Enhancing Movie Comprehension

For Individuals Who Are Visually Impaired Or Blind

Through Haptics

by

Lakshmie Narayan Viswanathan

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2011 by the
Graduate Supervisory Committee:

Sethuraman Panchanathan, Chair
Baoxin Li
Terri Hedgpeth

ARIZONA STATE UNIVERSITY

August 2011

ABSTRACT

Typically, the complete loss or severe impairment of a sense such as vision and/or hearing is compensated through sensory substitution, i.e., the use of an alternative sense for receiving the same information. For individuals who are blind or visually impaired, the alternative senses have predominantly been hearing and touch. For movies, visual content has been made accessible to visually impaired viewers through audio descriptions—an additional narration that describes scenes, the characters involved and other pertinent details. However, as audio descriptions should not overlap with dialogue, sound effects and musical scores, there is limited time to convey information, often resulting in stunted and abridged descriptions that leave out many important visual cues and concepts. This work proposes a promising multimodal approach to sensory substitution for movies by providing complementary information through haptics, pertaining to the positions and movements of actors, in addition to a film's audio description and audio content. In a ten-minute presentation of five movie clips to ten individuals who were visually impaired or blind, the novel methodology was found to provide an almost two time increase in the perception of actors' movements in scenes. Moreover, participants appreciated and found useful the overall concept of providing a visual perspective to film through haptics.

DEDICATION

This work is dedicated to my father, Mr. Viswanathan Athmanathan, my mother, Ms. Rajeswarie Viswanathan, and my sister, Ms. Narayanee Viswanathan.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my committee chair, Dr. Sethuraman Panchanathan, for being an excellent guide and role model towards developing my intellectual skills. I sincerely thank him for his patience in allowing me to search for my passion and continue along the lines of haptics research.

I would also like to thank Dr. Morris Goldberg and Dr. Terri Hedgpeth for their invaluable advice and suggestions with shaping the experiment conducted as part of this work.

I would like to convey my sincere thanks to my mentors, Dr. Sreekar Krishna and Mr. Troy McDaniel, for their invaluable guidance from the first day of this work. I would also like to thank Troy for his advice and suggestions during the writing of this thesis.

I would also like to thank Dr. Baoxin Li for serving on my thesis committee and for his role in the completion of this work.

I would like to thank Dr. John Black for his invaluable advice in shaping my career path and for making me realize my strengths.

I would also like to thank Mr. Jacob Rosenthal and Mr. Nathan Edwards for their contributions to the development of the haptic belt.

Lastly, I would like to thank all my colleagues, fellow researchers and friends at the Center of Cognitive Ubiquitous Computing for their help and encouragement.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

According to the World Health Organization (WHO) [**1**], about 284 million individuals in the world are visually impaired, out of which 39 million are blind. WHO estimates that about 65% of people who are visually impaired are aged 50 years or older. The National Federation of the Blind [**2**] states that there are 1.3 million legally blind individuals in the United States. Scientists and engineers have explored two research tracks, namely medicine and technology, in an effort to cure or circumvent this disability; specifically:

- Advancements in medicine toward prevention and/or cure.
- Accessibility of visual content via sensory substitution.

This work follows the second path, attempting to address some of the problems encountered by individuals who are visually impaired or blind during the common activity of watching a movie.

1.1 Media Accessibility

Within the context of this work, the term 'accessible' does not merely mean something that can be reached, used or attained. According to Joe Clark [**3**], "accessibility involves making allowances for characteristics a person cannot readily change". He further elucidates this definition through the following examples:

- A person, who cannot hear, cannot prevent himself or herself from being in this condition when confronted with a soundtrack.
- A person, who is blind, cannot suddenly begin perceiving visual information when confronted with visible words and images.
- A person with a mobility impairment cannot suddenly begin to move when confronted with a navigation task.
- A person, who does not know French, for instance, cannot suddenly begin comprehending the language when it is confronted.

Considering the last example, to resolve this problem, the person might be provided with a manual or automated tool for translation. An example of the former would be a dictionary, or a human translator well versed in both languages. If this translation is viewed as a concept, it can also be used to address the other three examples. The provision of a wheel chair is then a translation of the naturally available legs for an individual with mobility impairments in the context of movement. In closed captioning, spoken dialogues and key sounds such as “music playing" are translated into text and displayed at the bottom of the screen during a television show or movie. This translation, though, is to a large extent inherent, as the alphabet of almost any known language has a textual and verbal form, e.g., a unique sound is associated with the pronunciation of the graphical letter ‘a’. Another tool available to address hearing impairments is sign

language. If we analyze the previous translations used to overcome the loss of hearing, it becomes evident that a loss of one of the senses (hearing) was circumvented by utilizing and providing information through another sense (sight). This is the essence of *sensory substitution*. The following section provides examples of various problems that were addressed through the use of sensory substitution techniques and technologies to assist individuals who are visually impaired or blind.

1.2 Sensory Substitution for the Visually Impaired

This section presents several significant contributions in the area of sensory substitution techniques and technologies toward improving the accessibility of content for individuals who are blind or visually impaired.

1.2.1 Braille

Braille is a system of reading and writing text that is widely used by individuals who are visually impaired or blind. Each character is assigned a pattern of raised dots contained within a cell of size 3 x 2 dots. This provides a possible of $2^6 = 64$ characters to be represented. Hence, an individual who is visually impaired or blind can write and read words and sentences using tactile Braille letters and punctuations. To read Braille, a reader feels the raised dots using his or her fingers as he or she scrolls through the text. The user then

associates the raised dot patterns in a cell to the character that it represents, which is then combined with other cells to form words and sentences. Hence, it employs a system that represents graphical letters as a pattern that can be understood by readers through touch. Braille was devised by Louis Braille in 1825 [**4**].

1.2.2 Tactile Vision Substitution System

The Tactile Vision Substitution System, or TVSS, was designed to allow individuals who are blind to "see" visual objects through touch. The original system used a chair whose back rest consisted of vibration motors arranged in a 20 x 20 grid with one vibrator per cell. With the use of a camera, captured images were literally displayed to the skin of the back through vibrations. As the resolution of the vibrotactile display cannot match the resolution of the captured images, the latter resolution needed to be reduced. Here, each vibrator cell is assigned to a block of pixels within the captured image. The average intensity of the pixel values is computed, and if the value is above a threshold, then the respective vibration motor is actuated. Users explored objects in front of them through use of a camera, moving it as they swiveled the chair. It was observed that, after extensive training, such a system allowed individuals to recognize common objects through their gross shape, but participants found it harder to recognize internal details of objects [**5**].

1.2.3 Tadoma

Tadoma is a method of communication used by individuals who are both blind and cannot hear. In this method, the individual places his or her hands on the face of the other person with whom they are interacting. They place their "thumb on the speaker's lips and their fingers along the jawline. The middle three fingers often fall along the speaker's cheeks with the little finger picking up the vibrations of the speaker's throat" [**6**]. Such a positioning of the fingers allows an individual to feel "the movement of the lips, as well as vibrations of the vocal cords, puffing of the cheeks and the warm air produced by nasal sounds such as 'N' and 'M' " [**6**].

1.2.4 Audio description

Whenever a person talks about an event that took place in the recent past, it involves discussion of where it happened, what happened, who were involved and what they said. Audio description (AD) applies this concept in the realm of movies to assist individuals who are blind or visually impaired in its comprehension. Whenever there are no dialogues exchanged between actors, and there are no background scores, an additional narrator talks about what is happening visually in the scene, describing its location, time of day, actors involved, and other pertinent visual cues that allows the movies to remain coherent and cohesive to an individual who does not have access to the visual

content of the film. Hence, in this domain, audio is used as a substitution medium to provide information about visual cues.

1.3 Enhancing Movie Comprehension

Audio description in its current state leaves a lot of information in the visual domain untold, while summarizing others, due to lack of time (see section 2.1). It also uses a single mode of communication (audio) to deliver multimodal content: a typical movie is comprised of both audio and video. This work explores the usage of touch as a secondary communication channel to deliver complementary information along with audio. Such an approach would reduce the over-reliance of a single medium to convey information, and hence, should reduce the chances of overwhelming the end user. This work hypothesizes that such an approach will allow the communication of more information within the available time, providing a richer movie experience, without overloading the user. This work looks at a movie as a collection of scenes and embeds touch information in selected movie clips.

In the following sections, the background of audio description, situation awareness and haptics (the science of touch) is presented in section 2; related work with respect to haptics in movies, in particular, and situation awareness, in general, is covered in section 3; our proposed methodology is presented in section 4, and its validation through a user study is presented in section 5; and lastly, possible directions for future work are presented in section 6.

CHAPTER 2

BACKGROUND

This work began with an extensive literature survey on audio description, a mechanism that is used to assist individuals who are visually impaired or blind with comprehending the visual content in movies. Research in this domain has attempted to understand why it improves a person's capability to understand a scene, what are its constituents, and provides guidelines for describers to follow. Section 2.1 elaborates this aspect of linguistics and concludes with its shortcomings, some of which were addressed in this work. Evaluating how aware a person is about a situation during its presentation falls within a field in psychology called *situation awareness*. This domain provides mechanisms to assess how well a person has comprehended a situation. Section 2.2 briefly discusses what situation awareness constitutes of and presents its various perspectives in literature. This chapter concludes with a detailed account on the field of *haptics*, the science of touch. Since haptics is a broad domain, section 2.3 elaborates on just one of its many facets, namely tactile communication; in particularly, vibrotactile communication, which has been made use of in this work.

2.1 Audio Description

In literature, such as novels, an elaborate and imaginative account of a scene, in terms of the location, ambience, presence of characters, their expressions and emotions, and so on, is presented prior or during the conversation of those characters in the scene. In movies, similar information is portrayed, but through the use of visual cues. For individuals who are blind or visually impaired, the visual content of a film is largely inaccessible, which makes film interpretation difficult without additional aids. To assist with film visualization and comprehension, an additional audio track of a narrator, who describes the scene and events, may be incorporated. These narrations are most commonly known as *audio descriptions*, *video descriptions, described video information (DVI)* or *descriptive video service* (DVS). Whitehead defines audio description (AD) as "an additional narration that fits in between dialogue to describe action, body language, facial expressions, scenery, costumes – anything that will help a person with a sight problem follow the plot of the story" [**7**]. Film directors typically do not account for these descriptions, which are added to the audio track of a movie after it is completed to avoid overlap with the audio of the original movie, such as conversations, certain musical scores and sound effects. Since a verbal description of a scene typically takes more time to communicate than its corresponding visual depiction, only pertinent cues are verbalized.

Audio description as a technique was developed in the United States [**8**] [**9**]; it was first suggested to be provided on films by Chet Avery in 1964, and was also the subject of Frazier's Master's thesis in 1974 [**9**]. Its usage is not limited to movies; it is available in television programs in many countries including the USA, UK, France, and Germany [**8**]. The descriptions are broadcasted using the Separate Audio Program (SAP) channel that is available with most stereo televisions [**9**] [**10**]. They are also employed in theatres, museums, galleries, and heritage sites, as well as sports venues [**7**]. Audio description, combined with color commentary techniques, have also been employed to describe live fashion shows [**11**].

After observing the impact of audio described television programs on 111 legally blind adults, Schmeidler and Kirchner [**10**] concluded that participants gathered and remembered more information through audio described content. They also found that adding description to a program made it "more enjoyable, interesting, and informative" [**10**]. Their experiment involved the usage of two programs, each from a different science series, only one of which was shown with audio description in a session, where presentation of the programs was counter-balanced across sessions. In another experiment, Frazier and Coutinho-Johnson [**12**] determined that students who are visually impaired or blind, when presented a video with audio description, comprehended the video at least as well as

sighted students, who viewed the video both with and without AD. Peli, Fine and Labianca, on conducting an experiment that involved low vision participants [**13**], concluded that some of the visual information contained in the descriptions could be gathered through the original audio of the programs used in the experiment. In a case study, Peskoe observed that audio description enhanced interest and introduced new pieces of information to a five year old female child [**14**] when she was allowed to watch two one hour children's programs with descriptions. Snyder [**9**] suggested utilizing AD techniques for narrating picture books to not only improve accessibility for children who are visually impaired, but to help all kids to develop advanced language skills.

2.1.1 Elements

In his work, Piety [**15**] observed that individuals who are visually impaired or blind, unlike individuals who cannot hear, do not have a special language (e.g., sign language) to communicate; rather, the same words and phrases, as articulated by their sighted peers, are used. With the objective of furthering research in the field of audio descriptions, he analyzed the language by transcribing the audio descriptions of four English movies, a technique that he suggested is normally employed in analyzing spoken discourses, and developed a "set of core structural and functional definitions":

- **Insertions:** Piety defined an *insertion* for audio descriptions as "a contiguous stretch of description that is uninterrupted by

other significant audio content such as dialogue" [**15**]. Though this definition seems complete, it fails to define what is considered as significant. It has been observed that sometimes audio descriptions are delivered even in the presence of background scores. Moreover, an insertion can have one or more utterances.

- **Utterances:** Piety suggests that any verbal communication usually does not involve the use of complete sentences. Referring to both philosophy and linguistics, he defines an *utterance* as "the unit of language that is actually spoken" [**15**]. These utterances are used to fill the available duration of the insertions, but can be much shorter than the available time. It is through these utterances that visual components are described.
- **Representations:** Though it is clear that utterances are used to deliver visual components through descriptions to individuals who are visually impaired or blind, it does not provide an understanding of the content to be delivered. Piety suggests that *representations* delve into the "meaning and what the describer is attempting to communicate" [**15**]. Insertions and utterances are more inclined towards the form, whereas representations are inclined towards the functional aspects of AD. They are further explored in section 2.1.2.

- **Words:** After analyzing the audio descriptions, Piety reflected that unlike other forms of language use, where information is largely tailored in past and future tenses, AD delivers information in the present tense and corresponds to events occurring at that time on the screen. Salway extended this observation to suggest that present continuous tense is used as well [**16**]. An exception to this is when the describer reads aloud text that appears on screen, which might also be past or future tense (e.g., text that appears in a newspaper or titles that appear in a movie). Therefore, the choice of words is largely constrained as compared to other discourses [**15**].

2.1.2 Commonly Described Visual Cues

The practice of using real world text or samples for analyzing a language is known as corpus based analysis [**15**] [**17**]. From analyzing the corpus of 23,000 words, obtained by transcribing four audio described movies, Piety categorized the representations into a taxonomy, depicted in Table 1.

Table 1. Taxonomy of Audio Description. By analyzing the corpus of AD scripts of four English movies, Piety classified the information included in the script. He suggested that his classification is extensible and incomplete [**15**].

Taxonomy	Description
Appearance	How a person looks in terms of clothing, facial features, hair style and physique; the location of a scene; description of an object.
Action	A description of anything that is moving or changing. It includes "gestures, movements, and activities, and they can act as the core representation around which other representations are clustered".
Position	Where characters are located in the scene.
Reading	Literal reading of information present either in a visual component of the scene or a graphic displayed on the screen.
Indexical	Associating sounds such as footsteps or speech with an actor or an object.
Viewpoint	Indicates a change in the scene, and describes special graphics present in the scene.
State	The uses of words that better reflects the state of a character, location or object that is delivered in correspondence to visual information.

Salway, in his corpus based analysis of audio description scripts from 91 movies [**16**], expanded the information types to 15: physical description of characters, facial and corporal expressions (i.e., body language), clothing, occupation and roles of the characters, attitudes of the characters, spatial relationships between characters, movement of the characters, setting, temporal indicators, indicators of proportions, décor, lighting, action, appearance of titles, text included in image. He examined the frequency of the words used in this corpus, and determined the common types of utterances where they were used:

- Characters' appearances;
- Characters' focus of attention;
- Characters' interpersonal interactions;
- Changes of location of characters and objects;
- Characters' emotional states.

In another study, Salway, Vassiliou and Ahmad [**18**] analyzed the screenplays and AD scripts of 75 and 45 movies, respectively, and found frequent words common to both. They then observed where the most common phrases (the phrases that included the frequent words) appeared in the two texts. They found look/looking/looks/looked at, turn/turning/turns/turned to, smiles at, and "open*/close* (the) door or (the) door open*/close*, where * refers to ø, -s, -ed or –ing" [**18**], to be the most frequent.

2.1.3 AD and Translation Studies

Braun [**19**] and Orero [**20**] suggested that audio description consists of two segments: (1) the process of understanding the source which is audiovisual in nature, and making a script to be narrated by a describer, i.e., the conversion of the source into a narrative; and (2) the reception or comprehension of this narration by the end user. (In linguistics, any message which can be recorded using a medium is called a text, i.e., even an audiovisual recording is a text [**21**]). In this context, Braun suggested in [**22**] that AD is a specific form of translation: intersemiotic, intermodal or cross-modal translation. (Semiotics is the study of how meanings are made with signs, where signs are the building blocks of texts [**21**].) He used, in [**19**], a discourse-based approach to provide a theoretical framework that can be used to understand this translation. (A discourse is the medium used to deliver a story such as a movie, short film or novel [**23**].) His work was based on three dimensions:

- **Mental Modeling Theory**

  Braun suggested that whenever a situation is described to a person, either verbally or visually, he or she develops a mental model, i.e., a "big picture", from the cues provided (bottom-up processing), the world knowledge, and from the context and previous cues (top-down processing). He further suggested that this mental model is responsible for end user anticipation and

expectation on the course of the situation. He also observed that when cues are obtained from more than one medium, some of them are redundant though confirmatory, while others are supplementary or conflicting. He also noticed that these cues vary in their reliability and importance. Within the realms of audio description, Braun suggested that the task of a narrator is to enable individuals who are visually impaired or blind to build mental models that are similar to the ones created by their sighted peers. Since audio description has to fit within pauses, he noted that AD describers need to be highly selective in the cues that are to be described, but suggested that the precise contribution of each cue in the audiovisual source is difficult to establish. He further suggested that an audio describer, after assessing the contribution of a cue, should describe it only when the cue is visual and cannot be inferred from other complementary or redundant cues from the auditory medium.

- **Relevance Theory**

  Braun observed that the mental modeling theory fails to describe "how the cues and knowledge are integrated and how only relevant knowledge is activated in the creation of mental models". He suggested that inferential models focus on how individual cues are processed and understood. In order to understand the meaning of an utterance, he suggested that it is

essential to identify the factual content in the utterance and the assumptions manifested with the utterance by the speaker. Relevance Theory is one of the inferential models for communication and Braun reasoned its usage in his analysis rather than other inferential models given (1) this model claims that the derivation of the propositional content and the assumptions are highly inferential processes; and (2) it distinguishes between the assumptions that the speaker intended to communicate explicitly, known as explicatures, with those that were intended to be implicit, known as implicatures, both of which create meaning. The Relevance Theory is guided by the Principle of Relevance which suggests that the receivers of an utterance are "entitled to believe that an utterance or cue is presented to them in the optimally relevant way" and that an utterance is not further processed once the addressees have derived a sufficiently relevant meaning. He suggested that one of the strategies employed in AD is to verbalize the explicatures so as to reduce the processing load of the end user as well as conform to the timing constraints in AD. Also, he suggested that such a narrative leaves scope for interpretation of the implicatures.

- **Coherence Theory**

  An interesting observation by Braun is that when AD is produced as a result of a translation of the audiovisual source, it becomes part of the source, i.e., it is processed along with the utterances and the background scores of the source. He suggested that the notion of connectivity across AD sections and between AD and other elements of the source is, therefore, essential. He defined coherence as the connectivity of content that is created in the receiver's mind "based on their general assumption that utterances make sense in the context in which they appear". Coherence is observed to exist at two levels: local and global. He suggested that local coherence exists within a scene while global coherence exists across scenes. Through specific examples, he observed that the narratives ensure that both local and global coherence is preserved between and across AD, and other elements in the audiovisual source.

2.1.4 AD and Narratology

Salway and Palmer, in their work [**23**], defined narratology as a branch in literary study that deals with the study and analysis of the narrative and its structure; it "studies the nature, form and functioning of all narratives irrespective of their mode or medium of representation". They suggested that audio description follows a flavor of narrative known as behaviorist narrative. According to them, this

form of narrative provides information about a character's actions and words but only implicitly tells about the character's emotional state. They argue that one of the objectives of audio description is to help the audience understand the mental state of the characters on screen. Referring to Palmer's earlier work, they suggest that utterances in AD follow a concept called the Thought-Action Continuum. They explained the concept through the following example:

They hid behind the curtain.

They stood behind the curtain.

When the two sentences were compared, it was inferred that, though both the sentences meant the same when it came to action, the word *hid* in the first sentence suggested also about the mental state of the characters implicitly. The thought-action continuum can be imagined as a scale, where one end is marked as action and the other end is marked as thought. They suggested that *stood* in the second sentence is at the action end of the scale, while *hid* in the first sentence is at the middle of the scale.

