Let me listen to where you are
Spatial dimension resources in audio stories can increase imagery, transportation, attention, and recall

Emma Rodero. Department of Communication. Pompeu Fabra University (UPF). Roc Boronat, 138. 08108, Barcelona, Spain. emma.rodero@upf.edu

Laura Romero. UOC. Universitat Oberta de Catalunya. Rambla del Poblenou, 156, 08018 Barcelona, Spain. lauraromerovall@gmail.com

Abstract

This study analyzes the influence of the audio spatial dimension on the listener's cognitive processing. Participants listened to different audio stories with four sound verbal and nonverbal resources (dialogue; dialogue and sound effects; dialogue and sound filters; dialogue and sound shots) representing the space and perspective of the scene. They rated the stories concerning imagery, transportation, attention, and performed a cued recall test. The results showed that the audio stories with nonverbal sound resources achieved higher imagery and transportation levels than stories with verbal information. The stories with nonverbal resources, except sound effects, obtained the highest levels of attention and recall. These results reveal the crucial role of the spatial dimension in audio processing.

Keywords: spatial dimension; audio stories; imagery, transportation; attention, recall

Thanks to the possibilities that digital technology offers, we are in a period of boom in audio fiction formats. Listening through smartphones and headphones has contributed significantly to the development of podcasting and audio fiction series (Bull, 2005). The last survey about podcast consumption reveals podcast listeners spend 5 hours and 39 minutes per day (on average) listening to audio (Edison Research, 2019). Another recent study by BBC Story Works demonstrates the persuasive power of branded podcasts. Podcasts have a vast and growing audience, but also audiobooks. Half of the US population has listened to an audiobook, according to Edison Research (2019). With podcasts and audiobooks, in homes, the use of smart speakers has become more usual. A recent survey shows that one-quarter of North American adults have a smart speaker in their home (Pew Research Center, 2019). This resurgence of audio formats opens the possibilities to produce more extensive narratives in the transmedia universe (Dann, 2014). Hence, now it is an excellent moment to study how some audio production resources can improve the listeners' cognitive processing.

Literature about visual processing and media is vast, but there are few studies about audio (auditory system, processing, or audio in media), and even less about the role of the spatial dimension in audio production. This research defines the spatial dimension as the sound references that identify the space or location in an audio story. But why is this spatial dimension so important in audio processing? Imagine a woman who walks down the street, climbs stairs, and enters an ample indoor space with high ceilings. A man is waiting for her and starts a conversation from the center of the room. When we listen to this scene, our auditory system scans for two types of information: *what* the sound is -recognition- (who is talking and what the objects in the scene are), but also *where* the characters and objects are placed -space- and where they came from in the space -localization-. In the previous example, the spatial dimension should be indicated, apart from verbal information, using a reverberation (large room), sound effects (steps, doors, etc.), and sound shots (distances between the two characters). If spatial information is a crucial feature in human perception, this information should also be indicated in a fiction audio story to process sound as it would occur in a real situation.

At least theoretically, we can deduce that this spatial information is critical to placing the listener inside the scene and stimulating mental images about this specific place. However, this spatial information identified by nonverbal features such as sound effects, filters or reverberations, or distances usually is not taken into account in audio production. Rodero (2009) found that only 30 percent of award-winning audio stories used nonverbal sound resources to identify the location. Therefore, the goal of this study is to analyze the influence of the spatial dimension in the listener's cognitive processing by studying whether resources such as sound filters, sound effects, and sound shots, added to the verbal information, may stimulate imagery, induce transportation, and improve the levels of self-perception attention and recall on the part of the listener. The conclusions of this research can help advance in the study of audio processing in media and improve the sound design in media contents.

Auditory Spatial Dimension and Processing

In an audio story, the use of nonverbal sound resources, as sound effects, sound shots or filters, may improve the listeners' cognitive processing for two main reasons. The first one is that a stimulus that combines verbal and nonverbal resources can be dually coded in memory. The second one is that nonverbal resources add spatial references, which may be relevant information for processing stories.

The first idea is based on a dual-processing paradigm. The dual-coding theory (DCT) posits that information is stored in memory in a dual way –visual imagery and verbal associations (Paivio, 1991; Sadoski & Paivio, 2004). The stimulus can be encoded and stored in memory in two separate codes, one devoted to nonlinguistic objects and events (non-verbal code) and another for language processing (verbal code). First, the visual or imagery system is formed by objects and images spatially organized and represented as mental images. Secondly, the verbal system represents the structural and functional properties of the language (Paivio, 1991). In line with the dual-stream auditory processing models (Rauschecker & Scott, 2009), different brain parts are activated when verbal and nonverbal auditory information is processed in parallel (Herrmann et al., 2011).

The second idea is based on the relevance of spatial information for processing messages. According to the Baddeley and Hitch working memory model (Baddeley & Hitch, 1974; Baddeley et al., 2015), language is processed by a phonological loop, whereas a visuospatial sketchpad processes visual information. The phonological loop is responsible for the temporary storage of verbal and acoustic information (Baddeley et al., 1998; Baddeley, 2000). The visuospatial sketchpad is defined as a repository where visual and spatial information is kept for short periods. Therefore, while the subject records the verbal material through the phonological loop, the visuospatial agenda records and maintains the visual image or mental representation of that object. The part relevant for our study is that the visuospatial sketchpad is essential to process the appearance of objects and create mental images of them (what), but also to locate the stimulus in space (where), and to identify the spatial and geographical orientation (wherefrom). The visual part of the message can be activated not only by images but also by sounds. Some authors have hypothesized that, in the initial recognition of audio, visual or mental imagery is activated (Lewis et al., 2004). If we listen to a dog barking, we will see the image of a dog in our minds, and depending on the sound features, we can also imagine his size and place it in a location to a certain distance. In narratives, the situation models determine that spatiality is one of the event indices to comprehend the action of the story (Zwaan, Langston & Graesser, 1995). In addition, some studies have shown that changes in the spatial dimension can modify information processing and memory of media stimuli (Meitz, Meyerhoff, & Huff, 2019). Therefore, all these theories reveal the influence of the spatial dimension to process information ("where" and "wherefrom" added to "what" data).

According to sound theory by Rodero (2005), the spatial dimension is the set of sound references that identify the space or location of an audio story and the distances among objects and characters. This spatial dimension is divided into two sub-dimensions – space (where) and perspective (wherefrom). In this approach, the sound space is the physical place or the action's location (Forrester, 2012). Acoustically, space can be represented using different sound resources.

