Telops for Language Learning: Japanese Language Learners' Perceptions of Authentic Japanese Variety Shows and Implications for Their Use in the Classroom

Eline Christina Sikkema

BA, MA

Thesis Submitted for the Award of Doctor of Philosophy

School of Applied Language and Intercultural Studies

Dublin City University

Supervisors:

Dr Joss Moorkens

Prof. Françoise Blin

December 2019

DECLARATION

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ (Candidate) ID No.: _____ Date: _____

(Eline Sikkema)

ACKNOWLEDGEMENTS

I would like to express my gratitude to the principal investigators of the overarching project *Impact Caption for Everyone – Reception Metrics for Creative Subtitling on TV*, Dr Ryoko Sasamoto and Prof. Minako O'Hagan, for taking me on as a doctoral researcher and to have given me this opportunity. This thesis could not have been realised without the financial support that Dr Ryoko Sasamoto afforded through the Careers Enhancement Award and the International Strategic Cooperation Award. Without her help I would not have been able to travel to Japan or to have availed of eye-tracking technology in order to conduct this study.

This research was also funded by the School of Applied Language and Intercultural Studies at Dublin City University. I would like to extend my appreciation for this financial support. I would also like to thank the School of Applied Language and Intercultural Studies and the Faculty of Humanities and Social Sciences at Dublin City University for the grants that have facilitated my attendance at trainings and conferences.

This PhD journey has been a most challenging yet rewarding venture. I am deeply grateful to my supervisors Dr Joss Moorkens and Prof. Françoise Blin for their support, constructive feedback, and belief in me throughout this process. Their encouragement and patience have played a pivotal role in the completion of this thesis and during personally challenging moments over the past few years.

I would also like to express my sincerest gratitude towards Prof. Dorothy Kenny and the support networks at Dublin City University for their guidance over the course of this PhD programme.

Thanks also to the participants who were involved in this study. It was a joy to work with you and to experience your enthusiasm for both the project and Japanese language learning. I would also like to thank our partners in Japan and Ireland for the valuable insights into eye-tracking technology, Japanese variety shows, and telop production. It was a real privilege to visit your offices in Japan and to learn more about your experiences with television broadcasts.

I am very grateful to my fellow doctoral students and friends at Dublin City University. Your support meant a lot to me throughout the years and I am glad to have shared the workspace in the Postgraduate Research Room with you.

I am deeply grateful to the external and internal examiners of this thesis, Prof. Jorge Díaz Cintas and Dr Patrick Cadwell, for their constructive feedback. I am also thankful for the exciting and positive viva experience!

Lastly, I would like to thank my dear friends and family back at home in the Netherlands and elsewhere for their support over the past five years. I would definitely not have been able to finish this thesis without you!

LIST OF PRESENTATIONS AND PUBLICATIONS

Sikkema, E.C., Sasamoto, R. and O'Hagan, M. (2016) 'Comprehension of communicative multimodal content in Japanese TV programmes by language learners', *EPICS VII: 7th international symposium on intercultural, cognitive and social pragmatics*. Pablo de Olavide University, Seville, 4-6 May.

Sikkema, E.C. (2017) 'Reception of Japanese captions: a comparative study of visual attention between native speakers and language learners of Japanese' (from the Proceedings of the 25th EUROCALL conference, University of Southampton, 23-26 August 2017), in Borthwick, K., Bradley, L. and Thouësny, S. (eds.) *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017*. Research-publishing.net, pp. 289-293. doi: 10.14705/rpnet.2017.eurocall2017.728.

Sikkema, E.C. (2018) 'Exploring visual attentional shifts of language learners of Japanese' (from the Proceedings of the 26th EUROCALL conference, University of Jyväskylä, 22-25 August 2018), in Taalas, P., Jalkanen, J., Bradley, L. and Thouësny, S. (eds.) *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018*. Research-publishing.net, pp. 296-301. doi: 10.14705/rpnet.2018.26.853.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ALA	American Library Association
AOI	Area Of Interest
AV	AudioVisual
AVT	AudioVisual Translation
CAQDAS	Computer-Assisted Qualitative Data Analysis Software
DCU	Dublin City University
FL	Foreign Language
ICE	Impact Caption for Everyone
I-VT	Velocity-Threshold Identification
JF	The Japan Foundation
JLL	Japanese Language Learner
JLA	Japanese Language Acquisition
JLT	Japanese Language Teaching
JP	Japanese Participant
L1	First Language
LC	Library of Congress
MMR	Mixed-Methods Research
NHK	Nippon Hōsō Kyōkai [Japan Broadcasting Corporation]
OCT	Open Caption Telop
P	Participant
post1	Post-task questionnaire (part 1)
post2	Post-task questionnaire (part 2)
pre	Pre-task questionnaire
Q	Question(s)
QUAL	Qualitative
quan	Quantitative
RQ	Research Question
RTA	Retrospective Think Aloud
SCMC	Synchronous Computer-Mediated Communication
SDH	Subtitles for the Deaf and Hard-of-Hearing
SLA	Second Language Acquisition

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF TERMS

Absolute gaze duration – The accumulated duration of all fixations of all selected participants on a chosen test element (Bojko, 2009, p. 32; Tobii AB, 2014b, p. 66).

Abusive subtitle – Abusive subtitling is a term coined by Nornes and refers to the creation of subtitles that do not follow conventional subtitling rules. Subtitles are abusive when they are not restricted by the conventions of the target language or the constraints posed by the target text and have their origin in the source text foregrounded by the subtitler (Nornes, 1999; Nornes, 2007).

Accuracy – Accuracy is an indicator of data validity in eye-tracking research (Tobii AB, 2018). It is a measurement of error between a participant's actual gaze point and the idealised gaze point a participant is supposed to look at or has been instructed to look at by the researcher. The closer the actual gaze point of a participant is to the idealised gaze point, the higher the accuracy of the eye-tracking data is (Sullivan, 2016). Accuracy is reported in degrees and can be measured in relation to spatial and temporal qualities (Blignaut and Wium, 2014, p. 67).

Area Of Interest (AOI) – An AOI is a user-defined region in a test element over which descriptive statistics are extracted from the gaze data. There are two types of AOI:

(1) Static AOIs that remain unchanged and active for the duration of the test element
(2) Dynamic AOIs which can deactivate for certain time intervals of the test element and change shape
(Tobii AB, 2014b, pp. 77-78)

Area Of Interest (AOI) group – An AOI group consists of several AOIs. The AOIs contained in a group have been labelled with the same identifier (Tobii AB, 2014b, p. 83).

Authenticity – Authenticity in language learning material is defined based on a material's purpose. If a material has not been created for language learning purposes or constructed in order to teach a language and its linguistic features or form then the material is considered to be authentic (Gilmore, 2011, p. 791; Tomlinson, 2012, p. 162).

Authorial title – Authorial titling is a term coined by Pérez-González and refers to the use of intralingual text in commercial media products which do not fulfil a translational purpose. Authorial titles facilitate complex narratives as they provide viewers insights into characters' internal thinking processes and other diegetic knowledge that would otherwise not be accessible to viewers (Pérez-González, 2012).

Bodily configuration – Bodily configuration relates to meaning that results from body spacing, movements and appearances, and gaze (Kalantzis *et al.*, 2016, p. 386). For the purposes of this thesis, bodily configuration is defined less broadly and refers to movements of body parts other than the hands or face.

Calibration – The calibration process is a procedure during which a participant is instructed to look at a dot moving on the screen (Olk and Kappas, 2011, p. 439). An estimate of actual points the participant was looking at on the screen during the calibration are shown in a calibration plot and used by the researcher to gauge whether the eye-tracker can record a participant's gaze points accurately. Participants are calibrated prior to an eye-

tracking experiment in order to reduce the error in accuracy (Blignaut and Wium, 2014, pp. 69-70; Sullivan, 2016).

Data extract – Data extract refers to the raw data contained in a data item or part of a data item that has been labelled with one or more initial codes, and extracted from the data item for further analysis. This means that a data item is broken down into one or more extracts after it has been initially coded. A representative selection of these extracts are featured in the final report of a thematic analysis (Braun and Clarke, 2006, p. 79).

Data item – A data set consists of smaller chunks of raw data which are called data items. How a data item is defined depends on the type of data collected and the type of analysis conducted for a study (Braun and Clarke, 2006, p. 79). For example, each of the completed questionnaires can be called a data item while each response to one particular question in these questionnaires can also be defined as a data item.

Error vector – An error vector is a green line in the calibration plot which shows the error in accuracy between the calibration point and the actual point a participant is looking at (Tobii AB, 2014b, p. 35).

Eye-movement classification – The eye-movement classification is also known as the "fixation filter". Such a filter defines how the raw eye-movement data is visualised in the eye-tracking software and how the eye-tracking measures are calculated (Tobii AB, 2014b, p. 51).

Eye-position sample – Raw gaze data consists of eye-position samples which comprise timestamps and coordinates of a participant's gaze points (Tobii AB, 2014b, p. 51).

Fansub – Fansubbing is the creation of unofficial subtitles by fan groups. Subtitles created through this practice are called "fansubs" (Pérez-González, 2012; O'Hagan, 2013).

Fixation – A fixation is an aggregation of eye-position samples in accordance with the eye-movement classification and the fixation threshold set by the researcher. The way in which eye-position samples are counted as fixations changes when these settings are altered (Tobii AB, 2014b, p. 51).

Gaze path – A gaze path is a visual record of a participant's gaze points and can be replayed to observe that participants' gaze behaviour.

Gaze plot – A gaze plot is a type of data visualisation in which the test element is overlaid with gaze points of a participant's gaze path. Gaze points in gaze plots can be numbered to show the order of allocated visual attention and the size of each gaze point can be made proportional to its duration (Tobii AB, 2014b, pp. 63-64).

Gaze replay video – A gaze replay video is a recording in which the test element is overlaid with a participant's gaze path to observe that participants' gaze behaviour (Tobii AB, 2014b, p. 38).

Gaze sample ratio – The eye-tracking software *Tobii Studio* automatically generates a gaze sample percentage for each recording. This percentage shows the extent to which a participant's eyes have been detected during an eye-tracking experiment. It is calculated by dividing the total number of correctly identified eye-tracking samples with the total number of attempts. Although the gaze sample ratio does not specify which eye is detected

throughout an experiment, the general rule is that high percentages show high levels of detection for eye-tracking experiments in which participants are not required to look away from the screen (Tobii AB, 2014b, p. 39).

Gesticulation – Hand or arm movements that support or accompany utterances (Kalantzis *et al.*, 2016, p. 388). Although such movements can include gestures that visually represent some of the contents of an utterance (e.g. giving the thumps up), they can also refer to those automatic gestures made by a speaker that do not add any meaning to an utterance but help the flow of speech.

Hawthorne effect – The Hawthorne effect entails deliberate changes in a participant's behaviour as a result of the awareness that his or her behaviour is being studied. This affects the validity of findings (Winke, Gass and Sydorenko, 2013, p. 269; Saldanha and O'Brien, 2014, pp. 31-32).

Heat map – A heat map is a type of data visualisation in which the test element is overlaid with colour gradients to show the distribution of participants' visual attention (Tobii AB, 2014b, p. 65).

High-level factor – High-level factors are cognitive factors that guide top-down processing of a test element (Dyer and Pink, 2015, para. 4). Such factors include a user's goal(s), expectation(s) and previous knowledge or the specific task that was given to the user (Holsanova, 2014, p. 289).

Impact caption – Impact captioning is a term coined by Park and refers to a captioning practice used in South Korean variety and talk-shows. Impact captions are intralingual texts that not only represent utterances; they also provide unspoken commentary on a show's contents in various ways. Such captions are typically shown in a variety of typefaces, sizes and colours. According to Park, a producer can exploit these features (i.e. impact captioning) to convey his or her own interpretations in a manner that seems objective and authoritative as the producer's intentions are incorporated into impact captions and therefore attributed to people appearing on the show rather than him or herself (Park, 2009).

Initial code – An initial code is a label attached to a data item or part of a data item. The raw data a researcher labels with an initial code represents an aspect of the data set of which he or she thinks could prove to be important for the identification of themes across the data items (Braun and Clarke, 2006, pp. 88-89). Coding schemes vary between studies but initial codes generally comprise descriptive or interpretive codes (Moran, 2017). Initial coding can be done manually or with the help of software (Braun and Clarke, 2006, p. 89; Moran, 2017).

Interlingual – Interlingual is a term that originates from Roman Jakobson's tripartite taxonomy of translation and refers to "translation proper" or rendition from one language into another (Jakobson, 1959, p. 233).

Intersubject bias – Intersubject bias affects the validity of research findings and occurs when participants influence each other or share details about the research with other subjects during data collection (Saldanha and O'Brien, 2014, pp. 32-33).

Intralingual – Intralingual is another type of translation in Roman Jakobson's taxonomy and refers to rewording within the same language (Jakobson, 1959, p. 233).

Japanese variety show – Japanese variety shows are Japanese television programmes aimed at entertainment. They comprise a great variety of genres which range from talk and game shows to variety shows that are focused on music, quizzes, comedy or travel to name but a few (O'Hagan, 2010, p. 79; Shitara, 2011, p. 3; O'Hagan, 2013).

Late measure – Late eye-tracking measures give indications on later stages of cognitive processing as opposed to early measures which give insights into the initial stages (Roberts and Siyanova-Chanturia, 2013, p. 217; Conklin and Pellicer-Sánchez, 2016, p. 3).

Likert scale – A Likert scale is a type of attitudinal rating scale in which participants indicate whether they (dis)agree with a statement and the degree to which they (dis)agree with a statement. This happens through the selection of a number on a scale (Brace, 2004, p. 62).

Low-level factor – Low-level factors in a test element concern the salience of stimuli. Such factors include colour, contrast and motion to name but a few (Holsanova, 2014, p. 289). Eye-movements guided by low-level factors are related to bottom-up processing (Dyer and Pink, 2015, para. 4).

McGurk effect – The McGurk effect is a term coined by McGurk and MacDonald, and refers to a multisensory illusion that occurs in audiovisual speech perception under particular conditions (Tiippana, 2014, p. 1; MacDonald, 2018, p. 10). These conditions involve the simultaneous presentation of a voice that articulates a consonant with a face that articulates another consonant. Exposure to such incongruent stimuli results in the user experiencing the McGurk effect which means that the user hears a consonant that is different from those articulated in the stimuli (Tiippana, 2014, p. 1).

Multimodality – Multimodality refers to the nature of communicative situations or acts. In a very broad sense, communication depends on the interrelationship between its forms of expression or components in order to be effective (Bateman, Wildfeuer and Hiippala, 2017, pp. 7-8). Utterances, for instance, not only comprise linguistic content; they also rely on intonation, stress and tone of voice. Multimodal analyses are conducted in order to gain a better understanding of the choices that communicators make in relation to such forms of expression and the ways in which recipients perceive the information contained in multimodal messages (Holsanova, 2014, p. 293). The arrangement of forms of expression in a communicative situation is called *multimodal landscape* within the scope of this thesis. Recipients' perceptions of such multimodal landscapes are called *multimodal perception*.

Precision – Precision is an indicator of eye-tracking data validity (Tobii AB, 2018). It measures the spread or variance of a participant's actual gaze point. Precision is high when less measurement noise is detected in the gaze point of the participant (Sullivan, 2016). Precision is given in degrees and can be assessed for both spatial and temporal properties (Blignaut and Wium, 2014, pp. 67-68).

Relative gaze duration – The normalisation of the absolute gaze duration in which the accumulated duration of fixations of all selected participants on a chosen test element is relative to the total viewing time, is called the relative gaze duration (Bojko, 2009, p. 33; Tobii AB, 2014b, p. 66).