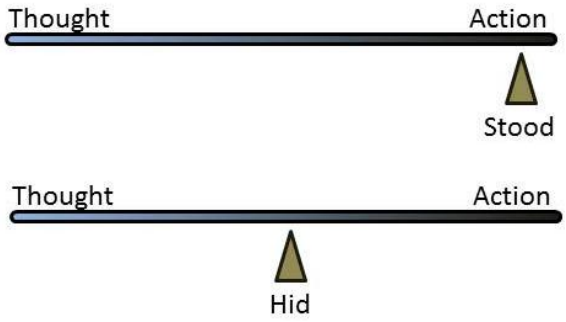

Figure 1. The words *hid* and *stood* as mapped to the thought-action continuum scale in a context in which they were used. This figure shows that the usage of the word *hid* provides more information about the mental state of a character.

Orero [**20**] suggested that recounting or narrating events "is a culturally determined, conventionalized activity, and that it remains so even when these events are recounted through new forms of communication". These cultural differences in narrating events were therefore used in her work to reason the different expectations of viewers of audio described content in different countries: Spain, Greece and the United States, in particular. She extended Tannen's work on the Pear stories project [**24**].

The Pear stories project involved a six minute video, called the pear film, which contained background scores but not dialogues. Tannen [**24**] showed this video to 40 participants, 20 each from the US and Greece, each of whom gave a verbal account of the video in response to the question, "What happened in the movie?", in English and Greek respectively. Tannen translated Greek narratives to English and found that:

- Americans tended to refer, either directly or indirectly (known as allusions), to the film as a film while providing an account, whereas Greeks tended to not do so. Direct reference to the movie was observed through the usage of the words "movie" or "film" in the recount, while indirect reference was observed through the usage of phrases such as "then we saw", "you could see", "it showed", "the camera pans", and so on.
- Americans tended to be more objective in their description of the events, while Greeks tended to interpret them. Tannen inferred that Americans were performing a "memory task" while Greeks appeared to be "telling a story", and that the Greeks weren't perturbed about the technique used in the scenes, rather the message that it conveyed.
- Americans predominantly used the present tense in their account while Greeks used the past tense.
- Greek narratives tended to not include information from the movie that did not contribute to their interpretations.
- Greek's interpretation sometimes showed signs of philosophizing.

Tannen, noting the changes in the narrative styles of working class, middle class and upper class, and those who received formal education, postulated that the Americans employed "strategies associated with the literate tradition of schools" while the Greeks

employed "strategies associated with the oral tradition of the family and peer group".

Orero [**20**] used the same video and asked 20 participants in Spain to provide a written narrative of "what they remembered" from the video. She observed that:

- Even though Spanish participants did use allusions, they weren't elaborative.
- They were neither interpretative of the actions that were recounted nor did their accounts involve elements of philosophy.
- Spanish participants tended to use present tense, akin to the Americans.

Hence, if the responses of the Americans and that of the Greeks can be visualized as two ends of a spectrum, Spanish tended to stay in the middle.

2.1.5 Different Styles of AD

An audio describer conventionally refers to the characters in the movie/television series, and describes the events taking place. The narrators themselves do not assume a character in the content; they use a third person narrative style. This form of narrative in AD is the most common, and is called conventional AD/DVI [**25**]. Fels, Udo, Diamond and Diamond [**25**] explored the usage of a first person narrative style in audio description and compared its performance with the conventional style. They suggested that in first person narrative

style, the narrator has to play a dual role of a) being a character in the actual content, and b) narrate the events from the character's perspective in a different time frame. They also suggested that this narrative style is "indicative of a more oral style of storytelling" and that it allows a person to think auditorily when used in AD. They classified the conventional third person style as extrovert or covert, and the first person style as introvert or overt. Table 2 compares and contrasts the two styles.

Table 2. First-Person versus Third-Person style narration in AD. Both styles of audio description provide information in a different perspective. The advantages and disadvantages of both these styles are provided below [**25**].

First-Person Narrative	Third-Person Narrative
Audience can identify with the character and comprehend the inherent subjectivity in the narrator's version of the story.	The narrative is objective in nature.
The audience can interpret and criticize the narrator's version of the events.	The narrator is not part of the story, and therefore, does not share the emotions as any of the characters.
The narrative is skewed and is influenced by a single point of view.	It is neutral and does not interfere with the story.
This skew may deem the narrative to be untrustworthy.	It is trustworthy and informative.

Fels et al. [**25**] confirmed the existence of these differences through their study which compared an episode of Odd Job Jack, an animated comedy series on Canadian television, when provided with different narratives. The first person style narrative was included by the makers of the series, while the conventional style narrative was provided by a third-party company. Since blind viewers more often view television or movie with sighted peers, they included an equal number of visually impaired and sighted individuals in the study, and made a group-wise analysis of their data. They suggested that their results were subjected to skews such as a) the level of comfort and familiarity of the conventional style narratives by individuals who are blind, and b) the desire of fun and entertaining content that they felt was apparent in the first person style. On finding that first-person style AD is more "engaging, entertaining and preferable, but less trustworthy than the conventional third person style AD", they concluded that one narrative style may not suit all types of contents or all viewers.

Udo and Fels [**26**] provided a case account of a live production - Shakespeare's play, Hamlet - where, after considering the results in [**25**], an alternate audio description strategy was implemented. They suggested that audio description should be considered by the director and should be part of a creative team headed by the director. This play

used the following techniques in delivering content to people who are visually impaired or blind:

- When describing the set, only the idea was communicated, not the actual layout.
- The approach involved more interpretation and less objectivity by the describer.
- Sensory-based images, or color commentary techniques, were used to describe events.
- At times, first-person narrative techniques were employed.

2.1.6 Audio Films

Lopez and Pauletto [**27**] suggested an alternative to audio description. They explored the usage of surround sound, enhanced with sound processing and spatialization, in conveying information about the visual elements on screen. Though this notion is similar to radio drama, they suggested that their work emphasizes the usage of sound and not narration that is common in radio dramas. They suggested that their aural approach is targeted to stimulate the imagination of a listener. They adapted and employed two different cinema languages:

- **Master Scene**

  This concept suggests the usage of a single shot to make the audience aware of the environment and the people involved,

followed by the actions and the events [**27**]. They used this concept to aurally establish the ambience of the scene.

- **Interpersonal Cinema Language**

  This concept suggests involving the audience with the emotions and expressions delivered by the actors. This was accomplished in this work through the usage of expressive voices and background scores/music.

They designed an audio film version of a 1954 movie, Lamb to the Slaughter. Their work involved two steps:

- **Production Stage**

  They emphasized the recording of soundmarks (which is similar to the notion of landmarks in the visual domain), footsteps, character-object interaction, and internal sounds such as breathing.

- **Sound Processing Stage**

  The production was enhanced by modifying character sound to better represent the environment in which the scene was taken: enhanced footsteps; sound heard through other rooms; sound heard through windows; modified dialogue sounds to suggest that one character is speaking with his/her back to the listener; and enhanced spatialization of content through the usage of a 6.1 surround sound layout as shown in Figure 2.

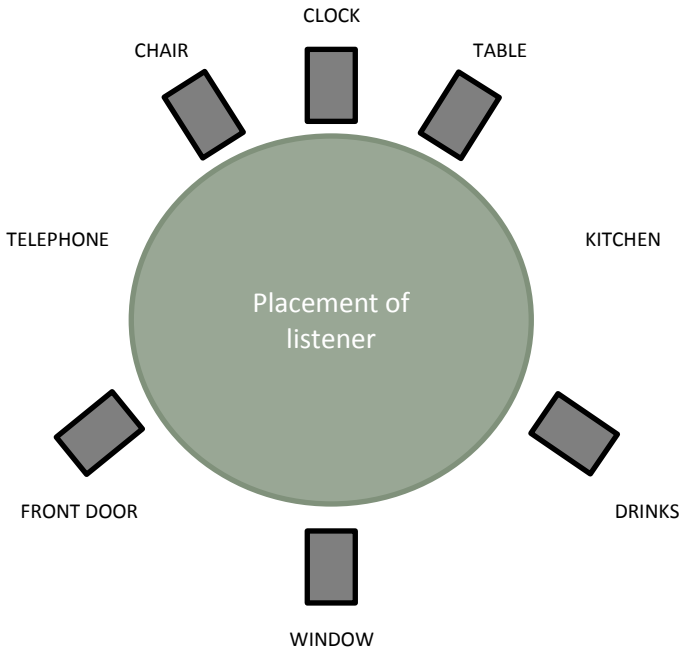

Figure 2. The layout of a living room with a 6.1 surround sound setup. Adapted from [**27**].

On analyzing the feedback of 13 sighted participants who listened to this audio film, they suggested that their format had the potential to convey a story successfully without the need of visual elements or descriptions. None of the participants were able to recognize all the characters. The presence of soundmarks and the use of sound effects in the film were suggested to be essential by the participants.

2.1.7 Computer Science and AD

Branje, Marshall, Tyndall and Fels [**28**] developed a system, called LiveDescribe, that allows adding near real-time audio description to online video content. It allows describers to define a small look-ahead

window, which could be anywhere between 30 seconds and 5 minutes, to determine the duration of silent periods in the live video stream. The describers only need to click on a record button to record their narrations with the stream. The authors suggested that, using their system, the audience of the live stream with the descriptions would experience a delay equivalent to twice the look-up window duration.

Salway, Graham, Tomadaki and Xu [**29**] worked on integrating audiovisual movies with audio descriptions, presented as a text, via representations of narrative, and the extraction of narrative information automatically from these descriptions. They built a system for AD called the AuDesc. This system takes the descriptions as time-stamped texts; time-stamped with respect to the video. These texts were used to automatically annotate video segments with keywords. These annotations were then proposed to generate higher level semantic media content through semantic networks and plot units. They also aimed to determine dramatically important sequences and emotion patterns, and use them to index the video. Through audio description scripts, they extracted various emotion words/tokens, and classified them as one of 22 emotion types. They then provided a visualization of the emotion types for the video analyzed. Through this visualization, they were able to determine the regions in a movie where there was a high degree of joy, distress, fear and other emotion types.

In an attempt to use eye-tracking technology in the area of media accessibility, Igareda and Maiche [**30**] aimed to determine the extent of personal interpretation by the describers in embedding descriptions in movies. They proposed using sighted participants and allowing them to watch movie clips without descriptions. Their eye movements were recorded and analyzed. Later, those participants who had similar scan patterns would watch the clips again with the descriptions. Issues, such as influence of the non-described content on the described video that would be shown again, were not addressed. They hypothesized that participants, when watching movie clips with the descriptions, would have a higher number of eye movements and saccades than in the non-AD version. They further suggested that participants' eye movements when watching movie clips with AD will be influenced by the descriptions. They targeted clips where emotions and gestures were described.

Lakritz and Salway [**31**] suggested a semi-automated mechanism to generate audio description from screenplays. Their objective was to build a system that could result in a reduction of man hours employed in embedding audio description in films. They conducted a corpus based analysis on 70 audio described movies and screenplays, mapped the utterances in the screenplay to the events in AD, each such mapping was called a SP-AD pair, and found that 60%

of the information conveyed in AD was available in the screenplays, and that 20% of the screenplay utterances were part of SP-AD pair. They further built a system that consisted of four modules:

- Find gaps in the dialogue.
- Extract pertinent information from screenplay.
- Convert to AD language.
- Compress to fit gap.

Each module was a computer program. On analyzing the performance of the extraction module on a set of three screenplays, they found that it provided a precision that was three times greater than that could be achieved when the whole screenplay was returned. They suggested that after sentence extraction using their algorithm (module 2), the percentage of sentences that could feature AD increased from 20% to 41%, and that after language conversion (module 3), it increased to 48%. They reasoned the usage of these modules because of the following differences between screenplay scripts and AD scripts:

- Screenplay scripts use first person perspective while AD scripts use third person perspective.
- Screenplays provide extensive information about camera movements and other information pertaining to how a scene should be shot, while, according to them, they are not included in AD.

- Screenplays provide descriptions of how a dialogue should be delivered, and refer to sound effects as well. They suggested that such information is not delivered in AD.

Since the sentences were made available from screenplays, they suggested that they may differ from the actual timeline in a movie due to re-ordering of scenes while editing.

Gagnon et al [**32**] used computer vision techniques to aid individuals in developing and rendering audio descriptions. They developed a system, called AudioVision Manager, which extracts key frames, people, actions, gestures, objects, texts and many others in a movie. It allowed describers to provide information via text about the extracted content and generated descriptions automatically. They also developed a player called AudioVision player that would include these descriptions, delivered through a synthetic voice, to the audio track.

2.1.8 Drawbacks

In an interview with more than 50 participants who were blind or visually impaired, Gagnon et al. [**32**] classified the difficulties/drawbacks of audio description into a) technical aspects, and b) informational aspects.

In terms of technical aspects, they found that individuals who are blind had:

- Difficulty in separating the description from the sound track in a movie.

- Difficulty in delineating information as that being part of narration or of audio description, when the original movie had a narrator.
- Trouble listening to the descriptions sometimes when they were overlaid on top of background scores, and found that the volume of the music was high and irritating. But at other times they found the scores to be enjoyable and information-bearing.
- Objection when the descriptions were interpretative in nature, and suggested that they should be neutral while at the same time maintain their interest in the movie.
- Consensus that the descriptions should provide some background/contextual information about the movie at its beginning such as genre so that they are prepared cognitively for the information that will follow.
- Different needs regarding the amount of information to be provided through the descriptions.

In terms of informational aspects, they found that individuals who are blind had:

- Difficulty in sorting and managing all the information received through the auditory medium.
- Difficulty in distinguishing a large number of characters, if they were introduced to them in a short period of time.

- Expressed importance in being aware of the environment and the context of the movie to comprehend the situation in that movie.

2.2 Situation Awareness

Endsley [**33**] defined *Situation Awareness* (SA) or situational awareness, as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". For example, consider a person driving a car. To successfully navigate from point A to B, the driver must accelerate, decelerate, halt, signal, blow horn and turn. All these acts were in response to the environment, such as traffic, speed limits, traffic signals, unruly pedestrians, and unruly drivers among others. This person is said to be situationally aware of the surrounding environment, enabling him or her to make the right decisions (such as applying brakes to avoid hitting a car that is directly ahead). From this example, it is clear that situation awareness requires two things: (1) an operator, and (2) the operated, or the environment where the operator acts. Being a well-established field in the domain of human factor studies in complex environments [**34**], Endsley, in [**33**], suggested that situation awareness is a state of knowledge that is different from the underlying processes involved to achieve that state. She termed these processes as situation assessment. The operator perceives and comprehends

(through a set of processes using his or her mind), while the environment (which is processed) provides a state. She suggested that these processes are used to achieve, acquire and maintain SA. Figure 3 presents Endsley's model of SA.

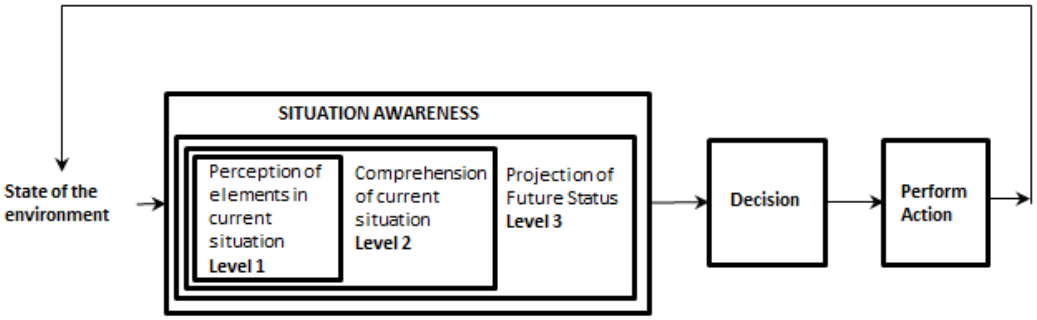

Figure 3. Though Endsley suggested that situation awareness is determined by the state of the environment, she provided three stages or processes involved in achieving, acquiring and maintaining SA, or in her words, situation assessment. Figure adapted from [**33**].

Rousseau, Tremblay, and Breton [**34**] suggested that every other definition in the literature on situation awareness deals with the duality of SA as a state or as a process. They suggested that in the literature, situation assessment has been a broader concept where situation awareness, the way Endsley had defined it, becomes a component. They further suggested that situation awareness is not the same as either the momentary knowledge of a situation that a person may be aware of or a verbal account about the content of a situation that the person may be conscious about. They suggested that there are two approaches to defining SA:

- **Operator-focused approach**

  This approach focuses on the cognitive processes, or mechanisms, that lead to the development of a mental representation of the state of the environment. They suggested that these processes are a general property of the human operator. They also suggested that this approach follows an information processing framework that describes the processes involved in providing cognition.

- **Situation-based approach**

  This approach is based on the viewpoint that SA is determined by the state of the environment or situation. They suggested that two principles in the study of direct perception are of interest to SA: (1) environment contains all the information necessary for perception, and (2) perception is immediate and spontaneous. They observed that though this approach provides a more factual basis to defining SA in terms of events, objects, other persons and systems, the situation itself is dependent on the domain and that the elements contained varies across situations and domains.

Patrick and James [**35**] introduced a task oriented perspective to SA. They derived techniques from Hierarchical Task Analysis (HTA) for "conceptualizing the status of the concept of SA and the processes associated with the acquisition of SA". In terms of HTA, the objective

of achieving a goal is broken down into tasks, with each task being broken further into sub-tasks. The tasks at each level are executed in an order determined by a plan, shown in Figure 4.

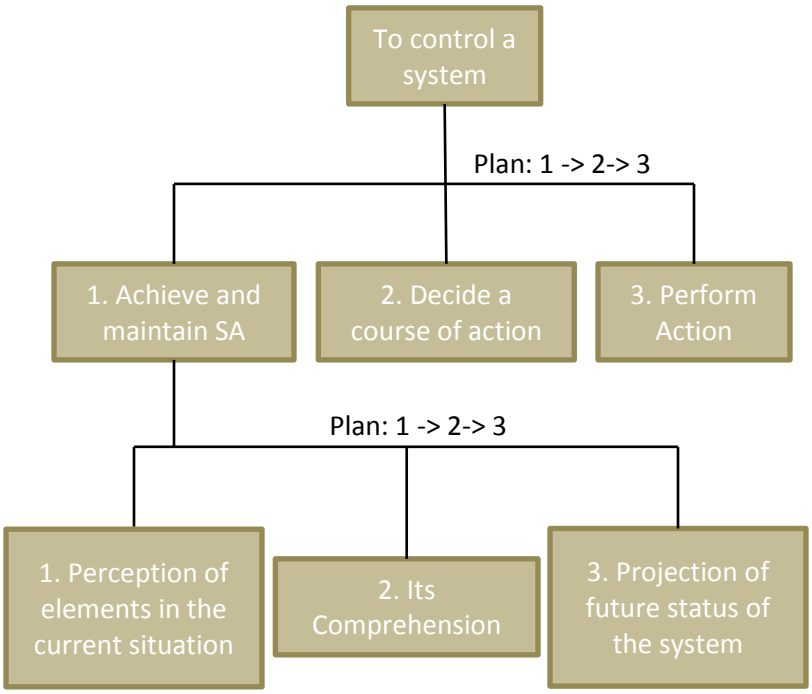

Figure 4. Hierarchical Task Analysis being applied to Situation Awareness. The plan specified at each level identifies the order of task execution. Figure adapted from [**35**].

This subdivision of tasks is governed by a *PxC* rule, where *P* stands for the probability that satisfactory performance can be achieved at a specified level of description, and *C* stands for the cost of system performance. If either of the parameters is adjudged as unacceptable, the tasks are broken down to another level of description.

2.2.1 Good and Bad SA

Situation awareness is about building a mental representation that resembles the actual state of a situation. Dekker and Luftzhoft [**36**] observed that based on the accuracy of this resemblance, a person may be termed as having good or bad SA (see Figure 5).

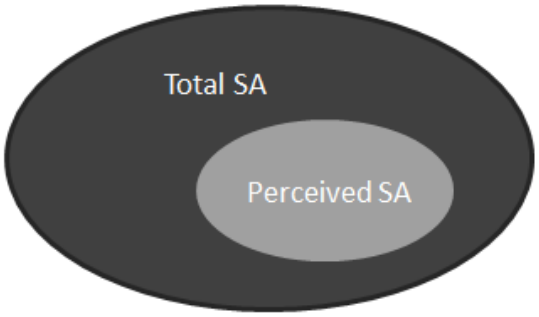

Figure 5. A person's SA is a subset of an ideal/total SA. In this Venn diagram, *Perceived SA* depicts that of a person's. Figure adapted from [**36**].