First, space can be identified by the narrator or in the characters' dialogue by explicitly mentioning "where" the characters or objects are (i.e., "I am very happy to visit Venice."). Using this verbal resource is the easiest way to represent spaces but eliminates dual processing, as it is only mentioned in the verbal part of the story. We will call this verbal information as "verbal sound resource." Secondly, music and sound effects can also describe locations. One of the functions that music can play in audio production is to identify places. For example, suppose we use a Mariachi song in an audio story. In that case, listeners can infer that there is something related to Mexico in the plot: the story is set in Mexico, the characters are there, or they are talking about this country. Sound effects are also powerful to recognize places and sound objects. Sound elements can describe objects (i.e., the sound of our keys) and spaces (i.e., the ocean's sound to identify the beach). Therefore, sound effects can provide "what" and "where" information. The advantage of music and sound effects is that we perceive an acoustic variation different from the verbal part. This contrast could increase the listener's attention and recall, thanks to the dual coding (Rodero, 2012). Third, the concrete space is also identifiable by the echoes or reverberations produced by the sound waves reflected on the surface of this location or by the absence of these elements when there are no physical obstacles. The existence of echoes or reverberations and the absence of them places the listener in small or large, open or closed space. There are no echoes in an open space, but we can find echoes and reverberations in closed places like a bathroom or a cathedral. Like music and sound effects, these elements also have an acoustic differentiation from the verbal part, but they have an important difference. In postproduction, reverberation is added to the audio message. Therefore, it is not an independent sound, like music or sound effects; reverberation is fused with the verbal message and the rest of the sounds.

Along with space, the sub-dimension of perspective indicates the distances among sound objects or characters in the scene, the “wherefrom” information (Beck, 1998; Chion, 2016). If we place the characters or some sound objects in different slots in the scene, we will create different distances among them. This information allows us to provide depth – or perspective – (Chion, 2016). Rodero (2012) called these distances *sound shots*. The sound shots (or different distances) locate each character and object in the scene. In this manner, listeners can get information about where the sound comes from and establish the distance and movement of these objects. The construction of this perspective depends on where we locate the listening point (Beck, 1998; Rodero, 2009), the reference place from which all other sound shots are formed from the foreground to the background. The characters and objects next to the listening point (or the listener) will sound louder; those that are far away will sound softer. As this space distribution simulates the structure of real space and normal hearing, sound shots provide a greater sense of reality. Like filters, sound shots are not independent sounds, but they are fused or melted with the verbal or the rest of the sound features. Listeners can perceive that characters or objects are close or far away because they sound louder or softer.

We will call all these sound elements (effects, shots, and filters) as “nonverbal sound resources” to differentiate them from the verbal resources (dialogue with the actors’ voices). Figure 1 illustrates the spatial dimension composition that we will test in this study.

[Figure 1 here]

Spatial Dimension and Cognitive Processing

Spatial Dimension and Imagery

Imagery is defined as the mental images created in a person's mind evoked by a stimulus. According to the dual-coding theory (DCT), the creation of mental images facilitates recall of information thanks to the benefit of being dually coded in memory (Paivio, 1991). Mental imagery is both visual and spatial. Mental images have spatial characteristics that depend on the spatial nature of the information processed (Mellet et al., 1996). Therefore, the different processing types derived from the kind of data (what and where) have proven to be very relevant for mental imagery. However, the creation of mental images can be derived not only from images but also from sounds. This potential of sound to stimulate imagery is supported on an empirical basis, thanks to different studies (Bolls, 2002; Bolls & Lang, 2003; Rodero, 2012). Goossens (1994) showed that high-imagery commercials could activate both systems, producing more in-depth cognitive processing. Miller and Marks (1997) compared three types of sound messages: a message with sound effects, a verbal message with a high level of vividness, and a message with instructions to imagine. The sound effects had the highest impact on imagery, while the verbal instructions had the lowest. Other studies concluded that sound effects increased the capacity to create mental images (Bolls, 2002). The sound effects directly activated perceptual structures in the memory that contained information, and this information helped the subjects to create more vivid mental images (DCT). Rodero (2012) analyzed how sound effects and sound shots could stimulate imagery and attention in audio fiction. This author used descriptive sound effects to represent the objects in the scene (i.e., the sound of a car or a rooster singing). The results showed that the stories with sound shots achieved the highest level of imagery, over sound effects and stories with verbal information. According to this study, sound shots achieved the best imagery and attention levels, thanks to the sense of reality through which to better imagine the story. In the second place, sound effects improved the stories' cognitive processing, thanks to the degree of novelty introduced by them in contrast to verbal information. Therefore, the few studies existing about this topic agree that sound effects and sound shots are potent in creating mental images, mostly sound shots. Yet, these studies have not explicitly used sound effects to develop the spatial dimension. Rodero (2012) studied these features indicating the description of objects and not specifically to create a spatial dimension. Also, as far as we know, there are no studies about filters in media production. We hypothesize that space “where” and perspective “wherefrom” of the spatial dimension should intensify the creation of mental images, as we can imagine the location and movement of the objects. These considerations allow us to formulate the first hypothesis:

H1: The stories with verbal and nonverbal sound resources, in particular, sound shots and effects, will achieve a higher listener's imagery compared to the stories with only verbal information.

Spatial Dimension and Transportation

Transportation is defined as an imaginative experience or mental journey inside the story (Gerrig, 1993; Sood, 2002). Listeners have the perception of being immersed or lost in a story (Nell, 1988) by adopting the characters' identity and experiences (Cohen, 2001). The term transportation is indicative of the importance of the spatial dimension: transportation to a particular place or space. Also, according to some authors, the level of transportation depends on the stimulation of mental images and, therefore, it should be related to the spatial dimension (Green & Brock, 2000; Green, 2004; Zheng, 2014). Some studies applied to radio advertisements found higher involvement levels when radio ads contained a high level of imagery, even when participants were doing a dual-task (Bolls & Muehling, 2007). We can hypothesize that if listeners can draw in their minds space and imagine themselves moving with the characters, then transportation should be high. Therefore, the spatial dimension in audio fiction, provided by the nonverbal sound resources, should help the listeners to submerge mentally in the space described and aid to "transport" the listeners inside the story. This is our next hypothesis:

> **H2:** The stories with verbal and nonverbal sound resources, in particular, sound shots and effects, will achieve a higher level of listener's transportation compared to the stories with only verbal information.

Spatial Dimension and Attention

According to the information processing approach, attention is a cognitive process that selects information to be processed preferentially (Chung, Golomb & Turk-Browne, 2011). When a stimulus is perceived, attention selectively allocates the cognitive resources on one aspect of the information and filters out the distracting stimuli (Johnson & Proctor, 2004). Filtering information is a necessary process, as attention is a limited resource for information processing (Lang, 2000; Oberauer, 2019). In this process, spatial attention has a strong influence on differentiating and filtering different sounds in our environment (Kong et al., 2014). There are two kinds of attention processes –controlled and automatic (Schneider & Shiffrin, 1977; Lang, 2000). Controlled processing is this that individuals use in a conscious way to process information; thus, directed by their own goals. This is a top-down process. In contrast, automatic processing or exogenous attention is influenced by external stimuli, there is not an active control, it is not intentional and, consequently, is faster and requires less effort to process the information (i.e., after learning to drive a car, many actions can be performed without being consciously aware of each one). This is a bottom-up process of attention or stimulus-driven attention (Oberauer, 2019).

