The 'hookability' of multimodal impact captions: A mixed-methods exploratory study of Japanese TV viewers

Ryoko Sasamoto, Stephen Doherty & Minako O'Hagan
Dublin City University, Ireland | University of New South Wales, Australia | University of Auckland, New Zealand

The use of captions has grown in recent years in both traditional and new media, particularly in terms of the diversity of style, content, and function. Impact captions have emerged as a popular form of captions for hearing viewers and contain rich multimodal information which is employed to capture viewer attention and enhance engagement, particularly in situations where there is competition for viewer attention. Drawing upon relevance theory, we argue how impact captions could effectively attract and hold visual attention owing to their balance between processing effort and contextual effects. This exploratory study employs a dual-task paradigm and uses authentic materials and viewing situations to further examine the ability of multimodal impact captions to attract and retain overt visual attention amongst a small sample of TV viewers. Our results provide novel insight into the apparent highly individualised efficacy of impact captions, where we identify several variables of interest in participants' viewing behaviours. We conclude with a discussion of the study's contributions, limitations, and an outline for future work.

1. Introduction

1.1. *Impact captions*

This paper seeks to understand the reception and the effect of intralingual open captions known as 'impact captions' (Park 2008) designed for hearing viewers. Originating in Japanese TV entertainment programmes in the 1980s (O'Hagan 2010), impact captions have evolved to become diverse in style, content, and function and have enjoyed widespread establishment across many genres of Japanese TV programmes (Shitara 2011; Koga 2013) and in other parts of Asia, e.g., China and Korea (Park 2008) enabled by ongoing development in language technology (see Doherty 2016). To date, scholars have examined the use of these captions as enhancing humour (O'Hagan 2010), capturing viewer attention (Shitara 2011), highlighting information (Sasamoto 2014), as well as their structure, terminology, and multimodality (Sasamoto 2014; Sasamoto, O'Hagan & Doherty 2017). Yet, there remain methodological and empirical gaps (e.g., see O'Hagan & Sasamoto; Orero et al 2018; Doherty 2020) in the literature that limit our inform the nature, use, and effects of impact captions, and, most significantly, into establishing an evidence-based best practice for their current ad hoc applications.

Distinct in both form and function from intralingual subtitles for the deaf and the hard-of-hearing (SDH), impact captions are multimodal and are designed to draw hearing viewers' attention to the screen visually and sometimes aurally with accompanying sound effects (Sasamoto, O'Hagan & Doherty 2017). They may display the verbatim rendering of speech content of TV programmes, similar to SDH, but they can be considered more as a subtype of multimodal text on screen. They have distinct features, as summarised as follows:

1. Contain verbatim, paraphrased, and editorialised content with an extra layer of information and meaning;
2. Can appear anywhere on the screen;
3. Are rich in multimodal sensory information;
4. Are used by producers to achieve a specific effect in viewer reception (e.g., humour, surprise, emphasis, incitement of bias, improved comprehension and narrative design);
5. Cannot be turned off, i.e., are 'open'.

Of particular relevance to the current paper is Koga (2013) who examined the composition and execution of impact captions on Japanese TV to show their specific role in ensuring the viewer receives the TV producer's intended message. Further to this, Sasamoto, O'Hagan & Doherty (2017) conducted a multimodal analysis on impact captions to illustrate how their application forms part of a larger media design in which they are expertly deployed to enhance and work in tandem with other messages and audio-video contents of the TV programme. O'Hagan & Sasamoto (2016) also examined the viewing behaviour of impact captions using the eye-tracking technology, where they found that viewers are indeed attracted to these captions even when they denied reading them. Their findings also show that the change in visual effects of impact captions (such as colour, change of font size, etc.) have impact on the viewing behaviour. Following these studies, we focus on the reception and effect of impact captions on viewers in terms of attracting and retaining viewer attention, or 'hookability' after Koga (2013). Furthermore, we wish to contribute to methodological discussions as our study involved a novel, albeit explorative, research design in eliciting relevant viewer behaviours akin to the holistic reception studies found in contemporary media production companies, e.g., MediaCityUK[1] .

1.2. *Theoretical framework*
Focusing on a particular intended effect of impact captions, we examine how they are used to attract the viewer's attention to particular elements of the TV programme and indeed to the TV screen itself (cf. O'Hagan 2010; Shitara 2011). While it has been reported that viewers attend, i.e. allocate cognitive effort, to impact captions once they appear on screen even when the content (related or unrelated to the programme content) and nature (static or dynamic) of the captions varies (Sasamoto 2014; O'Hagan & Sasamoto 2016; Sasamoto & O'Hagan 2020), it appears to be the case that viewers themselves can be unaware of reading subtitles and captions and processing their contents. Given that the human eye is physiologically attracted to movement (cf. Kennedy et al 2017), the very nature of impact captions appearing on screen in the form of striking colours, fonts, and styles, albeit for a limited time, makes it difficult, if not impossible, for viewers to ignore them in part or completely. A relevance-theory perspective can provide a theoretical account of such behaviours (Sperber & Wilson 1986; Shiota 2003; Sasamoto 2014, Sasamoto & Doherty 2016) and aligns with accounts from other approaches to such stimuli processing, particularly from cognitive psychology, as dicussed below (also see Doherty 2020).

Relevance theory has already been widely applied to audiovisual translation (AVT) research on subtitles to investigate translation strategies (e.g., Kovačič 1994), constraints (e.g., Bogucki 2004), viewer comprehension (e.g., Kovačič 1995), and viewer processing effort (e.g., see Gambier 2018). According to relevance theory, human cognition is designed in such a way that it pays attention to what is relevant to them. This Cognitive Principle of Relevance predicts that if someone tries to catch your attention by presenting an ostensive communicative stimulus, one could expect that whatever the person is trying to communicate would be worth their processing effort. This is called the Communicative Principle of Relevance (Sperber & Wilson 1995, 260). In this context, the prediction is that the viewer is to assume that any message being presented on the screen is worthy of being processed and is therefore worth the (cognitive) effort with the expectation that this expend of cognitive resources returns a reward, e.g., comprehension, additional meaning, humour, etc.

Cognitive processing, as described in relevance theory (Sperber & Wilson 1995; Wilson & Sperber 2002) is the effort of using cognitive resources to process a given input. This does not necessarily mean that there must be minimum processing effort in order to achieve relevance. According to Sperber & Wilson (1995), relevance is defined as a balance between processing effort and contextual effects. The less processing effort is required for interpretation, the more relevance the utterance is. In the context of the current study, this predicts that viewers have a trade-off between expending cognitive resources to attend to a message (e.g., an impact captions) and received a 'reward' for doing so whereby minimum effort is exerted for maximum gain, known as mini-max effect (Gibbs & Tendahl 2006; Kovačič 1994). Conversely, failure to return a gain for relative effort can lead to viewer dissatisfaction and other negative effects (Sasamoto 2014; O'Hagan & Sasamoto 2016).

---

[1] https://www.bbc.com/historyofthebbc/buildings/media-city/

Against this backdrop, we focus on the attraction of attention that can lead to the cognitive processing as described above. By measuring how viewers are attracted to, or not attracted to, the content of the TV programme, it is assumed that we can describe the role of impact captions in this attraction and retention process, given that these captions present to the viewer a static and dynamic stream of multimodal information, including verbatim utterances, summaries, sign posting within the programme (e.g., recapping of any previous utterance, previews of any future utterance), and explanations. Such information is both useful and interesting to viewers who may not be able to, nor want to, allocate their full attention to the TV programme, but may be performing other tasks, be in noise scenarios, etc. The next section presents the methodology adopted in light of the above aim, with findings and a wider summative discussion presented respectively in the following sections.