They suggested that this is a subtractive model where a person's SA is governed by a norm, hence terming it as normativism. In this approach, an objective outside world independent of the observer is essential. Such an approach, they suggested, coincides with rationalism. In literature, they observed that a precise representation of a person's environment in his or her mind was not essential as long it supports their action and enables achieving their goals. Hence the notion of global rationalism was replaced by local rationalism. They suggested that a person's good or poor SA should be analyzed by recreating his or her local rationality.

Three areas in psychology suggest as to what this representation is made of, and how meaning and orderliness is obtained. They observed that empiricism provides theories suggesting that a person perceives his or her environment as elements that are disjointed, but mediated though previously acquired knowledge. These elements are processed in our mind to create meaning. They also observed that gestaltism suggests that humans perceive meaningful wholes. Another area in psychology, called radical empiricism, debunks the separate notions of mind and matter, and suggests that humans perceive meaningful relations. According to this form of empiricism, there is no ideal or actual SA, but rather different rationalities, which vary from person to person, none of them being right, wrong or better.

2.2.2 The Effect of Existing Knowledge in SA

Croft et al. [**37**] suggested that implicit awareness about certain aspects of a situation influence a person's SA. They defined implicit memory as "the non-intentional, non-conscious retrieval of previously acquired information, and is demonstrated by performance on tasks that do not require conscious recollection of past experiences". They explained this influence through an example experiment. In a study, they allowed participants to read through a long list of words. They then asked participants to respond with the first word that strikes their mind to a series of word stems. They observed that participants tend to provide words that were from the list they had read through. This is

in contrast to explicit memory where only a partial picture on a situation's element and meaning is elicited. In the above experiment, explicit memory can be tested by asking the participants to provide all the words that they remember from the list. They suggested that similar behavior can be elicited through implicit learning, which enables participants to control a dynamic system that involves manipulating variables without knowing the underlying rules of the system. They defined implicit learning as acquiring "new information without intending to do so, and in such a way that the resulting knowledge is difficult to express". On analyzing implicit knowledge, they suggested that it has the following characteristics:

- Independent of the locus of attention.
- Temporal durability.
- Incidental testing.
- Independent of subjective confidence.
- Sensitive to the phenomenon of implicit expertise.

2.3 Haptics

Our sense of touch provides us with information from our environment; information such as the texture, shape, temperature and material of surfaces and objects we encounter and explore. In social interactions, our sense of touch is a component of nonverbal communication, and is vital in conveying physical intimacy [**38**]. The somatosensory system [**39**] is comprised of touch receptors and

processing centers that enable us to sense and perceive the following stimuli: tactile (pressure, vibration, temperature and pain) and kinesthetic (body movement, position and weight). As the number of touch receptors in the skin varies across the body, each part of our body varies in its spatial acuity to tactile stimuli. A two-point discrimination study [**39**] revealed the minimum distance across the body required for two point stimuli to be perceived as two points of contact rather than a single point of contact. This study showed that our hands and tongue have high spatial acuity compared to other regions such as our waist and thigh. The density of touch receptors in our body varies across its different regions. Each body part, therefore, has a different amount of cortical space in the somatosensory cortex of the brain depending upon the density of receptors in the skin at that location. The sensory homunculus is an exaggerated representation of the human body used to compare tactile sensitivity of different body regions where the size of a body part is correlated with the degree of its tactile sensitivity. A region with greater receptor density would therefore be enlarged in this representation, while a region with lesser receptor density will be comparatively smaller. An example of a body part with a high receptor density is our hand, while one of lesser receptor density is our arm.

Though the two-point discrimination study investigated pressure stimuli, rather than vibrations, it provides a starting point for vibrotactile spatial acuity studies. Researchers have conducted experiments to determine vibrotactile spatial acuity at specific body locations such as the waist [**40**] and arm [**41**], discovering the tradeoff between localization accuracy, and display size and spacing for these regions, along with useful design strategies for improving accuracy.

2.3.1 The Human Touch Receptors

Kandel et al. [**39**] suggested that tactile sensitivity in the human body is greatest at the glabrous, or the hairless skin, on the fingers, the palm of the hand, the sole of the feet, and the lips. There are primarily four receptors, known as *mechanoreceptors*, which sense tactile and kinesthetic stimulation applied to the skin and limbs of the human body: (1) Meissner's corpuscles, (2) Merkel disk receptors, (3) Pacinian corpuscles, and (4) Ruffini endings. Although the stimuli they are sensitive to, determined by their structure, are different, they share the following three steps when stimulated:

- First, a physical stimulus is delivered at or near the receptor.
- Then, the deformation of the receptor is transformation into nerve impulses.

- Finally, we perceive the sensation, where our perception depends on which neural pathways are activated and activation patterns.

Kandel et al. [**39**] suggested that all senses, including the ones that perceive touch, when stimulated, provide four rudimentary types of information:

- **Modality:** It specifies the type of stimulus that was received, the type of the impulses that were transmitted because of this stimulus and the receptors that are used to sense the stimulus.
- **Location:** Receptors of all four forms are distributed throughout the human body. The location of a stimulus is therefore represented through a set of such receptors that are currently active. Localized receptors, therefore, provide not only the body site of the stimulation, but also its size depending upon the number of receptors that were activated.
- **Intensity:** Since the generation of nerve impulses is in reality a transformation of the stimulus energy that was received, intensity, marked by the amplitude of these nerve impulses, suggest the total amount of such stimulus energy received by the receptor.
- **Timing:** The duration of a sensation is determined by the start and end of a response by a receptor, influenced by its adaption rate. Assuming receptors sensitive to light touch applied to the

skin, although sensory neurons will respond at a rate proportional to the pressure applied to the skin, as well as its speed of indentation, the sensation is lost if it persists for several minutes [**39**]. Depending on the receptor, adaption rates vary: Merkel disk receptors and Ruffini endings adapt slowly, whereas Meissner's corpuscles and Pacinian corpuscles adapt fast.

Kandel et al. [**39**] further suggested that each type of receptor only respond to a narrow range of stimulus energy, known as its bandwidth. Such a range is responsible for limiting the human perception of audio cues to the hearing range of 20 Hz and 20 kHz, as well as our perception of electromagnetic waves in the region of 790 to 400 THz as visible light.

Meissner's corpuscle lies just beneath the surface of the skin. It is a rapidly adapting receptor sensitive to light touches. They are highly sensitive to vibrations of frequency less than 50 Hz [**42**]. They were discovered by the anatomist Georg Meissner [**42**].

Merkel disk receptor is a slowly adapting receptor also present in the superficial layer of the skin. This receptor transmits and responds to compressing strains such as pressure and texture [**39**].

Pacinian corpuscle is a rapidly adapting receptor that lies in the subcutaneous tissues of the skin. It does not respond to steady pressure, but to rapid indentations, i.e., vibrations [**39**].

Ruffini endings are slowly adapting receptors, present in the subcutaneous tissue of the skin. These receptors are sensitive to skin stretch, and help us grasp and hold objects by preventing slippage [**39**].

2.3.2 Haptic Communication

There are essentially two types of haptic stimulation: kinesthetic and tactile. The former acts on the proprioceptors, and is responsible for exerting force through force feedback devices; while the latter acts on the surface of the skin. Both stimuli types are vital while exploring objects. For example, consider holding a rock or a tennis ball in your hand: tactile stimulation provides cues about the texture of the object's surface (smooth, course, rough, etc.), while kinesthetic stimulation provides cues about how heavy or light the object is. Both types of feedback are essential to our comprehension of the environment around us.

The most common force feedback devices are the PHANTOM®, by Sensable Technologies, and the Falcon®, by Novint (see Figure 6). Since this study focuses on tactile feedback, haptics will be synonymously used to refer to tactile cues.

Figure 6. The PHANTOM® (left) and the Falcon® (right) are both force feedback devices used to exert forces upon the human hand to simulate the shape, texture and/or material of virtual objects. The PHANTOM® is more commonly used as a research tool, whereas the Falcon® is designed for gaming.

2.3.3 Vibrotactile Communication

There are three main vibrotactile feedback devices discussed in the literature: C2 tactors, Tactaid actuators and pancake shaftless vibrators, each depicted in Figure 7.

Figure 7. C2 tactor [**43**] (left), Tactaid actuator [**43**] (middle) and pancake motor (right). All three motors have been widely used as part of research projects, but the pancake motor is most commonly used in commercial products.

There are essentially six parameters that can be used to design vibrotactile cues:

- **Frequency**: Frequency is the number of cycles that a waveform makes per unit time. Vibration frequency is the number of cycles a waveform, used in the creation of the vibration, makes per unit time. The higher the number of cycles per unit time, the greater the frequency. Figure 8 shows an example of high and low frequency waveforms. Similar to the limitation of our eyes to perceive electromagnetic radiations within a frequency range and interpret them as distinct colors, our skin has a limitation on the frequencies within which it can sense as vibrations. Studies indicate that humans are sensitive to vibrotactile stimulation that is within the range of 20 – 1000 Hz, but have been found to be most sensitive at 250 Hz [**44**]. The maximum number of distinct cues that can be formed by varying this parameter is unknown, but we are more accurate at relative comparisons compared to absolute comparisons, although the latter might be more useful for communication [**44**]. Moreover, varying this parameter affects our perception of amplitude, and vice versa. Thus, forming haptic cues based on this parameter is not recommended.

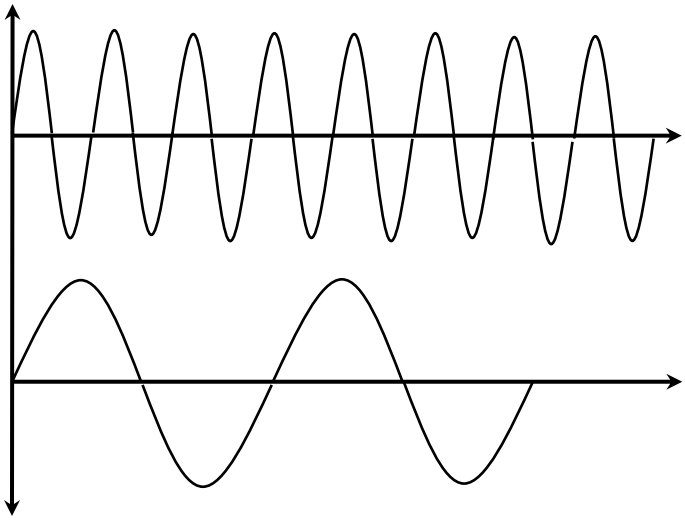

Figure 8. Example of frequency variations showing high frequency (top) and low frequency (bottom) waveforms.

- **Amplitude**: Also known as intensity, amplitude in a cycle is the magnitude of change with which a wave form oscillates in that cycle. For example, consider a sine wave having a crest value of 5dB and a trough value of -5dB, the magnitude of this wave for this cycle is 5dB. A pictorial representation is shown in Figure 9.

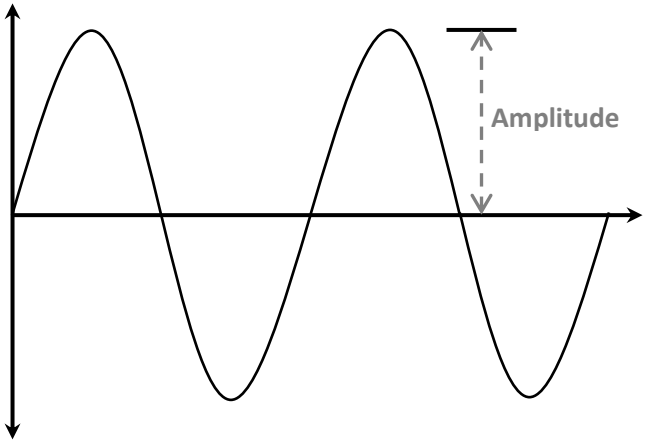

Figure 9. A constant frequency sine wave depicting amplitude.

In vibrotactile communication, studies indicate that 55dB above the threshold level of detection is the determined intensity range [**44**]; overshooting which results in pain [**44**]. Also, it has been determined that perception starts deteriorating at around 28 dB. The JND value for intensity has been determined to be 0.4 dB to 3.2 dB [**44**]. Brown et al. [**45**] conducted an experiment where intensity was used as a parameter to construct vibrotactile cues in pancake vibration motors. This was achieved by varying the frequency of the vibrations at three levels. They found that participants could recognize three different intensity values with 75% recognition accuracy, which provides a significant improvement over roughness (see below). This study suggested that cheaper tactors that cannot convey roughness, such as pancake motors, can still be used to provide complex tactile cues.

Waveform: Periodic signals can differ in the shape of their waves. The most common waveforms are made of sine, square and triangular waves (see figure 10). In the realm of vibrotactile communication, Brown et al. [**43**] suggested that the shape of a waveform was not a discernable dimension in making a distinction among various signals. A variety of waveforms can be generated by superimposing a collection of simpler waveforms (see figure 11).

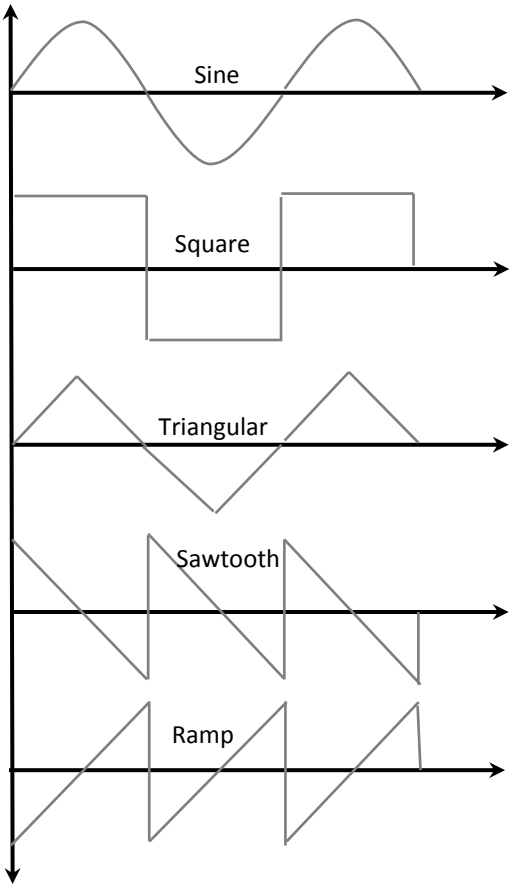

Figure 10. Common waveforms.

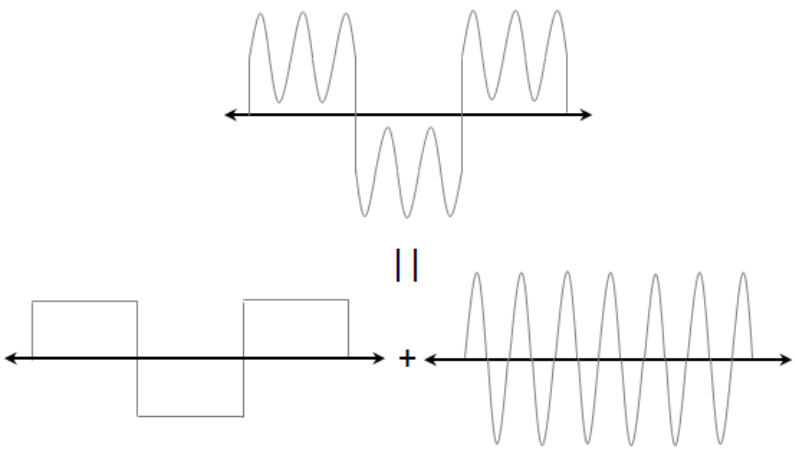

Figure 11. A square wave modulated by a sinusoidal wave creates a new waveform, shown in the top part of the figure. Figure adapted from [**46**]

Typically there would one base waveform superimposed by another waveform. This process is known as modulation and such waveforms are known as composite waveforms. In the same study, Brown et al. report that amplitude modulated sine wave (see Figure 11) was perceived by subjects as being rough. Brown et al. [**43**] investigated perceptual differences between the C2 tactor and the Tactaid actuator, and found that the former is better at conveying roughness; this was based on an experiment where four different amplitude modulated sine waves were presented along with an unmodulated sine wave with the aim being that all the amplitude modulated sine waves would be perceived as rougher than the unmodulated wave, and that sine waves with lower modulation frequencies would be perceived as rougher. Though pancake motors are traditionally viewed as a device incapable of sending multidimensional information, Brown et al. [**45**] have shown otherwise. They suggested creating an approximation of roughness in mobile phone motors by varying the speed of on-off pulses. These motors are increasingly being used in tactile devices such as vibrotactile belts, gloves, and suits among several others.

- **Timing**: A pulse is defined in vibrotactile literature as the duration for which the device producing the cues is switched on. Therefore, a pulse is a waveform having a certain frequency and

amplitude, and produced for the duration that defines the pulse. Geldard [**47**] suggested that users perceive pulses of less than 0.1 seconds as a tap or a jab. Different or similar pulses produced at periodic or aperiodic intervals and forming a cohesive whole is defined as a rhythm. In simple terms, it's a collection of pulses that occur in sequence, separated by a predefined off time. McDaniel et al. [**48**] [**49**] have used tactile rhythms as a means for communicating interpersonal distance to individuals who are visually impaired or blind through a vibrotactile belt. Other important contributions include aiding individuals who are visually impaired or blind in everyday social interactions [**50**], and in navigation [**51**].

- **Body Location**: Body location, or body site, can be used to send cues that may be comprehended based on the context of the presentation. Different body parts have been chosen to deliver tactile cues: these include the waist, arm, hand, and back. The delivery of the cues should take into consideration the sensitivity and the spatial acuity of the body part. The selection of a body part for cue delivery should be guided by results on vibrotactile spatial acuity. On the basis of these studies, a region of high spatial acuity can determine two vibrotactile cues placed very close to each other as distinct, and at the same time, localize the cue. Cholewiak et al. [**40**] explored vibrotactile

localization across the human waist, finding a mean localization accuracy of 74% (for 12 tactors around the waist), 92% (for 8 tactors around the waist), and 98% (for 6 tactors around the waist). Although this suggests that the *number* of tactors is the crucial factor in localization performance, tactor *spacing* is much more important. Given sufficient spacing, we are able to accurately localize different tactors, regardless of their number (obviously to some extent). They also explored the usefulness of anatomical reference points, such as the navel or spine, and artificial reference points, such as endpoints and odd sites, for further improving localization performance.

- **Spatio-temporal Patterns**: A one-dimensional or two-dimensional array of tactors can be used to convey spatio-temporal patterns [**52**], i.e., vibrations that vary spatial over time. Spatio-temporal vibrations are widely employed to elicit specific emotions [**53**] or convey high-level concepts. These are also commonly used [**54**] to elicit perceptual illusions such as *saltation*. Saltation [**55**] is evoked through the successive delivery of pulses in an array of tactors, one after the other in a particular direction, creating the sensation of apparent motion such that stimulations appear to be a continuous train of pulses from the start of the array, to its end, even appearing between tactors.

2.3.4 Vibrotactile Icons - Tactons

A graphical icon enables users to distinguish between the hundreds of applications installed in a computing device quickly. When the concept of such icons is designed in the tactile medium, we have *tactons*, or tactile icons. To formally define tactons, they "are structured, abstract messages that can be used to communicate messages non-visually" [**44**]. They are typically short, brief, and carry a distinct meaning within the context of a particular application. They could be used, for example, to signal an important event, convey the priority of a message coming in on a cell phone, identify interpersonal distances in social interactions, communicate the identity of an actor in a movie, and many other uses. Tactile rhythm, roughness, intensity and location have been explored to build tactons. Tactons fall into three categories:

- Compound Tactons
- Hierarchical Tactons
- Transformational Tactons

Consider a scenario where the most basic cues have been identified to be remote, key, open, close and car, and tactons have been created for them. A compound tacton can be formed by the concatenation of two or more of the above tactons, such as open car, close car, and so on. Hierarchical tactons add to the inherited properties of a base tacton. For example, consider Figure 12 which shows the hierarchical

structure for cars of different types. In this example, our base tacton (a simple rhythm) may vary in tempo depending on whether the car is a bus (slow) or race car (fast).

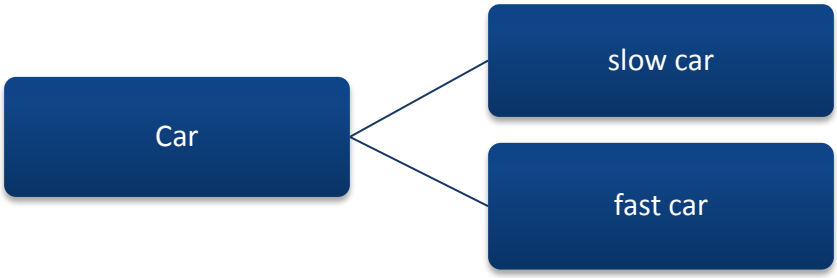

Figure 12. Hierarchical structure for a set of tactons used to represent slow car and fast car.