Lang, Potter, and Bolls (2009) showed that visual encoding is a relatively automatic process, whereas verbal encoding is a relatively controlled process; thus, visual information is processed faster than verbal and with less effort (Schneider & Chein, 2003). By creating visual images in the listeners' minds (Rodero, 2012), a sound can be processed faster than verbal information. If we hear the word "water," we have to associate this code (the oral language) with the concept and its image. When it is our mother language and, due to practice, verbal processing can become automatic (Schneider & Shiffrin, 1977). However, this encoding process can be challenging if the word is "wasser" and the individual is learning German. In this case, this task requires more effort, and it is slower than merely hearing the sound of water. Everyone can imagine a dog when listening to a dog barking, regardless of his/her mother language. Also, as sound conveys information about the object (location or size), imaging a concrete object can be a more automatic process than listening to the word. Therefore, if we include nonverbal sound resources in production, which may be easier to process than verbal information, we could achieve a claim of attention or an orienting response (OR), by eliciting an acoustic contrast in the message. A high level of stimulation and variation may increase attention during the stimulus (Olsen, 2002; Potter, 2006; Rodero, 2019). The nonverbal sound resources could provoke this level of variation. Some studies have demonstrated that the use of production effects and voice and music changes can provoke *orienting responses* (ORs) (Potter et al., 1997; Potter, 2000). Both studies by Rodero (2012; 2019) supported this idea. This author found that sound effects and sound shots increased the listener's attention compared to verbal information in audio stories. (Rodero, 2012). Also, but in commercials, Rodero (2019) studied the capacity of different sound elements, descriptive sound effects among them, to enhance the listener's attention and recall. In this research, the sound of the objects in the commercial (i.e., a computer) achieved the highest level compared to music and verbal messages. However, the level of attention has not been studied with filters or with sound effects indicating the spatial dimension. These ideas lead us to formulate the third hypothesis:

> **H3:** The stories with verbal and nonverbal sound resources, in particular, sound shots and effects, will achieve a higher level of listener's attention compared to the stories with only verbal information.

Spatial Dimension and Memory

Matlin (2005) defines memory as the process of maintaining information over time for our survival. Memory covers three important cognitive processes: encoding, storage, and retrieval of information (Baddeley, Eysenck & Anderson, 2015). Individuals process messages according to the filter of attention. If this encoding process is efficient, the individual stores this information in memory (Lang, 2000). The recovery will be a mental representation of the stimulus, a construction based on the individual's schemes (Barlett, 1995). This is why we can conclude that memory is closely linked to attention and mental imagery (Mazard et al., 2004). According to the dual-coding theory, we can better recall any information if verbal information is reinforced with visual images (Paivio, 1991). Cohen, Horowitz, and Wolfe (2009) have shown that visual memory is superior to auditory, so adding images (or mental images) to verbal information could improve recognition of information. A recent study applied to films demonstrated that memory for audiovisual dynamic scenes was superior to memory for visual dynamic scenes (Meyerhoff & Huff, 2019). Concerning sound, some authors have found (Bolls & Lang, 2003) that a message with a high imaginative load is more effective in long-term memory than a message with a little imaginative load. Also, memory is closely related to transportation. A study showed that short-term spatial memory improved when the immersion level increased in a complex virtual environment (Johnson & Adamo-Villani, 2010). Also, the distinct processing of "what" and "where" information is crucial for working memory (Lehnert & Zimmer, 2006); thus, nonverbal sound resources indicating the spatial dimension could help to this goal. Other experiments have demonstrated that the more complex a message was structurally (multiple voice changes, sound effects, music onsets, and/or production effects), the higher the memory (Potter & Callison, 2000; Lang, 2000; Potter & Choi, 2006). Recently, Lang et al. (2015) analyzed the number of cognitive resources that can be automatically allocated and required to process audio messages. Recognition accuracy resulted from a combination of automatic resources (depending on the number of audio changes in the message) and the required resources (depending on the information introduced by the audio changes). If the level of structural features was low, the information introduced by them (low or high) did not modify recognition. But if the level of structural features was low and the information introduced was high, then recognition was low, as more resources were required to process the message. The highest level of recognition was the combination of a high level of audio changes and a low level of information. As previous research has shown that the more audio changes, the more resources are automatically allocated (Lang, Zhou, Schwartz, Bolls, & Potter, 2000), it is logical to deduce that the stories with nonverbal resources (more audio changes) will get a higher level of recall compared to verbal stories. Also, Rodero (2019) compared several strategies applied to radio messages: verbal appeals, music, and sound effects and showed that sound effects achieved the best level of recall. Therefore, we can hypothesize that these effects with sound shots will get the best level of recall.

H4: The stories with verbal and nonverbal sound resources, in particular, sound effects and shots, will achieve a higher level of listener's recall compared to the stories with only verbal information.

Method

The experiment was a 2 (audio stories) x 4 (versions of each story) mixed factorial design. The same story was presented in four different ways, eight stories in total. Each group of participants listened to two different stories. These two different stories (story A and story B) had similar narrative characteristics but different places. The story was a within-subject factor. There were four versions of each story. On the one hand, there were verbal stories with only dialogue (stories only with the actors' voices). On the other hand, there were stories with verbal and nonverbal sound resources (voices + sound resources, such as sound effects, filters, and sound shots). The versions were a between-subjects factor. The stories were presented using different combinations and orders. Every participant listened to the two different stories (A and B). The dependent variables were imagery, transportation, attention, and recall. Group 1 listened to verbal stories A and B with only dialogue, Group 2 listened to stories A and B with sound effects, Group 3 listened to stories A and B with sound shots, and Group 4 listened to stories A and B with sound filters. The order of presentation of the stories changed in each group. Half of the participants listened to the stories in the order A-B and the other half in the order B-A.

Participants

The participants were 104 students at a major university. Fifty-four were females, and 46 were males (18-25 years old). 99.5% of participants had no prior experience in listening to audio stories. These participants were distributed in four gender-balanced groups. They were selected by combining the lists of students enrolled in four different courses. A statistical power analysis was performed for sample size estimation. The effect size (ES) in this study was .33, considered to be between a medium and large size according to Cohen's criteria. With an alpha = .05 and power = 0.80, the projected sample size needed with this effect size was 104. Thus, we recruited this number of participants (N = 104).

Materials and Stimuli

The different stories were specially written and designed for this research. There were two audio stories of two and a half minutes presented in four different versions: with only dialogue; dialogue and sound effects; dialogue and sound filters (reverberation), and dialogue and sound shots. Therefore, the stories were the same in all the versions. The two stories were written following a linear structure and based on a dialogue between a man and a woman. Both stories had similar narrative characteristics: the same number of actions (six), interrogative and exclamatory sentences (25), dialogue from each character in balance (27 sentences in Story 1, and 23 sentences in Story 2), turning points placed towards the second half of the story (from 1 minute and 20 seconds onwards), and the same length in each narrative block (between 40 and 50 seconds). We placed the same amount of nonverbal resources sound resources (six) in each version, balancing the number of appearances of new sound elements or sound changes in the four versions (six times). The sound resources were also distributed in the same timing during the narrative (every 15-20 seconds) and had the same duration in all the versions (between three and six seconds). All the nonverbal sound resources were congruent with verbal information, as some studies have shown that if auditory information matches visual information, memory increases (Meyerhoff & Huff, 2016). All the stories had the same general volume controlled by mastering the intensity levels to -3dB, except the stories with sound shots due to the difference in distances, indicated with different loudness. The space needed to have acoustic characteristics that allowed the addition of filters such as reverberation. Therefore, the space selected for Story A was a bathroom, and the space chosen for Story B was a parking lot. Both stories were produced with the voices of the same actors. The story told of a man and a woman casually meeting and living in a tense situation (in the bathroom of a nightclub - Story A; and in a parking lot - Story B). Throughout the story, their relationship becomes closer, ending in a friendship or love relationship.