2. Methodology

2.1. *Quasi-experimental design*
As has been described in the previous review of literature, the current study is informed by several intersecting areas given their appropriateness: reception studies, AVT, relevance theory, and cognitive psychology. Given the novelty and interdisciplinary nature of the research objective and its field application, it is necessary to take a holistic approach by means of mixed-methods, especially in the context of a small-sample pilot study. In this light, we implement a post-task psychometric questionnaire alongside unobtrusive behavioural observations via high-definition video cameras – these data sources provide us with quantitative and qualitative data in a convergent parallel design. Paas et al (2003) argue that subjective measures such as questionnaires are most often used in cognitive load studies, yet such measures are more easily distorted (Schultheis & Jameson 2004; Kruger & Doherty, 2016), thereby calling for a more objective data collection method to be used in tandem. A number of studies in AVT have also adopted similar approaches (e.g., Taylor 2003; Pérez 2007; Baugarten 2008; Caffrey 2008; Caffrey 2009), which also highlight the study of fonts and background as important (e.g., Karamitroglou 1998; Mangiron 2013; Karamitroglou 2000, Díaz & Remael 2021) a finding also reached in many usability studies (e.g., Knoblauch et al 1991; Legge 2006). In line with this interdisciplinary approach, we adopt a mixed-methods exploratory design in which we pose the following research questions:

1. Do impact captions attract the viewer's visual attention to the TV programme being shown?
2. Do impact captions support the retention of the viewer's visual attention on the TV programme being shown?

This design allows us to weigh both data sources equally and call for the data derived from each source to be analysed independently in order to form a more holistic interpretation overall, thus fitting the exploratory nature of this initial research. We also highlight the increase in mixed-methods approaches in interdisciplinary research that has traditionally focussed on case studies or single-discipline driven theories (Alise & Teddlie 2010).

As stated, we aim to ascertain the ability of impact captions to attract and retain viewers' attention, whereby we seek to identify the attraction, or the *hook*, and its efficacy in terms of the duration of attentional retention. We do so in a dual-task paradigm (Wickens 1991), which we argue is a more naturalistic environment than watching TV in isolation on a computer screen. There we include several other tasks to occupy participants while the TV programme 'competes' for their attention and retention thereof. In addition, such dual-task interference (see Pashler 1994) may be apparent whereby performance in either task is inhibited due to the need to divide resources on both tasks concurrently. As such, we operationally define attraction as the ability of the TV programme to draw the viewers' focus of attention away from other tasks to result in a visible fixation of attention on the TV screen, where retention is the duration of time the focus of visual attention is held there. In other words, we measure the frequency of attraction of visual attention to the TV and its subsequent duration. As behavioural frequency measures are both direct and sensitive, we hypothesise that, if successful, the impact captions will be able to draw the participants' attention away from their other assigned tasks to the TV programme. We then focus on a finer-grained analysis by analysing facial expressions and other related behaviours as well as the effect of

social TV viewing on participants via presence of a peer. Given the intricate link between visual and auditory stimulation in terms of assignment of attention (e.g., Quittner et al 1994; Fritz et al 2007) and the role of emotion in attracting attention (Lang 1990), we also classify the nature of the multimodal stimuli to capture such data and its impact and latency effects on the participants.

2.2. *Procedure*

We recruited participants from a cohort of a Japanese academic exchange programme who had recently arrived from Japan. Ten participants were recruited in a voluntary capacity based on convenience sampling given the rarity of native Japanese speakers in the research location. The criteria for inclusion as participants were: (1) the participant was a native speaker of Japanese; (2) they were accustomed to watching TV in Japan, i.e. they were not condition-naïve, and therefore represented a realistic general viewer of the experiment material, and, arguably, were less likely to exhibit novelty-related processing as a result of the new form of stimuli presentation posed by the impact captions; (3) they were willing to give consent to participate, on a voluntary basis, in a research project involving video recording and questionnaires; and (5) they were willing to participate in the research project as a pair so as to include a social interaction component in the experiment. In the final dataset, there was a sample of 10 participants of which 8 were female and 2 male, all aged between 19 and 22 years (mean = 19.8, median = 20).

In order to create a naturalistic setting akin to the reception studies carried out in contemporary media production companies, we created a living room set-up with comfortable furniture, including sofas, a coffee table, and a modern flat screen TV. This would not only allow participants to watch the TV as they would at home, but also to put them at ease in front of the recording equipment. Participants were informed that this was an observational experiment and that their behaviours would be recorded on the two video cameras at rostral and lateral angles, although it was made clear that the recorded conversations would not be used as part of the study – see Figure 1. They were also informed that they would be left alone in the room to watch the TV programme lasting about 45 minutes including time to fill in the post-stimulus questionnaire. Participants were asked to complete several tasks while the TV was on in the background, the emphasis on the task performance was clearly stated. These tasks were designed to keep the participants attention, including visual attention, on the tasks themselves rather than the TV. This set-up therefore tests the ability of the TV programme, containing impact captions and sounds, to attract and retain participants' attention away from their instructed tasks. The tasks were:

(i) wrapping a gift;
(ii) peeling and eating any of the selection of fruit on the coffee table; and
(iii) completing a quiz about Ireland.

No time limits were placed on the participants, and they were informed they would be left alone to complete the tasks and that the researchers would return in 45 minutes. Prior to leaving the room, the researchers started the TV programme which included 2 minutes of generic Japanese TV advertisements to allow time for participants to settle and ready themselves for the tasks. The recording equipment was calibrated for each session and was discreetly placed so as not to make sounds or show lights, etc.
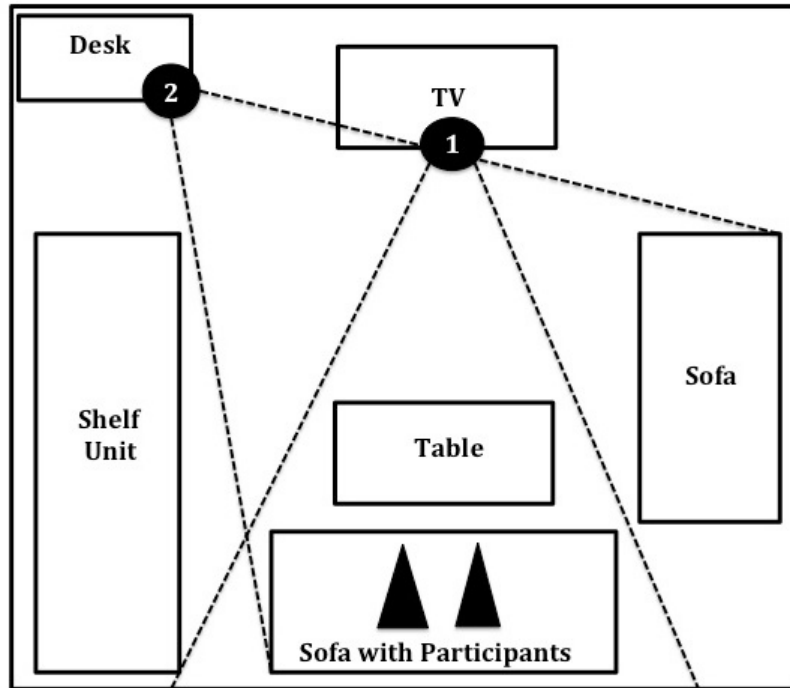
Figure 1. Experiment layout and camera locations (1 and 2)

Participants were shown the first half of a recorded episode of *Honmadekka* ('Is It Really True?' – Fuji TV) broadcast on the August 7th 2013 and 20 minutes in duration. It is a popular primetime variety show with a high weekly viewer rating since its launch in 2009 (Telebi Kiroku 2013). It uses a talk-show format with a veteran comedian as the host, who is joined by a popular female assistant, a panel of experts that change with the topic of the show, and a group of regular celebrity commentators. During the show, each expert panel member presents an expert opinion related to the theme of the week. The theme of the episode shown in our study was 'summer heat', which was embellished to develop an entertaining discussion. As indicative of the title of the programme, '*Is It Really True?'* in English, the unique characteristic of this programme is to surprise the viewers by revealing lesser-known scientific facts on a given topic. The didactic element focused on a particular theme is delivered in an informal and often humorous manner with the host together with the regular panel members constantly adding a comical touch to the otherwise serious topic. The clip used was fully transcribed for all multimodal input including utterances in order to gain an overview of the multiple semiotic resources involved (Kress & Van Leeuwen 2001) then subject to a fine-grained multimodal analysis based on Sasamoto, O'Hagan & Doherty (2017).

