviewed publications examined how the gender of participants had affected their affective responses to spatial sound. From all the scrutinized studies, only Cuadrado *et al.* [15] analyzed a possible effect of participants' age on experienced emotions. One of their conclusions was that an audio track mixed in a 3D format had a bigger emotional impact on children aged 12−13 than younger ones aged 9−10. Hence, the question of how much gender and age influence listeners' emotional reactions to spatial sound cannot be reliably answered, especially considering adult listeners.

D. Emotional Space

Thirteen out of 26 reviewed studies utilized custom non-standardized sets of multiple discrete emotional categories [13],[14],[17],[20],[22],[25],[26],[29]-[31],[33],[35],[38]. Examples of the discrete emotion categories investigated in the overviewed studies include: *happy* [26],[35], *sad* [14],[20],[26],[35], *calm* [14],[20],[26], *energetic* [13],[20],[26], and *tense* [14],[20]. In contrast, some studies, such those undertaken by Pätynen and Lokki [30] as well as Ekman and Kajastila [17], examined only a single emotional category (*emotional impact* and *scary*, respectively).

Alternatively, due to its simplicity, the standard valence-arousal continuous space is also commonly encountered in studies on emotions and spatial sound. It was employed in eight out of 26 inspected studies [15],[16],[18],[19],[23],[24],[27],[34]. Moreover, the work of Ramalho and Chambel [32] is an example of a study utilizing an extended continuous emotion model consisting of the three dimensions: valence, arousal, and dominance. The other studies examine only a single emotional dimension, either valence [36],[37] or arousal [21].

In conclusion, while emotional models differ significantly between the studies, particularly with respect to the types of the discrete emotion categories explored, the valence-arousal model seems to be the standard followed by many researchers examining the link between spatial sound and emotions.

E. Evaluation Methods

The emotional states of listeners were most commonly evaluated by *self-reports* (23 out of 26 studies) [13]-[20],[22]-[35],[38]. A prominent example of a self-report technique is the method employing Self-Assessment Manikins (SAM) [51], which consists of three rows of five pictograms each. It constitutes a standard graphical approach to 'measure' the emotional state of a listener in the valence-arousal-dominance space. A simplified two-row valence-arousal variant of this chart was utilized in six studies [15],[16],[18],[23],[24],[34]. Only one study made use of the full SAM chart in the valence-arousal-dominance emotional space [32].

An alternative method based on the valence-arousal space is an emotion map proposed by Barrett and Russell [52]. In contrast to the SAM technique, in this method, participants point out the position of their experienced feeling in a valence-arousal coordinate system. It was employed in the study of Gong *et al.* [19]. Another method of evaluating emotional states incorporates the so-called *Affective Slider* (AS) [53]. It was applied in the work of Moiragias and Mourjopoulos [27]. The remaining study-specific emotional models employed customized methods of self-reporting. They were utilized in twelve out of 26 reviewed studies [13],[14],[20],[22],[25],[26],[28],[29],[31],[33],[35],[38]. Pairwise comparison tests were uncommon. They were incorporated in only two studies [17],[30], examining a single emotional category (*scary* and *emotional impact*, respectively).

Since self-reports have limitations stemming from their subjective nature, information acquired using self-reports is often supplemented with 'objective' *physiological data*, containing information on bodily reactions taking place in response to a sound stimulus. Physiological data were acquired in eleven out of 26 reviewed studies [14],[15],[18],[21],[23],[24],[29],[30],[34],[36],[37]. Among those works, the following procedures were performed: electromyography (muscle activity) [23],[24],[34],[36],[37]; electroencephalography (brain activity) [18],[21]; electrocardiography (heart activity) [21],[29]; photoplethysmography (blood volume changes) [14],[36],[37]; inertial measurements [36],[37], heart rate estimation [36], respiration data acquisition [29],[36], as well as facial expressions recordings [36]. In the work of Warp *et al.* [36],[37], the physiological data were further processed by a proprietary emotion recognition system and converted to valence values. Filippi *et al.* [18], on the other hand, developed their own emotion recognition model based on support vector machines (SVM) trained with the collected electroencephalography data. Surprisingly, no other study made use of machine learning techniques to process the gathered physiological data.