Dependent Variables

The dependent variables were imagery, transportation, attention, and recall.

First, regarding imagery, we used the scale by Ellen and Bone (1991), with the revision by Babin and Burns (1998), and the scale by Sheehan (1967). Imagery activity was measured using two variables: vividness and quantity/ease. Vividness is the largest dimension of mental images and refers to the quality and intensity of the mental image (Morris & Hampson, 1983; Paivio, 1991). Participants answered the next questions for vividness: How clear and vivid are the images that you have created in your mind? (as the general level of imagery); How clear and vivid are the images about the space where the story takes place? (as the imagery of the space); How clear and vivid are the images of the characters of the story? (as the imagery of the characters). The alpha coefficient for this scale was α=.763, with high internal consistency. Quantity/ease refers to the number of stimulated mental images (Childers, Houston & Heckler, 1985; Ellen & Bone, 1991). Participants answered the next questions: how many images were stimulated; what degree of difficulty they experienced in generating the images, and how fast they created these images. The alpha coefficient for this scale was α=.878, with high internal consistency. The answers to these questions were measured according to the scale by Sheehan (1967) (see Table 1). The questionnaire also included two questions about if the participants had imagined the space where the story was placed and the story's characters. The questions were as follows: "How clear and vivid are the mental images about the space where the story takes place?"; and "How clear and vivid are the mental images of the female and male characters?"

In addition to the scales, participants answered open questions about the mental images they had experienced during the listening task. They responded in only two sentences the task "Describe the space where the story takes place and describe briefly what you have imagined." We observed that the descriptions made by the participants about the mental images they created could be about real or invented information, so the responses to the question *Describe briefly what you have imagined* were coded in these two categories. Real information was data mentioned explicitly in the story while invented information was data that have not been mentioned in the story but that the participants added to their descriptions. For example, some of the participants added adjectives about the characters, the space, or the objects of the scene, which were not in the stories. The answers to the open question *describe the space where the story takes place* were coded according to the number of adjectives mentioned about the space or the space objects. We coded these descriptions following this rule: 1 or 2 adjectives = 1 (i.e., a small bathroom), 3 or 4 adjectives = 2 (i.e., a small bathroom white and not very clean), 5 or 6 = 3 (i.e., a small bathroom white with grey tiles, not very clean with a big mirror on the wall), 7 or 8 = 4 (i.e., a small bathroom white with grey tiles, not very clean, a big mirror on the wall and a wooden and heavy door), 9 or 10 = 5 (i.e., a small bathroom white with grey tiles, not very clean, a white toilet brush, a big mirror on the wall, a dirty bin and a wooden and heavy door). A 5-point scale measured the extent and the level of detail of the descriptions, where 1 was the smaller number of adjectives used in the description and 5 the greatest number.

Secondly, concerning transportation, we used the Green and Brock (2000) scale, formulated for English audiovisual productions. The Narrative Questionnaire is a seven-point scale composed of 12 questions, in which the participants rated to what extent they felt inside the scenes of the stories. The mean of all the questions formed the transportation variable. The Cronbach's Alpha coefficient, $\alpha = .873$, showed a high internal consistency.

Third, the variable level of attention was measured with the Potter and Choi (2006) retrospective scale. Participants answered three questions depending on the level of attention that they had paid to the story: how much attention they paid throughout the story, to measure sustained attention; what degree of interest it had held for them, to measure interest; and how much they had concentrated on it, to measure concentrated attention. Participants received instructions to distinguish what was being asked in each question. Each answer was measured on a 7-points scale. The mean for these three dimensions formed the general attention variable. The Cronbach's Alpha coefficient, $\alpha = .852$, showed a high internal consistency. We completed this variable with two open questions: what caught your attention the most, and what you missed in the stories.

Lastly, the recall variable was measured using a set of cued recall questions about the information provided in the story (i.e., How much is the parking ticket?). As the example shows, the questions asked specific information about the story. Only the answers that were exactly the same, as mentioned in the story, were considered correct. The wrong answers were those where the participant did not correctly answer the requested information. The recall was scored as 1 for correct answers, and the mean was calculated. Table 1 shows the entire questionnaire.

Table 1

Questionnaire

Imagery

Vividness

1	I think of it, but do not have an image before me
2	Very vague and hardly recognizable
3	Vague un unclear
4	Not so clear and vivid but still recognizable
5	Generally, clear and vivid
6	Vivid and almost as clear as in reality
7	Very vivid and clear as in reality

According to the previous scale, rate the following questions from 1 to 7

-How clear and vivid were the images about the female character?

1 2 3 4 5 6 7

-How clear and vivid were the images about the male character?

1 2 3 4 5 6 7

-How clear and vivid were the mental images about the space where the story took place?

1 2 3 4 5 6 7

-Describe briefly what you imagined

-Describe the space where the story took place

Quantity/ease

-How many images were stimulated?

1 2 3 4 5 6 7

-How difficulty was to create the images?

1 2 3 4 5 6 7

-How fast you created these images?

1 2 3 4 5 6 7

Transportation

-While I was listening to the story, I could easily picture the events in it taking place
-While I was listening to the story, activity going on in the room around me was on my mind
-I could picture myself inside the scene of the events described in the story
-I was mentally involved in the story while listening to it
-After the story ended, I found it easy to put it out of my mind
-I wanted to learn how the story ended
-The story affected me emotionally

	-I found myself thinking of ways the story could have turned out differently -I found my mind wandering while listening to the story -The events in the story are relevant to my everyday life -The events in the story have changed my life -I had vivid mental images of the story Not at all 1 2 3 4 5 6 7 Very much
Attention	-How much attention did you pay throughout the story? Not at all 1 2 3 4 5 6 7 Very much -What degree of interest it had held for you? Not at all 1 2 3 4 5 6 7 Very much -How much did you concentrate on it? Not at all 1 2 3 4 5 6 7 Very much
Recall	Five open questions

Procedure

Participants were gathered in a room with optimal sound conditions. Prior to their arrival, participants were randomly assigned to one of the four groups of presentations. Each participant had a computer on which to listen to the stimuli with high-quality headphones. They were verbally informed that they would listen to two stories and then answer a questionnaire on the same computer. They listened to the first story and filled out the questionnaire and the same procedure for the second story. The only difference among both surveys was the recall questions, which changed depending on the story (A or B). Participants individually completed the experiment in about 20 minutes. Then the participants were thanked, given a gift for their participation, and were dismissed.

