# FILMS UNSEEN

Approaching sound design alternatives to enhance perception and mental imagery models among audio description audiences on Sci-fi films.

**Cesar Portillo** [1]

[1] *School of Audio Engineering (SAE Institute)*, London, United Kingdom
email:18661uk@saeinstitute.edu

**Abstract**

"Films Unseen" is a research conducted to analyse the nature of audio description on Science-fictions films. The paper explores the distinctive sound design features that could allow blind and visually impaired audiences to perceive and conduct an accurate mental imagery of the optical elements, presented within visually complex audio visual mediums, such as the film used as a case of study called *How To Be Human (Centofanti, 2017)*. Furthermore, the paper analyses how the implementation of verbal description played before the film, called audio-introduction, and how it benefited from the support of sound effects cues (SFX), which provided a fuller explanation and auditory recognition of characters, objects and ambiences portrayed within the film.

The sound design features, character's aural perspective, sound design enhancement and aural-audio-introductions were evaluated by undertaking a listening test among fifteen experienced audio description users. Correspondingly, the results demonstrated the efficiency for sound and immersive audio technologies to enhance the understanding of the visual features in correlation with audio description (AD).

## 1. Introduction

Within the current visual culture, films have prevailed as the audiovisual medium that introduces its audience to a story portrayed through a combination of complex interactive systems including sound and optical features. Sci-fi films, were used as the main focus throughout this research, due to the unconventional optical and aural features that embody futuristic environments, important to understand the messages behind its concept; which are interpreted from multiple perspectives depending on the audiences' background (social, religion and culture) [1]. Hence, the absence or misperception of particular visual/aural features could lead towards an unclear interpretation of the story behind a film's core message [4]. Such is the case of blind and visually impaired individuals (BVI), who are limited to visual perception or accessibility, construct a personal variant which relies on aural features such as sound effects, dialogue, verbal translation and music in order to impersonate the optical concepts of audio visual medium's (AV).

The integration of AD in AV mediums has proven to enhance the cognitive mental modelling of the optical elements through verbal description, allowing BVI audiences to construct a comparable narrative of the story presented [1]. Nonetheless, the challenges that audio describers face are deciding what, when and how to translate, within a limited time, two semiotic systems (visual and sound) into a comprehensive language that meets current AD guidelines from the Independent Television Commission (ITC) [2].

The functions of image and sound within films were not intended to be divided as the correlation of them both enhanced the narrative of a given story [1]. However, AD allocates its translation over the soundtrack of films which, in some instances, lead to the obscuring of predominant aural features that can picture as much as visual or verbal cues [3].

As stated by Remael [1] "[…]mental modelling approaches can also be applied to sound" as aural cues alone convey effectively specific visual events, allowing audiences to construct a narrative based on auditory stimuli. Hence, the main aim of this research is to analyse how sound design alternatives implemented on a Sci-fi film called *How to be Human (Centofanti, 2017)* could assist AD and enhance mental imagery among the BVI. The hypothesis studied within this paper are:

H1: Enhanced sound design cues could allow BVI audiences to recognise, perceive and build an accurate mental image of the visual features of the film in the study just by their sound.

H2: Audio-introductions could increase the interpretation and recognition of characters, objects and environments among AD users by the employing of detailed verbal cues reinforced by short-term representations of sound.

H3: Immersive sound design alternatives such as the character's aural perspective could translate visual events portrayed in specific scenes, allowing audiences to conduct a narrative by the use of first person perspective auditory cues.

# 2. Related Work

The definition of audio description has been debated across multiple pieces of literature, media and sciences which coined it as a service that translates audio-visual content into verbal commentary; following the audio describer's perspective of the optical events portrayed on a visual medium, which benefits a purely auditory audience [4].

Despite the advantages of AD, multiple inconveniences can occur, such as music, dialogue and sound effects being obscured by verbal translation due to the implementation of gain reduction setups; which allows this service to be understandable for its audience. Consequently, sound and description inconveniences are encountered due to the set boundary that AD has on the creative postproduction process [6]. Nonetheless, researchers within the EU offer findings that lead towards proposals that opened the lens of common translation guidelines; allowing new approaches to enhance the comprehension and perception of visual and aural elements on AV mediums for blind and partially sighted audiences (BPS) [5].

## 2.1 Sound and AD

Sounds in comparison with verbal cues proved to enhance mental imagery among visually impaired individuals when presented in conjunction with video-less mediums such as radio advertising (Bolls, 2002)[7]. Rodero [8] (2012) and López (2015) [6] support this statement by stating how the coverage of the soundscape of audio medium affects AD users' imagination as semantic information (verbal) is less immersive than perceptual stimuli (visual and sound) [7]. Consequently, Ibañez [10] notes that common AD scripting faults encountered across AV mediums compromise the soundtrack, in the following ways:

**Excess:** "Delivery of description that over-explains a given scene which covers, in abundance, the musical or sound effects information".

