Federico Schumacher,*
Vicente Espinoza,* Francisca Mardones,†
Rodrigo Vergara, Alberto Aránguiz,†**
and Valentina Aguilera††

*Department of Sound
Faculty of Arts
University of Chile
Compañía 1264, Santiago
Región Metropolitana, Chile
†Faculty of Psychology
Diego Portales University
Vergara 275, Santiago
Región Metropolitana, Chile
**Department of Kinesiology
Metropolitan University of Education
Sciences
José Pedro Alessandri 774, Ñuñoa
Región Metropolitana, Chile
††Department of Music
Faculty of Arts
University of Chile
Compañía 1264, Santiago
Región Metropolitana, Chile
{federico.schumacher,
valentinaaguilera04}@gmail.com,
vicente.espinoza@ug.uchile.cl,
rodrigocvergara@uchile.cl,
{francisca.mardones1,
alberto.aranguiz}@mail.udp.cl

Perceptual Recognition of Sound Trajectories in Space

Abstract: Sound spatialization is a technique used in various musical genres as well as in soundtrack production for films and video games. In this context, specialized software has been developed for the design of sound trajectories we have classified as (1) basic movements, or image schemas of spatial movement; and (2) archetypal geometric figures. Our contribution is to reach an understanding of how we perceive the movement of sound in space as a result of the interaction between an agent's or listener's sensory-motor characteristics and the morphological characteristics of the stimuli and the acoustic space where such interaction occurs. An experiment was designed involving listening to auditory stimuli and associating them with the aforementioned spatial movement categories. The results suggest that in most cases, the ability to recognize moving sound is hindered when there are no visual stimuli present. Moreover, they indicate that archetypal geometric figures are rarely perceived as such and that the perception of sound movements in space can be organized into three spatial dimensions—height, depth, and width—which the literature on sound localization also confirms.

Spatialization, Software, and Archetypal Figures of Spatial Movement

The use of sound spatialization methods is probably one of the most characteristic aspects of today's

music production. Spatialization is used in various musical genres, as well as in soundtrack design for films and video games. These methods have evolved from simply positioning the sound source or instrument at a fixed point in space during the mix (Gibson 1997) to creating complex multichannel arrays using algorithms like vector-based amplitude panning (VBAP, cf. Pulkki, Huopaniemi, and Huotilainen 1996), Ambisonics (Fellgett 1975), high-order

Ambisonics, and other algorithms based on psychoacoustic characteristics of human audition, mainly those based on head-related transfer functions. There is not enough empirical evidence to support that one spatialization method is better than another. Several studies conducted have not detected major differences between high-order Ambisonics, VBAP, and multiple-direction amplitude panning (cf. Frank 2014; Parsehian et al. 2015; Simon, Wuethrich, and Dillier 2017). Some studies do suggest that VBAP is a more appropriate method for low-density loudspeaker arrays, however, such as the one used in our study. According to Satongar et al. (2013, p. 2),

> Pulkki (1997) highlights the ability of VBAP to use the minimum number of loudspeakers necessary to reproduce a phantom source, contrasting with the tendency of ambisonic methods for all loudspeakers to contribute to reproduction. VBAP can also accommodate irregular loudspeaker layouts, a common problem for ambisonic systems and an active area of research.

Electroacoustic music in its various forms, and particularly acousmatic music, has made a more intensive use of spatialization techniques since the beginning of the 1950s. A classic example is the Pupitre d'Espace, designed by Jacques Poullin at the Groupe Recherces de Musique Concrète and used for the first time in July 1951 by Pierre Henry for the spatialized performance of his *Symphonie pour un homme seul*. On the other hand, the original version of *Gesang der Jünglinge* by Karlheinz Stockhausen (1955–1956) was reproduced by five groups of loudspeakers arranged around the audience (Emmerson 2007). To this day, concert devices composed of tens, or even hundreds, of loudspeakers in a wide variety of spatial configurations have been developed, which enable a feeling of spatial immersiveness within the listener, as well as the staging of a wide range of spatial sound trajectories.

The emergence of computer music and its widespread growth during the 1990s has been key to the development of several software products aimed at sound spatialization. As a result, most digital audio workstations today include at least one sound spatialization method, such as the automated panning of sound sources or the use of specific applications contained in plug-in packs offered to the user. Additionally, in recent decades, software has been developed that allows for a more-complex control of sound spatialization in settings with a larger number of loudspeakers. Some examples are Zirkonium (https://zkm.de/en/about-the-zkm/organization/hertz-lab/software/zirkonium), Spat-Gris (https://github.com/GRIS-UdeM/SpatGRIS/releases), GRM Spaces (https://inagrm.com/en/store/product/15/spaces), Spatium Panning (http://spatium.ruipenha.pt) and the more recent Spat Revolution (https://www.flux.audio). Thus, "composers could now incorporate spatial meaning into the lexicon of musical meanings they have traditionally employed" (Lennox and Dean 2009, p. 260).

A common aspect shared by these applications is that a series of spatial movements is offered by default in the form of presets, in addition to parameters that can be adjusted to achieve the desired spatialization. These presets consist of archetypal geometric figures, which are graphic representations of the trajectories that the sound is supposed to make. Among the most commonly used archetypal figures we can find circles, squares, spirals, ovals, pendulums, and stars. Although these figures may have an aesthetic value, there is, however, little empirical evidence regarding the listener's ability to perceive and recognize them.

Nonetheless, as Lynch and Sazdov (2017, p. 14) point out, "It is evident that composers develop and use many techniques to spatialize multichannel sound, but it is unclear what perceptual effect or sonic experience, if any, is created by these approaches." This means that the perceptual recognition of sound trajectories is not only a problem related to musical creation techniques, but is also of interest to various study and research fields. Consequently, studies from the field of psychoacoustics, both on the main physical and physiological aspects that intervene in the localization of sounds and on the abilities involved in this localization, have been particularly relevant (Searle et al. 1975; Perrott et al. 1990; Lakatos 1993; Küpper 1998; Moore and King 1999; Hüg and Arias 2009; Barton et al. 2013). A different approach to this problem relates to how individuals describe this perception beyond

the perceptual abilities put into play. Cognitive linguistics offers, although indirectly, some insights for studying and understanding these particularities (Lakoff and Johnson 1980; Talmy 1983, 2003). The particularity of our study—as opposed to the aforementioned research in the field of musical technologies, psychoacoustics, or linguistics—is that it focuses on assessing the perception of moving sounds in the acoustic space, an aspect on which there is little research. Thus, the main purpose of our research is to obtain evidence on the perceptual recognition of archetypal figures and other trajectories of sound in space developed for this study, which we present below.

Image Schemas of Spatial Movement

One typical aspect of electroacoustic music in general and acousmatic music in particular is the electroacoustic presentation device that these genres have developed. The loudspeaker setting allows for an immersive listening situation in 360 degrees that differs from traditional concerts, in which the sound sources are typically placed in the front.