Results

The results were analyzed using a variance analysis in a 2 (stories) X 4 (versions) ANOVA for the dependent variables: imagery, transportation, attention, and recall.

Hypothesis 1

The first hypothesis suggested that the stories with verbal and nonverbal sound resources, in particular, sound shots and effects, would achieve a higher level of listener's imagery compared to the stories with only verbal information. There was a significant effect of vividness for version, F (3, 205) = 36.28, $p < .001$, partial $\eta^2 = .352$, but not for story, F (1, 103) = 1.50, $p = .222$, partial $\eta^2 = .007$, or the interaction version by story, F (7, 103) = .861, $p = .462$, partial $\eta^2 = .013$. Also, there was a significant effect of quantity/ease for version, F (3, 205) = 18.15, $p < .001$, partial $\eta^2 = .214$, but not for story, F (1, 103) = 3.13, $p = .078$, partial $\eta^2 = .015$, or the interaction, F (7, 103) = 1.18, $p = .316$, partial $\eta^2 = .017$. Finally, the general level of imagery composed of vividness and quantity/ease was significant for version, F (3, 205) = 44.79, $p < .001$, partial $\eta^2 = .402$, but not for story, F (1, 103) = 3.73, $p = .055$, partial $\eta^2 = .018$, or the interaction, F (3, 100) = 1.68, $p = .171$, partial $\eta^2 = .025$. The nonverbal resources (effects, filters, and shots) achieved a higher level of imagery, vividness, and quantity/ease than the dialogue version. The stories with sound shots and filters obtained the highest level in the three variables followed by effects, but the Tukey's test only showed significant differences between the verbal stories and the nonverbal stories (Vividness: dialogue and effects, filters, and shots, $p < .001$; effects and filters, $p = .824$; effects and shots, $p = .953$; filters and shots, $p = .986$; Quantity/ease: dialogue and effects, filters, and shots, $p < .001$; effects and filters, $p = .244$; effects and shots, $p = .663$; filters and shots, $p = .886$; Imagery: dialogue and effects, filters, and shots, $p < .001$; effects and filters, $p = .422$; effects and shots, $p = .811$; filters and shots, $p = .918$).

The questionnaire included two questions about whether the participants imagined the space where the story was placed and the characters of the story. For the space, there were significant effects for version, F (3, 205) = 36.98, $p < .001$, partial $\eta^2 = .357$, but not for story, F (1, 103) = .271, $p = .603$, partial $\eta^2 = .001$, or the interaction, F (3, 100) = .654, $p =$

.581, partial η^2 = .010. The story with sound filters achieved the highest level, followed by sound shots, sound effects, and the stories with dialogue. The Tukey's test only showed significant differences between the verbal stories and the stories with nonverbal sound resources (dialogue and effects, filters, and shots, p < .001; effects and filters, p = .396; effects and shots, p = .824; filters and shots, p = .396). Concerning the characters, there were also significant differences among the versions, F (3, 205) = 16.31, p < .001, partial η^2 = .197, but not for story, F (1, 103) = 3.38, p = .067, partial η^2 = .017, or the interaction, F (3, 100) = .973, p = .406, partial η^2 = .014. The story with sound shots achieved the highest level, followed by sound effects, sound filters, and dialogue. The Tukey's test showed the same trend: significant differences between the verbal stories and the stories with nonverbal sound resources (dialogue and effects, filters, and shots, p < .001; effects and filters, p = .806; effects and shots, p = .977; filters and shots, p = .561). Table 2 shows the means and deviations of these results.

Table 2

Means and SD of Imagery

Version	Imagery		Vividness		Quantity/ease		Space		Characters	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Dialogues	2.94	.78	3.18	1.13	2.71	.69	1.88	.81	3.08	1.52
Sound effects	4.02	.64	4.65	.90	3.37	.75	3.35	.88	4.45	1.03
Sound shots	4.14	.56	4.75	.92	3.52	.70	3.38	1	4.54	1.20
Sound filters	4.22	.58	4.81	.73	3.62	.65	3.60	.93	4.24	1.01

Regarding the open question about the mental images participants had experienced during the listening task, the results showed significant differences in the invented data; F (3,

205) = 7.57, $p < .001$, partial $\eta^2 = .102$. Sound filters achieved the highest level ($M = 1.67$; $SD = .47$), followed by sound effects ($M = 1.67$; $SD = .47$), sound shots ($M = 1.52$; $SD = .50$), and, finally, dialogue ($M = 1.29$; $SD = .45$). The Tukey's test only showed significant differences between the verbal stories with dialogue and those with nonverbal sound resources (dialogue and effects, filters, and shots, $p < .001$; effects and filters, $p = .762$; effects and shots, $p = .845$; filters and shots, $p = .463$). Therefore, in the stories with these features, there was more invented information (more new mental images) than in the verbal stories.

Also, the participants who listened to the stories with some nonverbal sound resources wrote more detailed and imaginative descriptions, with significant differences, F (3, 205) = 37.31, $p < .001$, partial $\eta^2 = .359$. The descriptions about the space were more detailed and longer in the stories with sound filters ($M = 3.65$; $SD = .92$), followed by sound shots ($M = 3.38$; $SD = 1.01$), sound effects ($M = 3.29$; $SD = .91$), and, finally, dialogues ($M = 1.88$; $SD = .87$). These results support Hypothesis 1.

We represented in a word cloud the most frequent words that the participants recalled. Figure 2 shows the different images of each version, in which the size of each word indicates its frequency. The biggest ones are the most recalled. If we analyze this word cloud, we can see that there were more references to objects in the stories with sound effects (mirror, columns, sink, or water). In the stories with filters, there were explicit references to reverberation and adjectives about the space (large, empty, or full). There were more verbs with actions in the stories with sound shots (leaving, coming, helping, or vomiting).

[Figure 2 here]

Hypothesis 2

Hypothesis 2 established that the stories with verbal and nonverbal sound resources, in particular, sound shots and effects, would achieve a higher listener's transportation compared to the stories with only verbal information. The results for the analysis of transportation indicated significant effects for version, F (3, 205) = 21.38, p < .001, partial η^2 = .243, but not for story, F (1, 103) =.63, p = .802, partial η^2 = .775, or the interaction version by story, F (3, 100) = 1.16, p = .323, partial η^2 = .017. All the versions with nonverbal sound resources achieved a higher level of transportation than the dialogue version (see Table 3). The stories with sound filters obtained the highest transportation level, followed by sound shots and sound effects. The Tukey's test only showed significant differences between the verbal stories with dialogue and the stories with nonverbal sound resources (dialogue and effects, filters, and shots, p < .001; effects and filters, p = .763; effects and shots, p = .417; filters and shots, p = .941). Consequently, Hypothesis 2 is also supported.