**Timing:** "Description that overpasses the time stipulated, resulting in the overlap of significant dialogue".

**Patronising:** "Explanation of references that can be deducted by listening to the soundtrack."

The amplified information delivered through the soundtrack has a variety of tasks in which sound effects, dialogue and music work together to complement and detract the visual cues. Consequently, the recognition and effectiveness for SFX to represent visual features depends on the familiarisation and interpretation that BVI audiences have on identifying a sound's cause and source [1].

## 2.2 Audio Films

The sonic art that uses purely sound stimuli that remove the need for narration and visual elements to portray the messages behind a film's story is known as an audio film [9]. López et al (2010) explored the perception among sighted audiences on a film called the *Sound Machine* (Dahl, 1949). The study evaluated two version of the same piece, that aim to translate the layout of spaces, movements and plot elements presented in the film (V1: used sound design cues; and V2: used verbal narration that followed the concept of vococentrism). Within the research, the majority of the participants found the experience entertaining, were able to interpret the plot elements and to recognised the characters' location due to the spatialisation provided within a 5.1 speaker array [9].

In later studies, López (2015) concluded how the concepts behind audio films could be used to offer an entertaining experience that enhances the perception of the main plot elements, on a movie called, *Lamb to the slaughter* (Dahl, 1954) for BVI audiences [6].

## 2.3 Immersive sound:

For BVI audiences, sound corresponds the stimuli by which they recognise the distinctive elements presented on audio-visual mediums as it promotes the identification of spaces, characters and objects. Correspondingly, in visual mediums, such as games and films, sounds are used to reinforce and describe the images in unconventional environments which aid the audiences to navigate and interpret the actions and status of the characters [11] . Therefore, the immersion in which viewers are introduced by the use of sound design is "[…]characterised by diminishing critical distance of what is shown and increasing emotional involvement in what is happening" [12].

Mangiron and Zhang (2016) [13] studied distinctive "speech audio games" and "Non-speech audio games" that introduced auditory cues in order to facilitate the user's gameplay without the need of verbal description. This alternative has been implemented on "video-less" games such as *Os & Xs* (2003) and *Audio Defense: Zombie Arena* (2014) which establish the concept of binaural audio and repetitive sounds speeds to introduce players to an audio-only world [13]. Furthermore, educational games accessible to BVI and sighted audiences such as *La plug Leocadia* (2006) and *Onae, la aventura de Zoe (2007)* incorporated labelled sounds, accessible through keyboards, which indicates the keys to be pressed in order to receive audio and verbal cues necessary to proceed towards distinctive environments.

## 2.4 Radio Dramas

The practice of redeeming the narrative of a story by the use of purely acoustic performances, through radio signals is known as radio dramas [14]. The terms in which audio dramas base its content are mainly speech, dialogue, music and sound effects. Different type of speech such as indirect narration (information not evident on the events portrayed) and direct narration (description of the events perceivable through sound) are implemented to support the storytelling and immersion of its content [15]. Furthermore, sounds are used as an "amplification of the reality presented" which follows contextual and narrative functions [15]. The role of sound in this cases is to recreate realistic effects that confirm and advocate the information delivered by the narration. As an example, Sieveking (1999) [15] explored how the correlation

between SFX and verbal narration/dialogue provides an extra colouration to a specific sound, which provides multiple meanings depending on the context delivered through verbal cues [14]:

**[Sound of water]** + *[Character expressing intention to have a hot bath]* = sound of water interpreted as hot

**[Sound of water]** + *[Narration: broken river bank near the Arctic]* = sound of water interpreted as cold.

Rodero (2012), explored visual imagery responses among sighted individuals when presented with (1) verbal cues and (2) dialogue, sound setups and SFX that supported the narratives of the story presented in audio dramas. The conclusion of this study pointed at how sighted individuals' imagery and attention was superior upon the implementation of SFX, due to the immersion provided by the aural cues.

# 3 Methodology

The researchers and video-less content previously discussed provided an insight on how SFX alternatives are used to enhance auditory perception and mental imagery. Correspondingly, alternatives such as dialogue/narration supported by sound and 3D audio are used to promote the optical impersonation of the events presented in AV mediums. Hence, AD could benefit from SFX alternatives that aim at promoting BVI mental imagery and immersion within audio visual mediums.

## 3.1 AD on SCI-FI films analysis

Sci-fi films offer complex visual and aural information that helps to convey multiple features that offer a setup of modern futuristic environment that support the narrative of the storytelling. As current AD guidelines stipulate, the main focus of this service is to translate one semiotic system (visual) into a comprehensible language (verbal) that allows visually impaired users to conduct a narrative on the optical events portrayed in a medium [2]. Nonetheless, according to Whittington (2007) [16], sound effects are predominant to understand the concepts within Sci-fi films as these are set to stipulate reality, establish a location and to convey the concepts deliver by the director to create a mood.