Owing to the immersiveness of this listening situation, when describing their listening experience listeners will frequently use expressions such as "I had the impression that an object was moving towards me," "I felt the sounds closer to me," or "the sound moved behind my back, it circled around me" (Schumacher and Fuentes 2017). These are metaphorical expressions that describe the perception of the spatial movement of sound, and they seem to operate in acousmatic listening situations as well as in everyday acoustic spaces.

From this perspective, the listeners' descriptions of the perceived spatiality in acousmatic music could be construed as being closer to their experiences in everyday acoustic spaces than to the listening situations of nonacousmatic music (Kendall 2010). Additionally, during the acousmatic listening situation, the listener does not have visual access to the sound source, and the acousmatic composition generally does not intend to allude to such source in a representational way (Kendall 2010). Hence, we could conclude that although no link is established between the perceived auditory stimulus and the causal phenomenon, listeners still develop significant relationships with the stimuli to give meaning to their listening experience. These relationships could refer to general schemes of perception for the spatial movement of sound, accumulated in our everyday experiences.

These spatial behaviors could be conceived as a limited number of image schemas of spatial movement, mainly the orientation and container image schemas described in cognitive linguistics (Lakoff and Johnson 1980; Talmy 1983, 2003; Boroditsky 2000). The container schema has been particularly addressed by Kendall (Kendall and Ardila 2008; Kendall 2010) in the context of multichannel devices in electroacoustic music.

The aforementioned authors concur in an embodied understanding of these spatial schemas. According to Kendall and Ardila (2008, p. 230) "The feelings and thoughts that the listener associates with the experience of sound in space appear to arise from a deeply embodied knowledge of space and spatial movement." Lakoff and Johnson (1980, p. 15) argue that

> We are physical beings, bounded and set off from the rest of the world by the surface of our skins, and we experience the rest of the world as outside us. Each of us is a container, with a bounding surface and an in–out orientation.

Regarding the orientation schema, Lakoff and Johnson (1980, p. 30) point out that "These spatial orientations arise from the fact that we have bodies of the sort we have and that they function as they do in our physical environment." In other words, these schemas are profoundly rooted in the interaction and integration of an agent with the world, in this case, the auditory world.

Based on the foregoing, our contribution is to reach an understanding of the way in which we perceive the movement of sound in space as a result of the interaction between the sensory-motor characteristics of an agent or listener and the morphological characteristics of the stimuli and the acoustic space in which such interaction occurs.

Figure 1 shows a graphic representation of the two schemas. The orientation schema (in fine
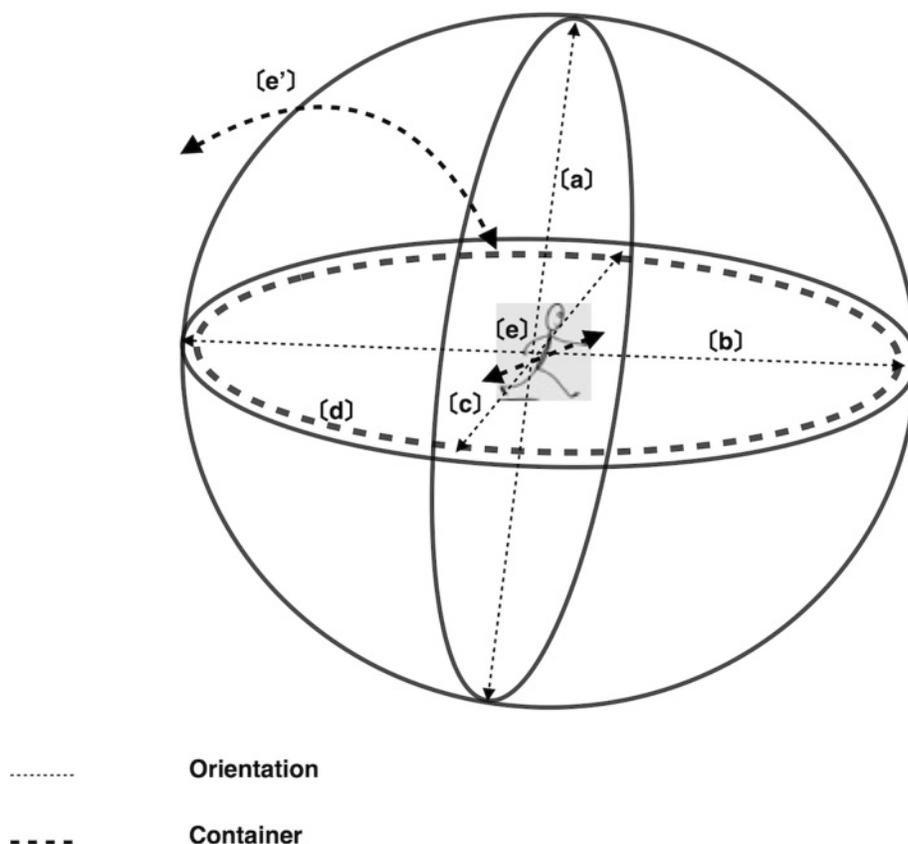
dotted lines) contains agents' descriptions of their perception of sound movement in space in three different dimensions or planes. Thus, the three categories that belong to this schema are Up–Down, Front–Back, and Left–Right (labeled "a," "b," and "c" in the figure). They share the same two values: Near or Far.

The container schema in Figure 1 (thick dotted lines, labeled "d") comprises descriptions related to the feeling of being immersed in and surrounded by the sound in a given acoustic space, which has both physical and acoustic limits (Rumsey 2002; Kendall 2010). We classify this perception of the acoustic space using the category Surrounded, which shares the variable Near–Far with the orientation schema. A second category associated with the container schema relates to the perception of sounds coming from inside or outside the acoustic space, which

includes the possibility for these sounds to cross the boundaries of the space (label "e" at the center of Figure 1).

A singular situation occurs during the acousmatic listening experience that rarely occurs in everyday acoustic spaces: Listeners often describe having the feeling that the sound went through their bodies (label "e'" at the top left of Figure 1). In this case, they identify the sound as being external and continuing its trajectory after going through their bodies, which is different from situations in which we hear sounds emerging from our bodies and going out into the acoustic space. In Table 1, we show a systematization of the two schemata, with their categories and variables.

Although the literature cited here tends to consider these categories as separate schemas, this research offers a theoretical proposal in which these

Table 1. Systematization of Sound Trajectories

Schema	Category	Variable
Orientation	Up–Down [a]	Near–Far
	Front–Back [b]	Near–Far
	Left–Right [c]	Near–Far
Container	Through [e][e′]	Inside–Outside
		Near–Far [e′]
	Surrounded	Near–Far

1. to evaluate the perceptual recognition of sound movements associated with these schemas and a group of archetypal figures of spatial movement, and
2. to estimate whether these schemas can be classified separately or as categories encompassed within the orientation and container schemas.

Methods

In the following we describe the participants, stimuli, experimental setting, protocol, and statistical analysis of this study.

Participants

Twenty-four individuals were included in this study, twelve men and twelve women, with an average age of 24.2 years (minimum 20, maximum 34, standard deviation 3.9). All participants had higher education and normal audition, and they signed an agreement of informed consent. They received remuneration for their participation in this study.

categories can be considered as part of a larger schema that encompasses them.