After viewing the TV programmes participants filled in a post-stimulus questionnaire which asked them several basic questions on demographics and several 5-point Likert scale items pertaining to typical TV viewing behaviours and preferences. In addition, this questionnaire contained a recall test of 5 items assessing to how much of an extent the participants had understood the programmes.

We collated data from the video recordings of both cameras, the post-exposure questionnaire, and the annotated multimodal corpus of the TV programme shown. This allowed us to examine the relationships between the stimulus shown and the participants' reactions or lack thereof. The video recordings were manually independently analysed by each of the three authors and areas of lower agreement were discussed and resolved in order to improve inter-rater reliability. Fixations on the TV and their duration were recorded on a per second basis for each participant and behavioural and facial reactions were also recorded using the Facial Action Coding System (Ekman et al 2002) with which the researchers were trained to code. Following data collation, three further researchers who were independent of the project verified the data on an individual basis.

Lastly, in terms of the multimodal information collated from the analysis of the programme and in keeping in line with the original multimodal transcription units used by Taylor (2004), we identified all elements that were both captions and caption related (e.g., sounds). We transcribed the

captions, including their attributes as meta-descriptions: colours, fonts, backgrounds, origin and type of utterance, related sounds, and the facial expressions and behaviours of the actors on screen. In this way, this fine-grained multimodal analysis used impact captions as the segmentation template for this bottom-up data collection for which NVivo (v10) was used in conjunction with SPSS (v20).

3. Results

The post-exposure questionnaire contained demographic items as well as items to ascertain the typical TV viewing habits and environments of participants. Six items then asked participants in detail about impact captions in general and then focussing on the TV programme they had just seen. These findings are outlined below, followed by the results from the video recordings as related to the multimodal analysis.

3.1. *Viewing behaviours*
Two participants lived alone, while the others lived with their family. Unsurprisingly, those living alone decided what they wanted to watch for themselves ($n = 2$) and the others who shared the family home had to negotiate somewhat: four shared this choice, two had their father choose, and two could decide on their own. All participants reported TV ownership with an average of 2.3 TV units in their household (median 1.5). The questions and the results are briefly reported below.

I. *How many hours of TV do you watch in an average week?*
On an average week, they reported watching between 11–19 hours of TV (average and median), with one participant reporting over 30 hours viewing time (Figure 2).
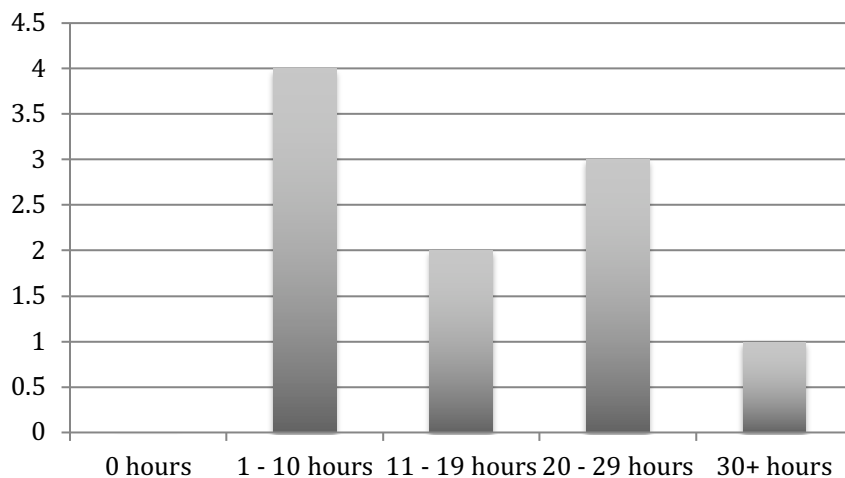

Figure 2. TV viewing in hours for an average week

II. *On which devices do you watch TV?*
Participants reported watching on traditional TV units in all cases, with two participants also using their smart phones.

III. *How do you choose TV programmes to watch?*
The motivations behind choice of programmes were completely mixed and ranged from finding interesting or educational programmes to keeping up with celebrities and the latest drama and comedy shows, where most participants responded that their choice of TV programmes is driven by their interest or anticipation about the programme (whether it is interesting, useful or of a good quality).

IV. *Is your TV on even when you are not watching?*

All but one participant switched their TV on only when they wanted to watch a specific programme, the exception being the participant who left the TV on whenever they were home. Then, having turned on the TV, the majority (mean = 4.3, median = 4) of the participants left the TV on even when they were no longer watching (Figure 3).
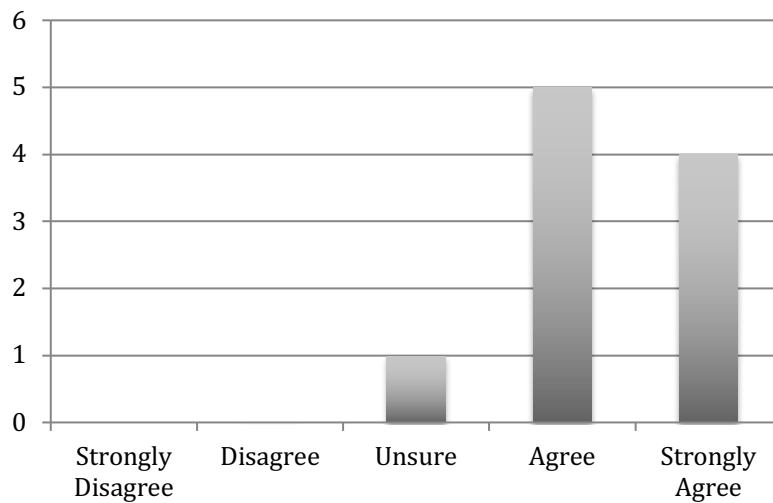

Figure 3. TV on when not watching

V. *When do you turn TV on?*
Contrary to their response, for IV above, the majority of the participants reponded that they turn TV on only when there is something they want to watch as opposed to TV being on all the time. This might be linked to the fact that most participants lived with their family with diffrent members making decisions about the programme of choice.
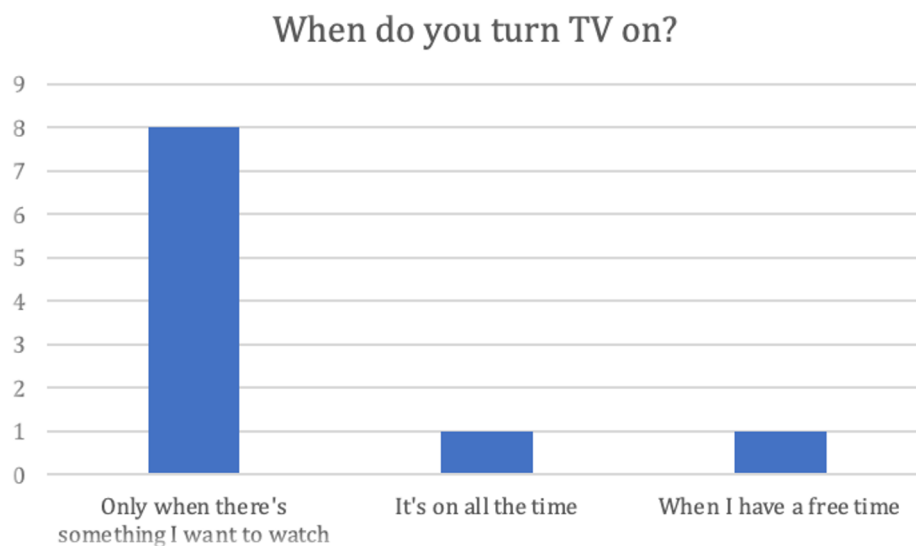

Figure 4. When do you turn TV on?

