ABSTRACT
The visually impaired increasingly demand more informative audio description (AD), but not an easy coding scheme for the describers and a lack of technological advances have limited this success. Our proof-of-design concept – affective cinematography and emotive vibration, is quick to complete the AD process and deliver the emotional information of the scene through the affective haptic device. A promising way to contribute to social TV viewing experience for the visually impaired is also discussed.

Author Keywords
Affective haptic, affective cinematography, audio description, visually impaired, social entertainment

ACM Classification Keywords
H.1.2 User/Machine Systems; H.5.2 User Interfaces; K.4.2 Social Issues

INTRODUCTION
Of the many activities carried out at home, viewing TV is arguably the most common user experience (UX) that should be enjoyable and pleasing. Further, TV viewing experience seems to encourage any people to interact with their families and/or friends, either talking about TV programmes they have viewed, or retrieving conversational motives from them to explain their own attitudes or social actions [16]. However, these benefits are escaping from the visually impaired.

According to the American Foundation for the Blind (AFB)'s Household Survey, 97% of the visually-impaired people reported they are missing information “very often” or “fairly often” in the TV viewing experiences. In particular, the most frustrating experience is, for example, when there are no verbal cues such as listening to a chasing scene in which they hear lots of tire squeals, crashes, and gun shots, but no dialogues that explain the scene. Others have watched a one-hour television mystery, enjoying it for 58 minutes until, at the very last scene, it turns dramatically silent and reveals the culprit with only a visual manner. Due to these limitations, many visually impaired audiences prefer to watch TV or movies along with a sighted friend or relatives [14].

Auditory vision substitution: The Audio Description
Many attempts to auditory substitution have been introduced to enhance accessibility of the visual contents for the visually impaired [17]. One of the most well-known forms of auditory substitution technique is the visual-to-auditory substitution - Audio Description (AD) (a.k.a., descriptive video service – DVS). In the UK, for instance, all TV broadcasters should provide audio description service at least more than 4% of their broadcasting schedules [3]. In particular, the leading, in this stance, BBC outperforms by producing more than 20% AD contents of its programmes.

Although audio description is important as an accessibility tool for the visually impaired audience, some limitations are still unanswered. First of all, the production of AD is time-consuming and labour intensive. On average, it consumes 60 man-hours to describe a 2-hour feature films [10], and costs around £400 per hour of AD production. Second, most AD guidelines stress that it should deliver factual information of the scene other than a describer’s emotional interpretations (i.e., “Karen picks up her clipboard and walks in front of the shattered dressing-table mirror. [Pause] She picks up a blood-stained lace mat. [Pause] She touches the corner of the red-spattered sheet on the bed. [Pause] She walks alongside the bed. [Pause] From the floor, she picks up a broken, silver-framed photograph of the missing woman smiling broadly with her arms around a teenage boy.” [4]). Hence, they are much limited only to...
narrate actor-focused viewpoints, and little attention to affective information that arises from specific camerawork (e.g., ‘dolly-in’ or ‘dolly-out’ is a film-making grammar to convey the mood of the scene) has been made [5]. Affective information for the visually impaired seems to be important, in particular, in watching romantic drama and films [1]. However, many worries exist in that it also interferes audience’s TV viewing experiences. Indeed, the BBC AD guidelines highlighted that “Audio describers do not need to pay attention to the elements of film direction, i.e., camera angles and etc.” [10].

In this regard, how the audio describer can effectively present affective information is still open to question. Our study develops a semi-automatic audio description coding scheme based on ‘Affective Cinematography’ to reduce the workload of audio describer, and suggests a proof-of-concept ‘Heart-to-Feel’ technique using emotive vibration. By these we mean that the visually impaired audiences can have subtle affective information of the scene, by which they can enjoy social TV entertainment with their loved ones.

HOW TO REVEAL AFFECTIVE UNDERSTANDING OF TV VIEWING EXPERIENCE FOR VISUALLY IMPAIRED AUDIENCES

Affective Cinematography: An easy coding scheme for Audio Describer

In producing films, the directors tend to make particular camera angles and movements with a specific intent. Figure 1 shows a conventional cinematographic technique that includes a dynamic camera movement shot - [TWO SHOT → CAMERA MOVES CLOSER → FILLS ENTIRE SCREEN]. By this, for instance, Alfred Hitchcock successfully elicited the mood of the scene and/or emphasised the subsequent emotional changes of the actors or actresses.

A well-chosen set of elicited emotions is thus critical for a consistent and principled coding scheme. Igareda and Maiche [7]’s study hinted a potential coding method using facial-to-emotional mapping of the screen, however it is little use when there are more than one character in the scene.

Instead, our approach in this study is to employ affective cinematography in developing a new coding scheme. Table 1 shows our proposed coding scheme with regard to seven emotions (joy, surprise, disgust, fear, anger, sadness, tender) and its relevant cinematography.