In conclusion, self-reports remain the most popular method of evaluating emotional states in the experiments exploring the relationships between spatial sound and emotions (23 out of 26 studies), with the SAM technique constituting a staple tool used to construct questionnaires. The SAM-based method, however, imposes the usage of the valence-arousal-dominance model (or a model comprising a subset of these three dimensions). In order to examine non-standardized emotional spaces, over half of the studies with self-reports employed custom questionnaires. Self-reports are often used in conjunction with the measurements of physiological responses (eight out of 26 studies). Only three out of 26 scrutinized studies incorporated machine learning techniques to process physiological data and only one of them resulted in a trained emotion recognition model. The last-mentioned observation indicates that the researchers in this area could benefit from better utilization of modern affective computing techniques (see [54] for a review of the state-of-the-art affective computing methods).

F. Audio Reproduction Devices

According to our review, 14 out of 26 scrutinized publications report using headphones in the listening tests [13]-[16],[19],[22],[25],[26],[29],[31],[32],[36]-[38]. Their portability grants the possibility of undertaking remote experiments, which were performed in three studies [16],[26],[38]. In five studies, binaural signals reproduced over headphones were acquired using a popular 'dummy head' recording technique [14],[25],[26],[28],[29]. In six other studies, binaural signals were generated by the convolution of monaural sounds with head-related transfer functions (HRTF) [15],[16],[22],[31],[33],[38].

Using headphones to reproduce spatial audio gives rise to the problem of the static nature of the rendered sound scenes, leading to well-known front-back confusion effects [55]. This is the consequence that the playback systems do not react to the listener's head movements. To counter this issue, head-tracking devices should be used to analyze head positions and dynamically adjust the binaural cues in the headphones. This solution is already built-in in most of the virtual reality (VR) headsets. However, head-trackers have been employed in only five studies investigating the influence of spatial sound on listeners' emotions [13],[25],[33],[36],[37], suggesting that the validity of the remaining nine headphones-based studies might have been compromised in this respect [15],[16],[19],[22],[26],[29],[31],[32],[38].

Loudspeakers were used in twelve out of 26 studies [17],[18],[20],[21],[23],[24],[27]-[30],[34],[35]. The number of utilized loudspeakers varied significantly among publications. More than half of the loudspeaker-based studies (eight out of twelve) employed less than eleven loudspeakers [17],[18],[20],[23],[24],[27],[34],[35]. Two studies utilized 24 loudspeakers [28],[30]. Studies with the highest number of loudspeakers were undertaken by Ooishi *et al.* [29] and Hyodo *et al.* [21], as they used 96 and 128 loudspeakers, respectively. Out of twelve loudspeaker-based studies, eight utilized loudspeakers arranged horizontally in one flat layer, approximately at the level of the subject's ears [17],[18],[21],[23],[24],[27],[34],[35]. The other four used more complex arrangements with multiple loudspeaker layers at different elevations [20],[28],[29],[30]. Only three out of twelve loudspeaker-based studies provided reports of applying well-established spatial audio rendering methods, such as *Ambisonics* [43] – used by Tajadura-Jiménez *et al.* [34], *Wave Field Synthesis* [56] – utilized by Ooishi *et al.* [29], and *Sound Field Synthesis* [57] (based on *Higher Order Ambisonics*) – employed by Hyodo *et al.* [21]. The techniques employed to record and reproduce spatial audio in the remaining loudspeaker-based studies are unknown or scarcely reported.

In summary, in the experiments exploring the relationship between spatial audio and emotions, spatial sound is typically reproduced either via headphones or arrays of loudspeakers. The validity of some headphones-based studies could be questioned due to the lack of a head-tracking device. While some researchers provide a comprehensive description of the recording and rendering techniques employed, a substantial number of reports lack a sufficient level of technical detail, making it difficult for other researchers to compare and verify their results.