VI. *Do you multitask while watching TV?*
All participants reported multitasking to some extent while watching TV (4 = Agree, 6 = Strongly Agree, where mean = 4.3, median = 5), and most participants found that TV distracted them from these other tasks (mean = 4.1, median = 5; Figure 4).
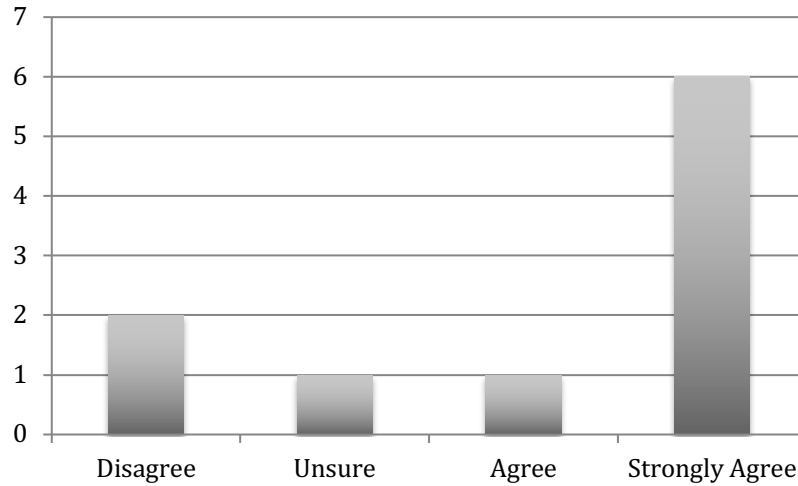
Figure 5. TV distracting from other tasks

3.2. *Impact captions*
VI. *Do impact captions help you to understand what is going on in the show?*
Most participants reported that captions help them to better understand the content of the TV programme (mean = 4.2, median = 4.5; Figure 5).
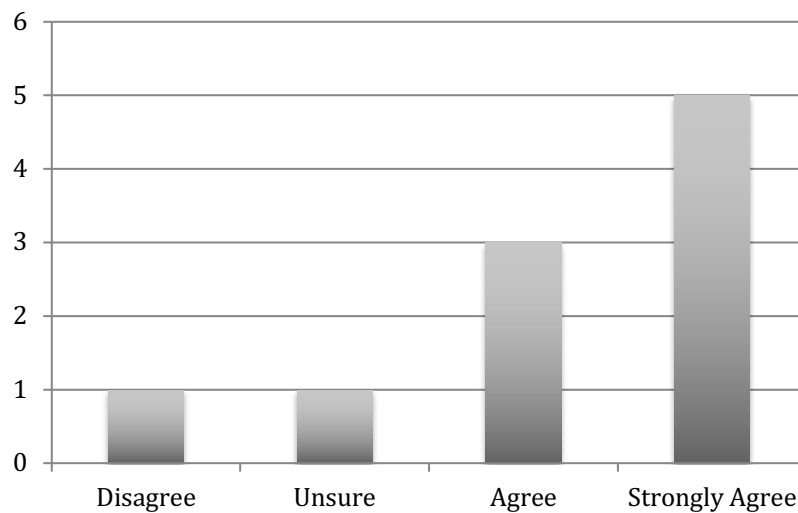

Figure 6. Improved comprehension

VII. *Do you find impact captions add to the comedic effect of the show?*
Most participants report that impact captions add to the comedic effect of TV programmes, where two participants disagreed strongly with this statement (mean = 4, median = 4; Figure 7).
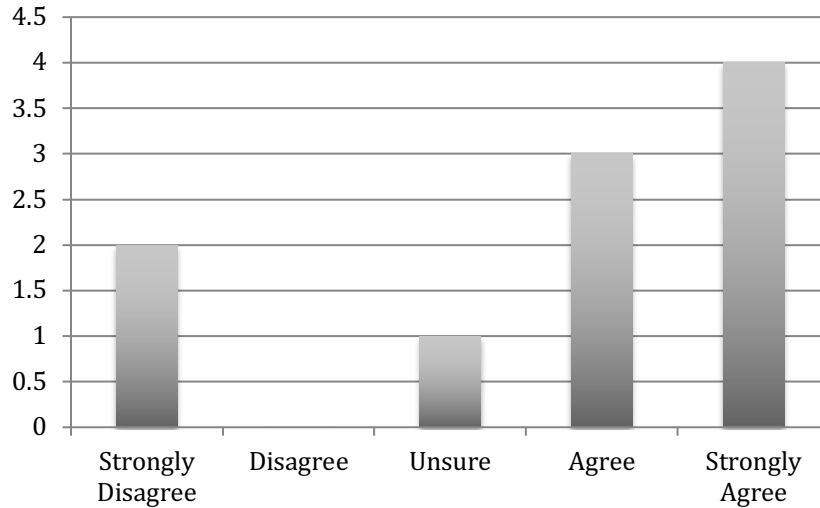
Figure 7. Enhanced comedic effect

VIII. *Do you find captions help you to keep your attention on the screen?*
A rather mixed result shows that participants were divided on whether or not captions helped them to keep their attention on the TV screen (mean = 3.2, median = 3, Figure 8).
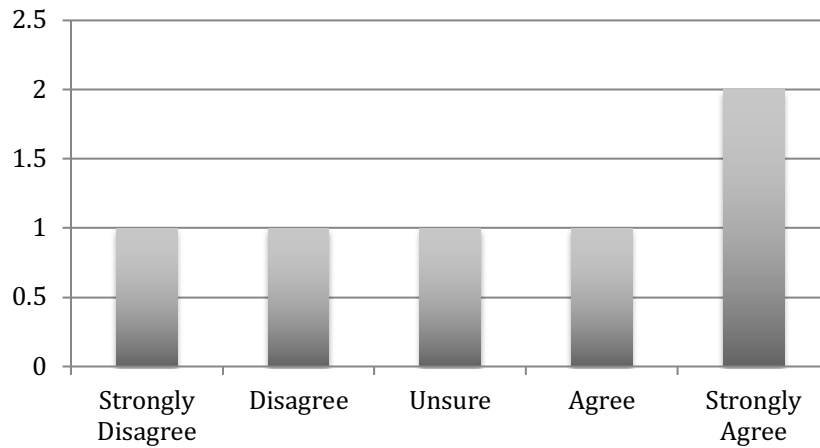

Figure 8. Helps visual attention

IX. *Do you find captions change your interpretation of what people say and do in the show?*
Slightly more participants reported that captions influence their interpretation of characters' utterances and actions on the TV shows they watched (mean = 3.7, median = 4; Figure 9).
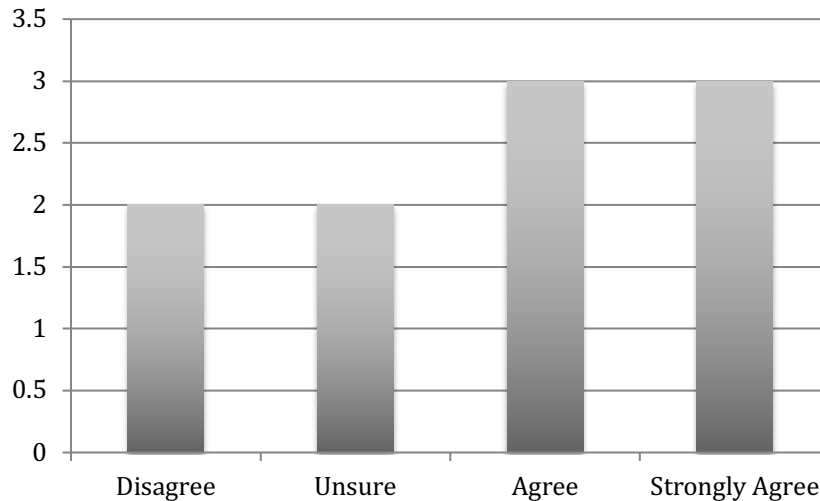
Figure 9. Influences interpretation