Fifteen Sci-fi films that incorporated AD were analysed in order to estimate the common scripting faults encountered among this genre, emphasising points which directly affect the soundtrack. A selection of the movies that were studied include: Jurassic Word (Colin Trevorron, 2015), The Hunger Games: Mockingly part 2 (Francis Lawrence, 2015), Terminator Genysis (Alan taylor, 2015), Ant-Man (Peyton Reed, 2016), Men in Black 3 (barry Sonnenfeld, 2012.

Upon the analysis of these films, it was noticeable how the intelligibility and timing were the most common scripting faults among this genre. Furthermore, the following points were deducted:

- Suppression was predominantly noticeable in the description of ambiences and characters. Correspondingly, the description of the main character's within these movies focused on facial features (been hair and age the most predominant) rather than shapes, height, clothing, personality and habits. The following table exemplifies the aspects by which a fuller description could have been accomplished taking into consideration Fryer and Fresno studies [5]:

| Age | Appearance (e.g hair, eyes,facial hair) | Clothing | Movement | Habits (speech) |
|---|---|---|---|---|
| 20% | 20% | 20% | 20% | 20% |

**Table 1:** Factors attributing a complete verbal description of characters.

- Patronising was noticeable on environmental changes and facial expressions, such as smiles and exhalations

The timing, patronising and suppression faults presented on the Sci-fi films analysed helped to estimate the factors that the AD In *How to be human* needed to avoid for the misperception and immersion of its content. However, it was crucial to create a description that follows the current ITC audio description guidelines, in order to estimate the main points by which sound could take the role of the describer and enhance the perception and immersion of BPS individuals.

## 3.2 Films unseen: How to be human

Set in the future, of a dystopian London, with a scarily relevant and frighteningly foreboding reality, *How to be Human*, a Sci-fi film directed by Bruno Centofanti (2017) explores the consequences of a world without humanity in which two sisters (Kimi and Adelphe) start a journey traveling from London to Dover and across the English Channel, Calais and beyond in order to reach a "better world". On completion of the mixing between the AD and the movie's original soundtrack, it was important to carry an analysis pointing at the percentage of elements that were not able to be described due to time restrictions. Hence, the following points were concluded:

1. Between 8 and 16 words were used to describe the characters' features (prioritising age, hair and skin colour) which constitute 40% for an accurate translation of the characters (exposed of table 1).

2. Within the 8 ambiences on the film, only 3 were described accurately represented by verbal cues.

3. 35% of the AD focused on translating information that could be deducted from the soundtrack (such as actions and location of the characters).

4. 40% of facial features (smiles/concerns) and 50% of visual elements aurally representable (ornamentation and ambiences) were not impersonated through SFX, making 30% of the verbal translation to focus on describing these points.

It was noticeable how conveying a translation that enhances mental imagery among BPS individuals would depend on the

information provided through verbal and sound cues, which, due to time limitations and gain reduction, might not embody as many details as the one perceived from the visual perception [4]. As time limitations are the main constraints for AD within this film, it was important to determine alternatives by which visual imagery and auditory translation could enhance the perception among the visual elements presented in *How to Be Human*.

## 3.3 Aural-audio-introductions:

As defined by Fryer and Romero [5] audio-introductions are "pieces of continuous prose, spoken by a single voice or a combination of voices... to provide a framework by which to understand the plot" before the performance/projection. AI has been an established feature of theatrical and musical live performances and included in a limited range of films, such as Die Wand (The Wall Pösler, 2013) [5]. A study by Resmael and Reviers (2013) [5] noted the main five functions for audio introductions: (1) Create a framework (2) Provide relevant details (such as cast, set and costumes) (3) fuller description of events missed in the original AD (4) Explain the nature of the performance)(5)Instructive function (gain setup indications).

Current models of AI [5], introduces its audience to a memory based method in which the verbal description helps to recognise and build visual features through detailed verbal cues. For this research, the approach undertaken introduced verbal description that is supported by the SFX of the event/ character or ambiences portrayed within the film (figure 1). Hence, this model aimed at SFX to be used as a recalling for the detailed translation provided at the beginning of the film, aiming at AD users to understand the visual events portrayed within the film. Furthermore, the AI implemented introduces an indirect narration in which the main character (Kimi) interprets from her perspective the visual elements of the film, creating an interactive experience that enhances the immersion and mental imagery from verbal and audio cues.

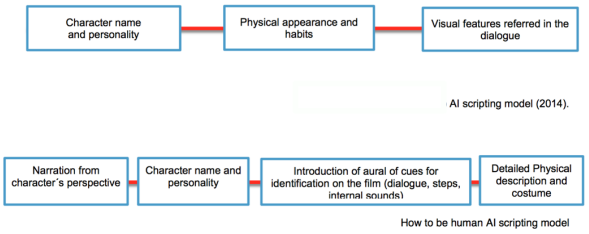


**Figure 1:** Comparison between conventional AI scripting models by Fryer and Romero (2014) and the AI supported by sound implemented on How to be human.