V. Key Findings

The key findings of the literature review are summarized below according to the category of the studies.

A. Technology-oriented Studies

- There is strong evidence that spatial sound, reproduced either using loudspeakers or headphones, evokes stronger emotions, with more positive valence and/or increased arousal, compared to mono [20],[27],[36]-[38].
- There is growing evidence that spatial sound induces in listeners stronger emotions compared to stereo [15],[22],[25],[32],[33]. However, some studies provide contradicting results, demonstrating that spatial audio brings little or even no benefit compared to stereo in terms of enhancing emotional responses [21],[28],[35].
- There is some evidence indicating that spatial sound reproduced over loudspeakers evokes stronger emotions compared to the same audio content reproduced over headphones [29].

B. Technology-agnostic Studies

- Side-arriving sounds evoke a stronger emotional response compared to front-arriving sounds [23],[24].
- Sound sources positioned behind a listener evoke more intense negative emotions than otherwise [34].
- When a sound source moves away from a listener's field of view, there is a tendency for arousal to increase and valence to decrease [16]. Moreover, a change in brain electrophysiological patterns related to emotional processing is observed [18].
- Fuzzy, difficult-to-localize sounds, especially arriving from the back of a listener, enhance 'scariness' [17].
- Spatial audio scenes exhibiting small room characteristics are perceived as more pleasant and safer

than those representing properties of a big room or outdoor setting [34].

- Spatial audio scenes representing concert halls with strong and lateral sound increase the emotional impact of orchestra music [30].
- Emotions experienced by humans may affect their localization capabilities while listening to spatial audio [31].

VI. Future Research Directions

As this topic is still relatively recent, the existing studies on spatial sound and emotions are still rare. However, some current trends can already be noticed. In particular, in recent studies, the technological aspects of spatial audio reproduction are commonly examined as factors potentially affecting listeners' emotions [13],[15],[20]-[22],[25],[27]-[29],[32],[33],[35]-[38]. It was mostly the older studies (with few exceptions) that focused on the characteristics of the spatial sound [14],[16]-[19],[23],[24],[26],[30],[31],[34]. As spatial sound technology matures and its effects are more understood, the research gap in that aspect might shrink in the future, causing the return to the more technology-agnostic approach. Further research directions, as identified by these authors, are as follows:

- Due to the contradicting results regarding the benefits of spatial audio compared to stereo in terms of listeners' emotional responses, more empirical work is needed in this area.
- The emotional impact of the spatial scenes with dynamically changing components requires further and more comprehensive research.
- Referring to our hierarchical taxonomy of scenes description, none of the reviewed papers investigated the influence of the spatial audio scenes at the medium level, indicating a specific research niche.
- Given that increasingly more researchers supplement subjective data with objective signals acquired using physiological sensors, the experimenters may benefit from the utilization of state-of-the-art affective computing techniques (reviewed in [54]).
- Out of 26 reviewed papers, only two studies [27],[28] resulted in regression models describing the relationship between spatially rendered audio and listeners' affective responses. Hence, there is a need for the development of universal and practical models in this field.

Conclusions

There is a growing body of research demonstrating the interactions between spatial sound characteristics and emotions. However, more data is required to build reliable, universal, and useful models explaining the above relationships. While there is conclusive evidence that spatial audio evokes stronger emotions compared to mono, with more positive valence and/or increased arousal, the outcomes of the studies comparing spatial audio to stereo are contradictory. The latter observation points to the need for further research in this area.

The two research trends on the topic of the relationship between spatial sound and listeners' affective responses were identified. Namely, the studies undertaken so far can be classified as either technology-oriented or technology-agnostic, depending on the research questions or experimental factors examined. The technology-agnostic studies could be further subdivided using a hierarchical three-level taxonomy of spatial audio scene description. Based on the currently existing research trends in this topic, in the future, a departure from a technology-oriented approach may take place. This may result in a shift to a more technology-agnostic methodology, focusing on the acoustic properties of spatial sound. In particular, mid-level acoustic scene description factors are worth researching, as they still remain unexplored. Another interesting research avenue may involve further and more comprehensive investigation of the influence of the dynamic spatial audio scenes on listeners' emotions. Moreover, considering the recent advancements in affective computing, researchers in the area of spatial sound and emotions may benefit from applying state-of-the-art machine learning algorithms to physiological data acquired from the listeners.