Table 3

Means and SD of Transportation, Attention, and Recall

Version	Transportation		Attention		Recall	
	Mean	SD	Mean	SD	Mean	SD
Dialogues	3.27	1.28	5.12	.99	3	.99
Sound effects	4.58	1.17	5.08	1.06	2.78	.99
Sound shots	4.83	1.13	5.65	.79	3.67	.90
Sound filters	4.96	1.23	5.95	.73	3.84	.84

Hypothesis 3

Hypothesis 3 suggested that the stories with verbal and nonverbal sound resources, in particular, sound shots and effects, would achieve a higher listener's attention than the stories with only verbal information. The results were significant for version, F (3, 205) = 11.15, p < .001, partial η^2 = .143, but not for story, F (1, 103) = 1.48, p = .225, partial η^2 = .007, or the interaction version by story, F (3, 100) = .304, p = .823, partial η^2 = .005. The sound filters' stories obtained the highest value, followed by sound shots, verbal stories, and sound effects (see Table 3). The Tukey's test showed significant differences between verbal stories and stories with filters (p < .001) and sound shots (p = .019). There were no significant differences between verbal stories and sound effects (p = .987) and sound filters and sound shots (p = .941).

The attention variable was completed with two open questions: what caught your attention the most and what you missed in the stories. Regarding the first question, 36.1% of participants in the stories with sound effects answered mentioning different effects (water, WC, cars…), 30% of participants in the stories with sound shots talked about distances (characters leaving, coming…), 25% in the story with filters answered mentioning echo and reverberation, and only 8.3% in the story with dialogues referred aspects related to the characters' voices. The chi-square test was significant (X^2 =19.03, p < .001). About the second question, what you missed, 54.5% of participants in the stories with dialogues, 30.3% in the stories with sound shots, and 9.1% in the stories with filters mentioned explicitly ambient sounds or sound effects. In the stories with sound effects, 6.1% missed echoes or revers. The chi-square test was significant (X^2 =23.73, p = .001).

These results partially support Hypothesis 3. Figure 3 represents all the means of imagery, transportation, and attention.

[Figure 3 here]

Hypothesis 4

Hypothesis 4 established that the stories with verbal and nonverbal sound resources, in particular, sound effects and shots, would achieve a higher level of listener's recall compared to the stories with only verbal information. The analysis showed significant results for version, F (3, 205) = 15.78, p = .001, partial η^2 = .191, but not for story, F (1, 103) = 3.20, p = .075, partial η^2 = .016, or the interaction version by story, F (3, 100) = 1.32, p < .269, partial η^2 = .019. The story with sound filters obtained the highest level, followed by sound shots, dialogue, and sound effects (see Table 3). The Tukey's test indicated significant differences between verbal stories and stories with filters and sound shots (p < .001). There were no significant differences between verbal stories and stories with sound effects (p = .480) and between filters and shots (p = .778). These data partially support Hypothesis 4.

Discussion

This study aimed to determine whether nonverbal sound resources such as filters, effects, and sound shots could stimulate imagery, induce transportation, and improve the levels of self-perception attention and recall. The results showed that these features stimulated imagery and transportation to a greater extent than the verbal stories, creating more vivid mental images about the stories, the spaces, and the characters, and "transport" the listeners inside the scene. The spatial nonverbal sound resources that obtained the highest level of imagery and transportation were sound filters and shots, followed by sound effects and verbal stories. Regarding attention and recall, the sound resources that obtained the highest values were sound filters and shots with no significant differences. Sound effects stories achieved a lower level than the only dialogue stories, both in attention and recall. These results confirm the crucial influence of the spatial dimension in audio processing, according to the dual-processing paradigms.

The stories with filters achieved the highest imagery level, followed by sound shots, effects, and dialogues. Sound filters, like the reverberation used in this study for identifying a closed parking lot and a bathroom, stimulated the listener's imagery by mentally representing these spaces ("where"). Filters acted by producing an immersive experience on the listener's part with high imagery (more references to space, more abundant information in the form of invented data, more adjectives about the location). Therefore, participants had very clear in mind the "where" information, thanks to the audio change present in all the story (reverberation). Verbal and nonverbal information were complementing each other, according to the dual-coding theory (Paivio, 1991), and few of the participants (9.1%) missed some sound effects. As filters, sound shots were benefited by dual coding, in this case, by providing the "wherefrom" information. Unlike filters, sound shots were not in all the story but at specific moments with volume changes. This "wherefrom" information allowed to create perspective in the scene, and consequently, the stories were more lifelike. The reality feeling can explain the imagery levels achieved by this resource with more descriptions of the characters and action verbs. Some participants explicitly mentioned this reality feeling, but 30% of participants missed some sound effects ("what" information). The sound effects stories also achieve high imagery levels, more descriptions about the space and objects, and more invented data than the verbal stories. Only 6% of participants missed sound resources in these stories (reverberations). These results are in line with previous studies showing that sound effects have an essential influence on imagery, as the "what" information helps create more vivid mental images (Miller & Marks, 1997; Bolls, 2002; Rodero 2012). Conversely, the verbal stories obtained the lowest level of imagery. We can suggest that the lack of spatial indicators using nonverbal sound resources ("where" and "wherefrom") reduced the creation of mental images about the story, and the descriptions were the poorest. The dual coding did not produce (Paivio, 1991). This is why more than half of the participants missed ambient sounds and sound effects.

Concerning transportation, participants felt especially inside the scenes in the stories with filters, followed by sound shots, effects, and verbal stories. Transportation is very linked to imagery, so, logically, the order was the same. The reason for the highest level of filters may be that these sound resources provided the "where" information, the critical information about the place. Moreover, they indicated the location throughout all the story (not only at certain moments). In consequence, this crucial information about the location and the "continuous sound cushion" placed the participants inside the space with total immersion. In sound shots, the "wherefrom" information was also a clear indicator of location. The spatial perspective introduced simulated a real space, and this reality feeling created more mental images and "transport" the participants to the scene. Sound effects also benefited from the dual-coding information processing but to the less extent. The reason may be that the information added by these elements was not so relevant. These effects represented the sound of objects identifying the space at specific moments, which may not be so important as an immersed sound (filters) or the perspective (shots). Regarding verbal stories, we can suggest that the lack of spatial indicators using nonverbal sound resources ("where" and "wherefrom") reduced the listener's transportation to the scene.

Regarding attention, filters were the resources that obtained the highest level, then sound shots, verbal stories, and effects. Theoretically, by introducing acoustic variations, all the nonverbal resources should have caught the participants' attention (Potter et al., 1997; Potter, 2000; Rodero, 2012). Effectively, sound filters and shots achieved this goal, but not sound effects. The self-perception attention of the different resources seems to be related to their different characteristics. The sound filters were applied to all the stories in a single layer. This uniformity could have avoided distractions while, at the same time, provided clear identification of the space "where"; thus, listeners focused more on the content. Something similar could have occurred with sound shots. In addition to the lifelike sensation, the fact that this element was fused with the voices could have improved the attention levels, as in Rodero's (2012) study. However, contrary to our hypothesis, the sound effects' levels of attention were lower than verbal stories. A possible explanation for this result is that these effects acoustically could have "interrupted" the story's flow. The sound effects used in these stories added *new* information about different objects in the space. There were sound effects not related to the action ("where" and not "what"), indicating the objects of the two locations. Therefore, these sound effects could have drawn too much attention to themselves (acoustic signals) and distracted the listeners from the story's primary information. 36% of the participants answered that some sound effects had caught their attention the most. Here, the level of structural features was high (sound effects at specific moments of the story), but the information added by them also was high (different sound effects); thus, this fact could have increased the required resources to process the message (Lang et al., 2015). In verbal stories, the lack of sound variation reduced the attention levels (Miller & Marks, 1992; Rodero, 2012). If listeners do not have additional elements than verbal information ("what"), they more easily experience a drop-in attentiveness.