A transformational tacton has meanings arbitrarily assigned to its different dimensions, where each dimension may have multiple values. For example, to communicate that the remote control has successfully unlocked the car door, a vibration with weak intensity may be delivered to the user via the remote control; whereas a strong vibration might convey the opposite, i.e., the car door has been successfully locked.

CHAPTER 3

RELATED WORK

Embedding haptics into audio-visual content (such as a television show or a movie) has been viewed as an avenue for enhancing its immersiveness or viewing experience. Gaw, Morris and Salisbury [**56**] suggested that such an approach is an extension to earlier enhancements such as higher-resolution formats, larger screens (such as with IMAX), greater sound fidelity and larger numbers of speakers (such as 5.1 surround sound). Section 3.1 elaborates on the usage of haptics in movies and other broadcast media.

This work views haptics as a sensory substitution medium, and uses it to deliver pertinent visual cues for individuals who are visually impaired or blind. It is viewed as a medium that supports and adds information to the auditory channel. Section 2.2 elaborates on this viewpoint and provides details of existing contributions.

3.1 Haptics in Movies and Broadcast Media

Gaw, Morris and Salisbury [**56**] provided an authoring environment to embed force feedback information into movies that can be experienced by a user through a force feedback device (such as a PHANTOM® device). Their objective was to develop an authoring system that allows users to feel, e.g., the forces experienced by one of the actors in a sword duel, thereby enhancing his or her viewing experience.

They suggested that there is a fundamental difference between haptic communication and traditional movies. They observed that haptics is an interactive medium while movies are not. They considered the idea of an authoring system that permits embedding information, allowing users to pause and probe the environment, but concluded that users would rather be guided by the movie, preserving its original form. Their authoring tool allows the recording of haptic feedback of a scene, which included the position and orientation of the device (known as trajectory) sampled at 1 KHz as well as force feedback defined through duration, magnitude, direction and shape of the forces. Force feedback was included to approximate sharp impulses. The tool allowed authors to record trajectories at reduced movie speeds as well as incorporate a mechanism to render it on-screen on top of the video.

O'Modhrain and Oakley [**57**] demonstrated the use of haptics in broadcast media. They embedded force feedback and vibrotactile cues in two cartoons. They re-housed a video game joystick, which used a two degree of freedom force-feedback actuator, to resemble a remote control. They encountered similar issues of active and passive interactions in delivering haptic media, and proposed a solution called presentation interaction. Through the haptic actuator, authored cues were felt by the users. They were also allowed to interact with the actuators and change the viewpoint of the scene. This resulted in experiencing varied haptic cues appropriate for this viewpoint. Users

could also drag a character or an object on screen to make it move faster, while experiencing enhanced friction and drag in their hands while doing so. The authors suggested that a user's interaction or non-interaction with their system does not alter the predetermined sequence or the predetermined duration of the broadcast.

The use of haptics in films or other broadcast media is not restricted to post production. Woldecke et al. [**58**] explored the possibility of using vibrotactile communication during the production of a television show which involved virtual elements or animated objects. Their work was motivated by the lack of perceptive aids for the actors to support a natural interaction with virtual objects. Their work involved the use of a wireless vibrotactile belt with four vibration motors that were triggered via Bluetooth. They evaluated it in the context of instructing actors to walk through a space filled with virtual objects, and compared their performance with traditional aids (visual and verbal instructions). A total of six participants were involved in this experiment. With one vibration motor at each side, and one at both the midline and middle of the back, they simulated a pull that indicated the direction of movement by projecting the force vector to the two adjacent tactors that were involved in this simulation. The magnitude of the projection determined the intensity of vibration of a tactor. This is illustrated in Figure 13. They used four paths, randomized the usage of an aid (visual, verbal or vibrotactile) with all

three employed for each participant for each path, and computed the accuracy of each aid. They observed that though no significant difference was noticed across the three aids, vibrotactile instructions resulted in less squared positional errors (squared error/m^2).

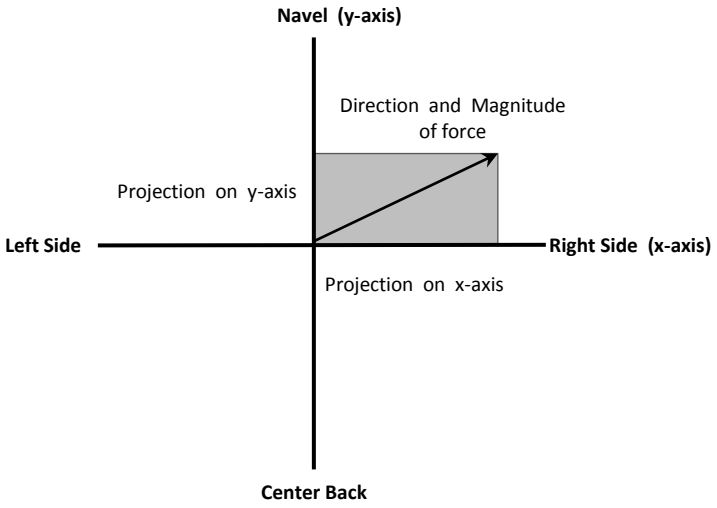

Figure 13. Determination of intensity of vibration of two adjacent motors. The figure shows the projection of the force vector to the two adjacent tactors (Navel, or front, and the Right side). The greater the magnitude of projection, the greater will be the intensity of vibration at that location (left, right, front, back). Image adapted from [**58**].

Cha, Oakley, Ho, Kim and Ryu [**59**] in their work suggested a mechanism for the transmission of haptic enabled audio-video content that involved both tactile as well as kinesthetic information. Taking note of two famous APIs available in virtual environment systems, Reachin and H3D, they observed that these APIs were insufficient for broadcasting temporal audiovisual content embedded with haptics. They also suggested that transmission of content in virtual environment systems through these APIs consist of a simple download

and play mechanism which may be time consuming and impractical for commercial broadcast media. They adapted the MPEG-4 BIFS format to carry haptic information along with the traditional audio-video content. Since haptic technologies are highly varied and more likely to be incompatible, they offered a framework that could work for a broad spectrum of devices. This framework categorized haptic media into two basic types: linear and non-linear haptic media. Linear haptic media corresponds to passive haptic playback where information is sequential and temporal, akin to the traditional audio and video content. Non-linear haptic media correspond to active haptic playback where the user probes the audio-video content in order to gain additional haptic information. They suggested that a tactile video, embedded with temporal vibrotactile cues, synched with the audiovisual content, and delivered to an array of vibration motors, falls under linear haptic media. Since BIFS does not have provisions for transmitting haptic content, tactile video is rendered as a gray scale video and is compressed using a video encoder. They also extended BIFS's shape node that is used to represent graphical 3D objects, and used to it to transmit non-linear haptic media which consisted of depth video, haptic surfaces and dynamic properties among many others; and were compressed either through an image encoder, video encoder or standard BIFS encoder according to the similarity of the information with traditional media. The final file is transmitted through a streaming

server. At the receiver end, they use the compositor process available in MPEG-4 to render the haptic cues.

Kim, Cha, Oakley and Ryu [**60**] designed a haptic glove and rendered tactile video delivered through the framework discussed in [**59**]. The glove consisted of four tactors attached to each digit of the inner glove. Through a haptic authoring tool, they manually embedded haptics in movie scenes consisting of three 30 second clips with each of the three different metaphors: first person, third person and background tactile effects, thereby producing 9 test clips. When shown to 10 participants, a majority of them favored the first person metaphor.

Lemmens et al. [**53**] developed a tactile jacket to enhance immersiveness in movies through vibrations that evokes or enhance emotions. Their suit consisted of 64 tactors distributed as arrays of four throughout the body: chest (2 arrays), stomach (4 arrays), arms (4 arrays total with 2 arrays for each arm, one on the front and one on the back), shoulders (2 arrays), neck (1 array), and back (3 arrays, with one for spine, and one each to its left and right). They restricted their study to seven emotions: love, enjoyment, fear, sadness, anger, anxiety, and happiness. They used one clip per movie, for a total of seven, with each clip targeting an emotion. They developed over 40 different vibration patterns, some suggesting events that take place on screen while others target emotions, derived based on "common

wisdoms and sayings". They conducted an experiment with this setup that involved 14 participants. Each participant viewed each clip twice, first without the vibrations, and second time with it; order was not counter-balanced. Through questionnaires and physiological sensors, they collected feedback from participants on each of these clips and observed the results to be promising.

Rahman, Alkhaldi, Cha and El Saddik [**61**] demonstrated adding haptics to YouTube videos. They used a vibrotactile suit and a vibrotactile arm band as the delivery medium for haptics. They provided a client web interface that can play YouTube videos on-screen, and render haptics through the suit and arm band. They also developed a web interface that can be used to author tactile content for YouTube videos. The authored content is an array of XML elements that describes intensity values of a set of tactors. The authored content acts as an annotation that is stored in an annotation server.

3.2 Haptics and Situation Awareness

Various scenarios where vibrotactile communication may be used in the cockpit, were suggested by van Veen and van Erp [**62**]. They provided four categories of information that can be delivered to pilots using a vibrotactile display:

- **Geometric information**: information about direction, air traffic borders, deviation from designated course, etc.

- **Warning Signals**: extending existing systems that involves visual and auditory mediums.
- **Coded Information**: certain graphical information such as speed and altitude can be conveyed through touch.
- **Communication**: certain covert communication among crew members such as information about directions of danger.

They also conducted an experiment on the effect of G forces on the perception of tactile information on the torso, and concluded that there is no substantial impairment at least up to 6G.

Rupert [**63**] presented the Tactical Situation Awareness System (TSAS) developed by the Naval Aerospace Medical Research Laboratory to help pilots become better aware of their spatial orientation through an array of vibration motors embedded in a torso garment.

Ferscha et al. [**64**] proposed a mechanism to amplify or readjust the perception of space, called space awareness, through vibrotactile communication. They developed a Peer-It architecture, which consisted of a Peer-It software framework and objects. Their framework is built around a descriptive model called Zone-of-Influence (ZoI), which is a circular boundary within which they can feel the presence of other objects. This model allows Peer-It objects to sense the presence of other Peer-It objects once an object crosses another's ZoI. Using this model, and a vibrotactile belt worn around the waist for

communicating information, their framework notifies the user through the belt about other objects, i.e., obstacles, through vibrotactile stimulation in a location that relates to the direction of the location of the object; vibration intensity is also utilized in that it is inversely proportional to the distance between the user and the object.

In another study, van Erp, van Veen, Jansen and Dobbins [**51**] demonstrated the usefulness of a vibrotactile belt as an alternative communication channel during waypoint navigation in demanding environments. Their belt was fitted with 8 vibration motors equidistantly placed around the waist – one at each side, one motor at the midline, one at the middle of the back, and one between each of the aforementioned pairs of motors. For conveying distance to reach a waypoint, they developed two schemes: monolithic and three phase scheme. In the monolithic distance scheme, the duration between two successive vibrations was dependent on either the absolute or the relative distance left to reach a waypoint. The three phase distance scheme consisted of vibrations being frequent in the beginning, less frequent in the middle, and more frequent, again, when a user is in close proximity of a waypoint, with the end of the first phase and close proximity distance defined in absolute or relative terms. They also defined a control distance scheme where the interval between two successive vibrations was fixed at two seconds. With the two schemes in absolute and relative mode, and the control distance scheme, they

developed five schemes for evaluation. They found that the distance schemes did not influence the walking speed of the participant. They suggested that this might be because either the distance information was not needed or the participants could not interpret them.

McDaniel et al. [**50**] developed a system, called Social Interaction Assistant, that enabled individuals who are visually impaired or blind to access non-verbal communication cues used during a social interaction with a sighted interaction partner. They suggested that providing such non-verbal cues would enhance the social skills of those individuals who are blind or visually impaired. The system consisted of two components: a pair of normal sun glasses embedded with a discrete camera that is connected to a computer vision system and a vibrotactile belt. The computer vision system enables identification of a person, with his or her name delivered as audio. The belt conveys the location of the person with respect to the individual who is blind or visually impaired, i.e., the user of the system, as well as the interpersonal distance between the interaction partners. Their belt was fitted with seven vibration motors spread equidistantly as a semi-circle around the user's waist with the first and last tactors placed at the sides, the fourth tactor placed at the midline, two tactors between left side and midline, and two more tactors between midline and the right side. They conveyed five distances by altering the duration of the vibration, with longer durations inversely

proportional to a person's distance. On conducting an experiment with the belt for location and distance recognition accuracies, they found the belt to be an effective mode of communication.

McDaniel et al. [**48**] [**49**] provided improved mechanisms for communicating interpersonal distance through touch using four prominent distances based on proxemics - intimate, personal, social and public - over those used in [**50**].

Krishna et al. [**65**] developed a vibrotactile glove for conveying another non-verbal communication cue, namely facial expressions, to individuals who are visually impaired or blind. Their glove consisted of 14 tactors, with three tactors placed along the dorsal side of each finger, except the thumb, which had only two tactors along the dorsal side. The tactors were placed such that they did not obstruct the bending of a finger's joints. The glove was used to convey seven basic facial expressions: happy, sad, surprise, neutral, angry, fear, and disgust. Each of these expressions was conveyed through a dedicated haptic expression icon. These icons either mimicked the shape of the mouth, similar to visual emoticons, or were designed to evoke a sense of the expression. On conducting an experiment with one visually impaired participant and several sighted but blind folded participants, they observed that the glove had the "potential for enriching social communication" for individuals who are visually impaired or blind.

Pielot, Krull and Boll [**66**] analyzed the use of a vibrotactile belt in video games. They engaged participants in a team based video game, where they were allowed to use the belt half of the time. The belt conveyed information about the direction and distance of other participants in the game. They observed a decrease in verbal communication among subjects and an increase in their ability to keep track of others in the game, while devoting less effort in doing so. They suggested that participants felt team play to have improved with the belt. They also observed that the subjects spread a lot further in the game when they wore the belt.

CHAPTER 4

METHODOLOGY

A major drawback in the field of audio descriptions in movies is the lack of time in delivering the pertinent visual cues through audio. This is largely due to the presence of dialogues and background scores. Also, the visual medium is faster and more efficient at conveying information compared to audio descriptions. Reading or listening, and understanding the sentence, "He sprints across the hallway", can be conveyed at a comparatively lesser amount of time in the visual medium, and with additional details such as what the character was wearing, whether he was short or tall, whether the hallway was carpeted, whether the floor was slippery, etc. Though certain information can be redundant and irrelevant depending on the context, there still remains a lack of time to convey other cues through audio. This can be observed in audio descriptions of many movies where the narration leaves out the description of certain visual cues and/or the describer narrates the event at a high pace, thereby overloading the auditory channel. Novels have an advantage in this regard as a) the reader can read it at his or her own pace, and b) the author can narrate to the greatest extent about the scene before/during/after the portrayal of events. Such disparity and yet the necessity to narrate events forces an audio describer to stunt and abridge the narration.

They may sometimes interpret the events on screen as well, which is a major concern expressed by the visually impaired community.

It is therefore evident that there is a need to convey information in an alternative medium that can complement information delivered through audio. Such a medium may help with reducing the reliance on a single alternative channel for delivering visual cues. This work identifies haptics as a potential alternative medium.

4.1 Form Factor Selection

Haptics has so far been used in movies to enhance their immersiveness (see chapter 3). Vibrotactile devices of various form factors – suits, wrist bands and gloves – have been used to, e.g., allow a user to experience a punch as it might have been received by an on-screen actor on his or her body. This loosely relates to the notion of first person style narrative in audio description suggested in the literature. This work uses the notion of third person style narrative, which has been observed by individuals who are visually impaired as reliable, trustworthy and informative.

Literature on audio description (AD) suggests that, among many others, a character's location on the screen, his or her movements, interpersonal interactions with other on-screen actors and objects, and facial expressions and body language, are most frequently described. After observing the description of location and movement of characters in audio described movies, this work found that AD provides this

information with respect to other objects or characters on-screen, leaving the task of scene creation entirely as an act of imagination by the listener. Such a system has two drawbacks:

- Given the time constraints in a movie and an overloaded auditory medium, such a mechanism does not allow enough room for scene re-creation.
- It does not allow users to appreciate the director's perspective and portrayal of a scene.

Literature related to applications of vibrotactile communication reveals the usefulness of a haptic belt (also known as a vibration belt or vibrotactile belt), i.e., a belt fitted with vibration motors, for waypoint navigation, direction perception and situational awareness, as well as conveying interpersonal distances of other individuals with respect to an individual who is visually impaired or blind (see chapter 3.2). These, in essence, capture the notion of tapping a person's back for immediate attention, acknowledgement, and response from him or her.

As a first step towards enhancing movie comprehension for individuals who are visually impaired or blind, this work conveys the location of on-screen actors as well as their movements during the course of a scene through a haptic belt. Though this work could not find previous contributions on the usefulness of such a belt in portraying movements of other individuals, it can be derived as a

collection of location information delivered in quick succession and in a specific order that could closely match to the ones actually performed by those individuals. This work does not suggest the use of haptics as an alternative to audio description, but identifies it as a potential complementary and additive medium that can offset the over reliance of a single medium for information delivery to individuals who are visually impaired or blind. Since the creation of editing software for embedding haptics, as well as broadcasting haptics information over a network, has already been explored in the literature, this work addresses the question of how information can be designed and delivered in coherence with the audio from the original movie as well as the descriptions through the chosen form factor of a haptic belt.

4.2 Information Delivery Design

Since a vibrotactile belt consists of a set of vibration motors, the first step in the design is to determine the total number of motors needed for this application. Cholewiak et al. [**40**] found that the use of anatomical reference points, such as the midline of the waist (which has greater spatial acuity compared to, e.g., the sides of the waist), can improve vibrotactile localization when exploited. Moreover, endpoints (which may be considered artificial reference points), that is, those vibration motors in an array that have only one neighboring motor, can improve localization accuracy as well. And as would be expected, they also observed that six vibration motors around the

waist resulted in a higher localization accuracy compared to a greater number of motors in similar configuration (12 motors was the upper bound in their study). They also suggested that in addition to the anatomical points, localization accuracy around these points improves. McDaniel et al. [**50**] [**48**] [**49**] in their study had only explored the usefulness of the frontal semi-circle of the waist for conveying interpersonal distances. Since a movie scene primarily conveys information that occurs in front of the camera, this work also uses this particular configuration of a frontal semi-circle of vibration motors. Also, since a person usually watches a movie while being seated, wearing the belt around his or her waist might result in vibrations spreading to, and propagating along, the legs. In order to avoid unwanted vibration conductance, users may wear the belt slightly below the navel, but above the waist line. The placement of the vibration motors around the frontal semi-circle of the waist was influenced by three factors:

- Exploiting the usage of anatomical reference points and endpoints for information delivery.
- Equal vibrotactile resolution for both the left and right halves of the screen, i.e., equal number of vibration motors placed on the left and right halves of the waist such that they are symmetric.
- Most movie scenes consist of a minimum of two actors.

In an effort to increase resolution, this work used areas near the midline, it's left and right, rather than the midline itself. Such a placement is hypothesized to enhance accuracy in two locations rather than one middle location. Also, in order to eliminate false interpretation of actors appearing from a person's left or right side, and to provide a more planar, rather than a curved, view of the screen, two motors, one each on the left and right end, were placed just slightly ahead of each side of the torso. In an effort to further expand the resolution, provide more opportunities for delivering finer positional information, and be aligned with previous contributions on the belt and their results, one vibration motor was placed between the left side of the torso and the midline, and another vibration motor was placed between the right side of the torso and the midline. Such a configuration, thus, arranges six motors on the frontal semi-circle of the waist. The movie screen is, therefore, divided to six equal columns along the width of the screen such that the six motors map one-to-one to each of these columns. This is illustrated in Figure 14. This work also named each column. Since there are three columns on the left and three more on the right, the former three columns start with the letter L for left, while the latter three columns start with R for right. Since, the L and R columns are symmetrical along a user's midline, this work numbered the columns at the center for both L and R as 1, while the sides were numbered 3. Hence the column names, when

provided as a sequence from the left side to the right, are: L3, L2, L1, R1, R2 and R3.

The second step involves the design of the vibrotactile stimulation. A scene provides information about how close or far an actor is with respect to the camera, providing a user an opportunity to perceive how the scene was shot and how actors move on the screen, as well as how they interact with other actors and objects present in the scene. An important question that was encountered related to the level of depth perception that should be enriched through haptics. Vibration patterns indicating a sense of distance to reach a waypoint did neither accelerate nor decelerate a person's walking speed [**51**], and hence, was inferred as not being useful. On the contrary, McDaniel et al. [**50**] [**48**] [**49**] found that providing interpersonal distances was both desired and showed highly promising results. Since this work closely relates to the concept of conveying such distances, but not as being egocentric as it is in their context, efficient information delivery necessitates building vibrotactile cues that are inspired by their results.