In a previous study (Schumacher and Fuentes 2017), participants listened to an acoustic piece and were then interviewed about their listening experience. Statements were identified regarding the spatiality of the piece that seem to associate with and exemplify the schemes described in Table 1. The testimonies seem to indicate that these schemes function at least on a linguistic level, as can be observed in Table 2.

In light of the theoretical proposals presented thus far, we designed an experiment taking into account the following objectives:

Table 2. Characterizations of Sound Trajectories

Schema	Category	Subjective Responses
Orientation	Up–Down [a]	The sound "was above me." (Participant 2)
		"There was a sound that felt very close, just above my head, as if it came from above." (Participant 7)
	Front–Back [b]	"I heard some acute sounds coming from behind rather than from the front." (Participant 3)
	Left–Right [c]	"The sounds came from everywhere." (Participant 2)
		"The sound came from behind me, and sometimes from the side." (Participant 7)
Container	Through [e][e′]	"The sound can go into you and then out. It was an enveloping sound. It was the most impressive experience." (Participant 19)
		"I don't know if I felt it inside of me, but the sounds that moved at my head's height felt like they were inside my head." (Participant 20)
	Surrounded [d]	"The sound moved like it was spinning, I felt surrounded by the speakers, as if the sound moved from one speaker to another in circles." (Participant 14)
		"I would describe it as a sphere that spins concentrically around the listener." (Participant 5)

*Figure 2. Waveform of
source audio fragment.*
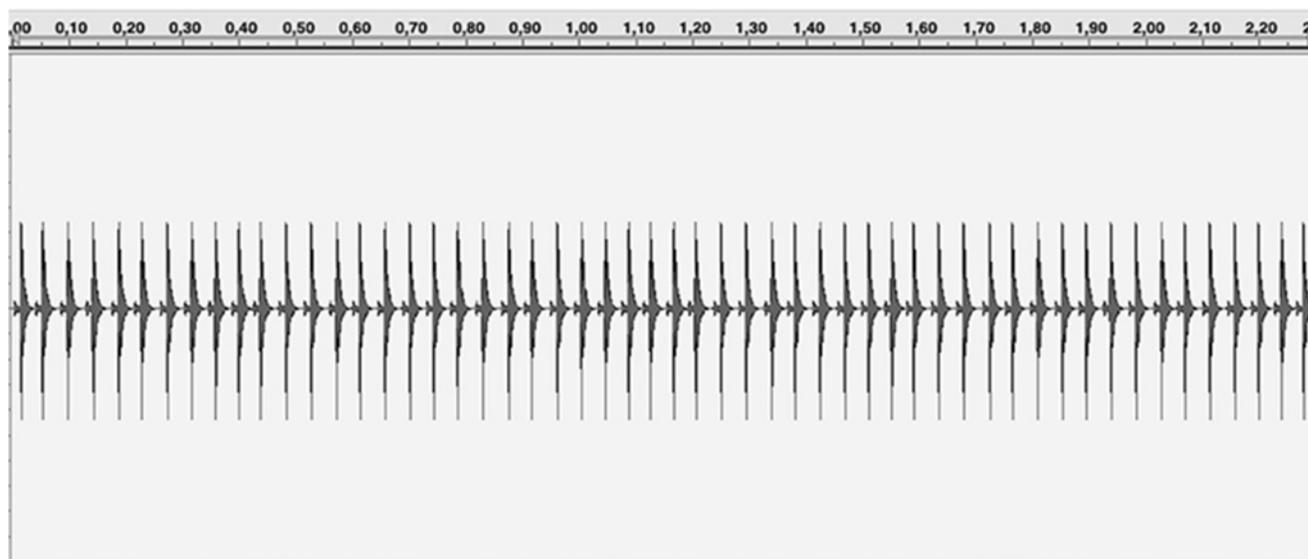
Stimuli

Eight auditory stimuli were created, each corresponding to one of the spatial trajectories identified as spatial movement schemas (Back–Front, Front–Back; Up–Down, Down–Up; Right–Left, Left–Right; Inside–Outside, Outside–Inside). Additionally, another four stimuli were created that represented spatial trajectories commonly used as presets in sound spatialization software, particularly the ones used in the GRM Tools application Spaces (Circle, Square, Spiral, and Star). Each audio fragment had an overall duration of 2.305 seconds and consisted of a sequence of evenly spaced click sounds (see Figure 2). Each of the clicks was of short duration (0.031 sec) and had a high frequency (1.2-kHz fundamental). Each click is preceded by a minor click who has a low amplitude and short duration (10 msec), so it is not heard separately from the main click (see Figure 2). Because all sounds were identical in amplitude and frequency, they could be understood as an iterative typology (Schaeffer 1966). The sound fragments were made using a special version of the Spatium Panning software in which the localization of each sound or spatial trajectory is defined by only two vectors using the VBAP technique. The Up–Down, Down–Up, Inside–Outside,

and Outside–Inside stimuli were created with the platform Reaper (https://www.reaper.fm) using the same spatialization technique described earlier. The reason for this is that each auditory stimulus had to physically trace the desired shape or trajectory by moving from one speaker to another. To do so, speakers were installed on the ceiling, on the floor, and outside the room where the experiment was conducted—a configuration that Spatium Panning does not consider within its possibilities.

The auditory stimuli were designed so that they would not have a recognizable source and could not be associated with any explicit semantic content or emulate real-world acoustic sources. Additionally, the experimental design sought to prevent other variables from intervening in the recognition of the sound trajectories, such as Doppler effect, filters, or spectromorphological variations. The auditory stimuli can be downloaded from https://github.com/AcusmaLab/SiSAudioFiles/blob/main/Tacs:Interleaved.zip.

Experimental Setting

The setting involved a multichannel device composed of eleven A5X loudspeakers from Adam