References

[1] B. Wu, A. Horner, and C. Lee "The Correspondence of Music Emotion and Timbre in Sustained Musical Instrument Sounds," Journal of the Audio Engineering Society, vol. 62, no. 10, pp. 663-675, 2014. https://doi.org/10.17743/jaes.2014.0037

[2] C. Chau, B. Wu, and A. Horner, "The Emotional Characteristics and Timbre of Nonsustaining Instrument Sounds," Journal of the Audio Engineering Society, vol. 63, no. 4, pp. 228-244, 2015. https://doi.org/10.17743/jaes.2015.0016

[3] J. Guo, J. Liu, Z. Li, J. Zhu, and W. Jiang, "A Study on the Relationship Between Timbre Perception Features and Emotion in Musical Sounds," in 2020 International Conference on Culture-oriented Science & Technology (ICCST), Beijing, pp. 22-27, 2020. https://doi.org/10.1109/ICCST50977.2020.00010

[4] C. Chau, R. Mo, and A. Horner, "The Emotional Characteristics of Piano Sounds with Different Pitch and Dynamics," Journal of Audio Engineering Society, vol. 64, no. 11, pp. 918-932, 2016. https://doi.org/10.17743/jaes.2016.0049

[5] C. Chau, S. J. M. Gilburt, R. Mo, and A. Horner, "The Emotional Characteristics of Bowed String Instruments with Different Pitch and Dynamics," Journal of Audio Engineering Society, vol. 65, no. 7/8, pp. 573-588, 2017. https://doi.org/10.17743/jaes.2017.0020

[6] R. Mo, B. Wu, and A. Horner, "The Effects of Reverberation on the Emotional Characteristics of Musical Instruments," Journal of Audio Engineering Society, vol. 63, no. 12, pp. 966-979, 2015. https://doi.org/10.17743/jaes.2015.0082

[7] R. Mo, R. H. Y. So, and A. Horner, "An Investigation into How Reverberation Effects the Space of Instrument Emotional Characteristics," Journal of Audio Engineering Society, vol. 64, no. 12, pp. 988-1002, 2016. https://doi.org/10.17743/jaes.2016.0054

[8] R. Mo, G. L. Choi, C. Lee, and A. Horner, "The Effects of MP3 Compression on Perceived Emotional Characteristics in Musical Instruments," Journal of Audio Engineering Society, vol. 64, no. 11, pp. 858-867, 2016. https://doi.org/10.17743/jaes.2016.0031

[9] Y. Hong, C. Chau, and A. Horner, "An Analysis of Low-Arousal Piano Music Ratings to Uncover What Makes Calm and Sad Music So Difficult to Distinguish in Music Emotion Recognition," Journal of Audio Engineering Society, vol. 65, no. 4, pp. 304-320, 2017. https://doi.org/10.17743/jaes.2017.0001

[10] W. L. Sin, X. Ma, and A. Horner, "The emotional characteristics of rain sound effects," in 2018 International Computer Music Conference (ICMC), Daegu, pp. 344-349, 2018. https://hdl.handle.net/1783.1/95682

[11] W. L. Sin, B. Y. Chang, X. Ma, and A. Horner, "The Acoustic Features and Their Relationship to the Emotional Characteristics of Rain Sound Effects," in 45th International Computer Music Conference (ICMC) and International Computer Music Conference New York City Electroacoustic Music Festival (NYCEMF), New York, pp. 84-89, 2019. https://hdl.handle.net/1783.1/98625

[12] C. Gafni and R. Tsur, "Some experimental evidence for sound-emotion interaction," Scientific Study of Literature, vol. 9, no. 1, pp. 53-71, 2019. https://doi.org/10.1075/ssol.19002.gaf