X. *Which do you prefer, programmes with impact captions or without them?*
Participants were exactly split on this item and 5 report preferring programmes with captions, while the other 5 prefer them without. All participants report reading captions regardless, with the majority reading all types of captions indiscriminately (Figure 10).
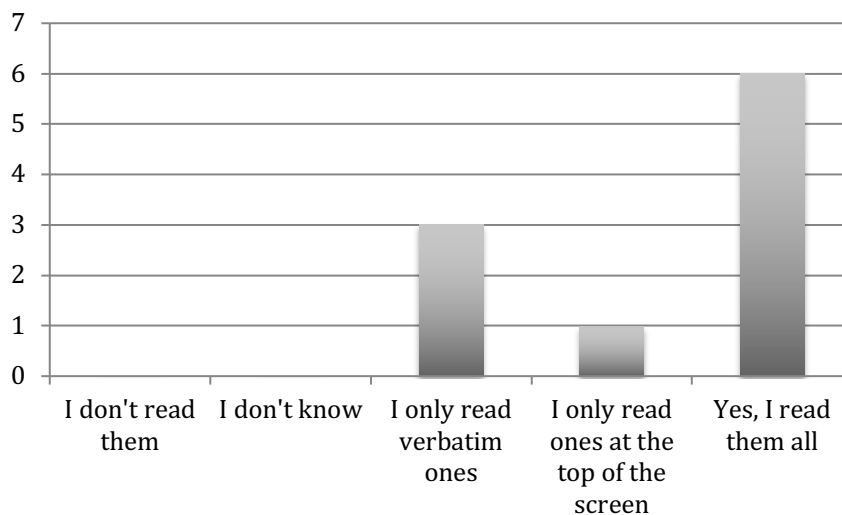

Figure 10. Reading captions

XI. *Do you ever just read captions while muting the sound?*
Lastly, none of the participants report reading captions while watching a programme that has been muted.

In summary, we have found that our participants watch TV regularly and mostly on TV screens. They mostly have the TV on in the background when they are not watching and they also multitask while watching TV. Coupled with this, they find the TV distracting when they want to switch their attention to concentrate on one task, in particular. Relating to captions, participants report that impact captions add to their comprehension of the programme and its humour, and help them to keep their attention on the screen to some extent. They state also that captions have an influence on their interpretation and that they indeed read all types of captions, and do not view the TV with captions when the sound is off.

3.3. *Visual attention*

On average, all participants spent 4.21 minutes overtly attending to the TV programme (median = 192, SD = 162.19, min. = 66, max. = 487, range = 421). Taken as percentages of the programme time, Figure 11 shows that in all cases, less than 40% of their time was spent attending to the TV (mean = 20%, SD = 11.65%, median = 17.03%, min. = 5.22%, max. = 38.58%, range = 33.36%).
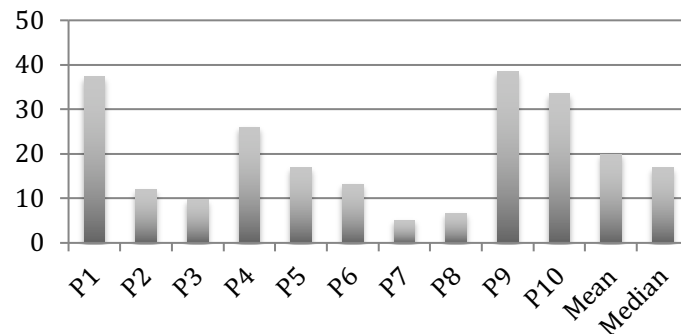
Figure 11. Participants' percentage of attention on TV

The TV programme contained 432 captions, of which an average of 60 were visually attended to by participants (SD = 43, median = 36, min. = 22, max. = 127, range = 105) – Figure 12 illustrates these instances as percentages of all captions.
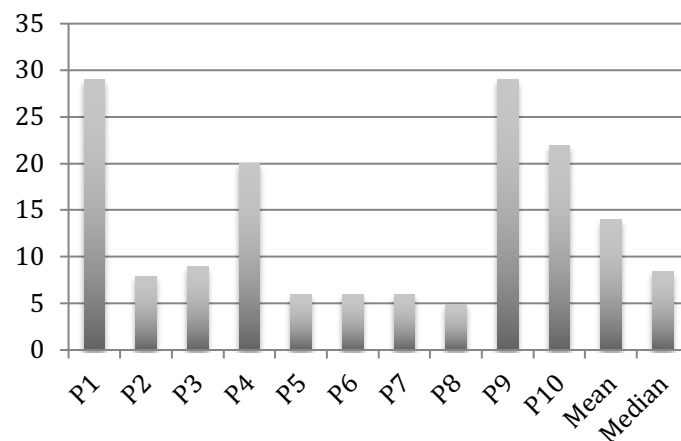
Figure 12. Participants' percentage of attention to impact captions

In addition, Figure 13 provides the average duration of participants' on-screen visual attention (mean = 6.5, SD = 2.83, median = 6, min. = 2, max. = 11, range = 9). Of note is the frequent short glances from the tasks to the TV programme which typically occurred in bursts rather than at fixed intervals over the duration of the programme.

Figure 13. Average duration of each participant's visual attention (seconds)

3.4. *Correlations between impact captions and visual attention*

In order to examine the relationships between the impact captions and the participants' visual attention, we study each capt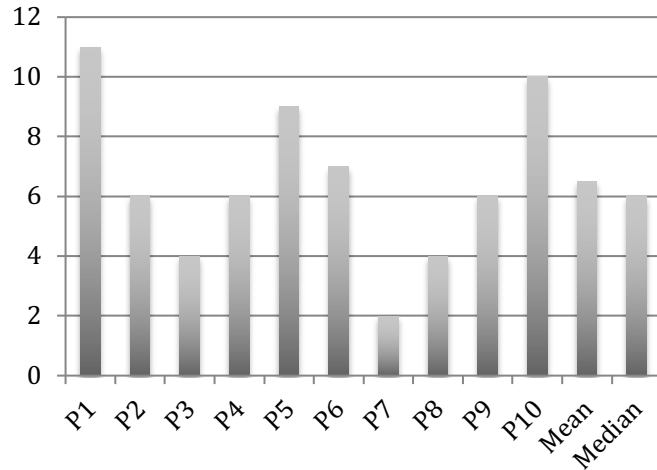ion in terms of its timing and various characteristics, namely: Captioned Speaker; Caption Content; Caption Type; Caption Colour; Caption Background; Caption Font; On-Screen FACS[2]: and Caption Sound. These stimulus variables are detailed for each of the 432 instances of impact captions that each of the participants was exposed to thus yielding 432 x 10 combinations, i.e., 4,320 data points. Table 1 details the significant correlations identified between the stimulus variables and each of the participants on an individual basis. Building upon this the table also shows the same test for correlations between the stimulus variables and the duration of visual attention participants spent on- and off-screen.