## 3.4 Increasing mental imagery and perception through sound:

One of the primary functions of sound is to derive its source from the same world presented within visual mediums, in order to reinforce a sense of realism [11]. Within movies, the majority of auditory information is applied to follow the actions portrayed within frames, in which aural environmental and visual perceptual features work as one [15]. Nevertheless, some visual elements are not aurally portrayed due to the prioritisation of dominant sound sources and the self-explanatory visual features that do not require sound to be interpreted by sighted audiences. The approach undertaken for this research inducted aural features that translated visual elements that produce an action on each scene, such as internal sounds and non-speaking characters, aimed at AD audiences to recognise the temporal optical attributes on the film in the study.

Within the 8 scenarios presented on *How to be Human*, it was noticeable how the induction of sounds was predominant on close spaces such as "the cleaner's tent" (3:28 minutes). In this scene, four characters perform key events and interact with multiple objects that define the understanding of the film's concept. Correspondingly, the complexity of the sounds in this scene worked as the case of study to evaluate how sound could be used to enhance the translation of visual elements. Table 2 presents a spotting list that analyses both: the optical elements that were not aurally represented on "the cleaner's tent" stereo released format and the sounds implemented to promote the perception of the visual elements aurally represented:

| Sounds implemented | Time | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0:00 | 0:20 | 0:40 | 1:00 | 1:20 | 1:40 | 2:00 | 2:20 | 2:40 | 3:00 | 3:20 |
| Dialogue | | | | | | | | | | | |
| Heavy wind | | | | | | | | | | | |
| Moving fabric | | | | | | | | | | | |
| Cleaner Steps | | | | | | | | | | | |
| Fireplace | | | | | | | | | | | |
| Pot | | | | | | | | | | | |
| Necklace | | | | | | | | | | | |
| Dense liquid | | | | | | | | | | | |
| Wood sound | | | | | | | | | | | |
| Lamp | | | | | | | | | | | |
| Candles | | | | | | | | | | | |
| Metallic hit | | | | | | | | | | | |
| Water pouring | | | | | | | | | | | |
| Girl´s breathings | | | | | | | | | | | |
| Kimi´s Breathings | | | | | | | | | | | |

**Table 2:** Timeline for sound features implemented on the cleaner´s tent scene (1:07:28:00) where Yellow represents dialogue from characters, Green: aural elements implemented on the film´s original version and Purple: additional sound design elements.

The complexity of several sounds been displayed within a limited stereo range interposes a challenge to establish the distance, direction and position of the distinctive elements presented on *How to be Human*. Hence, binaural recording techniques were applied by using a KU80 microphone within a 5.1 system array in order to provide a layout of the characters' position within the film. Furthermore, binaural redeemer software (Dolby Production VR Suite) were used to provide information of the objects positioned within multiple locations , as exposed on figure 2, in order to enhance the details delivered by AD.

The appliance of such techniques followed the redeeming of BBC Sci-fi radio dramas, such as Stone Tape (Peter Sasdy, 2016) in which monaural to binaural conversion was applied to enhance the narratives within non-visual-dependent mediums [17]. Furthermore, Sci-fi TV shows such as "Doctor Who" presented a similar approach on its "Knock Knock" episode in which binaural sound techniques were

implemented, in order for its audience to understand the distinctive locations of the unseen character [18]. Correspondingly**,** *How to be Human*'s 3D sound redeeming aimed at replacing forms of verbal translation referring to spatial temporal settings; while focusing the AD on visual elements not translatable through sound design.

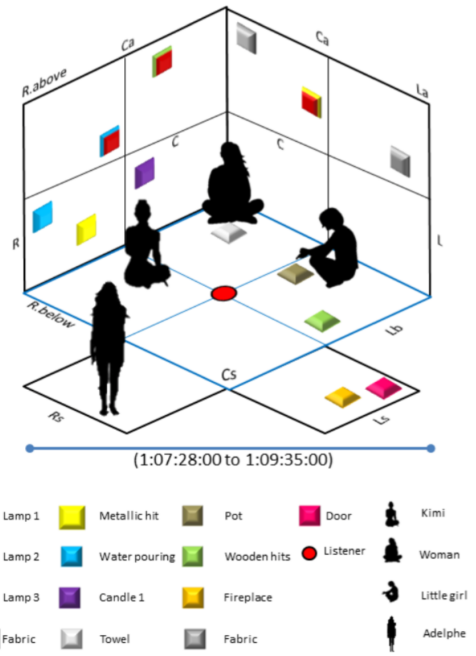


**Figure 2:** sound spatialisation mapping applied to sounds on table 2 using Dolby Production VR Suite.

## 3.5 Character's aural perspective:

Glassner (2004) notes how one of the levels of immersion within games, namely sympathy, conveys aural and visual information, in which the players start to see through the protagonist"s eyes and "empathise emotional bonding with the character" [19]. Consequently, Ermy et al (2005) [19] argues that imagery immersion, one of the tree-part-divisions of immersion, "introduces the player towards the possibility to use their imagination to impersonate the character or to enjoy the features within the game". On films, one alternative that enhances the immersion between visual mediums and sound, namely IMAX mixes, comprises of the optical and aural properties, in which "a sensory overload and emotional assault complete the immersion of its audience" [19].