Sampling frequency – The sampling frequency is given in Hertz and shows the speed of an eye-tracking device and the rate at which samples are recorded per second (Holmqvist *et al.*, 2011, pp. 29-31).

Semantic differential scale – A semantic differential scale is a type of attitudinal rating scale in which participants can express their attitude towards specific aspects of a research phenomenon. A semantic differential scale includes opposite statements for these aspects at each end of a scale. Participants can indicate with which of these statements they (dis)agree and to what extent by selecting one of the numbers on the scale (Brace, 2004, p. 65).

Subtitles for the Deaf and Hard-of-Hearing (SDH) – SDH are subtitles that not only represent the dialogue; they also incorporate brief descriptions of sounds, and use colours and other techniques to convey phonological information to viewers (De Linde and Kay, 1999, pp. 11-12; Neves, 2008, p. 178; Díaz Cintas, 2013, p. 280; O'Hagan, 2013).

Subtitling country – A country in which subtitles are the primary means for making films in foreign languages available to the public. Countries that resort to dubbing instead are called "dubbing countries" (Koolstra, Peeters and Spinhof, 2002, p. 326; Gottlieb, 2004, pp. 83-84; Gambier, 2013, p. 46).

Telop – A telop is a type of intralingual text that is superimposed on a Japanese television programme (O'Hagan, 2010, p. 73). A telop does not need to consist of text alone; it can also comprise imagery (O'Hagan, 2013).

Thematic map – A thematic map visually represents how the (sub)themes identified in a data set relate to each other (Braun and Clarke, 2006, p. 89). The relations shown between a theme and its corresponding subthemes in a thematic map is a "thematic network".

Theme – A theme encapsulates a pattern or meaning a researcher has identified as prevalent in a data set (Braun and Clarke, 2006, p. 82). It is the "central organising concept" for a cluster of initial codes that cohere meaningfully and are distinct from other code cohorts (Braun and Clarke, 2013, p. 224). Themes that are used to structure a larger theme are called "subthemes" (Braun and Clarke, 2006, p. 92). Tentative themes are known as "candidate themes".

Track status – Prior to initiating the calibration process the track status of a participant is checked in the track status box. The track status box shows whether both eyes of a participant are detected and how far a participant's eyes are removed from the eye-tracker (Tobii AB, 2014b, pp. 34-35).

Velocity-Threshold Identification (I-VT) filter – The I-VT filter is a type of eye-movement classification in which the classification of eye-movements depends on the velocity of the eye's directional shifts. For this particular fixation filter eye-position samples are classified as a fixation when the calculated velocity is below the threshold (Olsen, 2012; Tobii AB, 2014b, pp. 52-55).

Vertex – A vertex is the point where two sides of a polygon intersect. The shape and size of an Area Of Interest (AOI) are defined by vertices (Tobii AB, 2014b, pp. 77-80).

NOTE ON SYSTEM OF ROMANISATION

This thesis uses Hepburn romanisation to represent Japanese in Roman script. The chosen system of romanisation follows the guidelines stipulated in the ALA-LC Romanisation Table for Japanese[1], which is a variant of Hepburn romanisation that is approved by the American Library Association (ALA) and the Library of Congress (LC).

Hepburn romanisation is applied throughout this thesis for Japanese references and terminology. In such cases, the Roman script is followed with a translation in square brackets. However, there are instances where the original Japanese script is presented instead of romanisation. Original Japanese script is shown when emphasis is put on the visual form of katakana characters, and when reference is made to sections of questionnaires or written entries in Japanese participants' open-ended data items.

[1] For further details, please see Library of Congress (2012).

ABSTRACT

Telops for Language Learning: Japanese Language Learners' Perceptions of Authentic Japanese Variety Shows and Implications for Their Use in the Classroom

Eline Christina Sikkema

Research on the use of leisure-oriented media products in foreign language learning is not a novelty. Building further on insights into the effects of audiovisual input on learners, recent studies have started to explore online learning behaviour.

This research employed an exploratory design to examine the perceptions of a Japanese variety show with intralingual text, known as *telops*, by Japanese Language Learners (JLLs) and native Japanese speakers through a multimodal transcript, eye-tracking technology, questionnaires, and field notes. Two main objectives underlie this study: (1) to gain insights into participants' multimodal perceptions and attitudes towards the use of such authentic material for language learning, and (2) to gain a better understanding of the distribution of participants' visual attention between stimuli.

Data from 43 JLLs and five native Japanese speakers were analysed. The JLLs were organised into a pre-exchange, exchange and post-exchange group while the native Japanese speakers functioned as the reference group. A thematic analysis was conducted on the open-ended questionnaire responses and Areas Of Interest (AOIs) were grouped to generate fixation-data.

The themes suggest that all learner groups feel that telops help them link the stimuli in the television programme although some difficulty was experienced with the amount and pace of telops in the pre-exchange and exchange groups. The eye-tracking results show that faces and telops gather the most visual attention from all participant groups. Less clear-cut trends in visual attention are detected when AOIs on telops are grouped according to the degree in which they resemble the corresponding dialogue. This thesis concludes with suggestions as to how such authentic material can complement Japanese language learning.

Keywords: multimodality, Japanese language learning, eye-tracking, visual attention, telops

1. INTRODUCTION

Technological advancements of the past few decades have made authentic AudioVisual (AV) materials in Foreign Languages (FLs) more accessible to users. This is especially true for leisure-oriented media products such as film and television. Nowadays, access to such materials is not confined to (foreign) television channels, DVDs, Blu-ray discs or the cinema. They are also readily available on the internet through, for example, streaming services and can be watched on demand via many devices such as laptops, smartphones and tablets. Moreover, in most cases these platforms further enhance the accessibility of these leisure-oriented media products through captions, subtitles or both. These developments have not only impacted everyday life. They have also opened up opportunities for FL teaching and learning.

Most contemporary language classrooms provide FL teachers and learners easy access to the beforementioned materials. Nowadays, classrooms are often equipped with devices such as PCs, projectors, and interactive whiteboards. It is also not uncommon for language classes to be supported by software applications like virtual learning environments. These technologies have broadened the range of FLs in which authentic AV materials can be accessed from class as they are connected to the internet. This means that materials other than those broadcast on the TV set or stored on DVDs and Blu-ray discs have become available to FL instruction as well. This thesis focuses on authentic television programmes in an FL that used to be mostly inaccessible to language learners in Europe prior to these developments: Japanese.

For some FLs the use of authentic AV materials can expose language learners to unfamiliar formats of television programme. The Japanese media landscape includes many such formats that presumably most European learners of the Japanese language are not accustomed to. Take, for example, Japanese game shows (O'Hagan, 2010). Not only the Japanese language is vastly different from the languages spoken in Europe; the composition of Japanese television broadcasts differs as well. This is due to one factor in particular: superimposed intralingual text.

Honmadekka!?TV (2013)

Figure 1.1 Example of Japanese variety show

Figure 1.1 includes a still image that shows a typical scene from a Japanese variety show. The white circles have been added to the still image in order to make the intralingual text more distinct. As can be seen in Figure 1.1, Japanese intralingual text comprises various shapes, sizes, colours and fonts. The still image illustrates that the use of more than one colour in a single strand of text is not unusual. Furthermore, the intralingual text is located in different regions of the screen and occupies a large surface area. The superimposed text in this particular scene also appears on screen in a variety of ways and remains visible for varying lengths of time.

While the design of Japanese intralingual text differs between genres of television programme and broadcasting companies (Sasamoto and Doherty, 2016; Sasamoto, O'Hagan and Doherty, 2016, p. 2) and may look similar to some designs employed in other Asian countries (Park, 2009), the type of text used in Japanese television broadcasts is rather specific to Japan. As noted, their presence in television programmes is more pronounced as compared to most types of intralingual text broadcast outside of Japan. They also cannot be turned on or off according to the needs of viewers as opposed to most counterparts employed in Europe (O'Hagan and Sasamoto, 2016, p. 33; Sasamoto and Doherty, 2016; Sasamoto, O'Hagan and Doherty, 2016, p. 2). In other words, Japanese intralingual text is integrated and therefore inextricable from the programme contents. This specificity has granted such captions their own name: *telop*.[2]

[2] The term *teroppu* [*telop*] is used for both its singular and plural forms in the Japanese language. However, most research written in English use "telops" to indicate the plural form. This thesis follows the English usage; "telop" for singular use and "telops" for plural use.

1.1 Background

Even though technological advancements have created favourable conditions in which Japanese television broadcasts can be incorporated into Japanese language classes, principles on the appropriate use of such materials in Japanese Language Teaching (JLT) are less established. Research on the pedagogic use of leisure-oriented media products that feature captions or subtitles is mostly available on European languages (Araújo, 2008). The same applies to studies that focus on incidental FL learning as a by-product of watching AV materials with subtitles (Koolstra and Beentjes, 1999; Kuppens, 2010; Gambier, 2015, p. 69). Anecdotes from FL learners who claim to have acquired an FL through such materials are also often related to languages spoken in Europe, especially English (Koolstra and Beentjes, 1999, p. 52; Danan, 2004, p. 67; Gambier, 2015, p. 68). Evidence on how this translates to a Japanese context, however, is minimal.

In the case of Japanese television programmes the FL is not the only challenging aspect for language learners. As noted, the Japanese media landscape employs formats of television programme that may require adjustment from learners. Furthermore, telops comprise a variety of scripts: *kanji* (i.e. Chinese characters), *hiragana* and *katakana* (i.e. two syllabaries), and the Roman alphabet (Masuji, 2013; Maree, 2015, p. 126). It is therefore difficult to ascertain to what extent suggestions on the use of captioned and subtitled media products in European languages can be adopted for JLT. Further research is thus needed on the appropriateness of Japanese television programmes to the Japanese language classroom. This thesis is an attempt at developing suggestions as to how such materials can be effectively used by teachers and learners alike.

My interest in this research topic followed from my own experience with learning English and Japanese. Growing up in a subtitling country like the Netherlands strengthened my belief that captioned and subtitled leisure-oriented media products can supplement FL learning; at least in case of English. Exposure to such materials in everyday life has become inevitable in the Netherlands. Generally speaking, foreign language materials are not dubbed which means that the original soundtrack remains intact and is often accompanied with captions or subtitles. Think, for example, of video games, film and television programmes. My experience with learning English is therefore similar to the beforementioned anecdotes. However, my study of the Japanese language did not involve the type of materials shown in Figure 1.1. Such leisure-oriented media products were also not available to the extent that English materials were at the time. I first saw telops while

on exchange in Japan for my BA degree. Seeing such multicoloured captions on television made me wonder whether there is a place for such AV materials in Japanese language learning. This curiosity inspired me to conduct this study.

1.2 Research context

The research context in which this thesis has developed is twofold. Firstly, it evolved from the established literature on three research phenomena in particular. These relate to the use of authentic materials for FL learning with a specific focus on television programmes that feature intralingual text, FL learners' online learning behaviour when using learning materials, and the role of multimodality in the viewer experience and viewer reception of authentic AV materials.

Studies have shown that research on the measurable effects of exposure to intralingual text on learners' linguistic skills in a FL is well underway. It has been maintained that the simultaneous presentation of aural and visual information in a FL through the use of intralingual text has the potential to improve language learners' listening skills and ability to segment or chunk speech, to recognise word boundaries and to retune perceptual processing (Mitterer and McQueen, 2009; Zarei and Rashvand, 2011; Montero Perez, Peters and Desmet, 2013; Charles and Trenkic, 2015). Although it has been argued that more empirical evidence is needed to rule out any doubts on the benefits of intralingual text, such empirically-based findings that have already been published suggest that the use of intralingual text for FL learning as a pedagogical tool looks promising (Danan, 2004; Díaz Cintas and Fernández Cruz, 2008, p. 205). Issues that need further investigation relate to the ways in which such authentic AV materials should be integrated into the language classroom or FL learning in general (Danan, 2004; Díaz Cintas and Fernández Cruz, 2008).

Recent studies have called for new approaches to research that build on the insights gained into the effects of intralingual text on FL learners in order to deepen our understanding of its appropriate use by FL teachers and learners for language learning purposes (Vanderplank, 2010; Montero Perez, Peters and Desmet, 2013, p. 38; Vanderplank, 2015; Vanderplank, 2016). One of these new approaches that has proven to yield valuable data uses an eye-tracking method to analyse FL learners' visual attention while watching videos with intralingual text (Winke, Gass and Sydorenko, 2013; Montero Perez, Peters and Desmet, 2015). However, despite the fertile ground provided through such an approach, the adoption of this new perspective has raised points for discussion as to how an eye-

tracking method can be appropriately used in Applied Linguistics considering its original application is situated in more cognitive and quantitative disciplines (Spinner, Gass and Behney, 2013, p. 390; Conklin and Pellicer-Sánchez, 2016; Stickler and Shi, 2017).

Recent studies on online learning behavior in Synchronous Computer-Mediated Communication (SCMC) and viewer reception of television programmes, animation films and movies have started to experiment with new methodological approaches to move beyond a purely cognitive or quantitative perspective (Redmond and Batty, 2015; Stickler, Smith and Shi, 2016; Stickler and Shi, 2017). These studies maintain that interdisciplinary or mixed-methods approaches help deepen our understanding of users' viewer experience and viewer reception of screen-based multimodal environments (Redmond and Batty, 2015; O'Rourke *et al.*, 2016; Stickler and Shi, 2017). Examples of such approaches combine eye-tracking technology with stimulated recall interviews, interpretive analysis of gaze paths and multimodal analysis of the information that is contained in the moving images and visually attended to by users (Batty, Perkins and Sita, 2015; Dyer and Pink, 2015; Stickler and Shi, 2015). Findings have shown that such frameworks not only help foreground FL learners' personal viewer experience and viewer reception; they also demonstrate that visual attention is strongly linked with the multimodal nature of AV materials (O'Rourke, 2012; Dyer and Pink, 2015).

Secondly, this thesis developed as part of a larger project called *Impact Caption for Everyone (ICE) – Reception Metrics for Creative Subtitling on TV*.[3] This project is run by Dr Ryoko Sasamoto and Prof. Minako O'Hagan with Dr Stephen Doherty. The ICE project aims to gain insights into the multimodal nature of authentic Japanese variety shows that feature telops, to understand the viewer experience and viewer reception of such media products by disparate groups of users, and to create recommendations on the optimal use of telops based on the collated empirical data. This project combines multimodal analysis, eye-tracking technology and self-reports to collect evidence on its multimedia corpus and approaches the data through a relevance-theoretic framework.[4]

Studies that have already been completed as part of the ICE project focused on the use of telops in Japanese television programmes (Sasamoto and Doherty, 2016), the use of telops in Japanese variety shows to frame humour (O'Hagan, 2010) and to hijack viewers'

[3] For an introduction to the ICE project, please see Dublin City University (2016).
[4] For an introduction to Relevance Theory, please see Sperber and Wilson (1995).

interpretation processes (Sasamoto, 2014), the multimodal nature of an excerpt from the multimedia corpus (O'Hagan and Sasamoto, 2014; Sasamoto, O'Hagan and Doherty, 2016), and the viewer reception of excerpts from the multimedia corpus by native Japanese speakers (O'Hagan and Sasamoto, 2014; O'Hagan and Sasamoto, 2016). Findings from these studies showed that the use of telops has become a widespread phenomenon in the Japanese media landscape and that it is strongly associated with Japanese variety shows. The results also specify that telops play an essential role within the viewer reception of users and the multimodal nature of Japanese television programmes.

The ICE project combined with the beforementioned studies on intralingual text, FL learning and eye-tracking of moving images demonstrate that research has shifted to a vast repertoire of newly introduced topics and has started to test and improve innovative approaches to accommodate proper analysis of these new avenues of research. This thesis is an attempt at connecting these fresh perspectives.