Figure 3. 2-D representation of the experimental setting. The grid delimits the space where the experiment took place. The speaker labeled "Up" is located on the ceiling, and the speaker labeled "Down" is located on the floor. "Up" and "Down" speakers are in the exact center of the room, not literally where their icons appear in the figure.
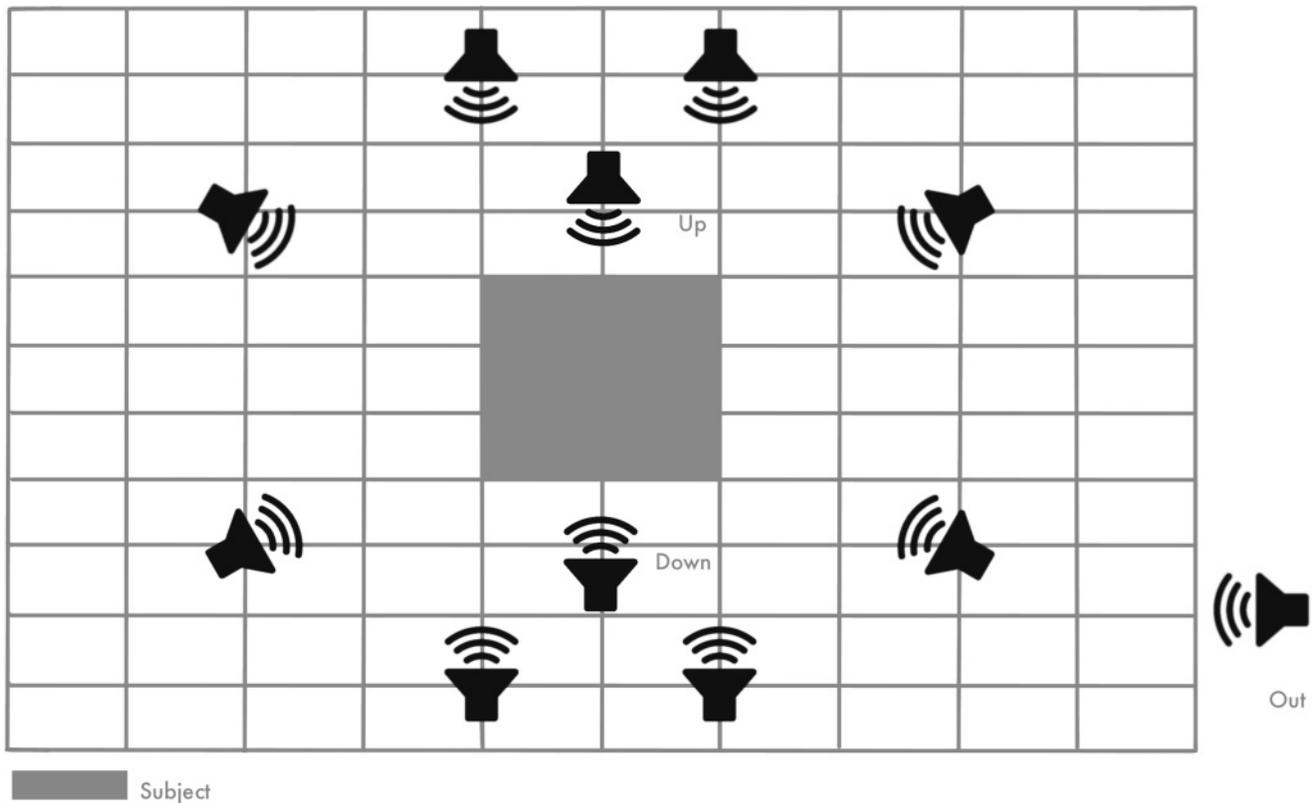
Audio. Eight of them were distributed in an oval shape, owing to the spatial characteristics of the room. Additionally, two loudspeakers were placed at the ceiling and on the floor, one directly above and the other directly below the participant. A further loudspeaker was placed outside the room (see Figure 3). The sound pressure levels for each loudspeaker were controlled so that the acoustic sensation would be similar at the listening focal point regardless of the location and proximity of the monitors. For the presentation of the stimuli, a patch was designed with the software Max, which was synchronized via OSC with a script written in Python using the PsychoPy 3 (https://www.psychopy.org) library. This, coupled with the randomized presentation of the stimuli, allowed for the twelve categories of spatial movement to be visualized on screen, for categories to be selected by clicking on them, and for responses to be stored. The experiment was conducted under controlled conditions at the Laboratory of Human Interaction of Diego Portales University.

Experimental Protocol

After reading and signing the consent agreement, participants had to pass a preliminary test to verify that they had normal binaural hearing. The test entailed listening to white noise at −20 dB through a pair of headphones, alternating between the left and right channels. Then instructions were presented to the participants on a computer screen and were read aloud by the researchers. To familiarize themselves with the stimuli, participants listened to each sound stimulus twice and were shown the corresponding category of spatial movement before the experiment began. This is a common practice in other studies, such as those by Lakatos (1993) or by Lee, Song, and

Horner (2021), which seek to mitigate the influence of any previous musical training the participants may have. As for the experiment itself, the task was to listen to an auditory stimulus and then to choose the category that best represented the trajectory of the sound heard. To this end, the twelve categories of sound movement, plus an additional category called "undetermined," were shown onscreen and from which participants could choose. The participants could choose the "undetermined" option if they thought the sound stimulus did not match any of the twelve categories. Each category of sound movement was played eight times in random order, so that each participant categorized 96 stimuli in total.

Statistical Analysis

To evaluate perceptual recognition we started by evaluating whether each schema and archetypal figure had been recognized with better than random probability. Given that participants answered from a forced-choice paradigm with 13 possible answers, random answers would be expected to produce correct answers 7.6 percent of the time. Using that threshold, we used a one-sample Wilcoxon signed-rank test with continuity correction (given the nonparametric nature of the variables) to evaluate whether answers were statistically above that chance threshold.

Once results were obtained, we proceeded to our second aim. To evaluate whether schemas can be categorized, we used factor analysis, a dimension-reduction procedure. This method evaluates the correlations between variables to then produce latent variables (i.e., categories). For a latent variable to be produced, it is necessary to obtain high correlations between a set of variables, while at the same time obtaining a low association with the remaining variables. In our case, we evaluated the associations between for each schema to explore potential latent variables and categories. If the performance on one schema is highly correlated to another, but not to the remaining ones, it was considered a category. To facilitate the comprehension of this procedure, we started by reporting the Pearson correlation between

the schemas so that readers may first interpret how the schemas are associated. We then continued on to factor analysis. For this procedure we started by evaluating the number of potential categories using parallel analysis (Hayton, Allen, and Scarpello 2004; Raîche et al. 2013). Once that number is defined, we use axis factorization as an extraction method (cf. de Winter and Dodou 2012; Costello and Osborne 2019). Because we expected that these categories are not strictly independent and that they can be slightly associated (having a good performance in one category may also imply having a similar performance in another category), we selected the oblimin rotation method (cf., e.g., Costello and Osborne 2005). A rotated loading matrix obtained from the factor analysis is presented that has loadings less than 0.3 pruned. Finally, we evaluated internal consistency using Cronbach's alpha (Cronbach 1951). Cronbach's alpha can be interpreted as a measure of how consistently correlated schemas that were contained in a category obtained from the factor analysis are.

Results

The main objectives of this research were (1) to assess the perceptual recognition of sound trajectories associated with the aforementioned image schemas and archetypal figures of spatial movement, and (2) to estimate whether these schemas can be classified separately or as categories encompassed within the orientation and container schemas. For this purpose, we first evaluated whether the participants' responses were significant and not attributable to chance. Once we identified which movements were effectively detected by the participants, we explored the associations among eight of these movements to determine whether there were any common perceptual abilities in the recognition of these categories.