![Figure 1. Storyboard with shooting script from "Lifeboat"-Hitchcock (1994): Uppercase indicates camera effects](image)

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Camera effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion</td>
<td>Framing</td>
</tr>
<tr>
<td>Joy</td>
<td>Whip-pan with a bouncy motion</td>
</tr>
<tr>
<td>Surprise</td>
<td>Vertigo, POV Object, Discovery</td>
</tr>
<tr>
<td>Disgust</td>
<td>Pan</td>
</tr>
<tr>
<td>Fear</td>
<td>Tilt, Inventory POV, Dark voyeur</td>
</tr>
<tr>
<td>Anger</td>
<td>Abrupt cuts, Tension to camera</td>
</tr>
<tr>
<td>Sadness</td>
<td>Character Dolly, Expand Dolly</td>
</tr>
<tr>
<td>Tender</td>
<td>Draw in, Long shot</td>
</tr>
</tbody>
</table>

Table 1. Coding scheme for affective cinematography

By reconnoitering previous literature, Ekman’s six basic emotions [6], which are thought to be universally identifiable across cultures, are firstly addressed. However, “happy” emotion is further detailed into two items – “joyful” and “tender”. “Joyful” represents both high in valence and arousal, which is often linked to the sense including comedic and merry-making. In comparison, “tender” tends to be a positive emotion but low arousal (heart-warming, sentimental, or relaxed). This is in line with the study of Wang and Cheong [18].

To reveal the seven emotions, we searched out all film production techniques and film grammar [5,18], finding that the three factors must be considered: Framing type (close-up shot, medium shot, and long shot), Camera angle, and Camera motion, as detailed in the three columns of Table 1.
Emotive vibration: Decoding the codes
Apart from the coding scheme for the broadcasters, how to present the visually impaired with the emotion-encoded description is key to success of our design proposal. A note of the decoding scheme of our concept, i.e., the ‘Heart-to-Feel’ technique, is thus needed.

A focus group study with several visually impaired people, we found the ‘4D’ cinema experience that gives them with haptic feedback is highly preferred, because it does promote more emotional feelings other than giving by any other modalities. Hence, we define the emotional vibration patterns of the seven emotions shown in Table 1, elaborating all relevant studies with empirical testing [12, 15]. Basic parameters of the vibration pattern of the seven emotions are summarised as ‘amplitude’, ‘duration’, ‘delay’ and the ‘number of times’. People well recognise changing the amplitude (voltage) and the rhythm (duration, delay and number of times) of vibration [2]. The seven emotion vibration pattern is shown in Figure 2, though it is still subject to scale up with a large-scale empirical test.

Tentatively, the TV broadcaster can insert relevant emotive vibration codes that are automatically converted from the affective codes given by the audio describer (see Table 1), and the ‘Heart-to-Feel’ device (a small hand-held device as shown in Figure 2) decodes the emotive vibration tag codes for the visually impaired to present certain emotional information.

TOWARD A SOCIAL USE OF TV CONTENTS: WHAT TECHNOLOGY CAN HELP THE DISABLED HAVE THE SAME SOCIAL USE?
The AD coding and decoding scheme briefly mentioned above might serve a way forward to a legitimate support for the visually impaired, but all TV programmes cannot be coded with it. In this respect, we believe that social altruism from friends and/or family members could cover another aspect of the TV viewing experience (i.e., social entertainment).

To our interest, past studies have observed that sports fans’ heart-beat rate seem to be synchronous when their supporting team is scoring [8]. A recent empirical study furthered this finding, putting heart rate monitors on 12 performers, 9 spectators who were relatives or friends of at least one performer, and 17 spectators with no social connections to the performers. Advanced statistical analyses revealed that the heart rates of relatives and friends followed similar patterns as those of the performers. No such effect existed in onlookers who did not know the performers, however. Research has also shown this ‘social entertainment’ is easily able to tune one’s emotions to the others.

In hindsight, Lull [11] defines the social TV viewing experience as crucial, emphasising that viewers employ their TV viewing experience to facilitate communication. Other studies have presented the TV viewing experience as an effective human-to-human communication tool. For instance, Kubey and Csikszentmihalyi’s study [9] revealed...
that TV viewers are talking with their family members at home and trying to share their emotions, for about 60% of their TV viewing time. In effect, one of the key aspects of the TV viewing experience is thus to be able to communicate and share TV viewing experiences between members of a social network. However, this “water cooler” chat is much limited to the visually impaired people [14].

As a future work of the affective cinematography (i.e., coding scheme) and the emotive vibration (i.e., decoding scheme) presented here, social interaction for the visually impaired seems to be a significant way forward that influences critical user experiences for them. The assumption underlying our proposal is that the ‘Heart-to-Feel’ technique can be made more usable by paying a careful attention to both the social use of TV and the current TV viewing experience for the visually impaired people and his or her friends and family members to have connected emotions (or social entertainment) with themselves. Of course, to confirm our technology, many other empirical studies that could be performed to increase our understanding of the behaviour of the visually impaired and their social connection are undoubtedly needed, and these are planned for the near future. It is to be hoped that, by applying the ‘Heart-to-Feel’ proof-of-concept to other conditions and contexts, researchers can provide the insights needed to make the visually impaired people of the new technology more useful and usable.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST) (No. 20110021398) and the Industrial Strategic Technology Development Program (10042694).

REFERENCES