1.3 Scope of thesis

This thesis intends to contribute to those areas of research that were identified as research gaps in the research context. These comprise the use of authentic AV materials that feature intralingual text in the language classroom, the analysis of FL learners' visual attention while watching such authentic materials, the application of an eye-tracking method to examine FL learners' visual attention, and the incorporation of multimodal analysis to examine moving images in relation to users' viewer experience and viewer reception. While this makes for a long list of research interests, this thesis has followed an interdisciplinary pathway that allowed for each of these topics to be taken up within its scope. This was achieved by taking a closer look at one issue in particular which appeared to be at the crossroads of these research areas.

This point of discussion revolves around the appropriateness of authentic AV materials with intralingual text to FL learners with differing backgrounds, linguistic abilities and motivations for learning an FL (Montero Perez, Peters and Desmet, 2013, p. 22; McMeekin, 2015). Concerns on the potential inappropriateness of such materials have been raised in relation to the leisure-oriented nature of television programmes (Vanderplank, 2015, p. 23), and the proficiency level that is required to understand such materials (Gambier, 2015, p. 68). Conversely, it has also been argued that authentic materials can cater for any stage when used appropriately (Danan, 2004, pp. 74-75;

Gilmore, 2007; Gilmore, 2011; Tomlinson, 2012). Studies that support this view maintain that input in authentic materials is highly contextualised which provides FL learners with opportunities to pick up on information that fits their capabilities and personal learning journeys. This has in large part to do with the multimodal nature of authentic materials. Now that is has been recognised in multimodal research that (AV) texts create meaning through a range of forms of expression rather than just linguistic input (Gambier, 2013, p. 47; Jewitt, 2014), it can be argued that FL learners can avail of both linguistic and non-linguistic information for their understanding of authentic AV materials with intralingual text.

This point of discussion is further investigated within the scope of this thesis. It is suggested here that a multimodal approach to FL learning, or the exploration of the multimodal nature of FL learning, can contribute to a deeper understanding of the beforementioned research areas. This thesis involves an exploratory study of FL learners' perceptions of an authentic Japanese variety show. An eye-tracking method is incorporated within this study's research design to inspect these perceptions further with the help of recorded eye-movements. The FL learners who are the focus of this research are at different stages of their language study and comprise a group of users that has not yet been researched by the ICE project: Japanese Language Learners (JLLs). This thesis differs from previous ICE research in that it does not take a relevance-theoretic approach. Rather, a mixed-methods approach is taken in which the emphasis is put on multimodality. This thesis concludes with suggestions as to how authentic Japanese variety shows can complement Japanese language learning.

1.3.1 Objectives and research questions

This thesis aims to make contributions to three communities in particular. These are teachers and learners of the Japanese language on the one hand, and researchers on the other hand. This study has two objectives. First, it aims to provide teachers of the Japanese language with empirically-based suggestions on the appropriateness of authentic Japanese variety shows to JLT. Second, it aims to determine the appropriateness of an exploratory mixed-methods approach to examine FL learners' perceptions of authentic AV materials. This thesis set out to answer the following three Research Questions (RQs) to meet these objectives:

RQ1 What are Japanese language learners' perceptions and attitudes towards authentic Japanese variety shows?

RQ2 To what extent does eye-tracking data allow us to interrogate and further examine those perceptions?

RQ3 What principles can be derived for the use of authentic Japanese variety shows in the Japanese language classroom?

1.4 Thesis structure

Chapter 2 delves deeper into the established literature that formed the research context of this thesis. It not only explains these and other studies in more detail; it also describes important concepts and points for discussion that carry implications for this study. Chapter 3 introduces the conceptual framework of this thesis. It explains the research design that was developed in response to the research questions and objectives, and describes the methodological approach of this study. Chapter 4 expands on this framework. It focuses on the methods that were used for the collection and analysis of the empirical data. Chapter 5 gives an in-depth account of the multimodal analysis of the video material that was chosen for the purposes of this study in preparation of the empirical data analyses. Chapter 6 reports empirical findings on JLLs' perceptions of the video material. Chapter 7 builds on these findings and presents results on JLLs' visual attention while watching the video material. Chapter 8 concludes this thesis. It answers the research questions, reflects on the ways in which this study was operationalised and conducted, and gives directions for further research.

2. LITERATURE REVIEW

Chapter 2 comprises the literature review of this thesis. It describes the research context of this study in more detail, and provides an overview of the main concepts and points of discussion from the established literature. The purpose of this chapter is to illustrate how this thesis fits into research on three phenomena. These are the use of authentic materials for FL learning with a specific focus on television programmes that feature intralingual text, FL learners' online learning behaviour when using learning materials, and the role of multimodality in the viewer experience and viewer reception of authentic AV materials.

Section 2.1 starts with a definition of authentic language learning materials. It thereafter takes a closer look at the nature of authentic AV texts. It describes the characteristics of such materials and discusses the implications these carry for FL learning. Section 2.2 then narrows down the discussion to Japanese television programmes. This section not only gives an account of the development of telops in the Japanese media landscape; it also provides an in-depth explanation of its use in authentic Japanese variety shows. Section 2.3 then gives an overview of mixed-methods approaches that focus on the research topic. It first reviews research on moving images before taking a closer look at online learning behaviour of FL learners.

2.1 Authentic materials and foreign language learning

2.1.1 Defining authenticity

Generally speaking, research on materials development distinguishes two types of language learning material. These are authentic and contrived materials. The original purpose of FL learning materials is often used as the main criterion by researchers to differentiate between the two (Gilmore, 2007, p. 98). The former describes materials that contain "language that has been produced to communicate a genuine message" (Gilmore, 2011, p. 791). These include television broadcasts, film and newspapers to name but a few. The latter refers to materials which comprise "language contrived by material writers to display particular lexicogrammatical items" (2011, p. 791). Textbooks are a typical example of such materials. These definitions demonstrate that the purpose of authentic materials is to communicate while contrived materials are created with the intention to use them for FL learning purposes (Gilmore, 2007, p. 98; Gilmore, 2011; Tomlinson, 2012, p. 162).

Both types of language learning material are useful to learning an FL (Gilmore, 2007, p. 98). However, considering that authentic and contrived materials are created for different reasons, they do not support FL learning in the same way. Contrived materials primarily focus on the linguistic aspects of an FL and are devised to deliberately draw the language learner's attention to such features. Authentic materials, on the other hand, put the emphasis on language use that is closer to the "real thing" and facilitates the integration of language in practice into FL instruction. It has therefore been maintained that authentic materials can develop a broader range of learners' communicative competencies as opposed to contrived materials as they provide FL learners with opportunities to consolidate (socio-)pragmatic competencies that add to a learner's knowledge of lexicogrammatical items (Gilmore, 2011; Tomlinson, 2012, p. 161).

The type of authentic language learning material that is central to this thesis, AV materials, is a topic of particular discussion in research on materials development (Gilmore, 2007, p. 103). The interaction between communicative competences of JLLs and the pedagogic use of Japanese AV materials has been described by The Japan Foundation (JF) in the JF Standard Tree (The Japan Foundation, 2012, p. 5). The JF Standard Tree depicts the relationship between communicative language competences and communicative language activities in a tree diagram where receptive activities, represented in the tree's branches, develop from underlying roots of communicative language competences (2012, p. 9). These receptive activities include watching television and film. While the JF Standard Tree indicates that the competences are the foundation for the activities, it can be argued that the reverse is also at play in the sense that such activities can strengthen competencies as discussed by the beforementioned studies on language learning materials.

2.1.2 Audiovisual materials in the language classroom

Research on materials development and the JF Standard Tree demonstrate that activities involving authentic AV materials such as television and film are established learning tasks. Nonetheless, recent research has started experimenting with new ways in which such materials can be integrated into the language classroom. These innovative applications involve the use of AudioVisual Translation (AVT). Learning activities that involve AVT are not only receptive by nature; they are also mediating (McLoughlin and Lertola, 2014, p. 74). Of such activities, interlingual subtitling in particular has been discussed by researchers (Williams and Thorne, 2000; Díaz Cintas, 2008; McLoughlin and Lertola, 2014). While the current research project takes a receptive activity as its research topic

these studies have provided important insights that need to be considered for any type of learning activity involving AV materials. These considerations relate to the nature of AV texts: multimodality.

Multimodality is a concept that refers to the nature of communication (Van Leeuwen, 2011, p. 668). It concerns the idea that communicative situations or acts rely on an interrelationship between a number of forms of expression in order to be effective (Bateman, Wildfeuer and Hiippala, 2017, pp. 7-8). Multimodality can be found anywhere but it has often been discussed in relation to AV texts as such materials provide some of the most conspicuous examples of multimodal messages. AV materials not only require users to process aural and visual information simultaneously; multimodal messages from AV texts also employ both verbal and nonverbal stimuli (Remael, 2010, p. 13). Take, for example, film. Films make use of imagery, soundtracks and dialogues to name but a few. The chosen combination of such stimuli shapes the filmic experience of users. Research on AVT categorises such types of input as follows:

1. Audio-verbal
2. Audio-nonverbal
3. Visual-verbal
4. Visual-nonverbal

   (Delabastita, 1989, p. 199)

Recent research on interlingual subtitling has recognised that these four forms of expression cannot be treated separately when dealing with multimodal messages and that the verbal component of a multimodal message should not take precedence over other types of input when subtitling (Zabalbeascoa, 2008; Sakellariou, 2012; Kaindl, 2013). McLoughlin and Lertola (2014, p. 72) argue that this multimodal nature of AV materials makes the traditional four-skill model of receptive and productive skills (i.e. reading, listening, speaking and writing) obsolete. As noted, AV materials comprise types of input across four forms of expression that are interdependent. McLoughlin and Lertola (2014, p. 72) explain that it is therefore necessary to have FL learners develop additional skills that will help them perceive multimodal messages. They call these skills *AV-watching* and *AV-listening*.

2.1.3 The case of television programmes with intralingual text

The absence of a language instructive nature in authentic materials and the emergence of a new set of skills with which to approach AV texts bring both opportunities and challenges to the integration of authentic AV materials into the language classroom. Pedagogic use of television programmes that feature intralingual text in particular has raised many points of discussion among researchers. Doubts regarding the appropriateness of such materials to FL learning concern mainly two issues.

First, research is divided on the implications of the leisure-oriented nature of television and the associations it has with entertainment for FL learning. Some studies argue that FL learners approach television programmes with a mind-set that is not focused on FL learning and that this could hamper language acquisition (Vanderplank, 2010, p. 12; 2015, p. 23; 2016, p. 64). Conversely, research has also endorsed television programmes for that very reason; the enjoyment and entertainment value of materials with intralingual text (Kothari *et al.*, 2002; Kothari, Pandey and Chudgar, 2004). Such studies argue that the mix of education and entertainment reduces the importance of a motivation to learn, and creates a low-anxiety setting that enables lifelong daily practice through repeated exposure and immediate feedback from intralingual text.

Second, the proficiency level needed to understand AV materials with intralingual text is another point of discussion. Some studies report that knowledge of the FL is required before learners can benefit from exposure to such materials (Vanderplank, 1988, p. 280; Gambier, 2015, p. 68; McMeekin, 2015). It has been argued that when taking a closer look at the material itself, it is not only the sheer volume of input that poses challenges to learners; dialogues in television programmes are rather different from those in contrived materials as they may include haphazard or inconsistent utterances, fast-paced and sometimes unclear speech, overlapping voices and disorderly turn-taking between speakers (McMeekin, 2015, p. 213). This implies that it has been called into question whether such materials are appropriate to all learners; especially those with lower or beginner proficiency levels in an FL.

Nonetheless, it has also been argued that familiarity with the formats of television programme eases the linguistic encounter for language learners as they do not need to grow accustomed to the medium itself and can fully focus on the programme contents (Mishan, 2005, p. 132). Furthermore, studies have maintained that the multimodal nature of

authentic AV materials with intralingual text can prove to be useful in an FL learning context as it includes semiotic redundancy (De Bruycker and d'Ydewalle, 2003, p. 672; McLoughlin and Lertola, 2014, p. 73). That is to say, such materials comprise information that is represented in more than one type of input or form of expression as is the case with intralingual text that visually represents the spoken dialogue.

The Dual Coding Theory of Paivio is often discussed in relation to such semiotic redundancy as it describes how the use of both verbal and nonverbal input can be helpful to users. This theory assumes that human cognition deals with verbal and nonverbal input through two independent yet interconnected subsystems (Paivio, 2008, para. 4). These subsystems can function on their own, in parallel or trigger activity in each other (2008, para. 5). This means that linguistic complexity or the difficulty of verbal input does not have to be the main criterion for assessing the difficulty level of an AV text, as its multimodal nature can scaffold FL learners' comprehension when they experience difficulty with particular types of input (Mishan, 2005, p. 137; Gilmore, 2011, p. 802). The simultaneous presentation of verbal and nonverbal stimuli (i.e. intralingual text, imagery and speech) thus helps learners with processing unknown FL input (Díaz Cintas and Fernández Cruz, 2008, pp. 207-208).

Further to the Dual Coding Theory, Krashen's Comprehension Hypothesis provides another explanation for the usefulness of video materials to FL learners. The Comprehension Hypothesis states that learners need to understand FL input in order to acquire it and that most language acquisition is the result of a subconscious process (Krashen, 2008, pp. 179-180). Díaz Cintas and Fernández Cruz (2008, p. 203) argue that, based on this hypothesis, video materials can provide FL learners with ample opportunities for language acquisition as they comprise many types of nonverbal input that learners can understand in addition to the linguistic contents.

Similar arguments have been made in recent research on materials development. According to such studies, the highly contextualised nature and "richness" of these materials can facilitate language acquisition at any stage (Mishan, 2005, p. 138; Gilmore, 2007, p. 111; Gilmore, 2011, pp. 803-804). Such studies have argued that authentic materials can benefit all language learners when accompanied with the appropriate learning goals, tasks and teacher's guidance (Gilmore, 2007; Gilmore, 2011; Tomlinson, 2012). Such learning goals, tasks and guidance are expected to alleviate learners from

feeling overwhelmed by the constant stream of input. Awareness-raising activities that allow learners to explore an FL have been proposed as suitable learning tasks to authentic materials as they do not approach the contents in an overly form-focused, informative or instructive manner (Gilmore, 2007; Tomlinson, 2012).

2.2 Japanese television programmes and telop

2.2.1 Defining telop

It was argued at the start of this thesis, in *Figure 1.1 Example of Japanese variety show*, that the Japanese media landscape includes formats of television programme that are presumably unfamiliar to most European JLLs. It was also explained that the type of intralingual text employed in such television broadcasts, telop, is integrated and designed in a way that is specific to Japan. The functions and pronounced, multicoloured appearance of telops in Japanese television programmes have often been discussed in relation to norm-defying subtitling practices. The established literature has made comparisons between telops and three types of text in particular: *fansubs* (O'Hagan, 2013), *impact captions* (Sasamoto and Doherty, 2016), and *authorial titles* (Sasamoto, 2014; Sasamoto and Doherty, 2016). The latter has most commonly been considered in relation to the British television series *Sherlock* (2010-2017). While such captions and subtitles are indeed examples of "abusive" or "transformative" subtitling practices (Nornes, 1999; Nornes, 2007; Pérez-González, 2012) it is important to further examine the characteristics of telops in order to gain a better understanding of their specificity.

Generally speaking, researchers use a number of parameters to describe captions and subtitles. These dimensions have been organised into more than one classification scheme. The framework presented by Bartoll (2004), for example, includes technical and linguistic parameters. The classification scheme proposed by Díaz Cintas and Remael (2014, pp. 13-25), on the other hand, uses five criteria to differentiate between types of captions and subtitles: linguistic parameters, time available for preparation, technical parameters, methods of projecting subtitles, and distribution format. When taking a closer look at these frameworks it becomes evident that telops hold an unusual position within these classification schemes as their characteristics cut across caption and subtitle varieties.