Stimulus Variable	Sig. Correlation with Participant	Correlation Coefficient (Rho)[3]	Sig. Correlation with Duration of Attention	Correlation Coefficient (Rho)
Speaker	P7	.0101*	-	-
Caption Content	P2	0.153**	P1 and P7	0.256**, 0.469*
Caption Type	P1	0.132*	P7	0.584**
Caption Colour	P2	0.214**	P3 and P7	0.432*, 0.596**
Caption Background	P3, P6, P9	0.157*, 0.206**, 0.213**	P3 and P6	0.544*, 0.770*
Caption Font	-	-	P1, P3, P7, P8	0.329**, 0.397*, 0.696**, 0.635**
On-Screen FACS	P1, P2, P8	0.198*, 0.249*, 0.169*	-	-
Caption Sound	-	-	P5 and P9	0.696*, 0.254*

Table 1. Correlations between stimulus variables and participants' visual attention and respective duration

3.5. *Summary of results*

These findings do not identify any particular variable as having a significant correlation with participant's visual attention or its duration, but they do serve to highlight the highly individualized viewing behaviour exhibited by the participants. Using multiple regression analyses, each participant's viewing behaviour (attending to the TV, sustaining attention to TV, not attending to TV) was analysed vis-à-vis the stimulus-related variables. This approach allows us to examine the role of

---

[2] Facial expression of those who appear on the screen, coded by Ekman's coding scheme Ekman et al 2002>

[3] * denotes significance at the 0.05 level and ** at the 0.01 level.

these variables, when they are all considered together, on each of the participants. Table 2 shows these results where the viewing behaviour of 6 of the 10 participants can be explained by at least one significant variable and each viewing pair is shaded. It is interesting to note the similarity between the pairs, e.g., Participants 9 and 10 were the only pair where their visual attention could not be accounted for (rho = 0.317, p < 0.01), where all other pairs had non-significant positive correlations less than 0.1. As evident from these results, none of the stimulus variables accounted for visual attention in all cases. On-Screen FACS (3 cases), Caption Sound (2), and Caption Type (2) all had significant influence on more than one viewer.

participant		predictors
1	caption sound**	
2	speaker**	caption type**
3	on-screen FACS*	
4	caption font*	on-screen FACS*
5	none	
6	on-screen FACS*	
7	caption sound*	caption type*
8	none	
9	none	
10	none	

Table 2. Participants' visual attention as predicted by stimulus variables

In summary, our results indicate a diversity in typical self-reported viewing behaviours, showing that participants still watch TV on a traditional TV unit, leave the TV on when not watching it, and multi-task while watching TV. Interesting, when trying to focus on one particular task, TV was identified as being a distraction. Impact captions, in general, were reported to improve viewers' comprehension, enhance comedic effects, and somewhat help sustain visual attention on the TV programme. Interestingly, most participants report that captions indeed influence their interpretation of the TV programme content thus acting as a moderator to some extent. This is coupled with the report that the majority of participants read all captions. In terms of visual attention, participants spent an average of 20% of the experiment time watching the TV programme, which means they focused more on the given tasks and the TV programme could only distract them on occasion, but it did indeed attract and retain their visual attention many times over the course of the programme – 60 times on average for all 432 captions throughout the programme with an average duration of 6 seconds. Lastly, correlations between stimulus variables and participants' visual attention show that that no particular aspect of the captions correlates or predicts visual attention for all participants in this sample. In most cases, however, On-Screen FACS, Caption Sound, and Caption Type were identified as having a significant influence on several participants.

3.6. *Retention of programme content*
As Chun & Turk-Browne (2007) argue, attention and memory are interdependent: memory has limited capacity and attention therefore dictates what is to be remembered, and, at the same time, past experience can guide what is attended. On this assumption, we included five items to ascertain the degree of retention to gain a further insight into the participants' retention of information, and hence the retention of attention, to the programme. The findings on information retention, together with the correlation between impact captions and visual attention, will enable us to capture the overall viewing behaviour of the participants. In what follows, we will present recall questionnaire results in relation to retention of programme content:

(1) *How many people participated in the variety show?*

Only two out of 10 participants answered this correctly, where the correct answer was 12 people (Table 3).

(2) *What was the main guest's name on the variety show?*

There were 12 guests on the programme. Most participants only managed to recall only a few names except for P04, who correctly recalled seven names. Table 4 shows the number of correct answer out of the number of names each participant listed:

(3) *What topic did they discuss on the variety show?*

Participants' responses gave only a limited number of topics with six out of 10 participants (P01, P02, P05, P05, P06, P09, P10) recalling incorrectly.

(4) *What did you find most memorable about the variety show?*

Some participants (P01, P02, P05, P06, P09, P10) responded with propositions that are not relevant to the programme contents while others' responses were related to the programme contents.

(5) *What was the best part of the variety show?*

Three participants' (P05, P06, P09) responses were irrelevant to the programme contents while other participants either did not find anything good, or listed the programme format or the MC of the programme who is a well-known comedian.

Overall, most participants (P01, P02, P05, P056, P09, P10) had generally poor recall results, where one outlier (P04) had considerably better results. P01, P09, and P10 had indicated higher Participants' Percentage of Attention on TV, Participants' Percentage of Attention to Impact Captions and above average in Average Duration of Each Participant's Visual Attention (Seconds). In addition, P02, was generally low for Participants' Percentage of Attention on TV, Participants' Percentage of Attention to Impact Captions and just below average in Average Duration of Each Participant's Visual Attention (Seconds). Lastly, P05, P06, were also generally low for Participants' Percentage of Attention on TV, Participants' Percentage of Attention to Impact Captions but above average for Average Duration of Each Participant's Visual Attention (Seconds).

Finally, Table 3 shows how we compared the overall picture of viewing behaviour between participants with poor recall results (in questions (1) to (5) above) and other participants. It is interesting to note that two groups seem to differ to the greatest extent in three questions 4, 7, and 9 [as highlighted in Table 3]. This suggests that participants with poorer recall results do not find impact captions adding humorous effects as much as other group does, while they see TV more as a distraction if they have other tasks at hand. Furthermore, most of the participants with poorer recall results prefer TV programmes without impact captions while participants with better recall results prefer programmes with impact captions.

| Q | participants with poor retention results | | | | | | | other participants | | | | |
	P01	P02	P05	P06	P09	P10	*M*	P03	P04	P07	P08	*M*
1	5	5	4	3	4	5	4.3	4	4	4	5	4.3
2	5	5	5	4	4	5	4.67	4	5	4	5	4.5
3	5	5	5	4	3	5	4.5	4	5	5	4	4.5
4	3	2	4	1	4	5	3.27	4	5	5	5	4.8
5	4	1	3	2	3	5	3	3	5	3	3	3.5
6	2	1	5	4	3	5	3.3	2	4	4	3	3.3
7	5	5	5	4	3	5	4.5	5	2	5	2	3.5
8	01–10	11–19	30+	20–29	1–10	20–29		1–10	11–19	1–10	20–29	
9	without	without	without	without	with	without		with	with	with	with	

	verbatim only	verbatim only	verbatim only	all	all	top-title only		all	all	all	all
10											
11	no	no	no	no	no	no		no	no	no	no

Table 3. Viewing behaviour comparisons between two groups based on recall results

4. Discussion

In the current paper, we examined the reception and effect of impact captions on TV viewers in terms of attracting and retaining viewers' attention. In addition to reporting general viewing behaviours of the sample of participants, i.e., regular viewers of TV with the TV usually on in the background while multitasking, our findings show that impact captions were reported to have added to the comprehension and humour of the programme and indeed helped to retain participant retention. It was also evident that the captions had a significant influence on the interpretation of the programme as perceived by the participants.

Further, our results did not identify any significant correlation between participant's visual attention and the caption stimulus, but revealed highly individualized viewing behaviour. While the caption stimulus predicted visual attention in most cases, not all participants' behaviours were accounted for owing to the limited sample size and degree of individual differences. Lastly, most participants had generally poor recall results. As such, an interesting trend became apparent in which those with poor retention results reported less of the comical effect of the captions, saw TV as more of a distraction when doing other tasks, and therefore prefered programmes without captions. On the other hand, those who had a positive perception of captions, paid more attention to the TV which resulted in better retention.

The fact that there was no strong pattern for general viewing behaviour identified in the findings, has some interesting implications. As we have seen in 3.2 (Figure 11), less than 40% of their time was spent attending to the TV (mean = 20%, SD = 11.65%, median = 17.03%, min. = 5.22%, max. = 38.58%, range = 33.36%). At the same time, the results of the recall test in general were somewhat poor as shown in 3.6. This suggests that they did pay some attention but their attention was not retained fully enough to process the contents to the level that would warrant good recall results. That is, our findings are contradictory in that, while on the one hand they do pay some attention to the programme despite having other tasks to complete, on the other hand the TV programme itself does not seem to be particularly memorable for our participants. The question is how and why this is the case.