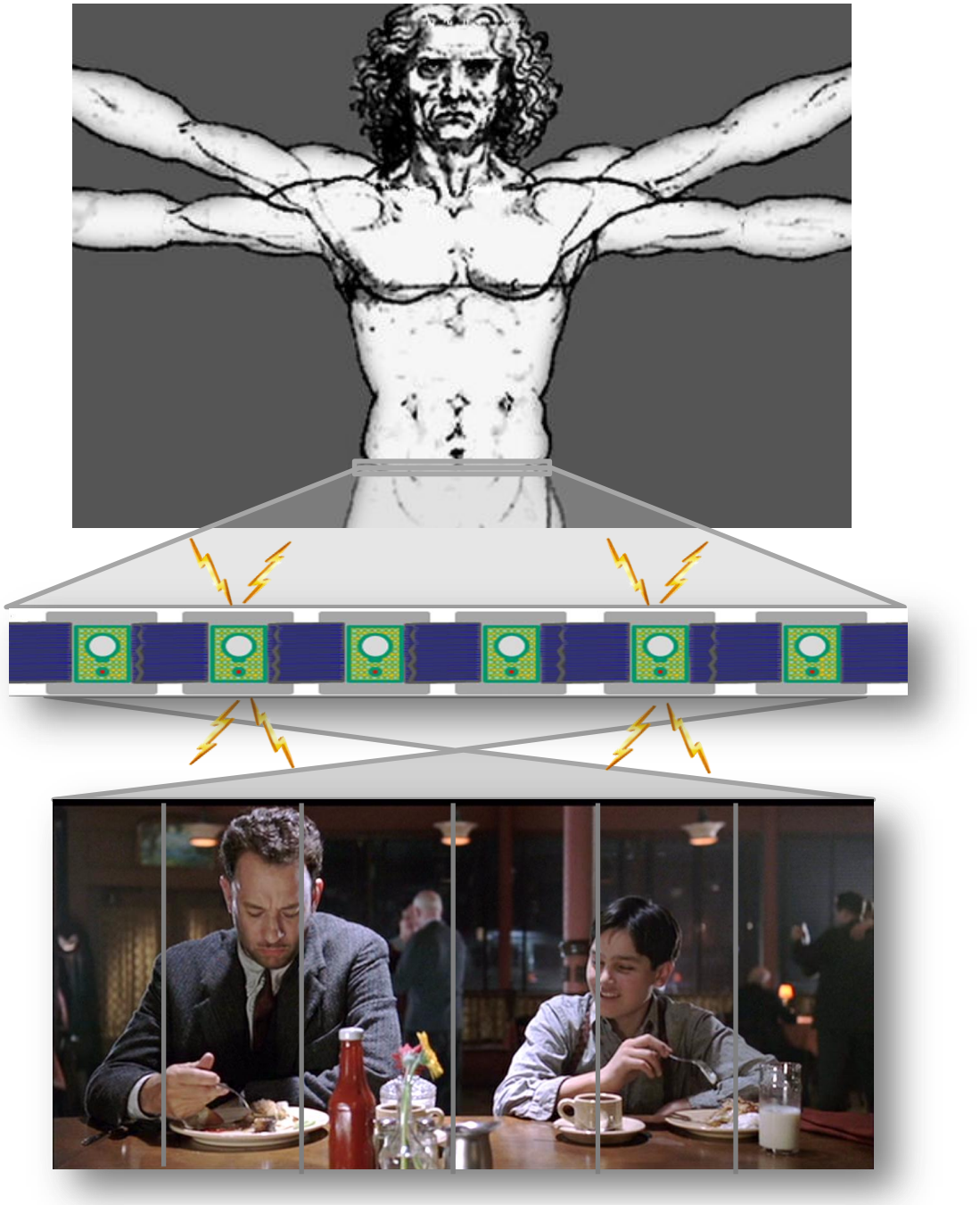

Figure 14. Placement of the vibration motors on the waist (top) and how it relates to a movie screen (bottom). The motors are housed using a plastic case and fitted to a belt (middle) such that they can move around and be placed at locations on the waist as specified in this work. The figure also shows the vibrations delivered for the actors on-screen.

Egocentric vibrotactile cues relate to their perception and interpretation with respect to the person himself or herself. In the context of the social interaction assistant [**50**], this is related to the perception of an interaction partner's interpersonal distance, as conveyed by the vibrotactile cue. McDaniel et al. [**50**] [**48**] [**49**] explored two vibrotactile designs for providing this information to individuals who are visually impaired or blind: a standard cue that varies in the duration for which it is presented, and different rhythms, each conveying a specific proxemics distance, that were presented for 10 seconds. Though participants were observed in their study to have performed better through the use of rhythms, this work found that their rhythms were unsuitable in the context of movies because:

- The presentation of the rhythms was for 10 seconds. This is unacceptably long for movies.
- These rhythms themselves were of variable duration. This was interpreted as a hindrance to deliver slow and fast movements of characters.

This work retains the concept of using rhythms for delivering distance cues. On observing the clips from several audio described movies, a duration of one second seemed reasonable for each vibrotactile cue; it was hypothesized that this time constraint will enable the haptic content to keep in synch with the audio described content. Considering the overload of the existing auditory medium in AD movies, this work

developed three distinct rhythms: one each for near, middle, and far distance from the camera. These rhythms resemble information delivery of threats in a radar system. A far away (but in visible range) threat is represented as faint beeps that have a greater pause between successive beeps. As the threat approaches the host system, this duration decreases until a sharp long beep, indicating that the threat is close. Figure 15 illustrates the proposed cue design based on this concept. Thus, the complete cue consists of two components – location and distance - that could be used to indicate the position of a character on screen to a fair level of detail.

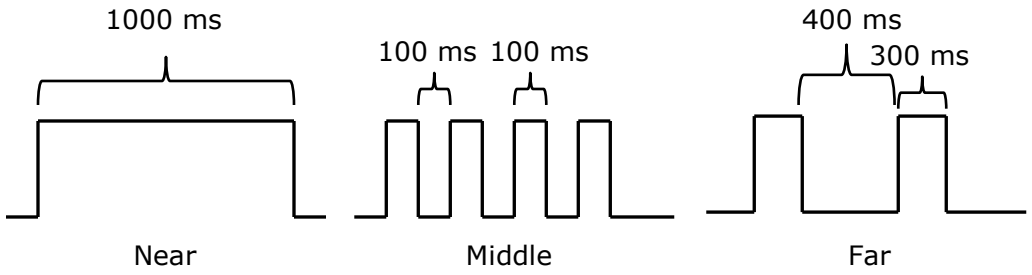

Figure 15. The three rhythms used in this work for near, middle, and far distances. Through pilot testing, all three rhythms were verified as being distinct and intuitive.

The third step involves extrapolating the positional vibrotactile cues to indicate movement. This may be interpreted as a series of positional cues portrayed through the belt in quick succession (see Figure 16). The limit on how fast a movement the vibrations can convey depends on two factors:

- The number of vibration motors involved.
- The usage of the same motor in succession.

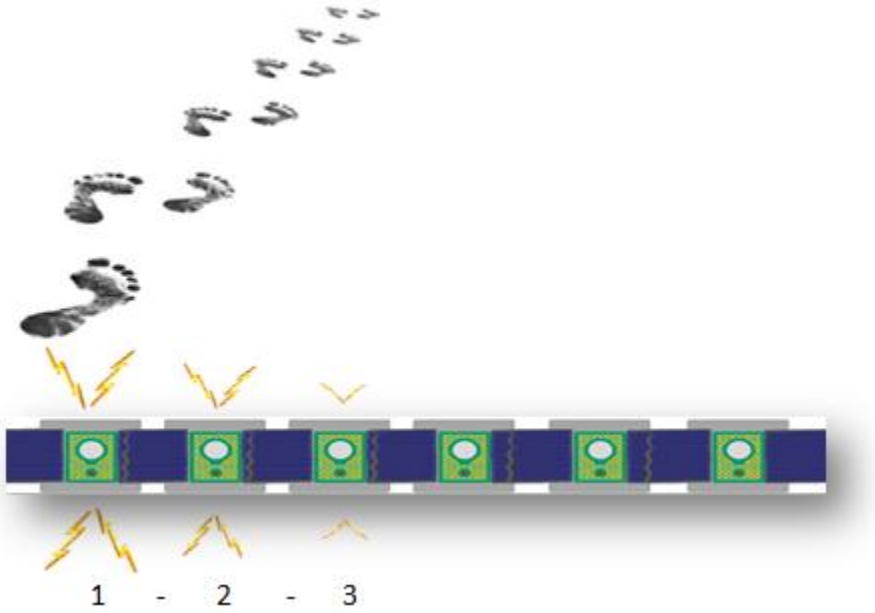

Figure 16. Movement of a person walking far away from the camera delivered as a sequence of near, middle, and far distance cues in the order 1, 2 and 3, where the numbers correspond to tactor locations on the haptic belt.

As an example, to convey a fast movement that involves three different motors, with not a single motor repeated in succession, would take a minimum of 3 seconds, since each rhythm presentation takes one second. On the other hand, if all the three positional cues had to be conveyed through the same motor or if different motors are involved but two successive positional cues have to be conveyed through a single motor, a minimum pause of 100 ms was observed in this study, though a formal study to explore this limit was not done. Thus, conveying a movement with three positional cues in this case

would involve a minimum time of 3.3 seconds. Movements need not necessarily involve successive motors, i.e., a movement can skip motors, e.g., a far jump, which would involve skipping 2 or more motors; these types of movements were not explored as part of this study and will be explored in future work.

Unlike proxemics, the study of interpersonal distances, the distance cues delivered through the belt are not based on rigid distance measurements. The cues presented are relative to how characters are positioned in front of the camera. Hence, for two characters, where one character, recognized as being at middle distance to the camera, may just be standing to the back of another character, recognized as being near the camera, even though the distance in reality may be minimal. Hence, the rhythm delivered is dependent on:

- The number of characters present in the scene.
- The number of distance cues available in the system.
- The viewpoint of the camera relative to the actors.

Also, gaps between the individual haptic cues need to be observed to prevent users from interpreting two cues as one. A gap equivalent to one rhythm time, i.e., one second, is observed here. A formal study exploring the tradeoff between recognition accuracy of haptic cues and gap duration was not performed.

4.3 Achieving Audio-Haptic Descriptions

Since a typical scene in a movie is composed of several shots, one important design decision is how haptic cues are integrated with audio. On observing various movie scenes, audio descriptions were found to avoid narrating various shots of the same scene. Since haptic descriptions are complementary to those descriptions received through audio, it is recommended that haptic descriptions also disregard multiple shots of the same scene. In this regard, a single representative shot of a scene is used from which all conveyed information is relative to. This approach enables haptic descriptions to track on-screen actors as they are engaged in the scene through multiple shots. Occasionally, a visual shot is conceptually zoomed out or zoomed in, or panned at the beginning of the scene to accommodate all its actors; it is recommended that these establishing shots be used as the representative shot for which haptic cues describe (see Figure 17). Such a scene portrayed through haptics constitutes a *haptic scene*. The act of zooming out results in a granular display of haptic information, while zooming in produces a more finely illustrated haptic scene.

To allow users to comprehend the haptic descriptions in addition to the audio of a movie, including audio descriptions, it is pertinent to maintain coherence with the existing information source. This will also help with maintaining the relevance of the haptic cues with the movie

scene. Braun [**19**], in his theoretical discussion on audio description, he suggested that there are two types of coherence that needs to be preserved: local and global. Since this work deals with only individual scenes rather than a set of scenes from a movie, only local coherence was felt as relevant. This work borrowed two important concepts from audio descriptions:

- Placement of the descriptions when there is no significant dialogues or background scores.
- The narration itself is placed either just before an event or immediately after it.

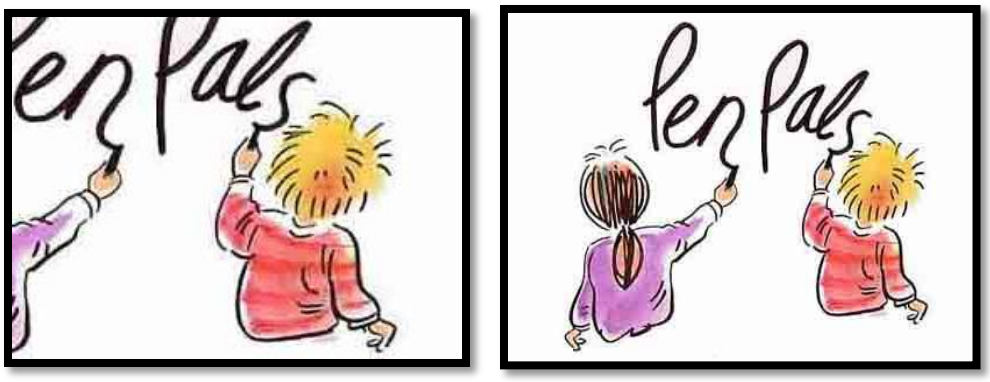

Figure 17. A representative shot of a scene (left) is zoomed out at the beginning of the scene (right) such that all actors are accomodated in the shot. This shot is then used to convey haptic information through the belt. Using such a shot as a reference and the haptic information conveyed for an entire scene constitute a haptic scene. Image adapted from [**67**].

Literature discusses utterances being shorter than insertions (see section 2.1), and that an insertion can consist of multiple utterances. The gaps between such utterances are useful moments for

providing positional information. As soon as an utterance recognizes an actor on the screen, or just before such an utterance, the positional cue for that actor is provided. Even if the describer fails to suggest the character who was about to talk, pauses in the conversation were chosen as moments for providing this information. For example, in the dialogue, "Merlynn, do you know…" the pause expressed by the actor after calling out the name of another character is an example of such a moment. Conveying movements through haptics may involve considerable overlap with either the original audio of the movie or the audio descriptions, but efforts should be made to minimize overlap. This may be done by either placing haptic information just before the actual movement is narrated, or conveyed via the visual medium. Long movements, however, may need to overlap in an effort to maintain the cues' relevance, and preserve coherence. Though a study was not done, the gap between haptic information and its corresponding audio cue was never allowed to become greater than 500ms.

CHAPTER 5

EXPERIMENT

Based on the proposed methodology (see chapter 4), an experiment was conducted to determine the efficacy of the proposed approach for augmenting audio described movies with positional information as conveyed through touch to visually impaired subjects. The experiment was conducted at both the Center for Cognitive Ubiquitous Computing and the Disability Resource Center at ASU. The aim of this experiment was to validate the idea of haptics as a sensory substitution medium in movies for individuals who are visually impaired or blind. Ten such individuals within the age group of 20-65 years participated in this study. There were an equal number of male and female candidates (five from each gender). Among these subjects, four of them were in the age group of 20-29 years, one was in the age group of 30-39 years, four were in the age group of 40-49 years, and one was in the age group of 60-69 years. Through their involvement in this study, each candidate spent two hours of their time, and as token of appreciation of their time and valuable feedback, they were awarded $25. This study was conducted after attaining approval from the Institutional Review Board at ASU.

5.1 Apparatus

This work used a haptic belt, clips from recent Hollywood movies, and software that enabled simultaneous video and haptic playback. In addition, participants also wore a pair of stereophonic headphones while listening to the audio described movie clips.

5.1.1 Haptic Belt

This work used a vibrotactile belt (see Figure 18) that was designed and developed by Edwards et al. [**68**] at the Center for Cognitive Ubiquitous Computing at ASU. It is a portable belt with Bluetooth capability that allows vibration motors to be actuated wirelessly. To fit any waist size, the belt is made with 1.5 inch flat nylon webbing, the length of which can be easily adjusted through a buckle. The tactors used in the belt are coin type shaftless vibration motors that are housed in a plastic case. Two status LEDs are present on each case that are primarily used for debugging. The belt is scalable and easy to re-configure, supporting a maximum of 128 tactors. In the present study, however, only six were used. Tactors may be moved and adjusted around the waist based on which locations need to be stimulated, and also to realign them for varying waist sizes. The tactors vibrate at a maximum frequency of 150 Hz. The belt is powered by a rechargeable 3.7 V lithium-ion battery.

Figure 18. The haptic belt that was used in this study [**68**]. The tactor modules (the black boxes) each contain a vibration motor, and are connected through an I^2C bus. The main controller (the white box) is also shown.

5.1.2 Movie clips

After consulting the list of movies that are available with audio description from [**69**], this work shortlisted 17 movies. The genre of the selected movies was predominantly drama, but action and comedy were also included. In order to curb any familiarity with a clip attained from previously watched clips within the experiment, only one clip from a movie was selected. This led to the creation of 17 clips for the experiment (see table 3). In order to not overburden a participant with long clips, each clip had an average duration of 2 minutes. All the clips selected for the experiment were conversational, and involved a maximum of three characters. The clips did not contain any action or sexual content, though mild language, if originally present in the clip, was not censored in order avoid disruption of continuity and comprehension of the clip. The chosen clips did not involve camera movements or complex presentation techniques such as those with split scenes shown side-by-side. Although care was taken to create

clips such that they began at the start of a movie scene, two clips, namely from *Iron Man 2* and *Public Enemies*, did not follow this rule. The video quality was degraded to that of a VCD as it served only as a status indicator to the experimenter. They were stored as standard WMV files, with the audio itself encoded in WMA format. These clips were produced using Microsoft Expression Encoder 3 software.

Table 3. The list of Hollywood titles that were selected for this work. In order to re-create the experiment, the start times as well as the duration of the clips are also provided.

Clip #	Movie Title	Start Time	Duration
1	Road to Perdition (2002)	01:15:47	1 min 10 sec
2	(500) Days of Summer (2006)	00:09:52	53 sec
3	The Ultimate Gift (2006)	00:24:04	1 min 7 sec
4	Cinderella Man (2006)	00:09:50	1 min 40 sec
5	The Bounty Hunter (2010)	00:12:58	1 min 22 sec
6	Munich (2005)	01:31:27	1 min 56 sec
7	Inside Man (2010)	00:52:29	1 min 16 sec
8	Iron Man 2 (2010)	00:36:10	2 min 12 sec
9	Public Enemies (2009)	02:08:54	2 min 24 sec
10	Evan Almighty (2007)	00:36:25	1 min 33 sec
11	Eat Pray Love (2010)	00:24:45	1 min 33 sec
12	Salt (Director's cut) (2010)	01:07:56	1 min 8 sec
13	The Karate Kid (2010)	01:11:51	2 min 11 sec
14	Wanted (2008)	00:23:17	1 min 42 sec
15	Blind Dating (2006)	00:12:02	1 min 54 sec
16	The incredible Hulk (2008)	00:25:21	1 min 46 sec
17	The Social Network (2010)	00:23:17	2 min 2 sec

5.1.3 Software

Two applications were used in this work. The first application (see Figure 19) was developed by Edwards et al. [**68**] as part of their haptic belt project. Through an interface within the application, specific rhythms, designed for this experiment (see chapter 4), were stored on the haptic belt. The same program can be used to vibrate a particular motor with a specific rhythm. This application also allowed the vibration strength to be of a specific magnitude and cycle. This work used a magnitude of 100% for all the rhythms, and the cycle selected was one. These settings provided non-repeatable vibrations at maximum intensity. A card shuffling algorithm, written in Java, was employed to randomize the type of rhythm, as well as its location around the waist.

Two spatio-temporal patterns were also used in this experiment and delivered using the same application (see Figure 19). One pattern indicated the involvement of two characters in a haptic scene, who initially were at the two sides of the movie screen (L3 and R3), approaching each other by moving to the middle of the scene (L1 and R1, respectively) one after the other. The actors were far away at the sides, and at a close distance to the camera when they were at L1 and R1. They both were at middle distance of the camera at positions L2 and R2, respectively. This, hence, provided the feel of a symmetric movement of the two characters. The second pattern involved just one

vibration motor, L1, which was used to indicate that a character was moving perpendicularly away from the camera. This movement was realized by sending the near, middle and far rhythms to L1 sequentially and in the same order. These two patterns served as examples of movements shown in a haptic scene.