[13] T. A. Alam and N. Dibben, “A Comparison of Presence and Emotion Between Immersive Virtual Reality and Desktop Displays for Musical Multimedia,” in Future Directions of Music Cognition, Virtual, pp. 97-101, 2021. https://doi.org/10.18061/FDMC.2021.0017

[14] A. Algargoosh, B. Soleimani, S. O’Modhrain, and M. Navvab, “The impact of the acoustic environment on human emotion and experience: A case study of worship spaces,” Building Acoustics, vol. 29, no. 1, pp. 85-106, 2021. https://doi.org/10.1177/1351010X211068850

[15] F. Cuadrado, I. Lopez-Cobo, T. Mateos-Blanco, and A. Tajadura-Jiménez, “Arousing the Sound: A Field Study on the Emotional Impact on Children of Arousing Sound Design and 3D Audio Spatialization in an Audio Story,” Frontiers in Psychology, vol. 11, pp. 737, 2020. https://doi.org/10.3389/fpsyg.2020.00737

[16] K. Drossos, A. Floros, A. Giannakoulopoulos, and N. Kanellopoulos, “Investigating the Impact of Sound Angular Position on the Listener Affective State,” IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 27-42, 2015. https://doi.org/10.1109/TAFFC.2015.2392768

[17] I. Ekman, and R. Kajastila, “Localization Cues Affect Emotional Judgments – Results from a User Study on Scary Sound,” in 35th Audio Engineering Society International Conference: Audio for Games, London, pp. 166-171, 2009. http://www.aes.org/e-lib/browse.cfm?elib=15177

[18] E. D. Filippi, T. Schmele, A. Nandi, A. G. Torres, and A. Pereda-Baños, “Emotional Impact of Source Localization in Music Using Machine Learning and EEG: a proof-of-concept study,” TechRxiv, 2022. https://doi.org/10.36227/techrxiv.21789866.v1

[19] J. Gong, Y. Shi, J. Wang, D. Shi, and Y. Xu, “Escape from the Dark Jungle: A 3D Audio Game for Emotion Regulation,” in Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry (VAMR 2018), Las Vegas, pp. 57-76, 2018, https://doi.org/10.1007/978-3-319-91584-5_5

[20] E. Hahn, “Musical Emotions Evoked by 3D Audio,” in Audio Engineering Society International Conference on Spatial Reproduction – Aesthetics and Science, Tokyo, 2018. http://www.aes.org/e-lib/browse.cfm?elib=19640

[21] Y. Hyodo, C. Sugai, J. Suzuki, M. Takahashi, M. Koizumi, A. Tomura, Y. Mitsufuji, and Y. Komoriya, “Psychophysiological Effect of Immersive Spatial Audio Experience Enhanced Using Sound Field Synthesis,” in 9th International Conference on Affective Computing and Intelligent Interaction (ACII), Nara, 2021. https://doi.org/10.1109/ACII52823.2021.9597435

[22] G. Kailas and N. Tiwari, “An Empirical Measurement Tool for Overall Listening Experience of Immersive Audio,” in IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, 2021. https://doi.org/10.1109/ICCE50685.2021.9427770

[23] P. Larsson and D. Västfjäll, “Emotional and behavioural responses to auditory interfaces in commercial vehicles,” International Journal of Vehicle Noise and Vibration, vol. 9, no. 1-2, pp. 75-95, 2013. https://doi.org/10.1504/IJVNV.2013.053818

[24] P. Larsson, A. Opperud, K. Fredriksson, and D. Västfjäll, “Emotional and Behavioral Response to Auditory Icons and Earcons in Driver-Vehicle Interfaces,” in 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV), Stuttgart, 2009.