The implementation of sound effects, equalisation, spatialisation and volume automation that places the listener into the main character's perspective, aimed at the visual information to be interpreted through sound design rather than verbal narration. Therefore, aural features were mixed and panned to replicate the reality within the physical impacts and proximity effects from the main character's location. An example of the cleaner's tent's approach is presented in table 3.

| Audio Description | Technique applied | Sounds recorded/used |
|---|---|---|
| The woman grabs a towel and gently cleans Kimi's face | Foley recording in which a towel rub a KU80 microphone in order to replicate cleaner's actions during musical events. | ➢ Towel<br>➢ Panning automation on singing |
| The child gives a tarot card to Kimi | Direct narrative from main character | "The hanging men tarot card". |
| The woman and the child proceed to tress Kimi's hair. | Positioning of wig over KU80 microphone that was brushed following the events of the film. | ➢ Hair been brushed |
| The cleaner helps Kimi to soak her hand in the water repeatedly. | Use of repetitive water drops and hand movements. | ➢ Water<br>➢ Hands friction |

**Table 3:** Character´s aural perspective approach.

## 3.6 Questionnaires:

The effectiveness of the alternatives previously exposed was justified by conducting a questionnaire, in which, quantitative and qualitative data acknowledged the perception and immersion of the plot through the enhancement of sound design cues. Therefore, 15 experienced AD users among different sight conditions (11 blind and 4 partially sighted) were contacted through the Royal National Institute for the Blind People (RNIB), Haringey Phoenix Group (HPG) and Noisy Vision (NV) in order to undertake a one-to-one listening test. The listening sessions aimed at correlating the following:

**Part 1:** this section gathered information regarding the participant's current sight condition and at acquiring figures on current their opinion on current AD faults; that affected their immersion and interpretation of the visual elements portrayed within audio-visual mediums.

**Part 2 and 3:** this section investigated frequently recalled visual/aural features of two clips of the film (Stereo vs binaural version) in which immersion, perception and understanding of the plot elements portrayed by SFX were analysed.

**Part 4:** Upon watching the fifteen minutes film, participants were asked free recalling questions in order to assess the following: (1) interpretation of the character and spatial features upon the inclusion of AI. (2) accuracy on representing the events portrayed through the character's aural perspective ; (3) the SFX and verbal features that enhanced participant's perception and comprehension of the film used as a case for the study.

# 4 Experimental Results

The present research explored the accuracy by which SFX and sound design techniques could translate the visual elements in order to assist with audio description on a Sci-fi film called *How to be Human*. Hence, a listening test consisting of four parts aimed at addressing H1, H2 and H3 by determining the readability of the sound design features implemented in two scenes of the film, namely "The Cleaner's tent" (X) and "The immigration point" (Y) without audio description. Furthermore, participants were asked to watch the fifteen minutes aural version of the film with AD in order to analyse their immersion and understanding of the visual features.

For this paper, the analysis will focus on the results obtained on X in which (X-1) symbolises the scene with the clip´s original soundtrack and (X+1) the version with the new alternatives.

The main plot elements of X are:

| Clip | The cleaner's tent (X+1) (X-1) |
|---|---|
| Plot element | Description |
| 1 | Outside the tent Adelphe knocks the door which later opens revealing a woman and a little girl |
| 2 | Adelphe pushes Kimi gently forward while she asks for the woman´s help |
| 3 | Inside the tent, the woman gives the child a pot with a dense liquid. |
| 4 | The child mixes the liquid |
| 5 | The woman sits and grabs Kimi´s hand and wipes it clean |
| 6 | The woman stands while asking Adelphe to come back for the little girl. |
| 7 | In the Conner, Adelphe mentions that they will commit to their promise as soon as they get their status |
| 8 | The woman sits and then spreads the liquid on Kimi´s forearm |
| 9 | The woman cleans Kimi´s head |
| 10 | Kimi washes her hands on water |
| 11 | Then the cleaner brushes Kimi´s hair |

**Table 4**: Key plot elements on How to be Human for (X-1) and (X+1)

The results obtained in the perceptual and recognition test were statistically analysed, by evaluating the recalling of the visual features aurally represented on the film. The discussions of the results per features are:

# 4.1 Sound design enhancement:

## 4.1.1 Objects recognition

Within the research, it was predominant to determine the accuracy by which participants could recognise the visual elements that comprised spatial-temporal settings.