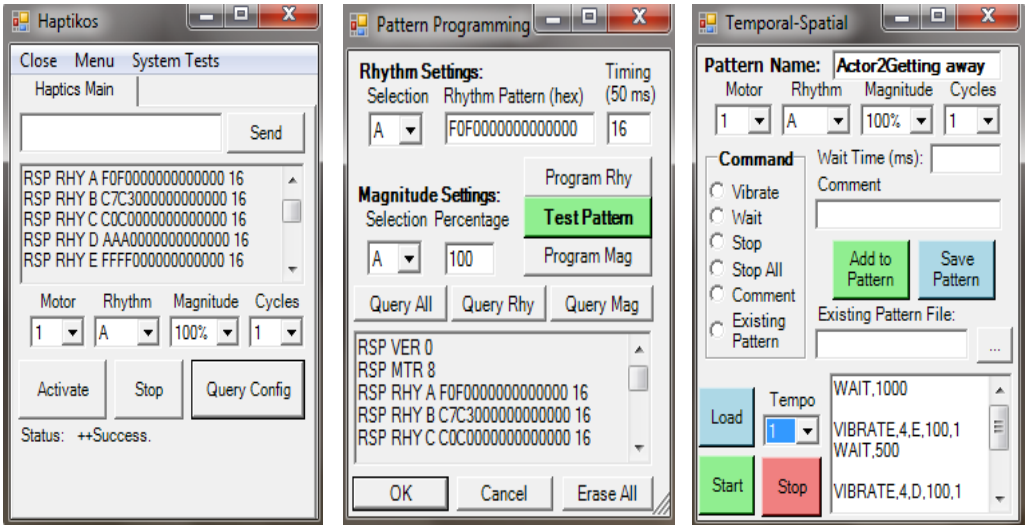

Figure 19. The Haptikos software that was developed by Edwards et al. [**68**]. Individual vibration motors could be actuated with a specific rhythm at a specified magnitude and cycles (left). The text area in the left image provides the current configuration of the belt which includes the coded rhythms. The rhythms are coded using an interface shown in the middle image. Hexadecimal values are assigned to rhythm holders (A, B, C, etc.). Spatial-temporal patterns for movements were delivered to the belt using the interface on the right. Predefined .pattern files were loaded with a tempo of one, where the tempo indicates the number of overlapping motors.

The second application was specifically developed for this experiment (see Figure 20). It was programmed in Visual C#, and used the Windows Media Player Library. It also used a DLL that was made available by Edwards et al. [**68**] to programmatically access the

haptic belt through a defined set of Application Programming Interfaces.

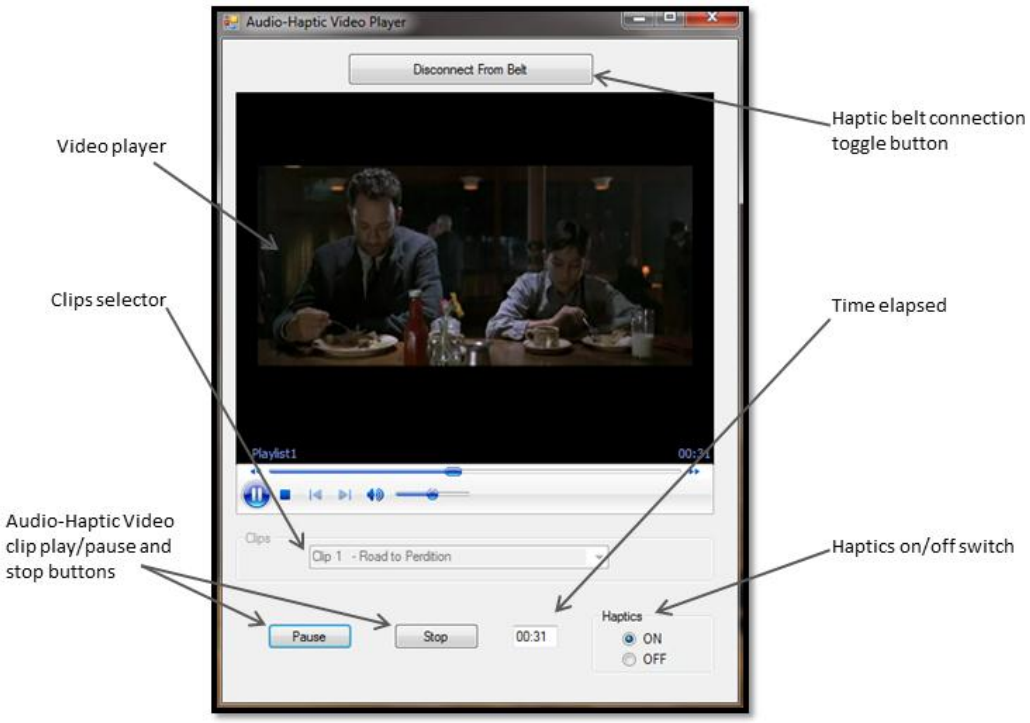

Figure 20. Audio-Haptic Video player that was specifically developed for this experiment. This was used during both the Audio-Only segment as well as the second part of the Audio-Haptic segment.

The application consisted of a Bluetooth module that allowed establishing a connection with the haptic belt. It used the media player API to play, pause and stop the clips. Through a drop down selector, a particular clip can be chosen. It also had an autosuggest feature to further ease the access of these clips. Manually authored haptic scenes were coded as modules. Each piece of haptic data for position, distance and movement, were mapped to specific elapsed times in the clip accessed through the media player API. The haptic data were sent

using a separate thread to the haptic belt, which rendered the vibrations. Lastly, the experimenter has the option to turn the haptic scene on or off through a pair of radio buttons.

5.2. Procedure

This experiment consisted of two segments – Audio-Only and Audio-Haptic. It compared the performance of the participants in traditional audio described movies, through clips played in the Audio-Only segment, with the proposed audio-haptic described movies, through clips played in the Audio-Haptic segment. Each participant was involved in both the conditions. To avoid any order effects, the sequence of these segments across participants was counter-balanced. Half the participants were first involved in the Audio-Haptic segment, while the other half was first involved in the Audio-Only segment.

5.2.1 Clips Selection

From the pool of 17 movie clips that were available, 12 were used for each participant. The selection of these 12 clips was determined through a Subject Information Form used to gather details of participants' movie watching frequency, exposure to audio-described movies, and which of the 17 movies they have seen. The latter information was collected by presenting a list of all 17 titles; if a participant had seen a movie, they were asked to suggest how well they remember the movie using a 5-point Likert scale, where 1 represented low remembrance, and 5, high remembrance. The clips

that a participant had not seen were selected. If this was insufficient for the experiment, clips were then selected from the already seen category, with the clip from the movie that was least remembered being selected first. This work considered providing new clips for assessment and, hence, in the event of insufficient new clips, the already seen clips were chosen for familiarization, thereby maximizing new ones for assessment. A clip counter was used for each clip in each segment, which provided a status of the number of times each clip was viewed across all participants. This measure allowed showing each clip almost the same number of times as other clips for each segment. At no time was the same clip shown in both the segments.

5.2.2 Audio-Only Segment

In the Audio-Only segment, participants were first introduced and familiarized with the audio described movies. This was achieved by playing one of the pre-selected clips, based on the subject information form, to the participant. The subject was then given an option to listen to the same clip again for a maximum of three times. This was followed by a testing phase involving a set of five additional audio described movie clips; again, each of which were pre-selected based on subject data. At the end of each of these clips, a standard question, "What happened in the clip?" or "Can you describe me the clip?" was asked by the experimenter. For this phase, the participant was suggested to provide the number of characters and the context of the

clip, i.e., the location of the scene, ambience, and the subject of conversation. Any inadequacy observed in their response was followed with clarification questions, as well as questions that would allow the participant to remember the scene better. The subjects were then asked if they could suggest the location and movement of the characters present on-screen either through the loudness of their voice and/or through its orientation with respect to the left or right speakers. The subjects were made aware of this process in the familiarization phase. The participants were then asked to respond to a SART-3 questionnaire (see section 5.2.4).

5.2.3 Audio-Haptic Segment

This segment involved use of the haptic belt. Since the participants needed to associate the vibrations with their intended meaning, they had to first go through a familiarization and training phase with the belt. Also, the distinctness of the haptic cues and the effectiveness of the proposed configuration had to be measured. The Audio-Haptic segment was, therefore, divided into two parts.

The first part consisted of familiarizing the different vibrotactile stimulation sites by actuating tactors at locations L3, L2, L1, R1, R2 and R3, presented in a sequence from the left end (L3) to the right end (R3). These positions were stimulated once more, if needed, after the first pass through. Each location was stimulated using a rhythm not present in any other part of this experiment. This was followed by

introducing the three distance rhythms, near, middle and far, which were presented at location L1 for simplicity. They were presented in either the sequence – near, middle, far – or in the sequence – far, middle, near. At this point, participants were instructed that distance rhythms were relative to the camera and did not carry a rigid distance measure in feet or inches. If needed, the distance cues were presented once more.

The training phase randomly presented 12 patterns where each tactor was actuated a total of two times, and each rhythm was presented a total of four times. Participants were asked to identify and respond with the position of the vibration (L3 through R3), and the rhythm used as part of the delivered vibration. The experimenter provided feedback, based on participants' responses, to correct wrong guesses pertaining to the dimensions of the vibration, or to confirm correct guesses. To pass training and move onto the testing phase, participants were required to achieve at least 80% accuracy within each dimension (at least 10 out of 12 locations and rhythms guessed correctly). The training phase could only be repeated once to ensure the duration of the experimenter was under two hours.

During the testing phase, each location was presented four times and each distance was presented eight times for a total of 24 randomly presented patterns. Participants' responses were recorded, and no feedback was provided by the experimenter during this phase.

Once the testing phase was completed, participants filled out a subjective questionnaire.

The second part of this segment began by familiarizing participants with the audio described movie clips with haptics through a single, pre-selected clip, and was followed by the use of five more pre-selected clips, based on the subject information form, for testing. Since the first part of this phase trained users to recognize positional information through the cues, and not movements, two sample movement-only haptic scenes (see section 5.1.3) were displayed through the belt. Each such haptic scene was presented for a maximum of three times. This preceded the familiarization with the movie clip. During the testing phase, participants listened to each of the five movie clips with its associated haptic scene delivered through the belt. At the end of each clip, the question, "What happened in the clip?" or "Can you describe me the clip?" was asked. Participants were suggested to provide the number of characters present in the clip, their location and movement through the course of the scene, and the context of the clip, i.e., the location of the scene, the ambience, and the subject of conversation. Any inadequacy observed in their response was followed with clarification questions, as well as questions that might allow participants to better recall the scene. Again, participants were made aware of this process in the familiarization phase. In order to determine the correctness of each position as

suggested by the participant, the haptic scene was drawn for each clip as a sequence based on time (see Figure 21), which was used as a guide. After each clip, participants were asked to complete a SART-3 questionnaire (see section 5.2.4). Once the segment concluded, participants were asked to fill out a questionnaire.

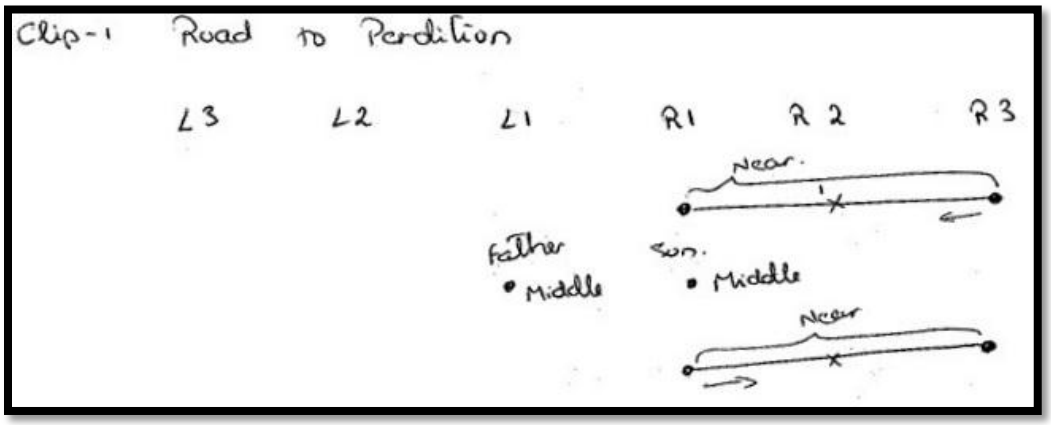

Figure 21. An example of a sequence diagram for the haptic scene of a clip from the movie, Road to Perdition. The individual columns indicate the locations of vibrations on a candidate's waist. Each row indicates the sequence of haptic information delivered through the belt with the first row delivered first. Movements are indicated by a line with the end and intermediate haptic cues marked. The direction of the movement is indicated through an arrow. Standalone positional cues are also represented in the diagram. The distance rhythms with which the individual cues are encoded are also marked.

At the end of the experiment, participants were allowed to make comments and provide suggestions on the approach. Participants were not allowed to watch the video of the clip (which in reality was applicable only to low vision candidates) in any of the segments, and were not allowed to make notes; participants were made aware of this

design during the consent process. During the experiment, participants were given short breaks between segments.

5.2.4 SART-3 Questionnaire

The Situation Awareness Rating Technique, or SART, was developed by Taylor [**70**] in 1990. This technique was originally used to assess a pilot's situation awareness. It is a subjective rating technique that uses a 7-point Likert scale, with 1 meaning low and 7 meaning high. Such a rating is sought on the following ten dimensions: (1) Instability of a situation, (2) Complexity of a situation, (3) Variability of a situation, (4) Spare mental capacity, (5) Arousal, (6) Concentration, (7) Familiarity of the situation, (8) Information quantity, (9) Information quality, and (10) Division of attention. Given that the proposed experiment requires an assessment of situation awareness for each clip, a condensed form of SART, called SART-3, was administered. This quicker version of SART determines a participant's (1) attentional demand, (2) attentional supply, and (3) understanding. In this work, attentional demand was measured through the question, "How complex did you feel the clip to be?"; attentional supply was measured through the question, "How much concentration did you employ on understanding this clip?"; and the understanding of the clip was measured through the question, "Rate your understanding of this clip?". Following [**66**], the ratings were sought on a 5-point Likert scale, with 1 meaning low and 5 meaning high. SART was chosen in

this work because of its applicability in a wide range of domains and its ease of use [**71**].

5.3 Results

5.3.1 Belt Configuration & Rhythm Design

Localization: For the first part of the Audio-Haptic segment (see section 5.2.3), participants achieved a mean localization accuracy of 91.25% (SD: 19.43%)—see Figure 22. The location L1 had the highest localization performance with a mean recognition accuracy of 100% (SD: 0%). Participants localized the other tactors, R1, L3, R3, L2 and R2, with a mean recognition accuracy of 92.5% (SD: 12.07%), 92.5% (SD: 16.87%), 92.5% (SD: 23.72%), 90% (SD: 24.15%) and 80% (SD: 25.82%), respectively.

Rhythm Design: For the same segment, participants achieved a mean distance recognition accuracy of 91.25% (SD: 14.37%)—see Figure 23. Far rhythm had the highest recognition accuracy at 95% (SD: 8.7%). Participants recognized near rhythm with a mean accuracy of 95% (12.08%), and middle rhythm with a mean recognition accuracy of 83.75% (SD: 18.68%).

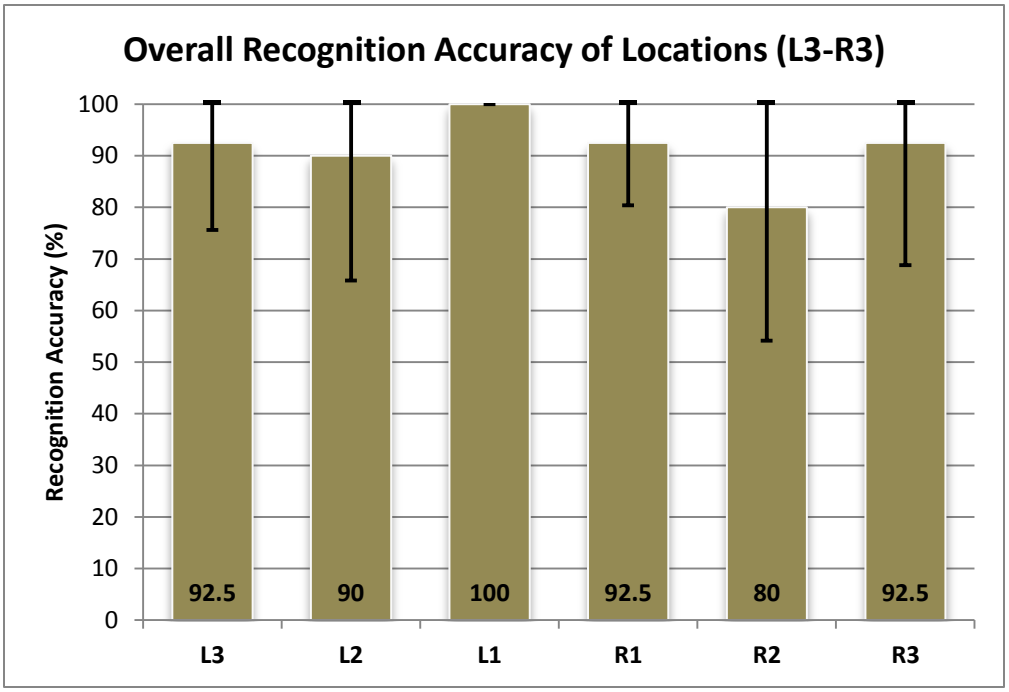

Figure 22. The localization accuracy observed for the configuration proposed in this work. The error bars indicate the standard deviation on the accuracy attained at each of the locations.

5.3.2: Audio-Haptic Versus Audio-Only

Location and Distance: When the haptic cues were presented along with audio described movie clips, recognition accuracy in two conditions were observed: (1) when participants associated the vibrations to the correct actor for whom they were presented for (known as *with association* or WA), and (2) when such an association was not considered (known as *without association* or WoA). In the WoA and WA conditions, participants achieved a mean localization accuracy of 71.73% (SD: 9.23%) and 66.73% (SD: 13.61%), respectively; and a mean distance recognition accuracy of 66.75% (SD: 9.81%) and 61.75% (SD: 13.62%), respectively. These are

illustrated in Figure 24. In the Audio-Only condition, subjects were not able to provide such information except for a few rare occasions. As audio is encoded differently in a movie compared to haptics, the perception of 'left' or 'right' in the audio of a clip (i.e., location) could actually refer to near or far (distance). Given this observation, a direct comparison was not performed for location and distance between Audio-Haptic and Audio-Only conditions.

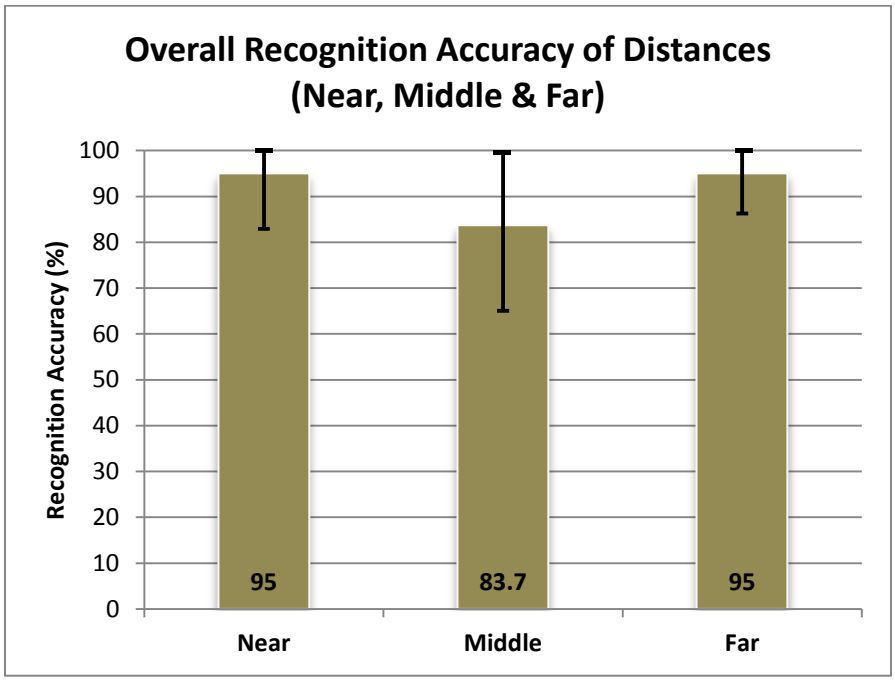

Figure 23. The recognition accuracy for each of the rhythms used as part of this work. The error bars indicate the standard deviation for each rhythm accuracy.