Díaz Cintas and Remael (2014, p. 14) include five types of intralingual subtitles in their classification scheme: Subtitles for the Deaf and Hard-of-Hearing (SDH), subtitles for

language learning purposes, subtitles for karaoke effect, subtitles for dialects of the same language, and subtitles for notices and announcements. As noted previously, telops are used in authentic materials from which they cannot be removed at a user's discretion. In other words, they are an open type of text on screen (Díaz Cintas, 2013, p. 279; Díaz Cintas and Remael, 2014, p. 21). Telops are intended for hearing users and the design of such intralingual text is largely based on the intuitions of the production team (Sasamoto and Doherty, 2016; Sasamoto, O'Hagan and Doherty, 2016, p. 2). Telops include bold and bright colours and can be superimposed on different areas of the screen as was shown in *Figure 1.1 Example of Japanese variety show*. Furthermore, as will be discussed in the following subsection, telops perform many functions within a Japanese television programme. This list of characteristics demonstrates that telops do not fit in any of the beforementioned categories despite the commonalities they share with some of them.

Researchers have given various names to telops as a result of their unusual position in such classification schemes. These include Japanese terms such as *jimaku sūpā* or *jimaku* [*subtitle*] for short (Shitara, 2008, p. 29; Sakamoto, 2009a)[5], *sūpā'inpōzu* [*subtitle*] (Shitara, 2005, p. 14), *moji teroppu* [*telop*] (Shitara, 2006; Shitara, 2009; Shitara, 2012), and *teroppu* [*telop*] (Matsukawa, Miyata and Ueda, 2009; Koga, 2013; Masuji, 2013; Fukunaga, Katayama and Shoyama, 2014a; Fukunaga, Katayama and Shoyama, 2014b). The established literature written in English calls them *Open Caption Telop* (OCT) (O'Hagan, 2010), *impact caption* (Sasamoto and Doherty, 2016), and *TELOP* (Dwyer, 2015, para. 2). Some studies employ one or more of these terms interchangeably with the name *telop* (Sasamoto, 2014; O'Hagan and Sasamoto, 2016; Sasamoto, O'Hagan and Doherty, 2016). The word telop is a contraction of *television opaque projector,* which is a device that allowed for direct insertion of text and images into television broadcasts before the advent of computerised systems (Shiota, 2003, p. 63; Kawabata, 2006, p. 211; Suto, 2008, p. 13; Sakamoto, 2009b; O'Hagan, 2010, p. 73; O'Hagan, 2013; Sasamoto, O'Hagan and Doherty, 2016, p. 2). It is considered to be the most common term used in both English and Japanese, and has therefore been adopted within the scope of this thesis.

2.2.2 Overview of telops

Nowadays telops are deeply-rooted in the Japanese media landscape. They not only appear in the genre of television programme they are most strongly associated with, i.e. Japanese

[5] For the original source, please refer to Sakamoto, M. (1999) 'Hanransuru jimaku bangumi no kōzai [The good and bad points of the plethora of subtitled programmes]', *GALAC*, April, pp. 30-35.

variety shows, they are also integrated into news programmes and drama series (Kawabata, 2006, pp. 211-212; Matsukawa, Miyata and Ueda, 2009, p. 194; Yoshizawa, 2015; Sasamoto and Doherty, 2016). Nonetheless, telops have not always been this widespread. Their elaborate designs and manifold applications are the result of gradual changes that have occurred over the past few decades. The functions and appearance of telops in current Japanese television programmes are a manifestation of this development.

Shitara (2011) discusses the development of telops for the period 1960-2000 in relation to 29 Japanese variety shows of the Japan Broadcasting Corporation (NHK). According to Shitara, this time span has seen several shifts in the usages, designs and amount of telops. At their onset, telops played a minor role in Japanese variety shows as they were only used to give indications on the composition of the programme contents. They were handwritten rather than projected onto the screen, which Shitara (2011, p. 4) largely ascribes to the production circumstances at the time. The role of telops changed in the period between 1970 and 1990, when they started to provide more detailed information, and were used to stage effects, to "hook" viewers, and to strengthen the individuality of Japanese variety shows (Sakamoto, 2009a; Shitara, 2011, pp. 4-7). This means that telops became prominent from the 1990s onwards in particular (Shitara, 2011, p. 1).

The 1990s proved to be an important time period in the development of telops as it brought the technological advancements that enabled the most conspicuous changes. These technological developments led to the appearance of those telops that are prevalent in current Japanese variety shows (Shiota, 2003, p. 70; Shitara, 2011). From that moment onwards telops became increasingly decorative and were employed to give users the impression that they are addressed directly by the programme contents (Shitara, 2011, p. 6; Maree, 2013, p. 113). These telops are also known as *impact captions* as it has been argued that such telops are designed to impact users (O'Hagan, 2010; Sasamoto, 2014). Generally speaking, the term *impact caption* is used by researchers when particular emphasis is put on this "impactful" nature of telops.

The amount of telops increased in Japanese news programmes as a result of their popularity in Japanese variety shows (Matsukawa, Miyata and Ueda, 2009, p. 194). Japanese news programmes adopted an approach that combined informative and entertaining elements, known as *infotainment* news, to draw viewers in a way similar to

that of Japanese variety shows (Kawabata, 2006, p. 210; Matsukawa, Miyata and Ueda, 2009, p. 194).

This overview demonstrates that the development of telops over the past few decades is mostly characterised by an increase in the amount and detail of linguistic contents, diversification of functions, and the embellishment of telop designs (Shitara, 2011, p. 7). Telops in current Japanese variety shows have diverted far from their original application. Nowadays they not only represent dialogues of people appearing on a television programme; they also include metalinguistic commentary (Shitara, 2006, p. 39; Arai, 2014; Sasamoto, 2014). Figure 2.1 shows examples of such commentary. White circles are shown around the telops in question.

Honmadekka!?TV (2013)

Gyōretsu no Dekiru Hōritsu Sōdanjo (2014)

Figure 2.1 Examples of telops that provide metalinguistic commentary

The still image on the left in Figure 2.1 depicts a scene where the host of the variety show makes a joke regarding two of his guests. He makes a comment that they, i.e. the people shown in the still image, look like twins even though they are not related. He then gives the person on the left a turn by saying “Dōzo otōto” [Go ahead, little brother]. The words otōto [little brother] and ani [big brother] thereafter appear on screen as new designations for these guests. This particular scene illustrates that telops can be used in response to a joke rather than spoken dialogue. The still image on the right in Figure 2.1 shows a scene from another Japanese variety show. The word chūmoku [attention] is superimposed on the screen. This telop is used to draw viewers’ attention to a particular person visible on screen in preparation of a joke that is made soon after the depicted scene. Figure 2.1 shows that telops from the 1990s onwards behave according to the “intention” or “scheme” worked out by the production team (Kawabata, 2006, p. 212; Sakamoto, 2009a). It also suggests

that telops can vary greatly in the degree to which they depend on dialogues and imagery of current Japanese television programmes (Shitara, 2005, p. 17; Arai, 2014).

Recent research on telops has proposed a number of taxonomies of telop for the classification of its functions and designs. The variation in classification schemes can largely be ascribed to genres of Japanese television programme and analytic approaches of researchers. Generally speaking, three taxonomies of telop exist in the established literature. The taxonomy of Kimura *et al.* (2000)[6] is most often consulted by researchers (Kawabata, 2006; O'Hagan, 2010; Sasamoto and Doherty, 2016). This taxonomy is based on telops encountered in ten Japanese variety shows that were voted top 10 for their frequent use of telops by 183 survey respondents. It provides a common framework on the functions of telop and its forms in terms of special effects and typographical features. A translated and slightly adapted version of this taxonomy has appeared in a study of O'Hagan (2010, p. 76). The taxonomy of Kimura *et al.* (2000) informed the second classification scheme, defined by Kawabata (2006, p. 217), on telops in Japanese news programmes. This taxonomy is based on telops encountered in six news programmes and is also arranged according to function and form. The third framework is the taxonomy of functionality created by Shiota (2003, p. 72). It does not specify the genre of television programme or the forms of telop; rather, it categorises telops from a relevance-theoretic perspective. It has therefore often been discussed in relation to Relevance Theory and impact captions (O'Hagan, 2010; Sasamoto, 2014; Sasamoto and Doherty, 2016).

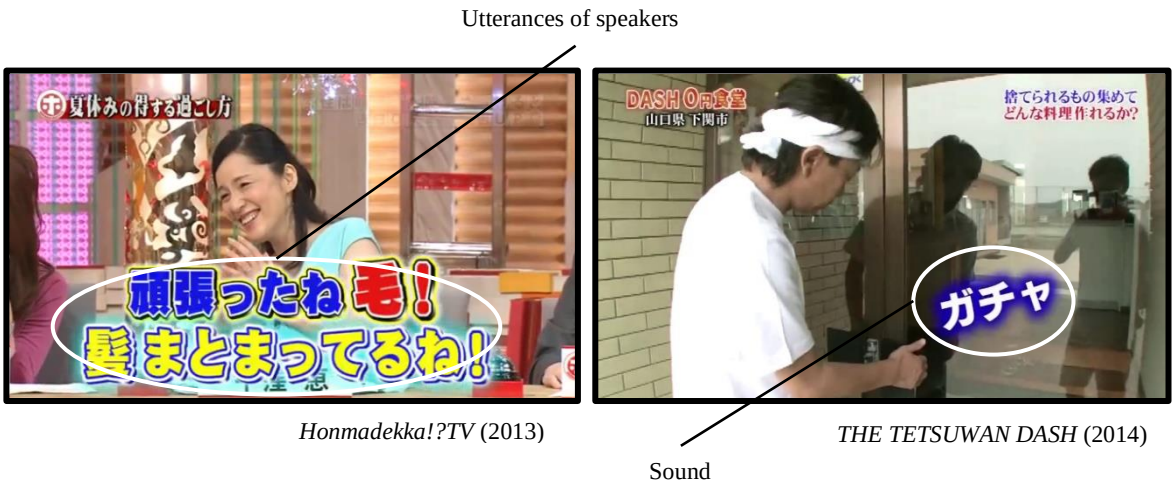
For the purposes of this study, a closer look is taken at the taxonomy developed by Kimura *et al.* (2000). It provides the most comprehensive overview of telops in Japanese variety shows and details both their functions and forms. The taxonomy included in Table 2.1 was translated by the researcher from the original version as cited in Sakamoto (2009b). As can be seen in Table 2.1, the taxonomy categorises telops according to three functions which comprise a total of ten varieties. It also includes eight types of form. According to Kimura *et al.* (2000), telops are a combination of varieties from the first typology of function and the second typology of form.

[6] For the original source, please refer to Kimura, T., Hosoi, A., Suzuki, E., Watabe, K., Koizumi, A., Kawamura, F., Ōsawa, Y., Katō, Y. and Honda, S. (2000) 'Terebi gamen ni odoru "mojitachi" no seitaigaku [The ecology of the words that dance on the TV screen]', *GALAC*, June, pp. 36-39.

Table 2.1 Taxonomy of telops (translated and adapted from Sakamoto, 2009b)

Typology	Type	Variety
Function	**1.** Direct rendering of dialogue	**1.1** Utterance of speaker
		1.2 Narration
		1.3 Sound (e.g. sound effect or natural sound)
	2. Facilitation of comprehension	**2.1** Circumstance
		2.2 Psychological state
		2.3 Time lapse
		2.4 Theme or title
	3. Bridging of scene change	**3.1** Connective or conjunction
		3.2 Foreshadowing
		3.3 Screen filled with text
Form	**4.** Standard	-
	5. Expansion or contraction	-
	6. Kanji character	-
	7. Background colour	-
	8. Punctuation	-
	9. Special effect	-
	10. Picture or emoticon	-
	11. Other	-

The first type of function included in Table 2.1, direct rendering of dialogue, comprises two varieties of telop that are intralingual renditions of audio-verbal forms of expression. These audio-verbal types of input are utterances of speakers and narrations. The first variety of telops is linked to utterances of people appearing on the show and includes speech that is both clearly audible and difficult to hear. The second variety, called *narration telops* within the scope of this thesis, are linked to speech of a narrator who is not visible on screen. The third variety, sound, comprises telops that visually represent audible sounds or speech from a source that is not visible on screen. Figure 2.2 gives examples of two of these varieties: utterance of speaker, and sound.



Honmadekka!?TV (2013) *THE TETSUWAN DASH* (2014)

Figure 2.2 Examples of telops that directly render dialogue

The still image on the left in Figure 2.2 shows telops that visually represent utterances of speakers. The telop displayed on the right shows an onomatopoeic expression. It reads as

"gacha" and refers to the sound of someone rattling a doorknob. This coincides with the scene that is depicted in the still image as the person tries to open a door but finds out that it is locked.

The second type of function, facilitation of comprehension, supports users' understanding of the programme contents. Telops included in this category represent four types of information. The variety of circumstance provides information on the setting and describes the context for scenes during which no speech is audible. The variety of psychological state includes comments on mental states of speakers and narrators. Such telops usually comprise speculations of the production team which do not always support the imagery. The variety of time lapse gives numerical indications of the passage of time. These include, for example, countdowns to future events. Themes or titles comprise headings that display topics to users. Three of these varieties are shown in Figure 2.3. These are psychological state, time lapse, and theme or title.

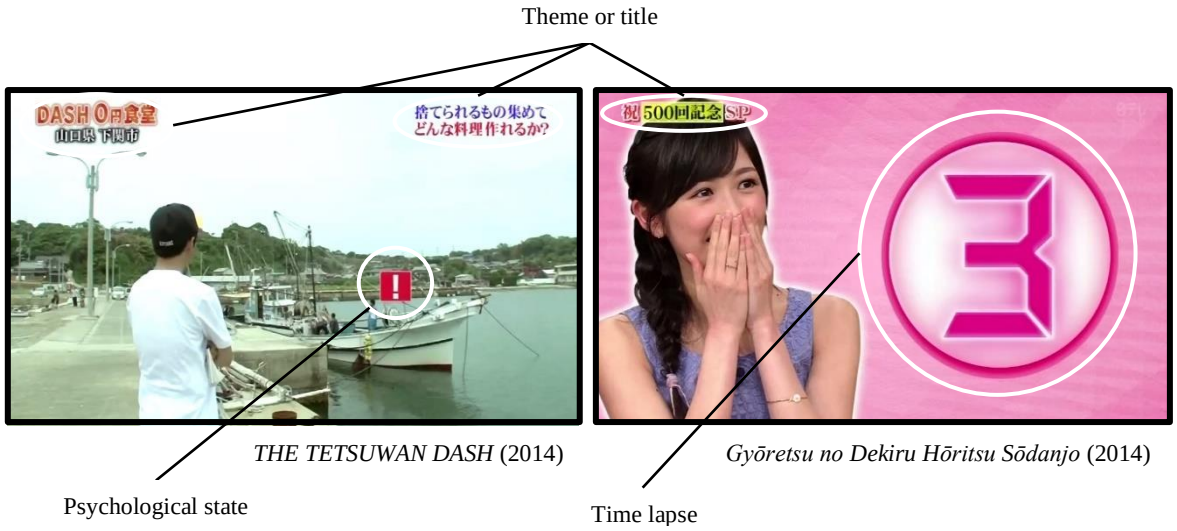


Figure 2.3 Examples of telops that facilitate comprehension

The telop that displays a psychological state is shown in the still image on the left in Figure 2.3. It comprises an exclamation mark, which is used to give viewers an indication of the mental state of the fisherman on the boat. He is depicted as being surprised by the unexpected visit of the camera crew. The time lapse in the still image on the right shows a countdown from three to one. The title of the programme or programme item is shown in the top left corner of both still images. A theme that represents the topic of conversation, visible in the still image on the left, is displayed in the top right corner.