Recall the Cognitive Principle of Relevance, discussed in §1.2, which describes the tendency that human cognition is designed in such a way that it pays attention to what is relevant to them. Information is deemed relevant if it interacts in a certain way with one's existing assumptions (Sperber and Wilson 1995). Further, as we discussed in §1.2, failure to return a suitable reward for the relative effort in processing the stimuli can lead to viewer dissatisfaction, additional cognitive load and other negative effects (Sasamoto 2014; O'Hagan & Sasamoto 2016; Kruger & Doherty 2016; Doherty 2020). We argue that this might have some impact on the viewing behaviour observed in this study. Cognitive and Communicative Principles of Relevance predict that the viewer is to assume that any message being presented on the screen is worthy of being processed and is therefore worth the effort with the expectation that this expend of cognitive resources returns a reward such as comprehension, humorous effects, etc. In this context, we would normally expect them to pay more attention to the TV programme as ostensively presented stimulus. We should also consider the participants' general TV viewing habits and their perception of programmes that are simply on when the participant did not actively choose to watch them. As we saw in the survey results, most (8 out of 10) participants responded that they switch on TV only when there is something they want to watch, where the majority (9 out of 10) of the participants responded that they leave the TV on even after they have finished watching it. This could indicate how TV is part of their 'background' ambience, which they have learned to co-exist with, while in the same room. For this experiment, the TV programme being displayed was not of the participants' choosing and, as the responses to our survey show, the participants are the ones that normally have control over what they choose to

watch, along with family members in some cases. This could mean that when they were simply asked to be in the room when a programme chosen by someone else was put on, they did not have any existing assumption relating to the TV programme in question that was salient in their cognitive environment. TV might be on in their environment and is sending ostensive stimuli to the viewers, but not enough existing assumptions are made salient and hence it was too costing for participants to process the programme (and hence they paid little attention). If this is correct, then it could mean that unless viewers already have some existing assumptions activated prior to watching the programme, they do not willingly engage with the programme, unless, of course, the programme manages to grab their attention enough to engage them fully (and hence achieve relevance). That is, the participants' viewing pattern is driven by their personal interest, rather than something guided by the TV producer's intention in instances where it wasn't the participants' choice to watch something. Even if something is ostensively communicated and viewers pay some attention to it, they simply could not or did not follow through the inferential process to recover the full intended meaning. This suggests that the multimodal design of TV programmes itself is not enough to attract or retain viewer attention. It might be the case that viewers need to be prepped to pay attention at the point of choosing to watch something–further research would be necessary to determine if this is indeed the case.

Given the above results and limitations, the weaknesses of the research design are evident in the artificial nature of the experiment in that it may not be sufficiently similar to participants' typical viewing behaviours. Naturally, a sample size of ten and use of one TV programme is rather small and limits the generalizability of these findings as well as the application of more robust quantitative analyses. In terms of data collation and manual analysis, while the complexity of the authentic materials of the study called for a completely manual analysis of the recordings (of the participants and of the TV programme) the risk of human error in annotation is apparent, hence the use of independent analyses conducted by the three authors in an effort to maintain interrater reliability and overall validity.

Future work could develop upon this design and expand the sample size and variety of TV programmes used in order to collate more representative data which will lead to further insights in the relationships between impact captions, their characteristics, and resultant effects. While this study did not identify one type of caption that consistently results in drawing or sustaining visual attention in all participants, several variables have been found to have significant correlations and predictions on several participants. We also note the individualised differences in watching TV, which are evident in our findings and may help identify specific attributes of viewers, who are not a homogenous group, affected by particular impact captions to a greater or lesser degree. Such findings will help ascertain more precisely the effects of impact captions on a wide variety of TV viewers across different genres and implementation of captions.

5. Conclusions

In this paper, we set out to examine the 'hookability', or the reception and effect of impact captions on viewers in terms of attracting and retaining viewer attention. Despite the limitation and weakness of the experimental design, this study has shown some evicence that impact captions do have some impact on hookability, while the long-lasting effect of such 'hook' is highly personal. While there is no strong pattern for general viewing behaviour, this study found that several variables, especially caption content and caption colour, have significant correlations on some participants and may be of interest in future work and applications. The most notable findings of this study is the individual differences in TV viewing pattern, and the influence of pre-conception about a given programme and impact captions in general on participants' responses. The results also suggest that viewers are likely to have developed a tacit viewing strategy over time in relation to manage a competing task demand with 'always on TV' in the background. Such strategies can be deeply personal, of which viewers themselves may not necessarily be aware. This confirms the methodological challenge in any reception study which cannot assume viewers as a homogenous group. Future research may focus on the individual differences further by detailed profiling of viewers and their viewing strategies.

References

Baumgarten, Nicole. 2008. "Yeah, that's it!: Verbal Reference to Visual Information in Film Texts and Film Translations." *Meta: Translators' Journal* 53 (1): 6–25. https://doi.org/10.7202/017971ar

Bogucki, Łukasz. 2004. The Constraint of Relevance in Subtitling. *The Journal of Specialised Translation* 1: 71–88.

Caffrey, Colm. 2008. "Using Pupilometric, Fixation-Based and Subjective Meastures to Measure of the Processing Effect Experienced when Viewing Subtitled TV Anime with Pop-Up Gloss". In *Looking at Eyes: Eye-tracking Studies of Reading and Translation* Processing, edited by S. Göpferich, A. Jakobsen, and I. Mees, 125–144. Copenhagen, Denmark: Samfundslitteratur.

Caffrey, Colm. 2009. "Relevant abuse? Investigating the effects of an abusive subtitling procedure on the perception of TV anime using eye tracker and questionnaire." PhD thesis. Dublin City University.

Chung, Marvin and Nicholas Turk-Browne, 2007. "Interactions between Attention and Memory." *Current Opinion in Neurobiology* 17 (2): 177–184. https://doi.org/10.1016/j.conb.2007.03.005.

Díaz-Cintas, Jorge & Remael, Aline. 2021. *Subtitling: Concepts and Practices.* London, UK: Routledge. 10.4324/9781315674278.

Doherty, Stephen. 2016. "The Impact of Translation Technologies on the Process and Product of Translation". *International Journal of Communication* (10): 947–969.

Doherty, Stephen. 2020. "Multisensory Integration in Audiovisual Translation". In *Multilingual Mediated Communication and Cognition,* edited by S. Halverson & R Muñoz Martín, 155–170. London, UK: Routledge.

Ekman, Papul, Wallace V. Friesen, and Joseph C. Hager. 2002. *The Facial Action Coding System CD-ROM*. Salt Lake City: Research Nexus.

Fritz, Jonathan, Mounya Elhilali, Stephen David, and Shihab Shamma. 2007. "Auditory Attention: Focusing the Searchlight on Sound." *Current Opinion in Neurobiology* 17 (4): 437–455.

Gambier, Yves. 2018. "Translation Studies, Audiovisual Translation and Reception." In *Reception Studies and Audiovisual Trannslation*, edited by E. Di Giovanni & Y. Gambier, 43–68. Ansterdam, Netherlands: John Benjamins. 10.1075/btl.141.05tuo

Gibbs, Raymond & Tendahl, Markus. 2006. "Cognitive Effort and Effects in Metaphor Comprehension: Relevance Theory and Psycholinguistics". *Mind & Language* 21 (3): 379–403. https://doi.org/10.1111/j.1468-0017.2006.00283.x.

Karamitroglou, Fotios. 1998. *A Proposed Set of Subtitling Standards in Europe.* [Online] Available at: http://www.translationjournal.net/journal/04stndrd.htm

Karamitroglou, Fotios. 2000. *Towards a Methodology for the Investigation of Norms in Audiovisual Translation.* Amsterdam, Netherlands: Rodopi.

Kennedy, Daniel, Brian D'Onofrio, Patrick Quinn, Sven Bölte, Paul Lichtenstein & Terje Falck-Ytter. "Genetic Influence on Eye Movements to Complex Scenes at Short Timescales". *Current Biology* 27 (22): 3554–3560. https://doi.org/10.1016/j.cub.2017.10.007

Knoblauch, Kenneth, Aries Arditi, and Janet Szlyk. 1991. "Effects of Chromatic and Luminance Contrast on Reading". *Journal of the Optical Society of America* 8 (2): 428–439.