About recall, the results were the same as attention. Filters were the resources that obtained the highest level, followed by shots, verbal stories, and sound effects. As filters and shots achieved the highest attention levels, it is logical that these stories were the most recalled. Sound filters and shots benefited from dual coding by introducing a reverberation effect or acoustic changes in the sound volume, respectively; thus, in both cases allocating more resources to encode the message (Lang et al., 2000). However, as the information introduced by these elements was not high, the listeners did not need many required resources to process. Although they added essential information for the spatial dimension ("where" and "wherefrom"), they did it without adding a load of new information. As a result, according to Lang et al. (2015), the level of required resources was low, there were enough resources to process the message, and, finally, the recall levels were the highest. Conversely, the low recall level of sound effects could have occurred by increasing the information to process (Lang, 2000). There were numerous sound effects introduced at specific moments of the story adding new information. Therefore, as mentioned, the level of structural features was high. The information entered by them was also high, more resources were required to process the message, and the recall levels were the lowest (Lang et al., 2015). In verbal stories, once again, the lack of sound information also reduced recall, as other studies have shown (Miller & Marks, 1992; Rodero, 2012). The dual coding did not produce (Paivio, 1991), the attention was lower, and recall also was affected. Therefore, the information processing was more complicated, and more controlled resources needed to be allocated.

In conclusion, the results of this study reveal the crucial role of the spatial dimension to process information in audio production, according to the dual-processing paradigms. The spatial dimension provided by space and perspective improved the listeners' cognitive processing thanks to a dual coding: "what," (provided by verbal information and partially by sound shots) processed by the phonological loop; and "where" and "wherefrom," (provided by filters and sound shots respectively) processed by the visuospatial sketchpad.

As this study aimed to analyze the effects of each sound resource, we designed the stories with only one element. However, it is clear that in audio production, sometimes different resources are used together. As some studies have shown that the more sound features in a message, the more resources are automatically allocated (Lang et al., 2015) and that all the elements together obtain high imagery and attention (Rodero, 2012), we may hypothesize that a story with all the resources should achieve higher levels of imagery, transportation, attention, and recall than a story only with one element, but this is an assumption that needs to be studied.

The findings of this study can be applied both to the academic and the professional field. In the academic area, this study allows us to delve into the research of specific nonverbal sound resources (in this case, filters, effects, and sound shots) to define the spatial dimension in audio processing. In this aspect, this study can be useful for audio or media teachers and academies. In the professional field, this study can help audio producers, sound designers, filmmakers, journalists, and sound creators, to mention only a few. The new platforms of podcasting and sound content nowadays offer more and more quality productions. This research can help them to improve the sound design. Although the spatial dimension is applied to audio fiction in this study, the results can extend to other genres, such as advertising or documentaries.

References

Babbin, L.A., & Burns, A.C. (1998). A modified scale for the measurement of communication evoked mental imagery. *Psychology and Marketing, 15,* 261-278. doi: 10.1002/(sici)1520-6793(199805)15:3<261::aid-mar4>3.3.co;2-x

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences, 4*(11), 417-423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Reviews, 63*(1), 1-29. doi: 10.1146/annurev-psych-120710-100422

Baddeley, A.D., & Hitch, G.J. (1974) Working memory. In *The Psychology of Learning and Motivation* (Bower, G.A., ed.), pp. 47–89, Academic Press.

Baddeley, A.D., Gathercole, S.E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173. doi: 10.1037/0033-295x.105.1.158

Baddeley, A.D., Eysenck, M. W. & Anderson, M.C. (2015). *Memory*. New York, USA: Psychology Press.

Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge, USA: Cambridge University Press.

Beck, A. (1998). Point-of-listening in radio plays. *Sound Journal*, *1*, 15-25.

Bolls, P.D. & Muehling, D.D. (2007). The effects of dual-dask processing on consumers' responses to high- and low-imagery radio advertisements. *Journal of Advertising, 36,* 35-47. doi: 10.2753/joa0091-3367360403

Bolls, P.D. (2002). I can hear you but can I see you? The use of visual cognition during exposure to high imagery radio advertisements. *Communication Research*, *29*, 537-563. doi: 10.1177/009365002236194

Bolls, P.D. & Lang, A. (2003). I saw it on the radio: The allocation of attention to high-imagery radio advertisements. *Media Psychology, 5*, 33-55. doi: 10.1207/s1532785xmep0501_2

Bone, P.F., & Ellen, P.S. (1992). The generation and consequences of communication-evoked imagery. *Journal of Consumer Research, 19*, 93-103. doi: 10.1086/209289

Bull, M. (2005). No Dead Air! The iPod and the culture of mobile listening. *Leisure Studies, 24*, 343-355. doi: 10.1080/0261436052000330447

Childers, T.L., Houston, M.J., & Heckler, S.E. (1985). Measurement of individual differences in visual versus verbal information processing. *Journal of Consumer Research, 12,* 125-134. doi: 10.1086/208501

Chion, M. (2016). *Sound. An acoulogical treatise.* Durham, England: Duke University Press.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual review of psychology*, *62*, 73-10. doi: 10.1146/annurev.psych.093008.100427

Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication & Society, 4*, 245–264. doi: 10.4324/9781315164441-14

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences, 106*, 6008-6010. doi: 10.1073/pnas.0811884106

Dann, L. (2014). Only half the story: radio drama, online audio and transmedia storytelling. *Radio Journal: International Studies in Broadcast & Audio Media, 12*, 141-154. doi: 10.1386/rjao.12.1-2.141_1

Ellen, P.S., & Bone, P.F. (1991). Measuring communication-evoked imagery processing. In R.H. Holman & M.R. Solomon (Eds.)*, Advances in consumer research* (806-812).

Provo, UT: Association for Consumer Research. doi: 10.1016/0010-0285(88)90012-6

Forrester, M. (2012). *Psychology of the image*. New York, USA: Routledge.

Gerrig, R. (1993). *Experiencing narrative worlds: on the psychological activities of reading.* New Haven, USA: Yale UP.