Telops categorised in the third type of function, bridging of scene change, facilitate shifts in scenes. This happens in mostly three ways. The first two varieties included in Table 2.1 (i.e. connective or conjunction, and foreshadowing) are employed to maintain users' attention and interest while changing scenes. The variety of screen filled with text, on the other hand, is not necessarily aimed at sparking users' curiosity during scene changes. Such telops can, for example, be used to summarise points that are made by people appearing on the television programme. This is shown on the right in Figure 2.4.

Honmadekka!?TV (2013)

Figure 2.4 Examples of telops that bridge scene change

Table 2.1 shows that the typology of form includes a type with no effects (i.e. standard) and seven types that each relate to a special effect or typographical feature. The type of expansion or contraction relates to telop designs in which word sizes change. The type of kanji character refers to the use of kanji characters in telops to emphasise words. Background colours are also used for emphasis and indicate the use of background colours to stress particular text. Punctuation comprises designs in which punctuation marks or symbols are used. They are incorporated into telops to facilitate a better understanding of the imagery. The examples given by Kimura *et al.* (2000) are exclamation marks, question marks and arrows. The type of special effect includes any effects added to telops such as shaking and shining. According to Kimura *et al.* (2000), pictures and emoticons are only used for telops that directly render dialogue. Any other categories of form that are not included in Table 2.1 are assigned to other.

Further to the taxonomies of telop, recent research has also proposed two typologies of particular types of telop. One of these classification schemes is the typology of utterance of speaker (Shitara, 2006, p. 39). The second framework is the typology of narration telop

(Shitara, 2005, p. 16). Table 2.1 shows that these typologies thus relate to the first two varieties of telops that directly render dialogue. As noted, these varieties comprise intralingual renditions of audio-verbal forms of expression. The proposed typologies of Shitara therefore take a closer look at the relationship between such telops and the audio-verbal types of input they represent.

Shitara (2006) created the typology of utterance of speaker based on telops that were encountered in 18 Japanese variety shows. This framework organises telops into two categories. The first category comprises telops that highly resemble corresponding utterances of speakers. These telops are described to be identical to corresponding utterances or to encompass identical representations of parts of such utterances (2006, pp. 37-38). The second category consists of telops that do not highly resemble corresponding utterances of speakers. Telops in this category are further divided into telops that either include or do not include a recurring expression. Shitara (2006, p. 51) argues that the fixed component in a telop provides the context of an utterance, as it reiterates information that has already been mentioned, while the non-fixed component summarises the utterance of a speaker. Shitara (2006, p. 52) further explains that telops without recurring expressions are different from the first category as they include indications on the viewpoint of the production team.

Shitara (2005) devised the typology of narration telop based on telops from five informative and three documentary-like television programmes. The overall categorisation of narration telops is similar to that of utterances of speakers. This classification scheme divides narration telops into telops that are identical to the corresponding narrations and those that are not. Four categories were detected for the latter: omission, recast, addition and other. Omission includes telops from which parts of narrations are omitted. They are condensed representations in which the core of corresponding narrations are kept intact (2005, p. 17). Recast encompasses telops that use synonyms or similar expressions as compared to the words used in narrations which may result in nuance differences (2005, p. 20). Addition refers to telops that contain additional information in order to disambiguate corresponding narrations. These often include recurring words and expressions that are known from the context and have therefore been left out from the narrations (2005, p. 21). The category of other includes telops that showcase alterations that do not fit the other three categories. These include different word orders and complete reformulations (2005, p. 22).

These taxonomies and typologies of telop illustrate that telops perform a range of translational and non-translational functions within Japanese television programmes. They are translational in the sense that they provide intralingual renditions of audio-verbal forms of expression and non-translational in that they visually represent metalinguistic commentary from the production team. Telops that are used for translational purposes correspond to the contents of utterances of speakers and narrations to varying degrees. Telops that include metalinguistic commentary from the production team include comments on the programme contents and speculations on speakers appearing on the show (e.g. psychological state). Further to these functions, telops can appear in Japanese television programmes in manifold ways by means of special effects and typographical features.

Recent research on telops has described the implications of these functions and forms for users' viewer experience and viewer reception of authentic Japanese variety shows. Such studies argue that production teams make use of telops to guide users' interpretation processes of the programme contents. This perspective on telops followed from a study conducted by Park (2009) on impact captions in South Korean variety and talk-shows. Park demonstrated that production teams can exploit the form of impact captions to incorporate their own interpretation into such intralingual text. This happens in a way that appears authoritative and neutral as these interpretations do not clearly belong to the production team considering that they are integrated into a type of input that is linked to speech of people appearing on the show. This means that the intertextual distance between audio-verbal forms of expression and impact captions provides leeway for manipulation (2009, p. 550).

Studies that linked this concept to Japanese variety shows argue that telops are used for the same purposes. It has, for example, been maintained that telops function as a device to frame humour (O'Hagan, 2010) or to highlight information (Sasamoto, 2014). Furthermore, colours and fonts of telops have been described as tools for incorporating attitudes of the production team towards people appearing on the show (Maree, 2013, p. 119; 2014, p. 62; 2015). These studies explain that such use of telops is aimed at creating common ground or cognitive mutuality between people appearing on the show and users (Maree, 2013, p. 128; Sasamoto, 2014).

2.2.3 Japanese speakers' perceptions of telops

The development of telops has stretched over a period of roughly sixty years. This means that different generations have gained familiarity with diverse usages and designs of telop. These experiences of Japanese users have resulted in different perceptions of those telops that appeared in the 1990s. Mainly two attitudes have been identified in the established literature.

Users who have been exposed to technologically advanced telops from a young age tend to view them in a positive light (Shitara, 2009, p. 2). They often not only see telops as useful or as an obvious and integral part of a Japanese television programme; they also claim to find telops indispensable and necessary in order to identify important information or jokes in dialogues of people appearing on the show (Shitara, 2008; Sakamoto, 2009b; Shitara, 2009). Production teams have developed a deep understanding of this group of users and let this awareness shine through in their telop designs. Research has argued that the salient appearance of telops highlights the most important information and jokes in speech, hooks channel-surfing users, and facilitates users who may have missed parts of the dialogue by doing several activities at the same time (Shitara, 2008; Matsukawa, Miyata and Ueda, 2009, p. 194; Sakamoto, 2009b; Shitara, 2009; Shitara, 2011, p. 7; Fukunaga, Katayama and Shoyama, 2014a, p. 26; Kami'iyu, 2014; Harada Senmonka, 2015). Some television programmes have made use of the latter point and link smartphones to telops in order to engage users (DENTSU INC., 2015). This suggests that telops function as a safety net to younger generations of Japanese users. That is to say, younger users seem to feel dependent on telops for their understanding of Japanese television programmes; especially when they are occupied with other activities while watching television.

Users who did not grow up with the new usages of the last three decades are more inclined to view them negatively. They find their use excessive or even feel that Japanese variety shows have become too busy (Shitara, 2009, p. 2; livedoor NEWS, 2015). This discontent is partly fuelled by recent incidents with telops (Kawabata, 2006, p. 214; livedoor NEWS, 2014; Tama-chan, 2014; livedoor NEWS, 2016; Sasamoto and Doherty, 2016). This suggests that a wish to reduce the amount of telops has been gaining traction in Japan as well.

2.3 Mixed-methods approaches and perception

2.3.1 Moving images

Research on multimodality has argued that a holistic approach to communicative situations is necessary in order to better understand multimodal messages, and that multimodal analyses of the highly contextualised nature of AV texts should focus on the interrelationship between types of input (Gambier, 2013, p. 47; Kaindl, 2013; Holsanova, 2014; Jewitt, 2014). This suggests that interdisciplinary approaches should be taken when analysing authentic AV materials.

Recent research on moving images has experimented with innovative approaches to the multimodal nature of authentic AV materials in relation to users' viewer reception (Redmond and Batty, 2015). Such studies have combined eye-tracking technology with disciplines that focus on narrative, screenwriting and screen theory (Batty, Perkins and Sita, 2015), anthropology (Dyer and Pink, 2015), and film (Robinson, Stadler and Rassell, 2015).

2.3.2 Online learning behaviour

Eye-tracking methods are increasingly used in research on SCMC and Second Language Acquisition (SLA) (Godfroid, Winke and Gass, 2013; Winke, Godfroid and Gass, 2013; O'Rourke *et al.*, 2017). Studies in these areas of research have recognised the benefits of using eye-tracking technology to study online language learning behaviour (O'Rourke *et al.*, 2016). More specifically, recent research on SCMC has maintained that eye-tracking technology, as part of a Mixed-Methods Research (MMR) design, is a helpful tool to explore personal learning experiences of FL learners in online settings (O'Rourke, 2012; Smith, 2012), distribution of visual attention during online interactions (Stickler and Shi, 2015), and reflection on such language learning by students and teachers (Stickler and Shi, 2017).

It has been argued in SLA research that attention and the selection of attended FL input is intertwined with individual differences between language learners (Robinson, 2003; DeKeyser and Juffs, 2005; Eskey, 2005, p. 571; Kasper and Roever, 2005). This makes language acquisition a largely personal and iterative process where learners constantly search for meaning by verifying their guesses, predictions and inferences during encounters with unknown FL input (Goodman, 1988; Robinson, 2003; Danan, 2004, p. 74;

Rost, 2005; Moreno, 2006, p. 64; Moreno and Mayer, 2007, p. 314). Taking into account that these processes yield manifold ways of perceiving AV materials (O'Rourke, 2012, p. 308; Holsanova, 2014, p. 288), it can be argued that the manner in which FL learners try to make sense of communicative situations becomes evident in their allocation of visual attention.

Eye-tracking technology enables researchers to record users' allocation of visual attention at the same time that a participant performs a task. Generally speaking, eye-tracking methods allow a certain degree of ecological validity (O'Rourke *et al.*, 2016, p. 286). Non-contact types of eye-trackers in particular facilitate experiment set-ups in which participants experience little interference from the eye-tracking system (O'Rourke, 2012, p. 313; Saldanha and O'Brien, 2014, p. 138; O'Rourke *et al.*, 2016, p. 287). Eye-tracking devices document detailed records of eye-movements (O'Rourke *et al.*, 2016, p. 286), which show what types of input are visually attended to by participants, how often and for how long subjects looked at certain stimuli, and in what order (Holsanova, 2014, p. 291). Researchers use such records to make inferences on participants' visual attention and cognitive effort (O'Rourke, 2012, p. 313; Saldanha and O'Brien, 2014, p. 137).

Studies that employ eye-tracking technology often refer to the *immediacy* and *eye-mind assumptions* for the interpretation of eye-movements, which are typically aggregated into fixations prior to analysis. Fixations are often considered to be the most important eye-movements as they represent moments during which users are processing visual information (Tobii AB, 2013b). The immediacy and eye-mind assumptions have been introduced by Just and Carpenter (1980) in their model of reading comprehension. In short, these assumptions state that there are two principles underlying the interpretation of fixations:

1. Content words are processed as soon as they are encountered by users
2. The eye remains fixated on a content word during processing which means that fixation duration equals the time it takes to process a content word
   (Just and Carpenter, 1980, p. 330)

The immediacy and eye-mind assumptions postulate that a direct link exists between visual attention and underlying cognitive processes, and that fixation measures can give insights into users' cognitive effort. While these principles have been instrumental in reading

studies in particular, recent research has questioned the applicability of the immediacy and eye-mind assumptions. It has been argued that users may not necessarily be consciously focused on a task as their minds may be wandering while fixating their eyes on a stimulus (Duchowski, 2003, p. 135; Stickler, Smith and Shi, 2016, p. 164). Furthermore, considering the multimodal nature of AV texts, it can be argued that processing does not only happen visually as users may process audio input when not fixating. While eye-tracking technology offers a robust method for collection of fixation-data, this critique on the two principles illustrates one of the main limitations of eye-tracking methods; they cannot provide insights into users' thinking processes, comprehension and reasons for visually attending to particular input (Holsanova, 2014, p. 292; Bateman, Wildfeuer and Hiippala, 2017, p. 161). This disadvantage of eye-tracking is often discussed alongside two other limitations: costs (O'Rourke *et al.*, 2016, p. 287) and the complexity of gaze data (Stickler, Smith and Shi, 2016, p. 178).

Considering that the original application of eye-tracking technology is situated in more cognitive and quantitative disciplines (Spinner, Gass and Behney, 2013, p. 390; Stickler and Shi, 2017), several additional challenges have been detected in relation to its integration into Applied Linguistics research. It has been argued that there is no standardised eye-tracking methodology available in SLA research and that standard operational definitions (i.e. codes) to describe or interpret FL learners' online viewing behaviour have not yet been established (O'Rourke, 2012, pp. 335-336; Spinner, Gass and Behney, 2013, p. 391; Winke, Godfroid and Gass, 2013, p. 207). There is also no general agreement on issues pertaining to what counts as a fixation, reading or noticing in research on SCMC (O'Rourke, 2012, pp. 335-336). Furthermore, it has been argued that processing of moving images is much more complicated to study than static images due to the competing forms of expression in an AV text (Kruger, Szarkowska and Krejtz, 2015, para. 2). It has been suggested that a better understanding of these issues will help inform proper use of eye-tracking technology in research on FL learning.

2.4 Summary

This chapter took a closer look at the research context of this study. It described important concepts and points of discussion from the established literature and demonstrated how this thesis fits into three research phenomena that relate to authentic materials, online learning behaviour and multimodality. It was shown that the use of authentic materials is a widely discussed topic in research on materials development. Of such materials, television

programmes that feature intralingual text were identified as promising language learning materials. This chapter also explained the format of Japanese television programmes of which variety shows in particular. It included an in-depth account of telops and demonstrated that such intralingual text plays an important role in the Japanese media landscape. It was also argued that a proper understanding of the composition of an AV text and the interrelationship between its forms of expression are essential for analyses of users' viewer experience and viewer reception of such materials. MMR designs that employ eye-tracking technology were identified as appropriate methodological frameworks for research on FL learning.

Furthermore, the literature review suggested that insights into JLLs' perceptions of an authentic Japanese variety show, multimodal analysis of such materials in relation to JLLs' viewer experience and viewer reception, and the application of an eye-tracking method would make for valuable contributions to the established literature.

The following chapter details the methodology and research design of this study. It explains the conceptual underpinnings of this research, takes a closer look at the chosen methods and mixed-methods approach, and describes the participants who took part in the experiment sessions.

3. RESEARCH DESIGN AND METHODOLOGY

Chapter 3 is the first methodological chapter of this thesis. It gives an account of the principles that underpin the research design and describes the participants who took part in the empirical data collection. The purpose of this chapter is to illustrate how this research builds on the research context and to demonstrate why the chosen methods and methodological approach are appropriate to the objectives of this study.

Section 3.1 brings the research questions and objectives into focus. This section not only gives an overview of the methods and approach deemed appropriate to answering the research questions; methodological considerations pertaining to their use are examined as well. Section 3.2 gives a thorough explanation of the research design, and details how the central concepts from the research questions, multimodal perception and visual attention, have been operationalised. Section 3.3 describes the research setting in which this study was conducted and provides a justification for the choice of video material. Section 3.4 takes a closer look at the participants of this study. This entails a description of their recruitment, organisation into participant groups and backgrounds.

3.1 Guiding principles

3.1.1 Research questions

It was explained in Chapter 1 that this thesis set out to gain insights into the appropriateness of authentic Japanese variety shows to JLLs at differing stages of their language study. This particular focus followed from research on materials development in which it was argued that authentic materials can be used by FL learners at any stage (Danan, 2004, pp. 74-75; Gilmore, 2007; Gilmore, 2011; Tomlinson, 2012). Three research questions were chosen to further investigate this claim:

RQ1 What are Japanese language learners' perceptions and attitudes towards authentic Japanese variety shows?