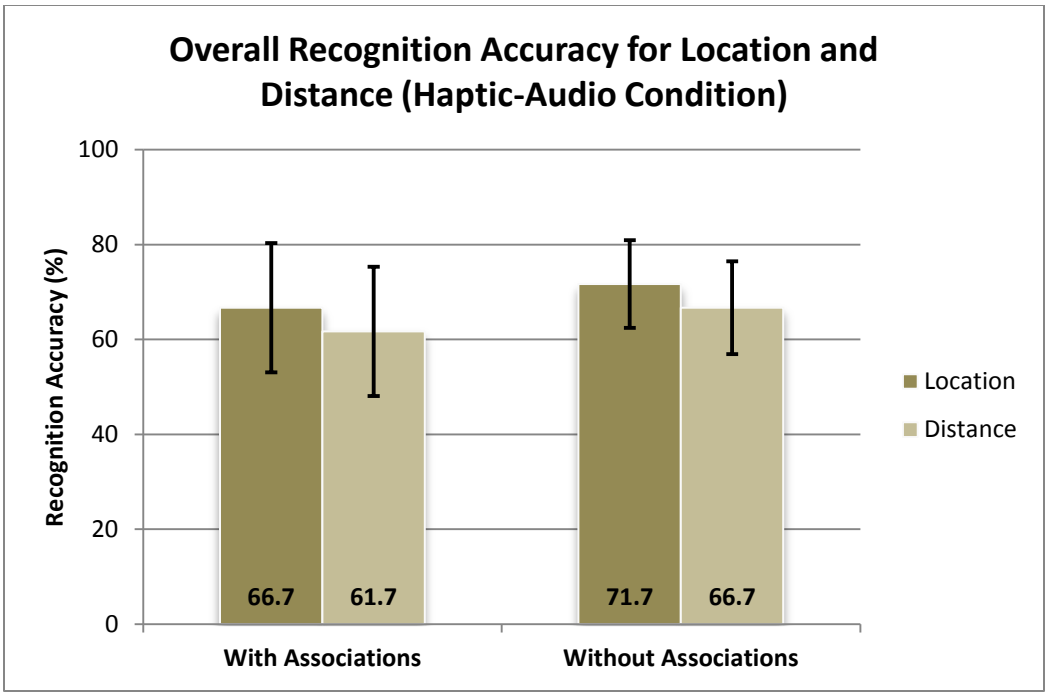

Figure 24. The overall recognition accuracy, with associations and without associations, of location and distance as presented through haptics during the movie clips.

Movement: In the Audio-Only condition, subjects were able to suggest the movement of actors through either the original audio of the movie clips such as footsteps, or through the audio descriptions. In the WoA and WA conditions, participants achieved a mean recognition accuracy of 49.98% (SD: 17.30%) and 48.69% and (SD: 18.01%), respectively. In the Audio-Haptic condition, for WoA and WA conditions, participants achieved a mean recognition accuracy of 89.40% (SD: 5.02%) and 85.90% (SD: 10.54%), respectively. These results are summarized in Figure 25.

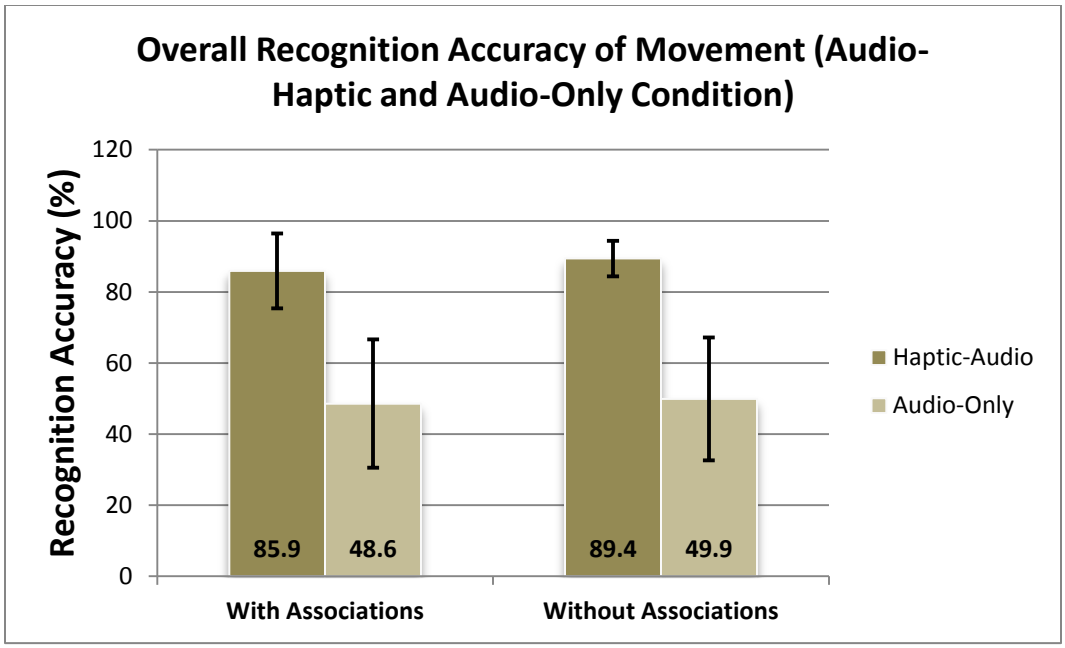

Figure 25. Comparison of the mean recognition accuracy, with associations and without associations, for movement of actors in both the Audio-Haptic and the Audio-Only conditions.

Understanding, Concentration and Complexity: In the SART–3, participants were asked to rate the understanding they attained from each clip, the concentration they had to devote (supply) in order to understand each clip, as well as the complexity of each clip that was perceived (demand) using a 5-point Likert scale with 1 being low and 5 being high. Also, the experimenter rated the understanding of each participant for each clip that was played using the same scale. In the Audio-Only condition, results revealed a mean participant-rated understanding of 3.92 (SD: 0.61), a mean experimenter-rated understanding of 3.56 (SD: 0.34), a mean participant-rated concentration of 3.06 (SD: 0.90), and a mean participant-rated

complexity of 2.56 (SD: 0.99). In the Audio-Haptic condition, results revealed a mean participant-rated understanding of 3.48 (SD: 0.94), a mean experimenter-rated understanding of 3.24 (SD: 0.47), a mean participant-rated concentration of 3.78 (SD: 0.84) and a mean participant-rated complexity of 2.96 (SD: 0.74). These results are shown in Figure 26.

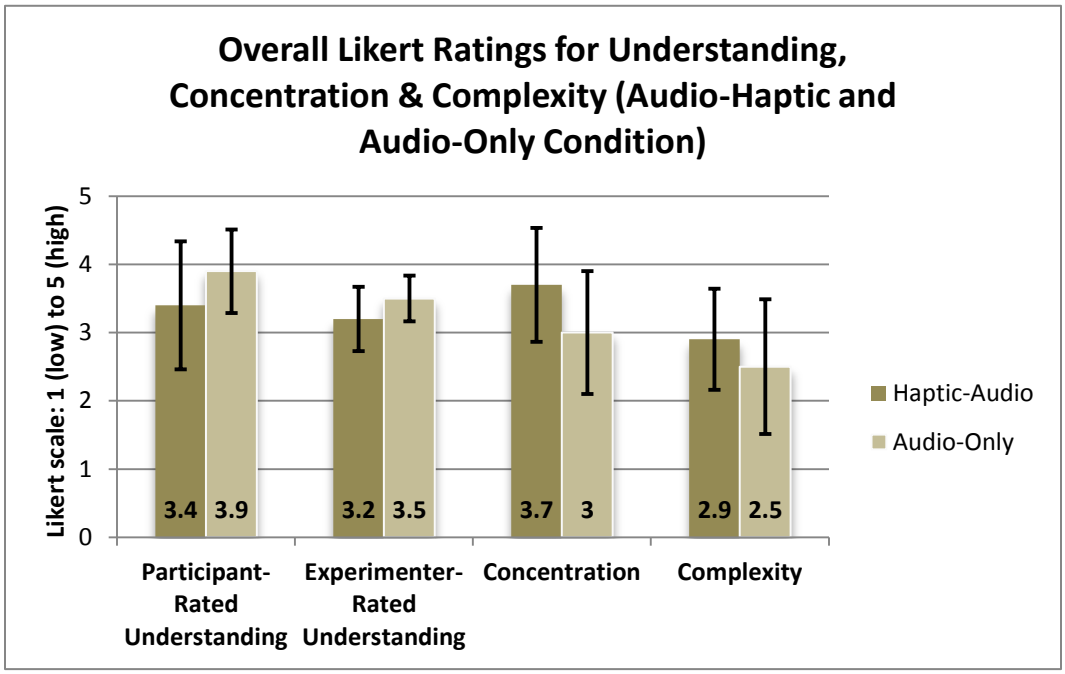

Figure 26. Mean participant rating of their understanding, concentration and complexity along with experimenter rated-understanding of the participants for each clip in both Audio-Haptic and the Audio-Only conditions.

5.3.3 Questionnaire

During the duration of the experiment, participants also completed a three-part questionnaire. They completed part one of the questionnaire after they were trained and tested on the haptic cues,

part two after the completion of the Audio-Haptic segment, and part three at the end of the experiment. The questions and the average scores for each of them are shown in table 4.

Table 4. The three-part questionnaire completed by participants with a score in the range 1 to 5.

Question	5-point Likert scale	Mean Score	SD
Part 1: After learning and testing the haptic cues			
How easy was it to learn the vibration patterns?	1 - very difficult 5 - very easy	3.7	1
How intuitive was the information about the location of a character presented?	1 - very difficult 5 - very easy	3.8	0.9
How intuitive was the information on the distance of a character presented?	1 - very difficult 5 - very easy	3.9	0.8
Part 2: End of video clips with Haptics			
How easy was it to wear the belt?	1 - very difficult 5 - very easy	4.2	0.9
How comfortable was the belt?	1 - very difficult 5 - very easy	4	0.9
When experiencing vibration(s) with the belt, how easy was it to associate them with an actor on screen?	1 - very difficult 5 - very easy	2.9	0.7

While listening to the movie clips, how easy was it to find the location of an actor across the breadth of the screen with the belt?	1 - very difficult 5 - very easy	3.4	0.9
While listening to the movie clips, how easy was it to find the distance of an actor from the screen with the belt?	1 - very difficult 5 - very easy	3.6	1.1
How easy was it to combine the information received through the vibrations with that of audio?	1 - very difficult 5 - very easy	2.8	0.9
How much were the vibrations obstructing your attention to audio?	1 - very little 5 – a lot	3.4	1
Part 3: End of experiment			
Do you think that the information presented through the belt added to the understanding of the clip? [Only those who answered 'yes' rated 1 to 5 on the Likert scale. Here, 8 out of 10 subjects answered 'yes']	1 - very little 5 – a lot	3.5	0.9

5.4 Discussion

5.4.1 Localization

Even though the location L1 seems to have been easiest to localize compared to other locations, and that R2 seems to have been least accurately localized, a one-way ANOVA between the overall localization accuracies of the six different locations revealed no significance differences [$F(5, 54)=1.12$, $p=0.3608$]. This suggests that no single location was more difficult to recognize compared to other tactors. Comparing these results with earlier work by McDaniel et al. [**48**] [**49**], whom used a similar semi-circular configuration but with seven tactors and arranged differently, their work found a higher overall localization accuracy of 95% with no significant differences observed. This disparity may be attributed to the following reasons:

- Cholewaik et al. have found that vibrotactile spatial acuity decreases with age [**41**] [**72**]. The average age of participants in this work was 37.7 years as compared to 32 years in [**48**] and 30 years in [**49**], suggesting that participants in this study were older. Indeed, most of the participants in the present study fell within two age groups: 20-29 and 40-49; compared to [**48**] and [**49**], which involved mostly subjects within the age group of 20-29 with the exception of several older subjects. Both of the aforementioned age groups in this study had only four participants, but within this sample, the localization

accuracy was much higher in the age group 20-29 (mean: 96.88% and SD: 3.99%) than in the age group 40-49 (mean: 83.33% and SD: 14.43%).

- The duration of vibration was one second in this work compared to 10 seconds in [**48**] and [**49**]. Therefore, the slight decrease in localization accuracy might be attributed to less time to localize the vibrations around the waist. In any case, given this large reduction in vibration duration, localization accuracy is very impressive.

Moreover, this study found that when misclassifications occurred during the localization of vibrations, these misclassifications were often off by just one tactor. This indicates that when a misclassification occurs, the perceived information, in the form of a location cue, is still useful to users as it conveys a rough estimate of the location.

When compared to the seven tactor semi-circular configuration of Cholewaik et al. [**40**], this work had a higher overall recognition accuracy. This was expected given that this work used fewer tactors, and Cholewaik et al. used a much shorter duration of 200 ms for each vibration as compared to the one second vibrations employed here. Although there are differences between the belt configurations employed in this work and Cholewiak et al.'s, comparisons are still useful; in their study, they found the anatomical reference point of the navel, and the endpoint tactors of the belt, which created artificial

reference points, to improve localization accuracy compared to configurations where these reference points were not available. Similarly, in our work, we found localization accuracy to improve, although not significantly, for the midline and endpoints. Using these reference points are also useful in that the localization accuracy of nearby locations increase as well.

5.4.2 Rhythm

A one-way ANOVA on the overall recognition accuracies of the three distance rhythms revealed no significant differences [$F(2, 27)=2.22$, $p=0.1285$]. This suggests that no single rhythm was harder to recognize compared to the other rhythms. On observing the accuracies, near and far distance rhythms were higher, 95% (SD: 12.07%) and 95% (SD: 8.74%), respectively, compared to middle distance rhythm, which was 83.75% (SD: 18.68%). Indeed, participants found the near rhythm to be distinct and easier to recognize in addition to being very natural in that the long constant vibration provided a sense of intimacy and closeness. This work observed that subjects had the tendency to classify any non-near rhythm as far, though this data was not recorded. This suggests that the middle and far distance rhythms could be further separated to make them more distinct. Participants found the far rhythm to be intuitive in that the short subtle pulses gave the impression of a person far away.

The rhythms designed in this work had a higher overall recognition accuracy to those observed in [**48**] (91.7%), but less accurate than [**49**] (94.3%). Moreover, both [**48**] and [**49**] had employed four rhythms as compared to just three used in this work. But, considering the fact that the rhythms were displayed for only one second in this study as compared to 10 seconds in [**48**] and [**49**], the accuracy achieved in this work is very impressive. Also, when considering the performance of the age group 20-29 and 40-49 in particular, which had four participants each, the age group 20-29 was far more accurate (mean: 97.91% and SD: 2.4%) than the age group 40-49 (mean 84.37% and SD: 15.73%) indicating an influence of age in accurate tactile rhythm recognition.

5.4.3 Audio-Haptic versus Audio Description

Participants perceived movements of on-screen actors more accurately when haptic cues were delivered in addition to the audio of audio-described movie clips. This observation was consistent in both WA and WoA conditions, i.e., associating movements to the correct on-screen actor and when such a requirement was relaxed, respectively. In the WA condition, the mean movement recognition accuracy for audio-haptic description (M: 85.9%, SD: 10.5%) was higher than the mean for the audio description (M: 48.6%, SD: 18%) providing a mean increase (M: 37.2%, SD: 16.3%) in recognition accuracy per participant. On performing a paired *t*-test on individual participant

accuracies, this work found this increase to be statistically significant, [$t(9)=7.2$, $p<0.01$, two-tailed]. In the WoA condition, the mean movement recognition accuracy for audio-haptic description (M: 89.4%, SD: 5%) was higher than the mean for audio description (M: 49.9%, SD: 17.2%) providing a mean increase (M: 39.4%, SD: 15.7%) in recognition accuracy per participant. A paired *t*-test performed on these accuracies suggested that they were statistically significant, [$t(9)=7.89$, $p<0.01$, two-tailed]. This suggests that audio descriptions along with the original audio from the movie clips were not capable of providing movement information to the level that audio-haptic descriptions portrayed this information to the participants. This can be attributed to the insufficient time available for describers to include this information. Moreover, since haptics is embedded in the same audio described clips, this result indicates its effectiveness and efficiency in portraying movements. Also, comparing the results of Audio-Haptic description between WoA and WA conditions, this study found a mean decrease (M=3.5%, SD=6.9%) in recognition accuracy per participant. This decrease was not significant, [$t(9)=1.59$, $p>0.05$, two-tailed]. This suggests that participants not only correctly guessed the movements portrayed though the audio-haptic descriptions, but they also were to correctly associate such movements to actors on screen without any difficulty. This further bolsters the earlier interpretation of the audio-haptic description being effective for

conveying movements. This work also observed in the audio-haptic descriptions that participants sometimes interpreted a single person's movement portrayed though the vibrations as movements for two characters (known in this study as *phantom movements*). Since the movement itself was perceived, and that the actor who had performed the movement was also mentioned to have performed the movement by the participant, such movements were assessed as correct by the experimenter. Such a perception suggests a requirement to refine the display of haptic scenes.

This work observed that participants were unable to provide even a vague idea as to where a character was located in a scene through audio descriptions. Exceptions to this observation include a couple of participants who were able to suggest for a few clips, like the one from Munich, vague location as provided through the loudness of a character's voice or its orientation through the left or right speaker. In contrast, participants, upon viewing the audio-haptic descriptions, had a mean localization accuracy of 71.73% (SD: 9.24%) and a mean distance recognition accuracy of 66.75% (SD: 9.8%) in the WoA condition. These accuracies slightly dropped in both location (Mean: 66.73%, SD: 13.6%) and distance (Mean: 61.75% SD: 13.62%). Even though the audio description does indicate location, they are always told in reference to another object in the scene and not with respect to the screen, making the user unable to suggest a location for the

actors. Though the accuracies for location and distance indicate considerable incorrect responses in both WA and WoA conditions, they are largely attributed to the high number of positional cues contained per clip (mean: 12.58, SD: 4.43, for both location and distance) as well as the positional cues for movements (different from movement as a concept that was discussed previously), therefore being more susceptible to incorrect interpretation. When localization accuracy was compared between the WoA and WA conditions in audio-haptic descriptions, a mean increase (M: 4.9%, SD: 7.5%) in recognition accuracy per participant was observed. This increase was not statistically significant, [$t(9)=2.09$, $p>0.05$, two-tailed]. This indicates that whenever a participant suggested a correct location, they also correctly associated it with its respective actor. Similar observation was drawn for distance information as well. Between WoA and WA conditions, a mean increase (M: 4.9%, SD: 7.5%) in recognition accuracy per participant was observed in the audio-haptic descriptions. This increase was not statistically significant, [$t(9)=2.09$, $p>0.05$, two-tailed].

5.4.4 Subjective Analysis

To analyze the collected data on the experimenter-rated understanding of each participant on a clip as well as the SART-3 questionnaires that recorded participant-rated understanding, his or her concentration exerted to understand a clip as well as participant-rated complexity of

each clip, a two tailed binomial sign test was employed. In this test, for each participant, an increase in their experimenter-rated understanding, and their self-rated understanding, concentration and complexity from audio-only to audio-haptic was recorded as positive while a decrease was recorded as negative. For experimenter-rated understanding, participant-rated understanding, participant-rated concentration and participant-rated complexity, the overall ratings of one, one, two, and one participant(s), respectively, were omitted from analysis as ratings did not change between audio and audio-haptic conditions. For participant-rated understanding, no significant difference was found between an increase or decrease in understanding between conditions, [S=2, p>0.05]. However, a decrease in understanding in the audio-haptic condition was observed to be approaching significance. For experimenter-rated understanding, a significant difference was found for a decrease in understanding, as opposed to an increase, from the audio-only to audio-haptic condition, [S=1, p<=0.05]. For concentration, no significant difference was found between an increase or decrease in understanding between conditions, [S=1, p>0.05]. However, an increase in concentration in the audio-haptic condition was observed to be approaching significance. For complexity, no significant difference was found between an increase or decrease in complexity between conditions,

[S=2, p>0.05]. However, an increase in complexity in the audio-haptic condition was observed to be approaching significance.

These results were expected given that haptics adds a new communication channel to audio-described movies. Although recognition accuracy of the location, distance and movement cues were impressive, there is still a learning curve involved especially when haptics is combined with another modality, namely audio. This learning curve was expected to increase concentration and complexity, as shown above. The added concentration and complexity of haptics may be, at times, distracting and divide attention. This is shown to be the case from the results of both participant-rated and experimenter-rated understandings. As a result of distraction and a lack of attentional resources for audio, participants might have missed information delivered through the audio-description and/or verbal cues of clips: dialogue, music, sound effects and audio descriptions including information about the scene, characters, etc. However, we hypothesize that concentration and complexity may be reduced, and in turn, improve understanding, through two approaches: (1) additional training and use of haptics in audio-described movies through which users will become more acquainted with the novel communication channel, thereby helping to reduce concentration and complexity; and (2) refinement of haptic cues to provide selected visual content that is crucial for comprehension of a given clip without overloading users'

haptic channel with information that is irrelevant and/or redundant, potentially creating distractions. The understanding of clips was recorded from both the participants' perspective and the experimenter's perspective, so that a comparative analysis could be performed. This relates to the different rationalities concept that is suggested in literature on situation awareness. Though this goes against SA literature by allowing the experimenter to evaluate the understanding of the participants, such a system allows this work to determine how much information, as conveyed participants, was comprehended in the intended manner.