Goossens, G. (1994). Enactive imagery: information processing, emotional responses, and behavioral intentions. *Journal of Mental Imagery*, 18, 119-150. doi: 10.14264/uql.2019.888

Green, M.C., & Brook, T.C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701-721. doi: 10.1037/0022-3514.79.5.701

Herrmann, B., Maess, B., Hahne, A., Schröger, E., & Friederici, A. D. (2011). Syntactic and auditory spatial processing in the human temporal cortex: An MEG study. *Neuroimage*, *57*(2), 624-633. doi: 10.1016/j.neuroimage.2011.04.034

Johnson, E., & Adamo-Villani, N. (2010). A study of the effects of immersion on short-term spatial memory. *Engineering and Technology*, *71*, 582-587.

Johnson, A., & Proctor, R. W. (2004). *Attention: Theory and practice*. London, UK: Sage.

Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B. G., & Somers, D. C. (2014). Auditory spatial attention representations in the human cerebral cortex. *Cerebral Cortex*, *24*(3), 773-784. doi: 10.1093/cercor/bhs359

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, *50*, 46–70. doi: 10.1093/joc/50.1.46

Lang, A., Potter, R. F., & Bolls, P. D. (1999). Something for nothing: is visual encoding automatic? *Media Psychology*, *1*, 145-163. doi: 10.1207/s1532785xmep0102_4

Lang, A., Zhou, S., Schwartz, N., Bolls, P., & Potter, R. F. (2000). The effects of edits on arousal, attention, and memory for television messages: When an edit is an edit can an edit be too much? *Journal of Broadcasting and Electronic Media, 44*, 94-109. doi: 10.1207/s15506878jobem4401_7

Lang, A., Gao, Y., Potter, R. F., Lee, S., Park, B., & Bailey, R. L. (2015). Conceptualizing audio message complexity as available processing resources. *Communication Research, 42*, 759-778. doi:10.1177/0093650213490722

Lehnert, G., & Zimmer, H. D. (2006). Auditory and visual spatial working memory. *Memory & cognition*, *34*(5), 1080-1090. doi: 10.3758/bf03193254

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral cortex*, *14*(9), 1008-1021. doi: 10.1093/cercor/bhh061

Matlin, M. W. (2005). *Cognition*. Crawfordsville: John Wiley & Sons, Inc.

Mazard, A., Tzourio-Mazoyer, N., Crivello, F., Mazoyer, B., & Mellet, E. (2004). A PET meta-analysis of object and spatial mental imagery. *European Journal of Cognitive Psychology*, *16*(5), 673-695. doi: 10.1080/09541440340000484

Mellet, E., Tzourio, N., Crivello, F., Joliot, M., Denis, M., & Mazoyer, B. (1996). Functional Anatomy of Spatial Mental Imagery. *Journal of Neuroscience*, *16* (20), 6504-6512. doi: 10.1523/JNEUROSCI.16-20-06504.1996

Meyerhoff, H. S., & Huff, M. (2016). Semantic congruency but not temporal synchrony enhances long-term memory performance for audiovisual scenes. *Memory & cognition*, *3*, 390-402. doi: 10.3758/s13421-015-0575-6

Miller, D.W., & Marks, L.J. (1992) Mental imagery and sound effects in radio commercials. *Journal of Advertising*, *21*, 83-93. doi:10.1080/00913367.1992.10673388

Miller, D.W., & Marks, L.J. (1997) The effects of imagery evoking radio advertising strategies on affective responses. *Psychology and Marketing, 14,* 337-361. doi: 10.1002/(sici)1520-6793(199707)14:4<337::aid-mar3>3.0.co;2-a

Morris, P.E., & Hampson, P.J. (1983). *Imagery and consciousness.* London, UK: Academic Press.

Nell, V. (1988) The psychology of reading for pleasure: Needs and gratifications. *Reading Research Quarterly, 23,* 6-50. doi: 10.2307/747903

Oberauer, K. (2019). Working memory and attention–A conceptual analysis and review. *Journal of cognition*, *1*, 36. doi: 10.5334/joc.58

Olsen, G. D. (2002). Salient stimuli in advertising: The effect of contrast interval length and type on recall. *Journal of Experimental Psychology*, *8*, 168-179. doi: 10.1037//1076-898x.8.3.168

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*, 255. doi: 10.1037/h0084295

Potter, R. F., Lang, A., & Bolls, P. D. (1997). Identifying structural features of radio: orienting and memory for radio messages. Chicago: Theory and Methodology Division of the Association for Education in Journalism and Mass Communication, Annual Conference.

Potter, R. F., & Callison, C. (2000). Sounds exciting!! The effects of auditory complexity on listeners' attitudes and memory for radio promotional announcements. *Journal of Radio Studies*, *7*, 29–51. doi: 10.1207/s15506843jrs0701_5

Potter, R.F., & Choi, J. (2006). The effects of auditory structural complexity on attitudes, attention, arousal and memory. *Media Psychology, 8*, 395-419. doi: 10.1207/s1532785xmep0804_4

Rauschecker, J.P., & Scott, S.K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12,* 718–724. doi: 10.1038/nn.2331

Rodero, E. (2005). *Radio production*. Madrid, Spain: Catedra.

Rodero, E. (2009). Point of listening in a radio fiction: the eternal problem. *OBS, 3*(3), 242-252.

Rodero, E. (2012). See it on radio story. Sound effects and shots to evoke imagery and attention on audio fiction. *Communication Research*, 39, 458-479. doi: 10.1177/0093650210386947

Rodero, E. (2019). The spark orientation effect for improving attention and recall. *Communication Research, 46*, 965-985. doi: 10.1177/0093650215609085

Romero, L. (2017). *The spatial dimension in an audio fiction story: analysis of imagery, attention, and recall on the listener.* Doctoral thesis. CEU Cardenal Herrera University, Spain.

Sadoski, M., & Paivio, A. (2004). A dual coding theoretical model of reading. In R.R. Ruddell & N. J. Unrau, (Eds.), *Theoretical models and processes of reading*, (pp. 1329–1362). Newark, DE: International Reading Association.

Sheehan, P.W. (1967). A shortened form of Betts' questionnaire upon mental imagery. *Journal of Clinical Psychology, 23,* 386-389. doi: 10.1002/1097-4679(196707)23:3<386::aid-jclp2270230328>3.0.co;2-s

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: II. Detection, search, and attention. *Psychological Review, 84,* 127–190. doi: 10.1037/0033-295X.84.1.1

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cognitive science*, *27*, 525-559. doi: 10.1016/S0364-0213(03)00011-9

Sood, S. (2002). Audience involvement and entertainment-education. *Communication Theory, 12*, 153–172. doi: 10.1093/ct/12.2.153

Meitz, T. G., Meyerhoff, H. S., & Huff, M. (2019). Event related message processing: Perceiving and remembering changes in films with and without soundtrack. *Media Psychology* published online. doi: 10.1080/15213269.2019.1636660

Treisman, A.M. (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-299. Doi: 10.7551/mitpress/2834.003.0019

Zheng, L. (2014). Narrative transportation in radio advertising: A study of the effects of dispositional traits on mental transportation. *Journal of Radio & Audio Media, 21*, 36–50. doi: 10.1080/19376529.2014.891213

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science, 5*, 292-297. doi: 10.1111/j.1467-9280.1995.tb00513.x