RQ2 To what extent does eye-tracking data allow us to interrogate and further examine those perceptions?

RQ3 What principles can be derived for the use of authentic Japanese variety shows in the Japanese language classroom?

It was also argued in Chapter 1 and Chapter 2 that this thesis intends to contribute to four emerging research areas based on the answers to these research questions. These areas of research relate to the use of authentic AV materials that feature intralingual text for FL learning purposes, the study of FL learners' visual attention while watching such materials, the integration of an eye-tracking method into Applied Linguistics research, and the multimodal analysis of moving images as an addition to the analysis of users' viewer experience and viewer reception. This thesis therefore speaks to teachers and learners of the Japanese language through its findings and researchers through its chosen methodology.

3.1.2 Methodological approach of this study

Mixed-methods approaches combine qualitative and quantitative methods to analyse research phenomena. Such an approach is often adopted in cases where either qualitative or quantitative methods alone are thought to provide inadequate answers to research questions (Riazi and Candlin, 2014, pp. 140-143; Bazeley, 2016). Although some would say that the mixing of methods from two distinct research paradigms is problematic or even unworkable, mixed-methods researchers argue that mixed-methods approaches combine the best of both worlds (Riazi and Candlin, 2014, p. 138; Johnson, 2018). Studies that maintain this view argue that the use of multiple methods allows for a researcher to triangulate or cross-check findings on a research phenomenon, to follow up on findings throughout the research process, and to offset weaknesses of the chosen methods (Creswell and Plano Clark, 2007; Riazi and Candlin, 2014; Johnson, 2018).

The rationale behind mixed-methods approaches is of particular importance to perception studies that employ eye-tracking technology. It has been argued in *Subsection 2.3.2 Online learning behaviour* that a limitation of eye-tracking is that it cannot provide insights into users' thinking processes, comprehension and reasons for visually attending to particular input as it is designed to generate accounts of users' gaze behaviour (Olk and Kappas, 2011, p. 442; Holsanova, 2014, p. 292; Bateman, Wildfeuer and Hiippala, 2017, p. 161). Consequently, the triangulation of eye-tracking technology with other methods such as questionnaires, comprehension tests, think-aloud protocols and (stimulated recall) interviews has been recommended in order to be able to examine these topics (Holsanova, 2014, p. 287; Leow *et al.*, 2014, p. 118; Saldanha and O'Brien, 2014, p. 144; Stickler and Shi, 2015). Conversely, it has also been suggested that eye-tracking is a good way to check such self-reports and test results as users may not be able to remember everything or may give a false account of their viewer experience (Sikkema, 2015; Stickler, Smith and Shi,

2016, p. 166; Bateman, Wildfeuer and Hiippala, 2017, p. 145). This illustrates that the mixing of methods strengthens the robustness of a perception study.

This study employs a mixed-methods approach that is guided by its research questions (Riazi and Candlin, 2014, p. 142; Johnson, 2018). The questions comprise two concepts that need to be investigated empirically, multimodal perception and visual attention, and one research question to which the answer follows from the empirical findings (Riazi and Candlin, 2014, p. 147). A combination of offline and online methods was determined to be best suited to answer the research questions. For the purposes of this study, offline refers to those methods that are not used concurrently with an experimental task while online means that the methods are used in real time during an experimental task (Leow *et al.*, 2014).

Three offline methods were chosen in order to examine the first research question on JLLs' multimodal perception. These methods include questionnaires and field notes with a comprehension test incorporated in one of the questionnaires. It was discerned that, from those methods that have been suggested as complements to eye-tracking, think-aloud protocols are least suited to this study's objectives. Considering that a think-aloud protocol is an online method, it was expected that it would be too big of an interference with the experimental task. It would not only complicate perception when participants have to think out loud about their perception at the same time; verbal reports would also interfere with the audio of the Japanese variety show (Olk and Kappas, 2011, p. 442; Leow *et al.*, 2014, p. 114). Stimulated recall interviews were the preferred method for the investigation of multimodal perception, but it was not feasible to implement such interviews due to time and financial constraints. As noted previously, one online method, eye-tracking technology, was chosen in order to investigate the second research question on JLLs' visual attention.

Although this means that methods were chosen that are commonly associated with quantification, the mixing of methods in this study becomes apparent in the analysis of the collated empirical data in particular. This study follows an exploratory MMR design. Data analysis in an exploratory MMR design consists of three stages. These are:

1. QUAL
2. quan
3. Interpretation based on QUAL $\rightarrow$ quan results
   (Creswell and Plano Clark, 2007, pp. 75-79)

Stage 1 comprises the study of multimodal perception. As noted, questionnaires were combined with field notes and a comprehension test to investigate this concept further. Although responses to closed-ended questions were counted and analysed, emphasis during this phase was put on participants' perceptions from responses to open-ended questions and field notes. The latter were analysed through a thematic analysis. It has been argued that eye-tracking is a useful tool to examine observations on perceptions (Bateman, Wildfeuer and Hiippala, 2017, p. 161). Stage 2 therefore continues the empirical data analysis with an investigation of participants' visual attention. This entailed the analysis of gaze data generated through eye-tracking technology. The gaze data was analysed in three different ways. Heat maps were checked to find out where on the screen participants' visual attention was concentrated, gaze videos were replayed to gain an impression of participants' gaze paths, and numerical fixation-data were exported. Stage 3 then interprets the empirical findings in order to derive principles for the use of authentic Japanese variety shows in the Japanese language classroom. Quantitative and qualitative procedures were thus mixed at every stage of this study's MMR design.

Further to the choice of methods and MMR design, there were some additional methodological choices that needed to be considered in relation to Stage 2. These related to the quality of gaze data that was required to meet the objectives of this study. Equipment features, participant characteristics, and experiment set-up are the main determinants that can directly affect gaze data quality (Blignaut and Wium, 2014, p. 67). These were therefore examined before finalising the research design.

Generally speaking, studies that implement eye-tracking technology assess gaze data quality based on how accurate and precise gaze points have been tracked in terms of their spatial and temporal properties (Holmqvist *et al.*, 2011, pp. 33-50). The level of precision and accuracy needed in the gaze data depends on the objectives of the study and the size of stimuli in the test. Smaller stimuli and examination of precise eye-movements require higher levels of accuracy and precision and powerful eye-tracking systems (Saldanha and O'Brien, 2014, p. 140; Sikkema, 2015; Sullivan, 2016). Another aspect that influences the quality of gaze data is the method used for data validation which can be done both quantitatively and qualitatively. Whereas quantitative validation involves computations to calculate exact eye-movements through the implementation of test stimuli in a test; qualitative validation happens mostly through an eye-tracker's user camera or the assessment of track status and calibration evaluation (Sullivan, 2016). This means that

quantitative validation affects the design of a test. It is often recommended to replicate or consult previous research in the relevant field of study when planning the procedures for empirical data validation (Sullivan, 2016).

Stage 2 of the MMR design focuses on a general impression of participants' gaze behaviour and deals with considerably large stimuli such as telops. This means that some leeway in the level of precision and accuracy is allowed and that the use of a portable eye-tracking device or an eye-tracker with a relatively low sampling frequency is accepted (Holmqvist *et al.*, 2011, pp. 29-31). This study therefore opted to work with an eye-tracker that has a 60Hz sampling frequency. In addition, it was observed that a qualitative approach to quality assessment of gaze data has been applied in previous studies of the ICE project (personal correspondence). Eye-tracking data in this study has therefore been validated by qualitatively assessing the spatial property of gaze points which centres on correctly estimating their position through assessments of track status and calibration. The remaining two determinants of gaze data quality, participant characteristics and experiment set-up, were further examined in a pilot study.

3.1.3 Pilot experiment sessions

The pilot of this research was modelled after studies that had already been completed for the ICE project. This means that the materials from these studies were used for the pilot experiment sessions. Considering that JLLs had not yet been researched by the ICE project, these materials were not focused on the research topic of this thesis. The main aims of the pilot were to test and refine the use of chosen methods, to gain familiarity with the eye-tracking hardware and software, to identify problems that could arise during the experimental task, and to find ways in which these problems could be troubleshot.

The Research Ethics Committee of Dublin City University (DCU) granted permission for the pilot study on 5th of November 2014 via an extension of the approval for the ICE project. Participants were recruited among JLLs and Japanese speakers. The latter were included as a point of reference for the analysis of learners' gaze behaviour. JLLs were recruited at DCU among BA students in the second and fourth academic year. The recruitment process of JLLs happened through email and by word-of-mouth. Eight JLLs were recruited of whom three were second-years and five were fourth-year students. Japanese speakers were also recruited at DCU via the DCU Japanese Society. They were approached in person and asked to enter their contact details and availability in a sign-up

sheet if they were interested in participating in an experiment session. Seven Japanese speakers were recruited for the pilot study. This means that a total of fifteen subjects participated in the pilot. Participants were not given any payments or incentives prior to participation in the study. However, they were offered chocolates at the end of experiment sessions. Experiment sessions were held later in November 2014 and April 2015.

An excerpt from the multimedia corpus of the Japanese variety show *Honmadekka!?TV* was used for the video stimulus. The excerpt was taken from an episode that was broadcast and recorded in Japan on 7th of August 2013 (Sasamoto, O'Hagan and Doherty, 2016, p. 6). The total duration of the video stimulus was approximately 22 minutes. This included Japanese commercials that were added at the start of the excerpt in order to ease participants into the experimental task and a shot added at the end of the excerpt in which participants were thanked for their involvement in the study. The ethics forms were prepared in both English and Japanese. These included a brief description of participants' involvement in the experiment sessions and explained that participation in or withdrawal from this study would not carry any consequences. The questionnaire was also prepared in both English and Japanese. Questions related to participants' perceptions of telops and their comprehension of the Japanese variety show. The questionnaire also included several questions regarding participants' demographic information and general habits when watching television programmes. An English experiment protocol was prepared for data collection and a stationary Tobii T60XL eye-tracking system was used for the pilot experiment sessions. Figure 3.1 shows the experiment set-up for the pilot experiment sessions. The eye-tracker device, incorporated into a 24 inch computer screen, is displayed on the right.

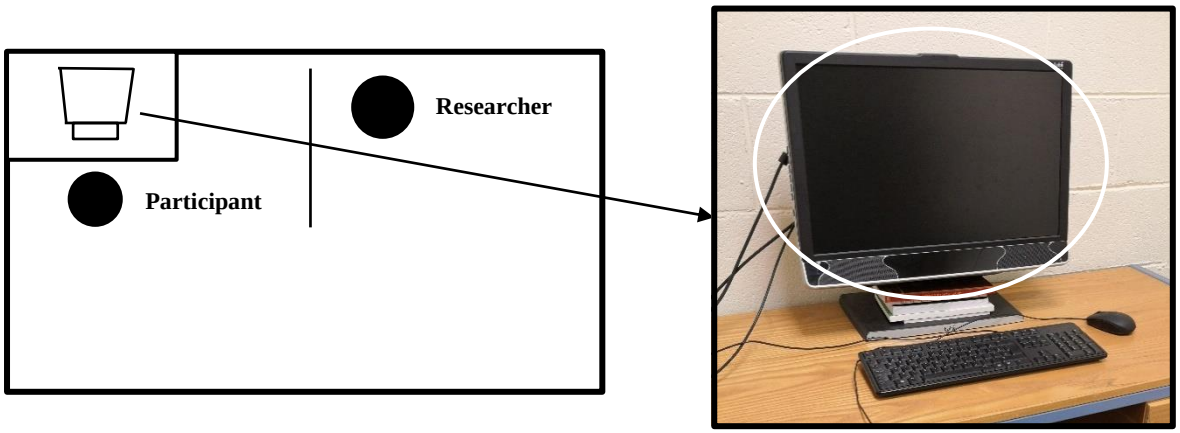


Figure 3.1 Set-up of pilot experiment sessions

The eye-tracking system, a plain language statement, two copies of the informed consent form, a questionnaire, pens, a screen and a token of appreciation were prepared for each experiment session. Participants were seated at a table with the eye-tracking system placed in front of them during the experimental task. The researcher was present in the same room but seated out of sight behind a screen. The ethics forms were filled out prior to the experimental task while the questionnaire was filled out directly after watching the video stimulus.

One observation in particular was made in relation to participant characteristics that needed to be taken into account for the main study. This related to the trackability of participants. There are several scenarios in which a participant's pupils cannot be found by an eye-tracker. Although this happens inevitably due to blinks, it may also be triggered by external factors such as make-up, light reflected off a participant's glasses, colour contacts or anything that resembles the shape of a pupil like facial piercings and earrings (Holmqvist *et al.*, 2011, pp. 116-128; Sikkema, 2015; Sikkema, 2016). It appeared that glasses were not an obstacle for the trackability of participants. What did seem to affect trackability was the cleanliness of glasses. It was therefore discerned that a cleaning kit for glasses should be available during the main experiment sessions and that participants with glasses would not have to be excluded from participation in the study. In addition, it was decided that an extra email was to be sent to participants in advance of the main experiment sessions to recommend them not to wear make-up, facial piercings, earrings and colour contacts (Saldanha and O'Brien, 2014, p. 139).

It appeared that the experiment set-up was appropriate for this study. This means that the order of steps, ethics forms, experimental task, questionnaire and debriefing did not have to be changed for the main experiment sessions. However, it was discerned that a list of precautionary measures needed to be prepared in order to make sure that environmental conditions were controlled for as much as possible.

Further to the determinants of gaze data quality, four additional observations were made with regards to the methods. First, some participants reported that the video stimulus was quite long and that it was hard to stay focused until the end. It was therefore decided that the duration of the video stimulus for the main study would have to be reduced to approximately ten minutes. Second, based on the answers to the comprehension questions, it appeared that either the format of the chosen excerpt of the Japanese variety show (i.e.

talk-show) or the type of questions (i.e. memory) might have been too difficult for JLLs. These choices were therefore reconsidered (*Subsection 3.2.2 Stage 1*) and revised (*Subsection 3.3.2 Video material* and *Subsection 4.1.2 Questionnaires*) for the main experiment sessions. Third, an extended version of the experiment protocol needed to be prepared in both English and Japanese for the main study in which consideration is paid to the beforementioned observations on participant characteristics and experiment set-up. Fourth, participants reported that the Tobii T60XL eye-tracking device might have been too big considering that participants had to sit quite close to it in order for the eye-tracker to be able to track them. One of the Japanese speakers commented that this made the telops in the video stimulus stand out more than usual. While different eye-trackers involve different strengths and limitations, it was determined that a smaller eye-tracking device would be more suited to the type of experimental task, and would improve viewing conditions and portability of the eye-tracking system (*Subsection 3.3.1 Context of this study*).

3.2 Research design

3.2.1 Overview of stages

The chosen mixed-methods approach formed the basis for the research design of this study. While the exploratory MMR design gives a good basis for the analysis of the empirical data it was preceded with several preparatory phases that enabled the empirical data analysis. These preliminary steps comprised video material selection, multimodal transcription and analysis, empirical data collection, and participant selection. As shown in Figure 3.2, the research design of this study was divided into three main stages.