For the first objective, one-sample Wilcoxon signed-rank tests with continuity correction (Wilcoxon 1945) were run to determine whether the participants' responses could be explained by chance. Figure 4 shows a box-plot diagram in which the dotted line represents a 7.6 percent performance, i.e., the chance level. The asterisks at the top of

Figure 4. Perceptual recognition of the stimuli. The box plot shows the median (horizontal line at the center of each box), two central quartiles (25%) above and below the median in the shape of a box, and the first and last quartile (22.5%) graphed as extension lines ("whiskers"). The dots are observations outside the 95% confidence interval. Asterisks at the top of each column denote the confidence of the result, with a single asterisk indicating 5% error (p < 0.05), a double asterisk indicating 1% error (p < 0.01), a triple asterisk indicating 0.1% error (p < 0.001), and the abbreviation n.s. indicating "not significant."
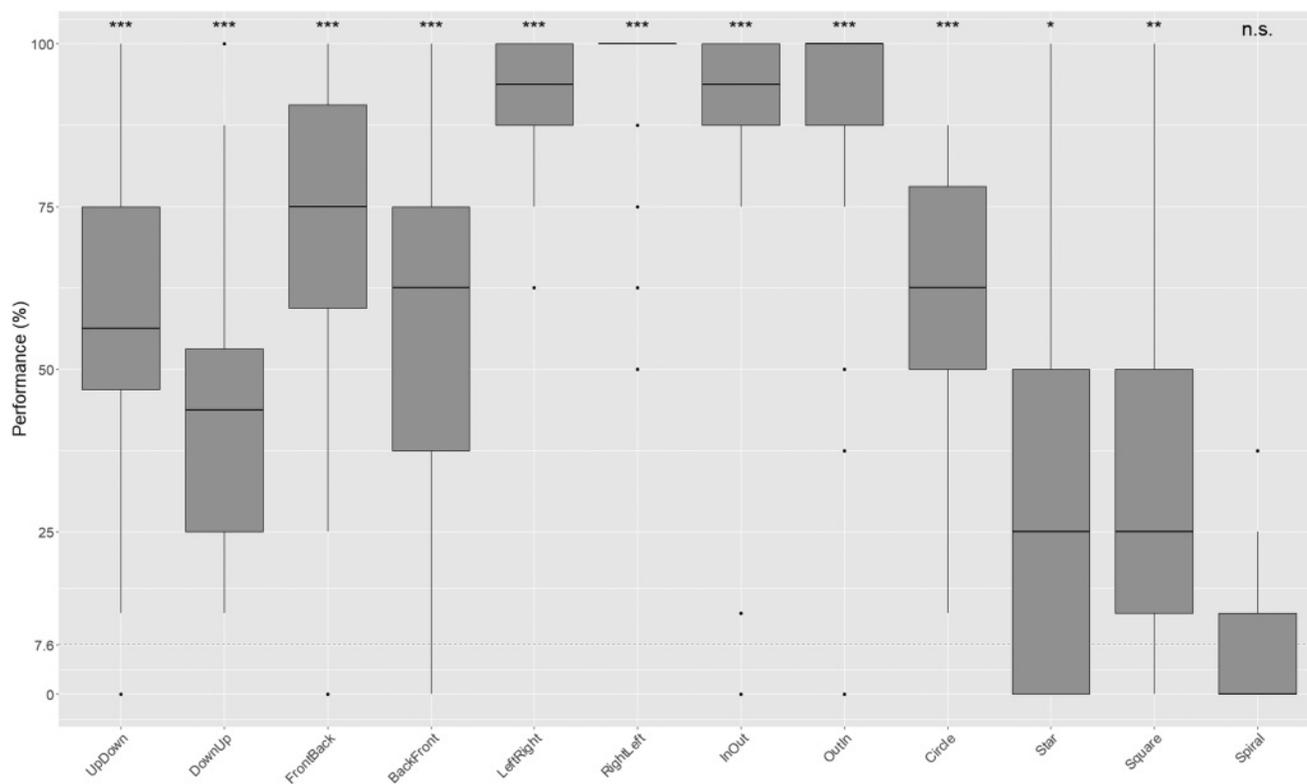
the box plot indicate statistically significant results ($p < 0.05$), meaning the confidence with which we can state that these answers are not attributable to chance. Lower *p*-values (i.e., more asterisks) represent more confidence in the results obtained. As shown, only the results for Spiral were significantly below the chance level. Nonetheless, considerable differences can be observed among the remaining conditions.

In Figure 4, it is possible to observe that certain abilities had more consistent results and a higher percentage of correct answers, such as Left–Right, Right–Left, Inside–Outside, and Outside–Inside (labeled InOut and OutIn in Figure 4). On the other hand, some categories, such as Up–Down and Down–Up, had an overall low performance and a greater dispersion of the data, despite being above chance level. In the case of Up–Down, 50 percent of the participants had 50–75 percent of correct answers, and at the extremes, there were
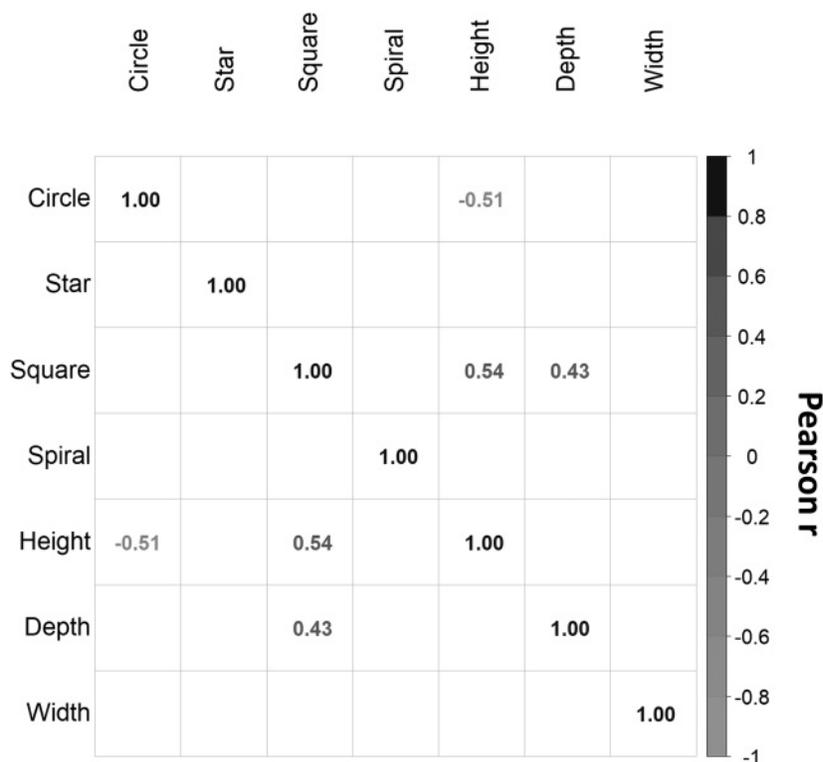
participants with a perfect score and others with a score of 10 percent. Down–Up had an even lower performance, and none of the participants were able to identify it 100% of the time.

As for Front–Back, its performance was higher than in the previous two categories, with half of the participants having scores between 60 percent and 80 percent, and with its minimum and maximum values being 25 percent and 100 percent, respectively. In contrast, Back–Front had a lower performance than the reverse movement, as well as a greater dispersion. Specifically, there were participants who could not identify the movement at all, scoring 0 percent for this category, and others who were able to identify it correctly 100 percent of the time. No dots are shown for Back–Front because there are no outliers, as the data are broadly but homogeneously dispersed. Therefore, we can conclude that there is a lack of consistency in the participants' ability to recognize this movement.