Koga, Tadamichi. 2013. "Research Report: An Attempt of Quantitative Analysis on Directing Elements in Information and Variety Programs: From an Analytical Survey on Primetime Television Programs." *The NHK Monthly Report on Broadcast Research* 63 (5): 68–81.

Kovačič, Irena. 1994. *Relevance as a Factor in Subtitling Reductions.* Amsterdam: John Benjamins.

Kovačič, Irena. 1995. "Reception of Subtitles. The Non-existent Ideal Viewer", Translatio (Nouvelles de la FIT/FIT Newsletter) 14 (3–4), 376–383. Audiovisual Communication and Language Transfer, International Forum, Strasbourg. 22–24 June 1995 Forum, Strasbourg 22–24.

Kress, Gunter, and Theo Van Leeuwen. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London, UK: Edward Arnold.

Kruger, Jan-Louis & Stephen Doherty. 2016. "Measuring Cognitive Load in the Presence of Educational Video: Towards a Multimodal Methodology". *The Australasian Journal of Educational Technology* (32): 19–31.

Lang, Annie. 1990. "Involuntary Attention and Physiological Arousal Evoked by Structural Features and Emotional Content in TV Commercials". *Communication Research* 17 (3): 275–299.

Legge, Gordon. 2006. *Psychophysics of Reading*. Mahwah, New Jersey: Lawrence Erlbaum Associates,

Mangiron, Carmen. 2013. "Subtitling in Game Localisation: A Descriptive Study." *Perspectives: Studies in Translatology*, 21(1): 42–56.

Orero, Pilar, Stephen Doherty, Jan-Louis Kruger, Anna Matamala, Jan Pedersen, Elisa Perego, Pablo Romero-Fresco, Sara Rovira-Esteva, Olga Soler-Vilageliu & Agnieszka Szarkowska. 2018. "Conducting Experimental Research in Audiovisual Translation (AVT): A Position Paper". *The Journal of Specialised Translation* (30): 105–126.

O'Hagan, Minako. 2010. "Japanese TV Entertainment: Framing Humour with Open Caption Telop". In *Translation, Humour and the Media*, edited by D. Chiaro, 70–88. London, UK: Bloomsbury.

O'Hagan, Minako & Ryoko Sasamoto. 2016. "Crazy Japanese Subtitles? Shedding Light on the Impact of Impact Captions with a Focus on Research Methodology". In *Eyetracking and Applied Linguistics*, edited by S. Hansen-Schirra & S. Grucza, 31–58. Berlin, Germany: Language Science Press.

Park, Joseph. 2008. "Ideological Functions of Subtitles in Multimodal Media Texts: Impact Captioning on Korean Television." Paper presented at the 4th International Conference on Multimodality: From Print to Interactive Digital Media: Technology, Multimodal Representation and Knowledge, Singapore Management University, Singapore, July 30–August 1.

Pashler, Harold. 1994. "Dual-task Interference in Simple Tasks: Data and Theory." *Psychological Bulletin* 116 (2): 220–244. https://doi.org/10.1037/0033-2909.116.2.220

Pérez González, Luis. 2007. "Intervention in New Amateur Subtitling Cultures: A Multimodal Account." *Linguistica Antverpiensia* (6): 67–80.

Quittner, Alexandra, Linda Smith, Mary Osberger, Teresa Mitchell, and Donald Katz. 1994. "The Impact of Audition on the Development of Visual Attention." *Psychological Science* 5 (6): 347–353.

Sasamoto, Ryoko. 2014. "Impact Caption as a Highlighting Device: Attempts at Viewer Manipulation on TV". *Discourse, Context, and Media* 6: 1–10. https://doi.org/10.1016/j.dcm.2014.03.003

Sasamoto, Ryoko & Stephen Doherty. 2016. "Towards the Optimal Use of Impact Captions on TV Programmes". In *Conflict and Communication: A Changing Asia in a Globalising World*, edited by M. O'Hagan & Q. Zhang, 210–247. Hauppauge, New York: Nova Science Publishers.

Sasamoto, Ryoko & Minako O'Hagan. 2020. "Relevance, Style and Multimodality: Typographical Features as Stylistic Devices". In *Relevance Theory, Figuration, and Continuity in Pragmatics*, edited by A. Piskorska, 194–228. Amsterdam, Netherlands: John Benjamins.

Sasamoto, Ryoko, Minako O'Hagan & Stephen Doherty. 2017. "Telop, Affect, and Media Design: A Multimodal Analysis of Japanese TV Programs". *Television & New Media*, 18 (5): 427–440.

Schultheis, Hoger. and Jameson, Anthony. 2004. "Assessing Cognitive Load in Adaptive Hypermedia Systems: Physiological and Behavioural Methods." In *Adaptive Hypermedia and Adaptive Web-based Systems,* edited by W. Nejdl and Paul de Bra, 18–24. Eindhoven, Netherlands: Springer Verlag.

Shiota, Eiko. 2003. "関連性理論とテロップの理解 [Relevance Theory and Understanding of Telop]." *The Graduate School Review of the English Language and Literature* 31: 63–91.

Shitara, Kaoru. 2011. "The Change of Telops on NHK Variety Shows." *Bulletin of Mukogawa Women's University, Humanities and Social Science* 59: 1–9.

Sperber, Dan and Deirdre Wilson, 1987. "Précis of Relevance." *Behavioral and Brain Sciences* 10 (4): 697–710.

Sperber, Dan and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Oxford: Blackwells.

Tan, Eduard Sioe-Hao. 2008. "Entertainment is Emotion: The Functional Architecture of the Entertainment Experience". *Media Psychology* 11 (1): 28–51. https://doi.org/10.1080/15213260701853161

Taylor, Christopher. 2003. "Multimodal Transcription in the Analysis, Translation and Subtitling of Italian Films." *The Translator* 9 (2): 191–205.

Taylor, Christopher. 2004. "Multimodal Text Analysis and Subtitling". In *Perspectives on Multimodality,* edited by E. Ventola, C. Charles and M. Kaltenbacher, 153–172. Ansterdam, Netherlands: John Benjamins.

Telebi Kiroku. 2013. Honmadekka TV Shichoritsu 2013. *Telebi kiroku* Available at: http://asobigokoro.info/honmadekka2013/

van Gerven, Pascal, Fred Paas, Jeroen van Merriënboer & Henk Schmidt. 2004. "Memory Load and the Cognitive Pupillary Response in Aging." *Psychophysiology 41* (2): 167–174. https://doi.org/10.1111/j.1469-8986.2003.00148.x

Wickens, Christopher. 1991. "Processing Resources and Attention." In *Multiple-task Performance,* edited D. Damos, 3–34. London, UK: Taylor & Francis.

Wilson, Deirdre & Dan Sperber. 2002. "Truthfulness and Relevance." *Mind* 111 (443): 583–632.

Address for correspondence

Ryoko Sasamoto
School of Applied Language and Intercultural Studies
Dublin City University
Dublin 9, Ireland

Biographical notes

Ryoko Sasamoto is Associate Professor in the School of Applied Language and Intercultural Studies, and a member of Centre for Translation and Textual Studies at Dublin City University. Her research expertise is in the interdisciplinary area, working across different disciplines such as pragmatics, audiovisual translation, and reception studies.
https://orcid.org/0000-0002-1644-6897

Stephen Doherty is Associate Professor in Linguistics, Interpreting, and Translation, and lead of the HAL Language Processing Research Lab at the University of New South Wales. With a focus on the psychology of language and technology, his research investigates human language processing and usage.
https://orcid.org/0000-0003-0887-1049

Minako O'Hagan is Associate Professor at the School of Cultures Languages and Linguistics, University of Auckland. She has research specialisms in translation and technology. Her edited publications include *the Routledge Handbook of Translation and Technology* (2020)*.
https://orcid.org/0000-0003-0870-0725