On (X-1) and (X+1) all participants were able to identify the three location changes presented in the scene (Outside—Inside the tent—walking towards destination), due to the complementary sounds applied that followed the position, dimension and surfaces on these environments. Furthermore, participants pointed at how the sounds inside the tent make them picture a small kitchen (partially accurate) in which the following visual objects, presented on table 5, were aurally identifiable:

| Sound Number | Element | X-1 | X+1 |
|---|---|---|---|
| 1 | Metallic Door | 90% | 100% |
| 2 | Towel | 10% | 80% |
| 3 | Pot | 70% | 90% |
| 4 | Moving Fabric | 0% | 10% |
| 5 | Water | 40% | 60% |
| 6 | Having lamps | 10% | 70% |
| 7 | Fre Place | 30% | 70% |
| 8 | Candles | 0% | 30% |
| 9 | Heavy wind | 50% | 80% |
| 10 | Dense Liquid | 70% | 80% |
| 11 | Metallic Hits | 30% | 70% |

**Table 5:** Perception of objects on (X-1) and (X+1).

It is noticeable how participants scored the lowest marks on element number 4 (Table 5) as their matching auditory representation misinterpreted the sound of heavy fabric movements with a storm (inaccurate). Nonetheless, participants´ perception of sounds number 1, 3, 9 and 10 was remarkably higher on (X+1), attributed to the additional SFX implemented and due to the familiarity that listeners had with these objects. Hence, participants stipulated their answers based on their previous auditory to memory representation, meaning that their narratives were built upon previous experiences rather than expectations.

The sound spatialisation implemented on (X+1) contributed towards the participants' interpretation of sounds number 2, 3, 6 and 7 (table 5), which were accurately recognised among the spatial locations exposed on figure 2. Furthermore, listeners were able to conduct a clear narrative of plot element number 3 (table 3) due to the recognition and spatialisation of sounds number 3, 8 and 12 presented on table 6.

| Sound number | Element | Location | Recognition score |
|---|---|---|---|
| 1 | Metallic Door | Right rear | 90% |
| 2 | Towel | Front | 80% |
| 3 | Pot | Left | 90% |
| 4 | Fabric | Top | **0%** |
| 5 | Water | Front | 60% |
| 6 | Hanging lamps | Top L and R | 70% |
| 7 | Fire Place | Right back | 70% |
| 8 | Necklace | Left to right, Front | **50%** |
| 9 | Heavy wind | L-C-R | 80% |
| 10 | Dense Liquid | Right | 80% |
| 11 | Metallic Hits | Left and right back | 70% |
| 12 | Cleaner steps | Right to left, Front | 80% |

**Table 6:** Recognition of visual objects aurally represented on (X-1) and (X+1). Visually presented in figure 2.

# 4.1.2 Characters recognition:

Candidates were asked to stipulate the number of characters on the scene and the factors by which these were identifiable, such as kind of voice and positioning. On the stereo version (X-1), none of the participants scored 100% on the recognition of the number of characters within the scene. However, 8 participants coined that they identified two out of four characters, in which Kimi and the Cleaner were the only recognisable performers. These results could be attributed by the centre panning of the dialogues and the similarities between two of the characters' voices (Kimi and Adelphe). On the other hand, 6 of the participants recognised the three speaking characters due to the dialogue conducted within previous environments.

On (X+1), 11 participants score 100% on the character recognition within the scene. The main factors that enhanced their understanding were the sound localisation applied on the character's dialogue and the inclusion of breathings on the non-speaking character (Little girl). Moreover, participants were able to provide an accurate representation of the characters' psychological state (due to internal breathings) and the actions that these were carrying, such as pouring the liquid (plot element 4) and whether they were sitting or standing (plot element 6).

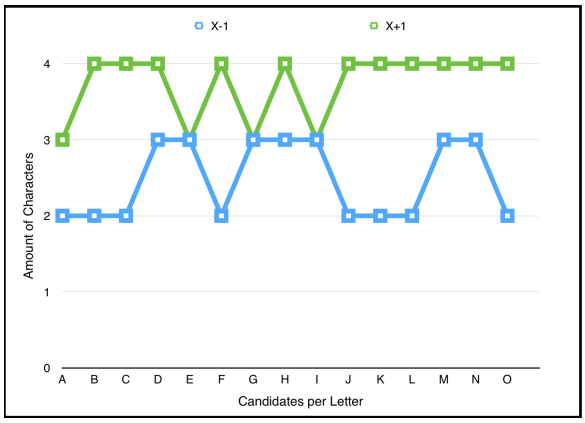


**Figure 3:** Number of perceived characters on (X-1) and (X+1) per participant, where 4 is highly accurate, 3 partially accurate and 2 or below inaccurate.