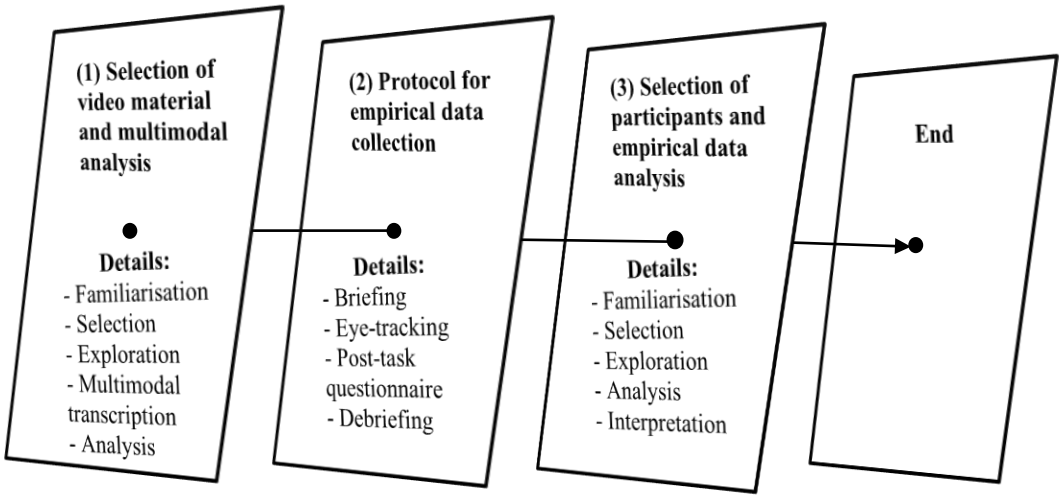


Figure 3.2 Structure of research design

Stage 1 consisted of two key steps, and focused on the selection and multimodal transcription of the chosen video material. It has been put forward in the established literature that it is important to report on the screen layout of a video stimulus as such a description not only makes the research design more transparent and replicable to other researchers; it also accounts for the potential influence of the multimodal landscape of a video stimulus on the findings from the eye-tracking data (Spinner, Gass and Behney, 2013). In addition, a clear and structured overview of the multimodal landscape of a video stimulus helps identify the context in which participants' perceptions are formed. The multimodal transcript was therefore consulted prior to and during the analysis of the empirical data. It helped structure the empirical data analysis and contextualise the findings from the empirical data.

Stage 2 followed from the selection of the video material and explains the experiment protocol. Although the experimental task focuses on participants' performance, experiment conduct is also shaped through interactions with subjects. An experiment protocol was therefore written up prior to the empirical data collection in order to make sure that participants are given the same instructions and information before and during the experiment sessions (Saldanha and O'Brien, 2014, p. 31), to ensure that the same order of steps was followed (Holmqvist *et al.*, 2011, p. 134), and to warrant transparency and consistency in the experiment design and conduct (Saldanha and O'Brien, 2014, p. 35). The protocol not only entails a description of the experiment set-up and procedures for empirical data collection; it also details precautions to mitigate risks to data validity.

Stage 3 comprised the last stage of this study and began after the empirical data had been collected. This stage can roughly be divided into two key steps. The first step concerned the selection of participants. Despite the careful consideration given to procedures in obtaining valid data, it is not uncommon for studies that employ eye-tracking technology to have their sample size affected at a later stage of the research process (Saldanha and O'Brien, 2014, p. 139). This can happen as a result of problems arising during recruitment and data collection. Benchmarks that have been set to assess the quality of gaze data need to be accounted for in order to guarantee that data of poor quality is excluded (2014, pp. 142-143). Participants who were not excluded were thereafter organised into participant groups that comprise subjects at similar stages of their Japanese language study. The second step of Stage 3 was structured according to the exploratory MMR design of this study and focused on the empirical data analysis.

3.2.2 Stage 1

The selection of video material was the first step of this study. It formed one of the most important procedures in the research design. The choice of video material not only shapes the multimodal analysis; it also defines the video stimulus for the experimental task on which the empirical data is collected and the multimodal blueprint with which JLLs' perceptions and gaze behaviours are analysed.

Findings from the pilot experiment sessions suggested that another excerpt would be appropriate for the main experiment sessions. The excerpt used for the pilot comprised of a talk-show in which the communicative situation is situated in a studio and depends heavily on dialogue and telops. It included conversations that some beginner learners might find too advanced and it appears to make little use of visual-nonverbal input. Considering that this study involves JLLs at different stages of their language study it was deemed appropriate to choose an excerpt that does not overwhelm learners at beginner level on the one hand and is stimulating to more advanced learners on the other. It was therefore discerned that it would be worth looking into authentic Japanese variety shows that are not based in a studio (e.g. travel programmes) or make more use of visual-nonverbal input such as kinesic action or imagery. The video material, further discussed in *Subsection 3.3.2 Video material*, was chosen from the multimedia corpus of the ICE project.

The choice of video material was followed with multimodal transcription of the chosen excerpt. Although a rough idea of the multimodal nature of the video material was formed by the time the excerpt was selected, a deeper understanding of the forms of expression, the communicative situation and the multimodal landscape was needed in order to make sense of the empirical data. It has been argued that it is not possible to predict in advance the forms of expression or types of input that will be encountered during multimodal analysis (Bateman, Wildfeuer and Hiippala, 2017, p. 11). An approach that takes this into account is therefore needed (2017, p. 76). The multimodal analysis in this study happened in two phases.

The first phase focused on familiarisation with and exploration of the multimodal nature of the video material and aimed to seek out those types of input that were expected to help JLLs' understanding of the programme contents. The taxonomy of Kimura *et al.* (2000) as described by Sakamoto (2009b), included in *Table 2.1 Taxonomy of telops (translated and adapted from Sakamoto, 2009b)*, was consulted during the exploration of the video

material. The definitions of gesticulation and bodily configuration as described by Kalantzis *et al.* (2016, pp. 386-388) were used for the definition of kinesic action. Considering that the video material has its own unique blend of forms of expression, these frameworks were tailored to the types of input encountered in the excerpt and altered where necessary. This is further discussed in *Subsection 5.2.4 Visual-verbal input* and *Subsection 5.2.5 Visual-nonverbal input*.

The second phase consisted of a more fine-grained analysis, and entailed the multimodal transcription of the selected types of input and the analysis of the multimodal landscape. A multimodal transcript of the chosen excerpt had already been prepared within the ICE project before the start of this research. This transcript was used as the basis for the multimodal transcript of this study. However, as it was not tailored to the objectives and research questions of this thesis it had to be adjusted considerably.

A combination of two transcription frameworks and the typologies of utterance of speaker and narration telop as defined by Shitara (2005; 2006) were consulted for this. The frameworks for multimodal transcription were those defined by Baldry and Thibault (2006), and Taylor (2003). Baldry and Thibault (2006) have defined a strategy for multimodal transcription of AV texts in which not only the visual image is described in great detail; such detailed descriptions are extended over six columns in which other forms of expression, frames and durations of frames are also included. Taylor has presented a simpler and modified version of this model in order to incorporate subtitles into multimodal transcription (2003, p. 194). The approach taken in this thesis is similar to that of Taylor (2003). The multimodal transcription of this study is discussed in further detail in *Subsection 5.3.2 Procedure for multimodal transcript* and *Subsection 5.3.3 Procedure for multimodal units*. In the same way that frameworks for the exploration phase needed to be tailored to the excerpt, the definitions from the typologies of Shitara (2005; 2006) included in *Subsection 2.2.2 Overview of telops* were altered for the scope of telops detected in the video material. These alterations are further discussed in *Subsection 5.3.4 Procedure for types of resemblance*.

3.2.3 Stage 2

The experiment protocol consisted of procedures taken in advance, during and after experiment sessions. Two of these happened before each session. First, an email was sent to participants. This email notified subjects of the potential negative influence of make-up,

facial piercings, earrings and colour contacts on the eye-tracker, and advised participants not to wear these during a session. The email also contained an .xlsx file with the pre-task questionnaire (*Appendix A. Pre-task questionnaire*). It was opted to send this questionnaire in advance of an experiment session in order to prevent sessions from exceeding the allotted one-hour time slot. Participants were requested to fill it in and to send it back by email.

Second, the experiment set-up was checked before the first experiment session on each experiment day. Personal advice from eye-tracking experts was followed in order to maintain experiment conditions as best as possible (Sikkema, 2015; Sikkema, 2016). None of the experiments were to be conducted directly underneath a lamp in order to avoid artificial light reflecting off a participant's glasses or the eye-tracker. Where applicable, participants did not sit close to a window in order to prevent changes in brightness of natural lighting. In addition, any automatic updates were disabled before running any experiments to prevent interruptions when recording. The volume of the video stimulus was kept at a setting of 30 for all experiment sessions. To check whether the experiment conditions were appropriate a trial evaluation of the track status and calibration was run on the researcher. The video stimulus contains several distinctive stimuli that were used to check whether the eye-tracker properly registered the gaze points by replaying the recording of the researcher. With these preparations finished the experiment sessions could start. The procedures that were taken during an experiment session can be divided into four main components as shown in Figure 3.3.

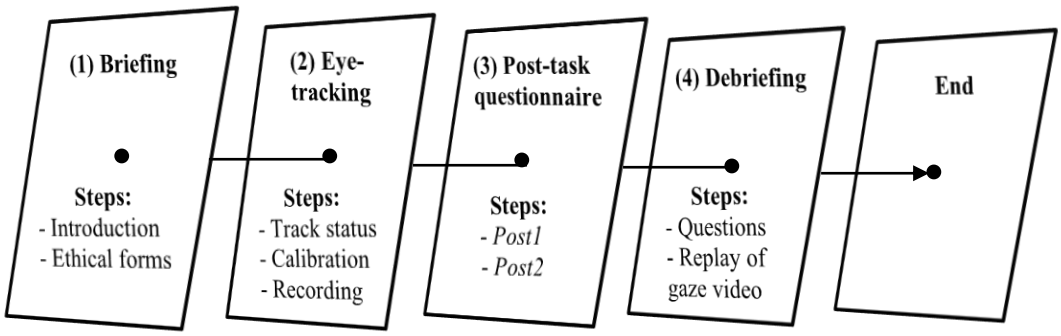


Figure 3.3 Structure of main experiment sessions

At the start of each session participants were informed about the structure of the experiment session and given a brief explanation of their involvement in each component. They were first presented with the plain language statement (*Appendix B. Plain language statement*) followed by the informed consent form (*Appendix C. Informed consent form*).

Each participant was given a copy of their informed consent form and subjects were reassured they could follow up on or withdraw from the research at any given time. Participants were asked whether they knew anything about the study or the experiment at the start of each session as an additional check to rule out any bias created between subjects.

After these steps were completed, the actual eye-tracking experiment could start. This procedure was initiated with an assessment of the track status and the calibration process for which a script was followed (*Appendix D. Script for main experiment sessions*). A 9-point calibration plot was used to check error vectors for as many points on the screen as possible. Stable detection of the eyes and an appropriate distance to the eye-tracker are enabled when both dots in the track status box are not flickering and when the bars on both axes are coloured green. The track status box was also used to show participants how much head movement is tolerated by the eye-tracking device (Saldanha and O'Brien, 2014, p. 139). A cleaning kit for glasses and make-up remover pads were available in case the eye-tracker cannot find the participant or shows a calibration plot with many error vectors. Participants were encouraged not to look away from the screen during recordings in order to reduce data loss and were requested to save any questions until after the recording had finished. After making sure that the calibration results were satisfactory and the participant was sitting comfortably the recording was started. The settings of the video stimulus were customised in such a way that each recording ended automatically after it had finished. The researcher stayed in the room to take field notes and to keep an eye on the experiment but was seated out of sight to alleviate the Hawthorne effect (Holmqvist *et al.*, 2011, p. 134). That is to say, it was hoped that participants would feel less inclined to change their behaviour (Winke, Gass and Sydorenko, 2013, p. 269; Saldanha and O'Brien, 2014, pp. 31-32) if the researcher was not visible during the experimental task.

A post-task questionnaire (*Appendix E. Post-task questionnaire*) was filled out after each experimental task. Participants were allowed to ask for clarification when certain questions were not understood; however, the researcher could not make any comments on answers to questions. The abbreviations in Figure 3.3 refer to the different parts of the questionnaire; *post1* stands for the first part and *post2* means the second part. There was no set time limit within which the questionnaire had to be completed. The researcher checked the questionnaire for any missing responses directly after it had been filled out.

Subjects were asked whether there was anything they wished to add to the questionnaire or whether they had any thoughts in relation to the video stimulus or experiment session they wished to share (Holmqvist *et al.*, 2011, pp. 139-140). Participants were given the opportunity to have a look at their gaze replay video together with the researcher and to elaborate on their gaze behaviour.

Three procedures followed the experiment session. First, participants were requested not to inform other students of the experiment session, video stimulus and research objectives. There was a strong possibility that participants were classmates or knew each other and that subjects who had already participated in the experiment could influence those who had not. Considering subjects were not made aware of the specifics of the study before participating and were given the opportunity to learn more about the aim of the study during the debriefing it was essential to be cautious of intersubject bias (Holmqvist *et al.*, 2011, p. 140). Second, a small token of gratitude was offered to subjects before they left the experiment room. Third, the interactions during the debriefing were briefly summarised within the field notes as soon as the experiment session had finished.

3.2.4 Stage 3

The empirical data analysis was preceded by the selection of participants. Subjects were selected based on their backgrounds and gaze data quality. This thesis acknowledges that FL learners have unique abilities, breadth of knowledge and experiences that shape their personal learning journeys. Strictly speaking, it is not possible to speak of FL learners who are at the exact same point of their language studies. For this particular reason, participant groups in this thesis represent cohorts of JLLs who are at roughly the same stages of their Japanese language study. The organisation of participant groups was to some extent based on responses to particular questions in the pre-task and post-task questionnaires. These questions were incorporated for the purposes of participant selection. The topics of these questions were drawn from the established literature:

1. Experience with studying Japanese and other FLs
2. Linguistic abilities in the First Language (L1) and Japanese
3. Media exposure to television programmes
4. Prior knowledge of the video stimulus and video material
   (Ortega, 2011, pp. 176-177; Gambier, 2015, pp. 75-76)

The organisation of participant groups was also based on the quality of gaze data. Participants' gaze videos were replayed to check their gaze paths as data quality may not only differ between recordings but also within portions of recordings (Holmqvist *et al.*, 2011, pp. 140-141; Saldanha and O'Brien, 2014, p. 139). It was noted in *Subsection 3.1.2 Methodological approach of this study* that a qualitative approach to data validation was taken in this study. The following three data quality indicators were therefore determined as most important for the selection of participants:

1. Quality of calibration
2. Gaze sample ratio
3. Prolonged sound interference during recording

Only calibrations without any error vectors or calibrations with very short error vectors were accepted in this study. It has been argued that ethnicity can influence the gaze sample ratio or trackability of participants (Blignaut and Wium, 2014). The benchmark was therefore set at 60% to allow for participants with different ethnicities to participate in the study without compromising on the quality of gaze data. This means that gaze sample ratios lower than 60% were not accepted. The last criterion was examined to ascertain whether participants had trouble hearing the video stimulus as this could potentially lead subjects to focus more on telops to make up for the loss of sound (Sikkema, 2015; Sikkema, 2016). This final criterion was used to flag participants but does not lead to exclusion unless the subject could not hear the video stimulus at all or if it occurred in combination with one or more of the other criteria.

The empirical data analysis started after the organisation of participant groups. The analysis of the empirical data followed four analytical phases for the investigation of multimodal perception and visual attention. These analytical phases comprised data familiarisation, data exploration, data analysis and data interpretation. Data familiarisation and data exploration happened prior to data analysis and data interpretation. They were conducted on both concepts simultaneously. These phases allowed for the extraction of data sets from the large volume of collated data, the preparation of data for analysis, and the identification of potential trends. They also provided opportunities to seek out potential links between the data sets or potential routes for further investigation. As shown in the MMR design, data analysis was thereafter applied to multimodal perception and then to

visual attention. The findings from these analyses were combined for the data interpretation.

As noted, the empirical analysis of JLLs' multimodal perception is based on answers to the comprehension test, counts of closed-ended responses and a thematic analysis on open-ended responses and field notes. Selections of questions that were expected to provide insights into JLLs' perceptions were chosen during the data familiarisation phase. Participants' responses to the selected questions formed the data sets. Whereas the data set of answers to the comprehension test and closed-ended responses was analysed straight after, the thematic analysis continued with more preliminary analyses.