Schumacher et al.   **47**

Figure 5. Pearson correlation matrix. The Pearson correlation coefficient r indicates how strong a linear correlation is, where an r of 1 means there is a perfect positive correlation between the X and Y values, graphed as a straight line, and an r of −1 means there is a perfect negative correlation. In contrast, a value of zero implies there is no correlation between the variables. Cells in which the correlations were not statistically significant were intentionally left blank.

	Circle	Star	Square	Spiral	Height	Depth	Width
Circle	1.00				-0.51		
Star		1.00					
Square			1.00		0.54	0.43	
Spiral				1.00			
Height	-0.51		0.54		1.00		
Depth			0.43			1.00	
Width							1.00

Pearson r scale: 1, 0.8, 0.6, 0.4, 0.2, 0, -0.2, -0.4, -0.6, -0.8, -1

Regarding the archetypal figures, Circle was the category in which the greatest number of participants had scores above 50 percent. There were still results close to 10 percent, however, suggesting that this archetype is perceived with difficulty. As for the Star and Square figures, they showed a similar pattern in which 75 percent of the sample performed below 50 percent for both. Simultaneously, a reduced number of participants were able to recognize them consistently or perfectly. This not only suggests they were hard to identify, but also shows there was little consistency among the participants. Finally, for the Spiral movement, none of the participants were able to score above 40 percent, and 97.5 percent of the sample performed below 25 percent. The Wilcoxon test indicates that these results cannot be considered as different from chance. Therefore, we can conclude that none of the participants, not even those with the highest scores in on the other movements, were able to truly identify the spiral.

Considering the heterogeneity of the results for the different spatial movements, we sought to estimate whether these schemas could be classified separately or as categories encompassed within the orientation and container schemas. For this purpose, we explored the potential associations among the participants' performances in each category. Figure 5 shows the results of a Pearson correlation for each category.

In Figure 5, we observe that each category or sound trajectory tends to correlate with its reverse movement (if any). For instance, Up–Down correlates with Down–Up, Front–Back with Back–Front, and Left–Right with Right–Left. In the case of Inside–Outside and Outside–Inside, however, they correlate with each other, but they also correlate with Front–Back and Back–Front. Up-Down also correlates with Back–Front. As for the rest of the movements classified as image schemas of spatial movement (ISSMs), they do not correlate with

other ISSM movements, only with their inverse movements. These results are apparently related to other studies on sound localization (Blauert 1985; Bregman 1990; Hüg and Arias 2009). For example, as Hüg and Arias state (p. 226):

> Sound localization refers to the perception of a sound source's position on a horizontal plane or azimuth (left–right), a vertical plane or elevation (up–down), and to the perception of the relative distance between participant and source.

We refer to these three axes as height, depth, and width, whereby the distance between participant and source corresponds to depth, the azimuth to width, and the elevation to height.

The results obtained in this study support the idea that each of these axes behaves independently and responds to different perceptual processes. Inside–Outside and Outside–Inside are the only exceptions to this model, as they are associated with the perception of depth.

To evaluate the robustness of this proposed model of perception in three axes, we conducted a dimension-reduction analysis (exploratory factor analysis). This analysis uses association measures among variables to determine whether they share one or more common aspects. For the spatial movement categories, our objective was to establish which performances or results were associated with each other to such an extent that we could conclude that they reflected a common perceptual ability.

Using only the sound trajectories classified as ISSM, which we could also catalog as basic movements, we began our analysis by establishing the most probable number of abilities that could be present in the detection of the eight basic movements. The results of factor analysis are presented in Table 3, in which rows contain the eight basic movements used for the analysis and columns show the three extracted attributes. The numbers in each cell, called loadings, indicate the degree to which these attributes are represented by the variables (the schemas in this case). A value of zero means the attribute is not represented by a given variable at all, and a value of one means it is perfectly represented. By convention, cells are left blank when

Table 3. Rotated Loadings Matrix

Movement	Distance	Azimuth	Elevation
Up–Down			1.00
Down–Up			0.64
Front–Back	0.73		
Back–Front	0.53		0.44
Left–Right		0.97	
Right–Left		0.88	
In–Out	0.88		
Out–In	0.76		

As in the correlation matrix, 0 means the variable is completely independent and has no relation to the spatial movements described in the columns of Figure 5. A loading of 1 implies a perfect association with the construct. Values lower than 0.3 were omitted.

their loading is below 0.3, because this is considered a negligible value. The names of the columns were assigned according to the previously described three axes of sound localization.

As illustrated in Table 3, the loadings suggest the existence of attributes that could be associated with abilities in the detection of sound trajectories in the three spatial dimensions height, depth, and width. The correlation between Inside–Outside and Outside–Inside would be consistently associated with depth. On the other hand, Back–Front has a loading above 0.3 in both depth and height, but the value is rather low.

To make sure that the found attributes are truly consistent, we conducted an internal consistency analysis, which determines how robustly the variables representing the detected attributes are associated with each other. The analysis was done using Cronbach's alpha, in which a value of one represents a perfect internal consistency and zero indicates there is no consistency. An internal consistency of >0.7 is generally considered sufficient to confirm the existence of an attribute and ensure it is being correctly measured. For depth and width, Cronbach's alphas were, respectively, 0.86 and 0.91, which is within the acceptable range. As for height, a Cronbach's alpha of 0.56 was found, which is too low to be considered a consistent attribute. After examining these results in more detail, however, it was

noted that Back–Front negatively affected the consistency of height, so it was consequently removed from this dimension. Once we removed Back–Front, the internal consistency of height changed to 0.78 and was therefore within the acceptable range.

These results suggest that the perception of sound trajectories could indeed be organized into the three axes height, depth, and width. It is important to note that these results do not imply that there is only one ability required for the recognition of every trajectory. On the contrary, they indicate that the participants may be adept at recognizing movements of one spatial axis while being quite unskilled at recognizing movements in others.

Discussion

The main objective of this research was to evaluate the perceptual recognition of different trajectories of sound in space, some categorized as ISSMs and others as archetypal figures, with the latter being frequently used in sound spatialization software in various production contexts. On the other hand, it was of interest to evaluate whether ISSM trajectories could be understood as part of two larger cognitive schemas, the orientation and container schemas.

Regarding the first objective, the results show that the participants' performance in the recognition of these figures are overall relatively low, except for the categories Left–Right, Right–Left, Inside–Outside, and Outside–Inside. The degree to which the remaining figures are recognized ranges from mildly consistent to completely dispersed and, in some cases, below chance level. The results suggest the existence of certain abilities that relate to two essential aspects, however: on one hand, the specialization of the auditory system, and on the other, the physical experience of the agent when interacting with the acoustic world.

The specialization of the auditory system seems to have played a role in the different results obtained for the Front–Back and Back–Front movements, which can be explained by the fact that our auditory system can detect and localize sounds with more accuracy when they come from a frontal horizontal plane than from behind (Moore and King 1999).