## 4.2 Character's Aural Perspective:

The character´s aural perspective was employed in order to allow participants to conduct a narrative based on sound cues instead of verbal narration. On (X+1) the following elements were translated and recognised through this feature:

| Element number | Description | Scores with SFX | Scores with SFX and AD |
|---|---|---|---|
| 1 | Woman cleaning Kimi's face with a towel | 0 | 11 |
| 2 | Woman brushing Kimi's hair | 7 | 12 |
| 3 | Hands been washed on water | 8 | 11 |
| 4 | Shirt been buttoned | 5 | 6 |
| 5 | Sneakers been pulled off | 8 | 8 |

**Table 7:** Interpretation rate of character´s aural perspective with and without AD per participant.

More than 50% of the sample were able to interpret elements number 2, 3 and 5, presented in table 7, within this feature based on sound design cues only. Nonetheless, plot number 1 was in its majority misinterpreted with the sound of cardboard/snow instead of the cleaner washing Kimi´s face, which could be attributed to the low-frequency nature of the sound implemented. However, upon the inclusion of AD, more than 70% of the participants scored highly accurate results on plot elements number 1, 2 and 3 due to the verbal cues provided before these events. Hence, higher results were encountered due to the correlation between AD and the character´s aural perspective, as verbal cues were used to reinforce the hypothesis stipulated by the participants (example: description of towel [sound number 2, table 6] served to identify plot element 1, table 7).

In order for this feature to fully immerse and translate aural features in future films the following aspects should be considered: (1) Introduction of verbal cues of object/character that could enhance their recognition within this feature ;(2) Implementation of audio cues that follows the same frequency spectrum as the one translated from verbal cues; (3) Sound spatialisation that monitors the character´s aural perspective rather than the camera frame.

Upon watching the fifteen minutes film, participants expressed how "the immigration point" (Y) was the scene in which they felt more immersed at interpreting the visual elements through the character's aural perspective. This could be deducted by the fact that the soundtrack presented on (Y) was segmented to sound design cues only, whereas (X+1) was presented in conjunction with more than one soundscape element (Music and SFX) which could have caused an overload on the cognitive load on the participants. Consequently, one of the participants described the following: "In the Immigration area, in particular, having the character's aural perspective is a very useful extra feature ... this had the additional consequence of adding tension and suspense to the scene, which enhanced it in itself ".

## 4.3 Aural Audio-introductions:

More than 80% of the participants encouraged the implementation of aural-audio-introductions on future films, as it allowed them to interpret and build fuller mental images of the elements essential to understand the play. Nearly 70% of the participants recognised and interpreted key visual components of the film, such as the physical appearance of the characters and elements number 4 ,6 and 9 presented in table 8.

| Audiointroduction | Element | Recognition | Features remembered | |
|---|---|---|---|---|
| 1 | Kimi physical features | **90%** | Scar over forehead and dirty clothing | |
| 2 | Adelphe physical features | **90%** | Clean, slim | Android |
| 3 | Map locations | **100%** | Cube shape | SFX |
| 4 | Cold city | **70%** | Skyscrapers | Drones |
| 5 | Security guards | **30%** | Blue makeup | X |
| 6 | Cleaner´s Tent | **80%** | Lamps | Red fabric |
| 7 | Woman | **40%** | Age | X |
| 8 | Little girl | **60%** | Age | Skinny |
| 9 | London | **70%** | Fire | Brick building |

**Table 8 :** Interpretation rate of the information delivered through AI and recalled features per description.

Participants were able to recognise elements numbers 1, 2, 3 and 6 (table 8) based on auditory cues only, without the need of further translation in the movie. On the other hand, it was noticeable how 60% of the participants remembered behavioural and physical descriptions, such as the scars on Kimi´s forehead, Adelphe as an android and the difference between these two characters, rather than their clothing. These results are consistent with Fresno et al. (2016) who coined how physical features and shorter descriptions were recalled efficiently by BVI audiences, as it allowed them to build a schematic image from basic information.

When asked about how the inclusion of AI contributed towards their understanding of the film, 14 participants mentioned that they pictured a clear representation of the character's appearance; 12 described how they could build a visual image of the layout of the ambiences presented on the film; 9 expressed how they recognised specific objects (Lamps, Drone and Map) and characters just by their sound. Additionally, one of the participants described the following: "The AI connected the film together by making it more coherent. Without this feature, I would be wondering the visual elements of the film and I would not be able to picture most of the landscapes in it ".

# 5 Conclusion

Verbal translations are indispensable features that allow blind and visually impaired audiences to experience and comprehend audio-visual mediums. However, the scarce time allocated for the AD to translate complex visual elements, such as the ones presented in Sci-fi films, incentives the need of new alternatives that provide a fuller immersion and interpretation of the visual features portrayed through visual mediums.

The quantitative and qualitative data collected from this research explored the perception and recognition of sound design cues that aimed at translating visual elements to assist AD. The participation of 15 experienced AD users facilitated the proper analysis of the features employed, in which 14 volunteers coined how their interpretation and perception of the visual features aurally/verbally represented increased due to the inclusion of the aural-audio-introduction.