In addition to above, participants filled a questionnaire at the end of first and second parts of the audio-haptic segment, as well as at the end of the experiment. For part one of the questionnaire, which pertains to the learnability and testing of the haptic cues prior to viewing the clips, results were satisfactory in terms of how easy it was to learn the cues, and the intuitiveness of the cues. Overall, participants found the cues for location and distance intuitive, and felt that they represented an actor's location across the screen, and his or her distance from the screen, respectively, reasonably well. Results might have been improved here through a more thorough training session involving more presentations of each cue. In any case, both the objective and subjective results are impressive given the short training time.

For part two of the questionnaire, which consisted of questions asked after all clips were viewed with haptics, it was found that:

- The usability of the belt was high in terms of how easy it was to wear the belt and how comfortable the belt was.
- Participants felt that they had difficultly associating actors with their respective haptic cues; this result conforms to the sign test employed on the SART-3 questionnaire. It is because of this difficulty that participants paid more attention to correctly comprehending the haptic cues, resulting in a division of attention and increased perceived complexity of the clips.
- The participants felt that when they were provided a vibration through the belt, it was satisfactorily easy for them to suggest the location and distance of an actor. This conforms to just the marginal decline observed when comparing the accuracies for location and distance without and with associations.
- They felt that it was a slightly difficult to combine the information received through the belt with that from audio, and that the haptic information considerably obstructed their attention to audio; this result again conforms to the sign test performed on the SART-3 questionnaire.

In part three of the questionnaire, which consisted of questions that related to the entire system and their suggestions and feedback, a promising eight of the ten participants felt that the proposed approach

added to their understanding of the clip. Out of the two participants who felt the system did not add to their understanding, one subject was constantly distracted by phone calls either made or received during the experiment, while the other participant was the oldest of the ten subjects (beyond 60 years). All participants, though, felt that they were improving as they advanced through the audio-haptic segment. Again, eight out of ten participants suggested that if such a system were available to use with audio described movies, they would like to use it. The participant, who was constantly distracted by the subject's personal phone, suggested that the additional information conveyed through the belt distracted his attention to audio, thereby, not preferring to use it. The other participant, who had some useful vision, suggested that although the belt provided a lot of additional information, his vision was sufficient to watch the video when the monitor is up close and hence, would not use the system. Two participants, who used to often watch movies, suggested that the system helped them to understand more about the scene and aided in providing a perspective of where characters were in the scene. One other participant found that it was difficult to watch the clip with the proposed system, but suggested that it would be a useful addition for subsequent viewings of the same clip. Two participants also raised concerns with the current audio system in movies that sometimes localizes actors on the left, while other times to the right, which is

confusing and that this system can eliminate that confusion. Again, another participant suggested that the system would be useful if the participant is familiar with the movie. Some participants, though, felt that it was difficult to "multitask" between comprehending the information received through the vibrations and the ones received though audio. Some participants felt that the system would be easier to use only when one actor was present in the scene. Overall, all participants liked the idea and felt that more work needed to be done on the system.

CHAPTER 6

CONCLUSION AND FUTURE WORK

This work proposed a sensory substitution method for augmenting audio described movies with positional information conveyed through touch to individuals who are blind or visually impaired. The proposed methodology was evaluated through a formal, IRB-approved user study with ten visually impaired or blind participants. Results showed the design to (1) enable participants to accurately localize vibrations around their waist, and interpret them as the location of an actor on the screen; (2) enable participants to accurately recognize tactile rhythms, and interpret them as the distance of an actor relative to the camera and/or other actors in the scene; (3) combine location and distance cues to accurately perceive movements of actors as they travel across the screen; and (4) combine haptic and audio cues to comprehend scenes in terms of both content and context. In the realm of audio-described movies, this work found that the inclusion of a new information channel of haptics enabled individuals with visual impairments to perceive a visual perspective of the scene. Specifically, it enabled users to accurately perceive the location and movement of actors in a scene in the 3D space captured by the camera. Overall, participants were pleased with the proposed approach, and found the idea useful and interesting.

As part of future work, there are several possible directions of research to pursue:

- Design guidelines will be derived and evaluated for optimal integration of haptic cues with the audio descriptions and audio content of films such that user concentration and distraction is reduced. It is hypothesized that a reduction in concentration and distraction will improve user understanding above audio-only interpretation.
- Optimal time gap between an audio cue and its corresponding haptic cue needs to be assessed. In this work, the time gap was not constant, but was limited to a maximum of 500 ms.
- Longitudinal studies will be conducted to learn about the effects of extended training on haptic cue perception, integration with audio, and scene understanding.
- Other non-verbal cues will be explored for the use in augmenting audio described movies. Possible non-verbal cues include facial expressions, body mannerisms and eye gaze. Form factors for haptic delivery of these cues will be explored, as well as useful designs. The proposed form factors and designs will be evaluated through a similar study as described here. User performance during the presentation of multiple non-verbal cues, in addition to audio descriptions and audio content

of a film, will also be assessed to learn about the bandwidth of touch as a communication channel.

- The proposed approach will be applied to other application domains, one of which will be radio plays.
- The concept of a single shot in a movie scene being used as a reference in the current work will not scale for scenes which cover a vast expanse of space. This work will be enhanced to allow the user to seamlessly relocate to a new visual segment of a scene.
- The concept of adding haptic descriptions in movies will not be limited to audio descriptions. As a part of future work, more interesting visual information will be determined and displayed though haptics.
- Future work will also explore if the existing semantics in movies can be used and exploited to provide information through haptics for individuals who are visually impaired or blind. Such a haptic description may eliminate the need of coupling haptics with audio description.
- The selection of the most significant shot for haptics in a movie scene will be automated through video processing techniques.
- Lastly, future work will explore and resolve overlapping actor movements, and accommodate more complex movie scenes that involve camera movements and split screens.

REFERENCES

[1] World Health Organization. (2011) WHO website. [Online]. http://www.who.int/mediacentre/factsheets/fs282/en/

[2] National Federation of the Blind. (2011) NFB website. [Online]. http://www.nfb.org/nfb/blindness_statistics.asp

[3] Joe Clark, "What is media access?," in *Building accessible websites*. United States of America: New Riders Press, 2002, ch. 4, pp. 37-42.

[4] Wikimedia Foundation, Inc. (2011, June) Wikipedia. [Online]. http://en.wikipedia.org/wiki/Braille

[5] Hong Z. Tan and Alex Pentland, "Tactual Displays for Sensory Substitution and Wearable Computers," in *Fundamentals of Wearable Computers and Augmented Reality*, Woodrow Barfield and Thomas Caudell, Eds. Mahwah, NJ, United States of America: Lawrence Erlbaum Associates, 2001, ch. 18, pp. 579-598.

[6] Wikimedia Foundation, Inc. (2011, June) Wikipedia. [Online]. http://en.wikipedia.org/wiki/Tadoma

[7] Jill Whitehead, "What is audio description," *International Congress Series*, vol. 1282, pp. 960-963, 2005.

[8] Bernd Benecke, "Audio-Description," *Translator's Journal*, vol. 49, pp. 78-80, 2004.

[9] Joel Snyder, "Audio description: The visual made verbal," *International Congress Series*, vol. 1282, pp. 935-939, 2005.

[10] Emilie Schmeidler and Corinne Kirchner, "Adding Audio Description: Does that make a Difference?," *Journal of the Visual Impairment & Blindness*, vol. 95, pp. 197-212, April 2001.

[11] J.P. Udo and Deborah I. Fels, "Re-fashioning Fashion: A Study of a Live Audio Described Fashion Show," *Universal Access in the Information Society*, 2009.

[12] Gregory Frazier and Ida Coutinho-Johnson, "The effectiveness of audio descrioption in providing access to educational AV media for blind and visually impaired students in high school," AudioVision,

San Francisco, 1995.

[13] E. Peli, E.M. Fine, and A.T. Labianca, "Evaluating Visual Information Provided by Audio Description," *Journal of Visual Impairment & Blindness*, vol. 90, no. 5, pp. 378-385, Sept-Oct 1996.

[14] Melanie Peskoe. (2009, May) Described and Captioned Media Program. [Online]. http://www.dcmp.org/caai/nadh237.pdf

[15] Philip J. Piety, "The Language system of Audio Description: An Investigation as a Discursive Process," *Journal of Visual Impairment & Blindness*, pp. 453-469, August 2004.

[16] Andrew Salway, "A Corpus-based Analysis of Audio Description," in *Media for All*.: Radopi, 2007.

[17] Wikimedia Foundation, Inc. (2011, May) Wikipedia. [Online]. http://en.wikipedia.org/wiki/Corpus_linguistics

[18] Andrew Salway, Andrew Vassiliou, and Khurshid Ahmad, "What happens in films?," in *IEEE International Conference on Multimedia and Expo*, Amsterdam, Netherlands, 2005.

[19] Sabine Braun, "Audiodescription from a discourse perspective: a socially relevant framework for research and training," *Linguistica Antverpiensia*, vol. NS6, pp. 357-369, 2007.

[20] Pilar Orero, "Three different receptions of the same film," *European Journal of English Studies*, vol. 12, no. 2, pp. 179-193, 2008.

[21] Daniel Chandler, *Semiotics: The Basics*, 2nd ed.: Routledge, 2007.

[22] Sabine Braun, "Audiodescription Research: State of the Art and Beyond," *Translation Studies in the New Millennium*, vol. 6, pp. 14-30, 2008.

[23] Andrew Salway and Alan Palmer, "Describing Actions and Thoughts," in *Advanced Seminar: Audiodescription – towards an interdisciplinary research agenda*, University of Surrey, 2007.

[24] Debora Tannen, "A Comparative Analysis of Oral Narrative Strategies: Athenian Greek and American English," in *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, Wallace L. Chafe, Ed.: ABLEX Publishing Corporation, 1980, pp. 51-87.

[25] Deborah I. Fels, John Patrick Udo, Jonas E. Diamond, and Jeremy I. Diamond, "A Comparison of Alternative Narrative Approaches to Video Description for Animated Comedy," *Jounal of Visual Impairment & Blindness*, vol. 100, no. 5, pp. 295-305, May 2006.

[26] John Patrick Udo and Deborah I. Fels, ""Suit the Action to the Word, the Word to the Action": An Unconventional Approach to Describing Shakespeare's Hamlet," *Journal of Visual Impairment and Blindness*, vol. 103, no. 3, pp. 178-183, March 2009.

[27] Mariana Julieta Lopez and Sandra Pauletto, "The Design of an Audio Film for the Visually Impaired," in *Proceedings of the 15th International Conference on Auditory Display*, Copenhagen, Denmark, 2009.

[28] Carmen Branje, Susan Marshall, Ashley Tyndall, and Deborah Fels, "LiveDescribe," in *Proceedings of the Twelfth Americas Conference on Information Systems*, Acapulco, Mexico, 2006.

[29] Andrew Salway, Mike Graham, Eleftheria Tomadaki, and Yan Xu, "Linking Video and Text via Representations of Narrative," in *Proceedings of American Association for Artificial Intelligence Spring Symposium on Intelligent Multimedia Knowledge Management*, Palo Alto, 2003.

[30] Paula Igareda and Alejandro Maiche, "Audio Description of Emotions in Films using Eye Tracking," in *Proceedings of the Symposium on Mental States, Emotions and their Embodiment*, Edinburgh, Scotland, 2009, pp. 20-23.

[31] James Lakritz and Andrew Salway, "The Semi-Automatic Generation of Audio Description from Screenplays," University of Surrey, Guildford, CS-06-05, 2006.

[32] Langis Gagnon et al., "Towards computer-vision software tools to increase production and accessibility of video description for people with vision loss," in *Univ Access Inf Soc*, vol. 8, 2009, pp. 199-218.

[33] Mica R. Endsley, "Towards a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, vol. 37, no. 1, pp. 32-64, 1995.

[34] Robert Rousseau, Sebastien Tremblay, and Richard Breton, "Defining and Modeling Situation Awareness: A Critical Review," in *A Cognitive Approach to Situation Awareness: Theory and Application*, Simon Banbury and Sebastien Tremblay, Eds.: Ashgate Publishing Company, 2004, ch. 1, pp. 3-21.

[35] John Patrick and Nic James, "A Task-Oriented Perspective of Situation Awareness," in *A Cognitive Approach to Situation Awareness: Theory and Application*, Simon Banbury and Sebastien Tremblay, Eds.: Ashgate Publishing Company, 2004, ch. 4, pp. 61-81.

[36] Sidney Dekker and Margareta Lutzhoft, "Correspondence, Cognition and Sensemaking: A Radical Empiricist View of Situation Awareness," in *A Cognitive Approach to Situation Awareness: Theory and Application*, Simon Banbury and Sebastien Tremblay, Eds.: Ashgate Publishing Company, 2004, ch. 2, pp. 22-41.

[37] Darryl G. Croft, Simon P. Banbury, Laurie T. Butler, and Dianne C. Berry, "The role of awareness in Situation Awareness," in *A Cognitive Approach to Situation Awareness:Theory and Application*, Simon Banbury and Sebastien Tremblay, Eds.: Ashgate Publishing Company, 2004, ch. 5, pp. 82-103.

[38] WikiMedia Foundations, Inc. (2010, November) Wikipedia. [Online]. http://en.wikipedia.org/wiki/Haptic_perception

[39] Eric R. Kandel, James H. Schwartz, and Thomas M. Jessell, *Principles of Neural Science*, 4th ed. New York, United States of America: McGraw-Hill, 2000.

[40] Roger W. Cholewiak, J. Christopher Brill, and Anja Schwab, "Vibrotactile localization on the abdomen: Effects of place and space," *Perception and Psychophysics*, vol. 66, no. 6, pp. 970-987, 2004.

[41] Roger W. Cholewiak and Amy A. Collins, "Vibrotactile localization on the arm: Effects of place, space, and age," *Perception & Psychophysics*, vol. 65, no. 7, pp. 1058-1077, 2003.

[42] Wikimedia Foundations, Inc. (2011, July) Wikipedia. [Online]. http://en.wikipedia.org/wiki/Tactile_corpuscle

[43] Lorna M. Brown, Stephen A. Brewster, and Helen C. Purchase, "A First Investigation into the Effectiveness of Tactons," in *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2005, pp. 167-176.

[44] Stephen Brewster and Lorna M. Brown, "Tactons: Structured Tactile Messages for Non-Visual Information Display," in *5th Australasian User Interface Conference*, Dunedin, 2004.

[45] Lorna M. Brown and Topi Kaaresoja, "Feel Who's Talking: Using Tactons for Mobile Phone Alerts," in *CHI '06 extended abstracts on Human factors in computing systems* , Montréal, 2006, pp. 604-609.

[46] Mario Enriquez, Karon MacLean, and Christian Chita, "Haptic Phonemes: Basic Building Blocks of Haptic Communication," in *International Conference on Multimodal Interaction*, Banff, Alberta, 2006, pp. 302-309.

[47] Frank A. Geldard, "Adventures in tactile literacy," *American Psychologist*, vol. 12, no. 3, pp. 115-124, March 1957.

[48] Troy McDaniel, Sreekar Krishna, Dirk Colbry, and Sethuraman Panchanathan, "Using Tactile Rhythm to Convey Interpersonal Distances to Individuals who are Blind," in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, Boston, MA, USA, 2009.

[49] Troy McDaniel, Daniel Villanueva, Sreekar Krishna, Dirk Colbry, and Sethuraman Panchanathan, "Heartbeats: A Methodology to Convey Interpersonal Distance through Touch," in *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, Atlanta, GA, USA, 2010.

[50] Troy McDaniel, Sreekar Krishna, Vineeth Balasubramanian, Dirk Colbry, and Sethuraman Panchanathan, "Using a Haptic Belt to Convey Non-Verbal Communication Cues during Social Interactions to Individuals who are Blind," in *IEEE International Workshop on Haptic Audio Visual Environments and their Applications*, Ottawa, Canada, 2008.

[51] Jan B. F. van Erp, Hendrik A. H. C. van Veen, Chris Jansen, and Trevor Dobbins, "Waypoint Navigation with a Vibrotactile Waist Belt," *ACM Transactions on Applied Perception*, vol. 2, no. 2, pp. 106-117, April 2005.

[52] Erin Piateski and Lynette Jones, "Vibrotactile pattern recognition on the arm and torso," in *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2005 , pp. 90 - 95.

[53] Floris Crompvoets Paul Lemmens, Dirk Brokken, Jack van den Eerenbeemd, and Gert-Jan de Vries, "A body–conforming tactile jacket to enrich movie viewing," in *Third Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Salt Lake City, UT, USA, 2009.

[54] Daniel Spelmezan, Mareike Jacobs, Anke Hilgers, and Jan Borchers, "Tactile motion instructions for physical activities," in *Proceedings of the 27th international conference on Human factors in computing systems*, Boston, MA, 2009, pp. 2243-2252.

[55] Frank A. Geldard and Carl E. Sherrick, "The Cutaneous "Rabbit": A Perceptual Illusion," *Science, New Series*, vol. 178, no. 4057, pp. 178-179, October 1972.

[56] Derek Gaw, Daniel Morris, and Kenneth Salisbury, "Haptically Annotated Movies: Reaching Out and Touching the Silver Screen," in *14th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2006, pp. 287-288.

[57] Sile O'Modhrain and Ian Oakley, "Adding Interactivity: Active Touch in Broadcast Media," in *Proceedings of the 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2004.

[58] Bjorn Woldecke, Tom Vierjahn, Matthias Flasko, Jens Herder, and Christian Geiger, "Steering Actors Through A Virtual Set Employing Vibro-Tactile Feedback," in *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction (TEI)*, Cambridge, UK, 2009.

[59] Jongeun Cha, Ian Oakley, Yo-Sung Ho, Yeongmi Kim, and Jeha Ryu, "A Framework for Haptic Broadcasting," *IEEE Multimedia*,

vol. 16, no. 3, pp. 477-491, July-September 2009.

[60] Yeongmi Kim, Jongeun Cha, Ian Oakley, and Jeha Ryu, "Exploring Tactile Movies: An Initial Tactile Glove Design and Concept Evaluation," *IEEE Multimedia*, vol. PP, no. 99, September 2009.

[61] Md. Abdur Rahman, Abdulmajeed Alkhaldi, Jongeun Cha, and Abdulmotaleb El Saddik, "Adding Haptic Feature to YouTube," in *ACM Multimedia*, Firenze, Italy, 2010.

[62] Henricus A.H.C. van Veen and Jan B. F. van Erp, "Tactile Information Presentation in the Cockpit," in *LNCS 2058*, S. Brewster and R. Murray-Smith, Eds.: Springer-Verlag Berlin Heidelberg, 2001, pp. 174-181.

[63] Angus H. Rupert, "An instrumentation solution for reducing spatial disorientation mishaps," *IEEE Engineering in Medicine and Biology* , vol. 19, pp. 71-80, 2000.

[64] Alois Ferscha et al., "Vibro-Tactile Space-Awareness," in *Adjunct Proceedings of the 10th International Conference on Ubiquitous Computing*, 2008.

[65] Sreekar Krishna, Shantanu Bala, Troy McDaniel, Stephen McGuire, and Sethuraman Panchanathan, "VibroGlove: An Assistive Technology Aid for Conveying Facial Expressions," in *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems* , Atlanta, GA, USA, 2010.

[66] Martin Pielot, Oliver Krull, and Susanne Boll, "Where is my Team? Supporting Situation Awareness with Tactile Displays," in *Proceedings of the 28th international conference on Human factors in computing systems* , Atlanta, GA, USA, 2010.

[67] Ratz. (2010, September) blogspot.com.

[68] Nathan Edwards et al., "A Pragmatic Approach to the Design and Implementation of a Vibrotactile Belt and its Applications," in *IEEE International Workshop on Haptic Audio visual Environments and Games*, Lecco, 2009, pp. 13-18.

[69] American Council of the Blind. (2011) The Audio Description Project. [Online]. http://www.acb.org/adp/dvds.html

[70] R. M. Taylor, "Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design," in *AGARD, Situational Awareness in Aerospace Operations*, 1990, pp. 23-53.

[71] Mica R. Endsley, Stephen J. Selcon, Thomas D. Hardiman, and Darryl G. Croft, "A Comparative Analysis of SAGAT and SART for Evaluations of Situation Awareness," in *the 42nd Annual Meeting of the Human Factors and Ergonomics Society*, Chicago, 1998.

[72] Roger W Cholewiak, Amy A Collins, and J. Christopher Brill, "Spatial Factors in Vibrotactile Pattern Perception," in *Eurohaptics 2001 Conference Proceedings*, 2001.