Given our specialization on the horizontal axis, this could explain the high recognition rate for the Left–Right and Right–Left trajectories. This specialization of the auditory system in the frontal plane had been previously detected in a study conducted by Stephen Lakatos (1993), in which individuals showed considerable ability in recognizing alphanumeric patterns (letters and numbers) traced with audio signals generated by a frontal array of loudspeakers The case of Inside–Outside and Outside–Inside is an exceptional one due to its high success rate, which suggests the presence not only of a perceptual ability, but also of a cognitive ability in the sense that there is an assessment of sounds that move in and out of the participant's visual field. This result must be observed with caution, however, because it is the only sound trajectory whose recognition could be mediated by an additional variable. Given that one of the loudspeakers was located outside the room, as shown in Figure 3, this sound trajectory was affected by a change in the frequency spectrum owing to the fact that the sound had gone through a wall. Several studies have demonstrated the importance of the visual factor in the detection and localization of sounds coming from outside our frontal area. In other words, the ability to localize and identify sound trajectories in space is primarily related to our visual sense, and only secondarily to our auditory system (Perrott et al. 1990; Marentakis and Mcadams 2013). Vision is therefore a determining factor in the detection and localization of sound trajectories. Moreover, as Lennox and Dean (2009, p. 259) state, "how accurately we hear all these items is not clear since hearing in real environments is normally accompanied by other senses; perception is inherently multimodal." In the case of our study, which focuses on the auditory detection of these sound trajectories, the absence of visual components suggests that there is a specialization of the auditory system when the stimulus occurs outside our frontal visual field. This was the case for Left–Right, Right–Left, Inside–Outside, and Outside–Inside, as opposed to the lower detection rates for the other evaluated movements.

In this regard, Marentakis and Mcadams (2013, p. 2) state that "similar to sound localization, auditory motion perception is more accurate for

horizontal versus vertical movements, frontal versus lateral incidence, and broadband versus narrowband sounds." Although the data collected in our study support the first statement regarding a greater accuracy of our perception on a horizontal plane rather than a vertical one, the data do not seem to support the claim that frontal detection is more accurate than lateral detection. Nonetheless, more evidence is required to confirm this.

On the other hand, previous experience and knowledge on the spatial behavior of sound in the physical world could have influenced the different recognition rates for Up–Down and Down–Up. In the physical world, we frequently interact with objects that follow a downward trajectory, rather than with objects that move in the opposite direction. This could explain why the results for Down–Up were so widely dispersed, ranging at one end from individuals who were close to the chance level and, at the other extreme, to others with an almost perfect score. This is the stimulus with the lowest recognition rate of all ISSM categories, closer to the results obtained for the archetypal geometric figures. Something similar, though in the opposite sense, occurs with the Circle movement. We frequently have the experience of sound stimuli moving around us. An example of this is how we perceive the movement of some insects, which we tend to describe as circular, even when the insect does not necessarily follow this trajectory. This phenomenon may have affected the higher recognition rate for Circle. It is therefore possible to conceive this category of movement as part of the container schema rather than as an archetypal geometric figure.

As for the archetypal figures, their recognition was low. Except for the previously mentioned case of Circle, the participants' performance was either below or close to chance level. Therefore, it appears that none of the elements we have so far considered to be relevant for the detection of sound trajectories in space play a role in the recognition of these figures. Moreover, the results indicate that, apart from Circle, none of these geometric figures are perceived as such by the participants. This doesn't invalidate their pertinence as technical and aesthetic resources for sound spatialization,

however, especially in situations in which there is no association between visual and auditory stimuli or no intention of generating one, as is generally the case in acousmatic music.

Regarding the second objective of our research, the results suggest that each sound trajectory is positively correlated with its inverse movement, as could be seen in Figure 5. In other words, the participants who were able to successfully identify the Back–Front category could also correctly detect Front–Back, for example. Apart from these correlations, no other correlations could be found to support the notion that ISSM movements are encompassed within two larger schemas, that is, the orientation and container schemas. Nonetheless, the analysis does suggest that the perception of these movements may be associated with the three spatial axes height, depth, and width. These three axes could be associated with the categories we had classified as part of the orientation schema, however, except for Inside–Outside and Outside–Inside, which are associated with the depth axis in Table 3 and the container schema in Table 1. This seems to support the idea that each spatial axis can be considered a schema of its own, but more evidence is required to explain this association.

As previously stated, these results coincide with various studies on the perceptual localization of sounds in space (Blauert 1985; Hüg and Arias 2009), but because much of the research in this area has utilized sound sources that are fixed in space, it is interesting to find that the three spatial axes are also relevant in the perception of moving sounds. This aspect has only been partially studied so far (Marentakis and Mcadams 2013; Barrett and Crispino 2018).

Conclusion

The results of this study indicate that it is generally difficult to recognize sound trajectories in space when there are no supporting visual stimuli. Perceiving the localization of sounds and trajectories in space is a multimodal ability (Lennox and Dean 2009; Marentakis and Mcadams 2013). This makes research into the perception of sound trajectories

in controlled conditions more difficult because, as Lennox and Dean (2009, p. 260) point out:

> Although elevation perception in the laboratory is solely due to the pinnae, perception in the vertical plane in the real world can utilize interaural differences. In addition, head movement and perceiver movement can improve the robustness of spatial hearing.

Still, there are some exceptions, such as the Left–Right, Right–Left, Inside–Outside, and Outside–Inside trajectories, which suggest that the auditory system specializes in the recognition of these types of movements. In the case of the archetypal geometric figures, they all had a low recognition rate except for Circle, which could imply that this trajectory should be understood or classified as within the container schema or as a schema in itself rather than as an archetypal geometric figure. Additionally, the recognition rate for this movement could possibly be influenced by the agents' experience in the acoustic world, where this type of sound movement is frequent, or at least described as such. Despite this exception, the participants' low performance in the detection of the rest of the archetypal geometric figures support the idea that these movements could be functional as aesthetic artefacts rather than as perceptible spatial movements. Therefore, in creative contexts, it could be taken into consideration that these figures will most likely not be perceived or recognized as such. This does not necessarily invalidate the aesthetic value associated with the exploratory use of this and other spatial movement figures. Our results also suggest that the other variables that were excluded from our experimental design, such as the morphological characteristics of sounds, frequency variations, dynamic envelopes, and Doppler effect, may be necessary for the perceptual recognition of these sound trajectories. The latter has already been pointed out in studies by Lakatos (1993), Schaeffer (1966), and Barrett and Crispino (2018), among others, where the shape of the amplitude envelope of the stimuli with different types of attacks could influence the recognition of spatial movement patterns.

On the other hand, our results indicate that the sound trajectories classified as ISSM correlate with each other individually rather than as part of a larger spatial-movement schema. Thus, the recognition of these trajectories seems to imply the existence of perceptual abilities that fall into three spatial dimensions, height, depth, and width, which have already been described in the literature concerning the localization of sound at fixed points in space.

As for future research, it seems necessary to produce qualitative information regarding the identification of sound trajectories, which could allow for a better understanding of how sound is perceived in the absence of visual stimuli.