Upon the analysis of the listening test, it was noticeable how the participant's matching auditory to memory representation allowed their mental imagery to build an environment that is partially accurate to the one presented in the film. The inclusion of internal sounds (such as breathings and distinctive steps) allowed 11 of the participants to translate the emotional status of the characters and provided a cue for their presence within specific scenes. Moreover, SFX cues proved to benefit from sound spatialisation as 60% of the participants were able to recognise the layout of environments and the location of distinctive objects within scenes. Nonetheless, the effectiveness and appliance of such techniques were superior on close spaces, due to the predominance of visual features that produce an action that were aurally represented.

The recognition of the events portrayed through the character's aural perspective was on their majority wondered rather than accurate. This feature represented severe variables within the research as more than translating the visual features of the film it allowed participants to have a greater immersion within tension scenes. Moreover, 60% of the participants expressed how in correlation with AD this feature helped them to empathise with the main character and to have a better understanding of the visual elements. Consequently, continuous auditory representations, that share similarities with verbal description, suggest that gained information encoded in short-term rendering enhances mental imagery.

Visual perception is highly exploratory, which should be the approach for auditory perception on audio-visual mediums. Hence, distinctive objects, characters and environments presented in *How to be Human* proved to benefit from fuller aural impersonation, enhancing the interpretation of additional optical events for BVI audiences. Further studies will be needed to analyse the perception and immersion among sighted individuals on the sound design features implemented; aiming at stipulating the factors that could create an experience that is equally immersive to both groups of audio-visual users. Nonetheless, it was noticeable how the inclusion of AD in the post-production process of *How to be Human* allowed its plot and visual features to be perceived by the use of auditory cues, within an immersive experience that compensates for the absence of visual perception.

# References

[1] Remael, „For the use of sound. Film sound analysis for audio-description: some key issues", Multidisciplinarity in Audiovisual Translation, 4, 255-276, (2012) doi: 10.6035/monti.2012.4.11.

[2] ITC, Guidance on Standards for Audio Description, (May 2000).

[3] Thom, Randy. (1999) "Designing A Movie For Sound." Online at FilmSound.org. Learning Space dedicated to the Art and Analyses of Film Sound Design. Full text-version at: http://www.filmsound.org/articles/designing_for_sound.

[4] Freyer, L. An Introduction to Audio Description: A practical guide, 1st ed. Oxon, 27-28 (2016) New york: Routledge.

[5] Benjamins Translation Library : Audio Description : New perspectives illustrated 2014, John Benjamins Publishing Company, Philadelphia, NL. Available from: ProQuest ebrary.

[6] López, M., 'Perceptual evaluation of an audio film for visually impaired audiences' in the Audio Engineering Society (AES) 138th Convention, Warsaw, Poland, 7-10 May 2015.

[7] Fryer, L. Putting it into Words: The Impact of Visual Impairment on Perception, Experience and Presence, 1, (2013) doi: https://research.gold.ac.uk/10152/1/PSY_thesis_Fryer_2013.pdf

[8] Rodero, E. (2012). See it on a radio story: sound effects and shots to evoked imagery and attention on audio fiction. Communication Research, 39, 458–479.

[9] López, M. and Pauletto, S., 'The Sound Machine: a study in storytelling through sound design' in Proceedings of the 5th Audio Mostly Conference, Piteå, Sweden, 1-17 September 2010.

[10] Marza Ibañez, A (2010) "Evaluation criteria and film narrative: a frame to teaching relevance in audio description" Perspective studies in Translatology 18, no 3: 143-153

[11] Stevens et al, (2011), "The Game audio tutorial: a practical guide to sound and music for interactive games" Focal Press, 1, 275-279

[12] Grau O., (2003) Virtual Art: From Illusion to immersion. Cambridge, Mass: MTI Press.

[13] Matamala, A. and Orero, P. (2016).Researching audio description. 1st ed. Barcelona: Palgrave Macmillan.

[14] Crook, T. (1999). Radio drama theory and practice. London: Routledge. 73-7

[15] Alten, S.R. & Quin, D., 1999. Audio in media 2nd ed., Australia: Wadsworth, 180-185

[16] Whittington, William. Sound Design and Science Fiction. Austin, US: University of Texas Press, 2007. ProQuest ebrary. Web. 31 August 2017.

[17] Anon, Binaural Sound - BBC R&D. BBC News. Available at: http://www.bbc.co.uk/rd/projects/binaural-broadcasting [Accessed August 4, 2017].

[18] Pike C, (2017) http://www.bbc.co.uk/rd/blog/2017-05-doctor-who-in-binaural-sound [Accessed June 15, 2017].

[19] Collins K., (2008) Game sound: an introduction to the History, Theory and Practice of video games music and sound design, MTI Press. 133-136