[25] S. Lepa, S. Weinzierl, H.-J. Maempel, and E. Ungeheuer, “Emotional Impact of Different Forms of Spatialization in Everyday Mediatized Music Listening: Placebo or Technology Effects?” in 136th Audio Engineering Society Convention, Berlin, 2014. http://www.aes.org/e-lib/browse.cfm?elib=17171

[26] J. Li, L. Maffei, A. Pascale, and M. Masullo, “Effects of the Spatialisation of Water-Sounds Sequences on the Perception of Traffic Noise,” Vibrations in Physical Systems, vol. 33, no. 1, 2022. https://doi.org/10.21008/j.0860-6897.2022.1.10

[27] G. Moiragias and J. Mourjopoulos, “A listener preference model for spatial sound reproduction, incorporating affective response,” PLoS One, vol. 18, no. 6, 2023. https://doi.org/10.1371/journal.pone.0285135

[28] S. Oode and A. Ando, “Estimation of Kandoh Degree with Emphasis on Spatial Sound Impressions,” in 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Kyoto, 2012. https://doi.org/10.1109/SNPD.2012.98

[29] Y. Ooishi, M. Kobayashi, M. Kashino, and K. Ueno, “Presence of Three-Dimensional Sound Field Facilitates Listeners’ Mood, Felt Emotion, and Respiration Rate When Listening to Music,” Frontiers in Psychology, vol. 12, pp. 650777, 2021. https://doi.org/10.3389/fpsyg.2021.650777

[30] J. Pätynen, and T. Lokki, “Concert halls with strong and lateral sound increase the emotional impact of orchestra music,” The Journal of the Acoustical Society of America, vol. 139, no. 3, pp. 1214-24, 2016. https://doi.org/10.1121/1.4944038

[31] A. P. Pinheiro, D. Lima, P. B. Albuquerque, A. Anikin, and C. F. Lima, “Spatial location and emotion modulate voice perception,” Cognition and Emotion, vol. 33, no. 8, pp. 1577-1586, 2019. https://doi.org/10.1080/02699931.2019.1586647

[32] J. Ramalho and T. Chambel, “Immersive 360º Mobile Video with an Emotional Perspective,” in ACM International workshop on Immersive media experiences (ImmersiveMe ’13), Barcelona, pp. 35-40, 2013. https://doi.org/10.1145/2512142.2512144

[33] M. Shin, S. W. Song, S. J. Kim, and F. Biocca, “The effects of 3D sound in a 360-degree live concert video on social presence, parasocial interaction, enjoyment, and intent of financial supportive action,” International Journal of Human-Computer Studies, vol. 126, pp. 81-93, 2019. https://doi.org/10.1016/j.ijhcs.2019.02.001

[34] A. Tajadura-Jiménez, P. Larsson, A. Väljamäe, D. Västfjäll, and M. Kleiner, “When room size matters: acoustic influences on emotional responses to sounds,” Emotion, vol. 10, no. 3, pp. 416-422, 2010. https://doi.org/10.1037/a0018423

[35] D. Västfjäll, “The Subjective Sense of Presence, Emotion Recognition, and Experienced Emotions in Auditory Virtual Environments,” CyberPsychology & Behavior, vol. 6, no. 2, pp. 181-188, 2003. https://doi.org/10.1089/109493103321640374

[36] R. Warp, M. Zhu, I. Kiprijanovska, J. Wiesler, S. Stafford, and I. Mavridou, “Moved By Sound: How head-tracked spatial audio affects autonomic emotional state and immersion-driven auditory orienting response in VR Environments,” in 152nd Audio Engineering Society Convention, The Hague, 2022. http://www.aes.org/e-lib/browse.cfm?elib=21703

[37] R. Warp, M. Zhu, I. Kiprijanovska, J. Wiesler, S. Stafford, and I. Mavridou, “Validating the effects of immersion and spatial audio using novel continuous biometric sensor measures for Virtual Reality,” in IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, 2022. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00058

[38] Y. Wycisk, K. Sander, R. Kopiez, F. Platz, S. Preihs, and J. Peissig, “Development of the Immersive Music Experience Inventory (IMEI),” Frontiers in Psychology, vol. 13, pp. 951161, 2022. https://doi.org/10.3389/fpsyg.2022.951161

[39] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, The PRISMA Group, “Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement,” PLoS Medicine, vol. 6, no. 7, pp. e1000097, 2009. https://doi.org/10.1371/journal.pmed.1000097

[40] ITU-R. BS. 775-4 Recommendation. “Multichannel stereophonic sound system with and without accompanying picture,” International Telecommunication Union, Geneva, 2022.