Acknowledgments

References

Barrett, N., and M. Crispino. 2018. "The Impact of 3-D Sound Spatialisation on Listeners' Understanding of Human Agency in Acousmatic Music." *Journal of New Music Research* 47(5):399–415. 10.1080/09298215.2018.1437187

Barton, B. K., et al. 2013. "Developmental Differences in Auditory Detection and Localization of Approaching Vehicles." *Accident Analysis and Prevention* 53:1–8. 10.1016/j.aap.2012.12.040, PubMed: 23357030

Blauert, J. 1985. "The Psychophysics of Human Sound Localization by Jens Blauert." *Journal of the Acoustical Society of America* 77(1):334–335. 10.1121/1.392109

Boroditsky, L. 2000. "Metaphoric Structuring: Understanding Time through Spatial Metaphors." *Cognition* 75(1):1–28. 10.1016/S0010-0277(99)00073-6, PubMed: 10815775

Bregman, A. S., 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, Massachusetts: MIT Press.

Costello, A., and J. Osborne. 2005. "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis." *Practical Assessment, Research and Evaluation* 10(7). 10.7275/jyj1-4868

Costello, A., and J. Osborne. 2019. "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis." *Practical Assessment, Research, and Evaluation* 10:Art. 7.

Cronbach, L. J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16(3):297–334. 10.1007/BF02310555

de Winter, J. C. F., and D. Dodou. 2012. "Factor Recovery by Principal Axis Factoring and Maximum Likelihood Factor Analysis as a Function of Factor Pattern and Sample Size." *Journal of Applied Statistics* 39(4):695–710. 10.1080/02664763.2011.610445

Emmerson, S. 2007. *Living Electronic Music*. London: Ashgate.

Fellgett, P. 1975. "Ambisonics: Part One—General System Description." *Studio Sound* 17(8):20–22.

Frank, M. 2014. "Localization Using Different Amplitude-Panning Methods in the Frontal Horizontal Plane." In *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, pp. 41–47.

Gibson, D. 1997. *The Art of Mixing: A Visual Guide to Recording*, vol. 236. Vallejo, California: MixBooks.

Hayton, J. C., D. G. Allen, and V. Scarpello. 2004. "Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis." *Organizational Research Methods* 7(2):191–205. 10.1177/1094428104263675

Hüg, M. X., and C. Arias. 2009. "Estudios sobre la localización auditiva en etapas tempranas de desarrollo infantil." *Revista Latinoamericana de Psicología* 41(2):225–242.

Kendall, G. S. 2010. "Spatial Perception and Cognition in Multichannel Audio for Electroacoustic Music." *Organised Sound* 15(3):228–238. 10.1017/S1355771810000336

Kendall, G. S., and M. Ardila. 2008. "The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata." In R. Kronland-Martinet, S. Ystad, and K. Jensen, eds. *Computer Music Modeling and Retrieval: Sense of Sounds*. Berlin: Springer, pp. 125–138.

Küpper, L. 1998. "Space Perception in the Computer Age." In F. Dhomont, ed. *L'espace du son I*. Brussels: Musiques et Recherches, pp. 59–61.

Lakatos, S. 1993. "Recognition of Complex Auditory–Spatial Patterns." *Perception* 22(3):363–374. 10.1068/p220363, PubMed: 8316523

Lakoff, G., and M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lee, D., W. Song, and A. Horner. 2021. "A Head-to-Head Comparison of Emotional Characteristics of the Violin and Ehru on the Butterfly Lovers Concerto." In *Proceedings of the International Computer Music Conference*, pp. 289–294.

Lennox, P., and R. T. Dean. 2009. "Spatialization and Computer Music." In *The Oxford Handbook of Computer Music*. Oxford: Oxford University Press, pp. 258–273.

Lynch, H., and R. Sazdov. 2017. "A Perceptual Investigation into Spatialization Techniques Used in Multi-channel Electroacoustic Music for Envelopment and Engulfment." *Computer Music Journal* 41(1):13–33. 10.1162/COMJ_a_00401

Marentakis, G., and S. Mcadams. 2013. "Perceptual Impact of Gesture Control of Spatialization." *ACM Transactions on Applied Perception* 10(4):Art. 22. 10.1145/2536764.2536769

Moore, D. R., and A. J. King. 1999. "Auditory Perception: The Near and Far of Sound Localization." *Current Biology* 9(10):R361–R363. 10.1016/S0960-9822(99)80227-9, PubMed: 10339417

Parsehian, G., et al. 2015. "Design and Perceptual Evaluation of a Fully Immersive Three-Dimensional Sound Spatialization System." In *Proceedings of the International Conference on Spatial Audio*. Avaliable online at hal.archives-ouvertes.fr/hal-01306631. Accessed November 2021.

Perrott, D. R., et al. 1990. "Auditory Psychomotor Coordination and Visual Search Performance." *Perception and Psychophysics* 48(3): 214–226. 10.3758/BF03211521, PubMed: 2216648

Pulkki, V. 1997. "Virtual Sound Source Positioning using Vector Base Amplitude Panning." *Journal of the Audio Engineering Society* 45(6):456–466.

Pulkki, V., J. Huopaniemi, and T. Huotilainen. 1996. "DSP Tool to 8-Channel Audio Mixing." In *Proceedings of the Nordic Acoustical Meeting*, vol. 96, pp. 307–314.

Raîche, G., et al. 2013. "Non-Graphical Solutions for Cattell's Scree Test." *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 9(1): 23–29.

Rumsey, F. 2002. "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm." *Journal of the Audio Engineering Society* 50(9):651–666.

Satongar, D., et al. 2013. "Localisation Performance of Higher-Order Ambisonics for Off-Centre Listening." BBC Research and Development White Paper WHP 254. British Broadcasting Corporation.

Schaeffer, P. 1966. *Traité des objets musicaux.* Paris: du Seuil.

Schumacher, F., and C. Fuentes. 2017. "Space–Emotion in Acousmatic Music." *Organised Sound* 22(3):394–405. 10.1017/S1355771817000449

Schumacher, F., et al. 2021. "Perceptual Evaluation of Sound Trajectories in Space." In *Proceedings of the International Computer Music Conference*, pp. 281–288.

Searle, C. L., et al. 1975. "Binaural Pinna Disparity: Another Auditory Localization Cue." *Journal of the Acoustical Society of America* 57(2): 448–455. 10.1121/1.380442

Simon, L. S. R., H. Wuethrich, and N. Dillier. 2017. "Comparison of Higher-Order Ambisonics, Vector- and Distance-Based Amplitude Panning Using a Hearing Device Beamformer." In *Proceedings of the International Conference on Spatial Audio*, pp. 131–137.

Talmy, L. 1983. "How Language Structures Space." In H. L. Pick and L. P. Acredolo eds. *Spatial Orientation: Theory, Research, and Application*. Berlin: Springer, pp. 225–282.

Talmy, L. 2003. "Commentary: The Representation of Spatial Structure in Spoken and Signed Language." In K. Emmorey, ed. *Perspectives on Classifier Constructions in Sign Language*. Abingdon, UK: Routledge, pp. 169–195.

Wilcoxon, F. 1945. "Individual Comparisons by Ranking Methods." *Biometrics Bulletin* 1(6):80–83. 10.2307/3001968