[41] F. Rumsey, “Spatial Audio,” 1st Edition, Routledge, London, 2001.

[42] J. Blauert ed., “The Technology of Binaural Listening,” Springer, Berlin, 2013. https://doi.org/10.1007/978-3-642-37762-4

[43] J. Paterson and H. Lee ed., “3D Audio,” Routledge, New York, 2021.

[44] J. A. Russell, “Affective space is bipolar,” Journal of Personality and Social Psychology, vol. 37, no. 3, pp. 345-356, 1979. https://doi.org/10.1037/0022-3514.37.3.345

[45] A. Mehrabian, “Relations among personality scales of aggression, violence, and empathy: Validational evidence bearing on the Risk of Eruptive Violence Scale,” Aggressive Behavior, vol. 23, no. 6, pp. 433-445, 1997. https://doi.org/10.1002/(SICI)1098-2337(1997)23:6<433::AID-AB3>3.0.CO;2-H

[46] P. Ekman, “An argument for basic emotions,” Cognition and Emotion, vol. 6, no. 3-4, pp. 169-200, (1992). https://doi.org/10.1080/02699939208411068

[47] R. Plutchik, “A psychoevolutionary theory of emotions,” Social Science Information, vol. 21, no. 4-5, pp. 529-553, (1982). https://doi.org/10.1177/053901882021004003

[48] F. Rumsey, “Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm,” Journal of Audio Engineering Society, vol. 50, no. 9, pp. 651-666, 2002. https://www.aes.org/e-lib/browse.cfm?elib=11067

[49] S. K. Zieliński, F. Rumsey, and S. Bech, “Effects of Down-Mix Algorithms on Quality of Surround Sound,” Journal of Audio Engineering

Society, vol. 51, no. 9, pp. 780-798, 2003. http://www.aes.org/e-lib/browse.cfm?elib=12208

[50] S. K. Zieliński, F. Rumsey, and S. Bech, "Effects of Bandwidth Limitation on Audio Quality in Consumer Multichannel Audiovisual Delivery Systems," Journal of Audio Engineering Society, vol. 51, no. 6, pp. 475 501, 2003. http://www.aes.org/e-lib/browse.cfm?elib=12222

[51] M. M. Bradley, and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," Journal of Behavior Therapy and Experimental Psychiatry, vol. 25, no. 1, pp. 49-59, 1994. https://doi.org/10.1016/0005-7916(94)90063-9

[52] L. F. Barrett and J. A. Russell, "Independence and Bipolarity in the Structure of Current Affect," Journal of Personality and Social Psychology, vol. 74, no. 4, pp. 967-984, 1998. https://doi.org/10.1037/0022-3514.74.4.967

[53] A. Betella and P. F. M. J. Verschure, "The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions," PLoS One, vol 11, no. 2, pp. e0148037, 2016. https://doi.org/10.1371/journal.pone.0148037

[54] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, and W. Zhang, "A systemic review on affective computing: emotion models, databases, and recent advances", Information Fusion, vol. 83-84, pp. 19-52, 2022. https://doi.org/10.1016/j.inffus.2022.03.009

[55] F. L. Wightman & D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement", The Journal of the Acoustical Society of America, vol. 105, no. 5, pp. 2841-2853, 1999. https://doi.org/10.1121/1.426899

[56] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," The Journal of the Acoustical Society of America, vol. 93, no. 5, pp. 2764-2778, 1993. https://doi.org/10.1121/1.405852

[57] Y. Mitsufuji, A. Tomura, and K. Ohkuri, "Creating a highly-realistic "Acoustic Vessel Odyssey" using Sound Field Synthesis with 576 Loudspeakers," in Audio Engineering Society International Conference on Spatial Reproduction – Aesthetics and Science, Tokyo, 2018. http://www.aes.org/e-lib/browse.cfm?elib=19648