The thematic analysis of this study followed the approach of Braun and Clarke (2006). After noting down some initial ideas and observations on the data set the data exploration phase started with the generation of initial codes (2006, pp. 87-89). A data-driven or inductive approach was taken to this initial coding (Frith and Gleeson, 2004; Braun and Clarke, 2006, p. 83), which happened in two rounds. The first round of initial coding was a tentative type of coding in which the nuances between initial codes were determined and a coding scheme was devised. The second round of initial coding applied the coding scheme to the data set (Attride-Stirling, 2001, pp. 390-391). This process resulted in cohorts of data extracts labelled with a number of initial codes. These cohorts of extracts were then used to search for (sub)themes. The (sub)themes in this study were generated over data items of participants across participant groups. The search for (sub)themes also happened in two rounds. The first round focused on the identification of relating concepts across data extracts to work out overlap between initial codes which could form the basis for the thematic map (2001, pp. 396-397). This process resulted in a number of candidate (sub)themes (Braun and Clarke, 2006, pp. 89-91) that were checked against the data extracts and data set in the second round. This involved alterations or revisions where necessary to make the thematic map representative of the data (2006, pp. 91-92). After finalising the thematic map the (sub)themes were analysed internally to work out their prevalence per participant group. This entailed the data analysis phase.

It is important to note that the thematic analysis of this study aimed to find "patterned response or meaning within the data set" (2006, p. 82). It was discussed in *Section 1.1 Background* that the researcher herself is a learner of the Japanese language and an enthusiast of authentic AV materials. This guiding principle was therefore established in

order to reduce the risk of confirmation bias and to make sure that data items are considered and treated in the same way.

The subsequent empirical analysis of JLLs' visual attention is based on the eye-tracking data. Heat maps, gaze replay videos and fixation-data formed the data sets. Version 3.4.8 of *Tobii Pro Studio* was used throughout the gaze data analysis. The Velocity-Threshold Identification (I-VT) filter was set as the eye-movement classification and applied to participants' eye-position samples at all times during the four phases of empirical data analysis. The default settings[7] were used in accordance with the advice given through email by Tobii AB Technical Support and the general recommendation of Tobii AB to use default settings when no particular requirements have been determined in a study for the definition of fixations (Olsen, 2012; Tobii AB, 2014b, p. 51).

The data familiarisation encompassed the replay of gaze videos to gain an initial impression of participants' gaze paths (O'Hagan and Sasamoto, 2016, p. 46). This phase was incorporated in the participant selection. The following phase in which the gaze data was explored started with the analysis of heat map findings. These data visualisations functioned as an aid to pinpointing areas of the screen where visual attention is concentrated. Nonetheless, it has been argued in the established literature that heat maps need to be used with caution as their application does not entail an analysis of the data; rather, they help visualise large quantities of data (Bojko, 2009; Tobii AB, 2014b, p. 65; Conklin and Pellicer-Sánchez, 2016, p. 5). In addition, considering that authentic Japanese variety shows consist of moving images, it needs to be confirmed that the input visible in the static images of the heat maps properly represent the type of input that is visually attended to. Further data exploration was therefore carried out with the help of gaze replay videos.

With a better understanding of the types of input that gathered the most visual attention from participants the data analysis was conducted. This phase was centred on the analysis of fixation-data that was extracted on different types of input in the video material. However, rather than analysing every eye-tracking measure available in the eye-tracking software, a selection was made that is appropriate to the purposes of this study. The choice

[7] The default settings are: (1) gap fill-in (interpolation) function enabled with max gap length of 75 milliseconds, (2) eye selection set at Average, (3) noise reduction disabled, (4) window length of Velocity calculator set at 20 milliseconds, (5) Velocity threshold of I-VT classifier set at 30 degrees per second, (6) merge adjacent fixations function disabled, and (7) discard of short fixations function disabled.

of eye-tracking measures not only depends on the objectives of a study; it is also determined by the size of the user-defined Areas of Interest (AOIs) and sampling frequency of the eye-tracking device (Rayner, 1998, p. 377; Rayner, 2009, pp. 1461-1462; Roberts and Siyanova-Chanturia, 2013, pp. 219-220; Conklin and Pellicer-Sánchez, 2016, p. 9). The eye-tracking analysis of this research involves gaze data generated from AOIs on large stimuli with an eye-tracker device that has a 60 Hz sampling frequency. The fixation-based analysis therefore focused on two late measures that give indications on later stages of cognitive processing (Roberts and Siyanova-Chanturia, 2013, p. 217; Conklin and Pellicer-Sánchez, 2016, p. 3) in order to gain insights into visual attention accumulated over the whole duration of the analysed video segments:

1. Fixation count
2. Total fixation duration

This thesis used dynamic AOIs that were organised into AOI groups to extract the fixation-data on these eye-tracking measures. Dynamic AOIs can transform to align with different types of input on the screen when changes to their vertices are entered in keyframes on the timeline. There are two advantages to the use of dynamic AOIs. First, they can move, expand and shrink at any given time. Second, they can be made "dormant" or inactive for certain time intervals during which no gaze data is collected. This allowed for a proper analysis of types of input in the video material that move on the screen and have different screen times.

It was discussed in *Subsection 2.3.2 Online learning behaviour* that recent research has questioned the immediacy and eye-mind assumptions of Just and Carpenter (1980). This thesis recognises that the possibility exists that participants may be thinking about something else during an experimental task. Nonetheless, due to the nature of the video material used in this study it was expected that participants' gaze behaviour would be characterised by a search for meaning in the multimodal messages (Batty, Perkins and Sita, 2015, para. 15).

3.3 Research setting

3.3.1 Context of this study

This study was mostly set at DCU and involved adult JLLs registered at this university. The JLLs targeted for this study were students of BA degrees in which linguistic skills in an FL of choice, in this case Japanese, had to be acquired. This means that this thesis was shaped according to the trajectory that JLLs follow at DCU. This trajectory includes an exchange year at one of several Japanese universities in addition to the three years of study at DCU. This exchange is organised for the third academic year. It was aimed to recruit JLLs in all four academic years. This means that this thesis is not a longitudinal account of students moving through their BA degrees; rather, it involves different students at different stages of their language study from one academic calendar year. Most of these JLLs did not know the researcher. However, at the time when the call of participants was sent out the researcher was teaching kanji classes for one semester to students in the second academic year. This study also involved adult native Japanese speakers registered for BA degrees at a Japanese university. They were targeted for this study as a point of reference for JLLs' perceptions.

The type of experimental task that is central to this study involves a video stimulus that consists of linear dynamic stimuli (Tobii AB, 2013a; Bateman, Wildfeuer and Hiippala, 2017, p. 109). This means that the video stimulus is composed of input that is in motion and follows the same sequencing and timing for all participants. It comprises a communicative situation depicted in an excerpt from an authentic Japanese variety show. The role of participants in the experimental task is that of an observer (Bateman, Wildfeuer and Hiippala, 2017, p. 109). Participants watch the video stimulus in a non-interactive way as communication cannot take place between the communicative situation visible on screen and the subjects.

3.3.2 Video material

The authentic AV material used for the video stimulus comprises an excerpt taken from an episode of the Japanese variety show *Honmadekka!?TV*. This is the same episode that was used for the pilot study but the chosen excerpt is different. As noted, this episode was broadcast and recorded in Japan on 7th of August 2013, and is included in the multimedia corpus of the ICE project (Sasamoto, O'Hagan and Doherty, 2016, p. 6). *Honmadekka!?TV* is aired by Fuji Television Network, Inc. and gathers high viewer ratings (Kanai, 2011;

@DIME, 2014). It is a talk-show in which jokes, questions and interesting facts on a variety of topics are exchanged between hosts, celebrity panellists and expert guests. The selected episode contains approximately 27 minutes of conversation on summer-related themes followed by roughly 18 minutes of dialogue in a Q&A format that consists of five short self-contained panel discussions. The latter programme item is called *Moshimo no toki ni yakudatsu Honmadekka!?* [*Useful Honmadekka!? for unexpected times*]. The first portion of the episode was used for the pilot study while the last portion was used for the main study.

There are several reasons behind the choice of video material. This last portion of the episode was considered most suited as it could be edited into an experiment video with an approximate duration of ten minutes. Each panel discussion was self-contained and did not include any references to the first portion of the episode or any of the previous episodes (Mishan, 2005, p. 134). Participants would therefore not need to know what had happened outside of these panel discussions. Three panel discussions were selected as they had the desired duration when combined and were considered to be most accessible to JLLs at differing stages of their language study. The selection of panel discussions included several visual-nonverbal types of input in addition to the dialogue and telops. Examples include drawings and different types of kinesic action. The visual-nonverbal input could help prevent a feeling of being overwhelmed for JLLs at early stages of their language study without making the video stimulus too easy for the more advanced learners (2005, p. 138). The Q&A format of these panel discussions was also considered to be appropriate for the experimental task as it was more centred on comprehension, which made the preparation of the comprehension test more straightforward.

3.4 Participants

3.4.1 Recruitment and ethics

Separate approval was sought from the DCU Research Ethics Committee for the main experiment sessions in relation to both the design of materials and the experiment set-up, as new ethics forms, experimental task and questionnaires had been developed. The approval for the main study was granted on 24th of November 2015. The recruitment of JLLs happened through email, a sign-up sheet and word-of-mouth. Students in the first, second and fourth academic years were approached through a recruitment email sent by their lecturers or the researcher. Students in the second and fourth academic years were

also approached in person through a sign-up sheet circulated in class. A different approach was taken for students in the third academic year who were on exchange in Japan. They were recruited via their lecturer at DCU who was in charge of them during their stay in Japan. Students who volunteered outside of these methods had been encouraged to participate by their classmates. Japanese speakers were recruited in class at a Japanese university. Their lecturer was contacted in advance by the lecturer at DCU who was also in touch with the JLLs on exchange.

A total of 46 JLLs and seven Japanese speakers responded to the call for participants. They were all recruited on a voluntary basis. One of the JLLs who had volunteered could not make it to the experiment session which means that 52 subjects participated in this study. Participants were not offered any payments or incentives prior to participation. However, upon completing the experiment session, JLLs were offered chocolates or crisps and Japanese participants were sent an *Amazon* gift voucher as a token of appreciation. Participants were reassured that there was no penalty if they wished not to participate or to withdraw from the study before its completion.

Full names of participants are not mentioned within this thesis in order to protect their anonymity. The empirical data has been anonymised by assigning each participant an identifier. JLLs have been named Participant (P) while Japanese speakers have been named Japanese Participant (JP). A number has been appended to these abbreviations to distinguish between subjects. This numbering follows the order of experiment sessions. The data set therefore comprises a JLL cohort consisting of P01-P46 in which P07, i.e. the identifier of the participant who could not make it to the experiment session, is missing and a Japanese speaker cohort consisting of JP01-JP07.

3.4.2 Meaningful selection of participants

It has been noted in *Subsection 3.2.4 Stage 3* that a list of selection criteria was established for the organisation of participant groups. This list comprises four topics (i.e. experience with studying Japanese and other FLs, linguistic abilities in L1 and Japanese, media exposure to television programmes, and prior knowledge of the video stimulus and video material) that were identified as factors that influence the coherence of participant groups and three gaze data quality indicators (i.e. quality of calibration, gaze sample ratio, and prolonged sound interference during recording) that ensure good quality data are included in the gaze data sample. Whereas implementation of the gaze data quality indicators is

straightforward, it appeared that the four factors needed to be narrowed down as not all of the included topics turned out to be appropriate for the purposes of this study. The majority of questions included in the pre-task questionnaire and sections of the post-task questionnaire were incorporated for the selection of participants. The responses to these questions were examined to gain a better understanding of participants' backgrounds and to work out a selection of questions that could form the basis for the participant groups. The following four questions were considered the most suitable bases for categorisation of the participant groups:

pre Q2 Which academic year are you enrolled in?[8]

pre Q3 Have you been to Japan?[9]

pre Q4 For how long and for what purpose were you in Japan?[10]

The abbreviation *pre* in this list refers to the pre-task questionnaire. These questions were chosen as main determinants as they provide insights into the number of years that JLLs have formally studied the Japanese language at university and the time JLLs have spent on exchange in Japan. This means that questions regarding the study of the Japanese language before entering university, self-assessments of proficiency level in the L1 and Japanese, test results on Japanese proficiency level, use of the Japanese language outside of class, methods for learning kanji, media exposure to English television programmes, preferences for captions or subtitles in television programmes, and general viewing habits have not been chosen as main determinants.

It appeared that all of the first-year participants were enrolled for degrees that offer Japanese at beginner level. It was discerned that the advantage of previous study may had levelled out by the time this study was conducted as data from these subjects were collected towards the end of the second semester. Furthermore, the 45 JLLs who had filled out the pre-task questionnaire did not tick the box in one of the multiple choice questions that could have made a key difference in language skills; the study through Japanese relatives. It was discerned that the self-assessments and responses on Japanese proficiency levels do not give an accurate reflection of participants' linguistic skills. Test results were incomplete or included scores that were based on a variety of proficiency tests that were

[8] This question is not included for JLLs on exchange in Japan. It is included as *pre* Q3 for Japanese speakers.
[9] This question is not included for Japanese speakers. It is phrased as "Have you been to Japan before going on your exchange year?" for JLLs on exchange in Japan.
[10] This question is not included for Japanese speakers. It is phrased as "For how long and for what purpose were you in Japan before going on exchange?" for JLLs on exchange in Japan.

not taken at the time of data collection. Considering that this study does not focus on crosslinguistic interference of the L1 with FL learning it was also decided that questions regarding linguistic skills in the L1 were not appropriate after all. It was observed that the remainder of the discarded questions was unrelated to the research focus or not properly phrased. Findings from responses to the discarded questions have not been taken up for further discussion in the remainder of this thesis.

3.4.3 Grouping of participants

Appendix F. Evaluation of collected data gives a complete overview of the empirical data collated for the main study together with indications on the quality and completeness of data sets, and comments from field notes on potential influences that may have affected gaze data quality. Missing data or aspects of data that do not meet the standards set previously are marked with red. *Pre* refers to the pre-task questionnaire while *post1* and *post2* are used to distinguish between the first and second parts of the post-task questionnaire. For the purposes of this thesis, subjects have been organised into four participant groups as can be seen in Figure 3.4.

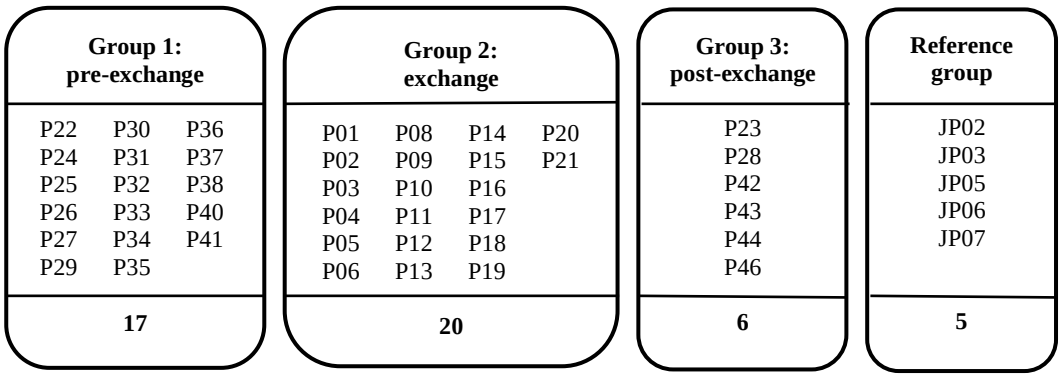


Figure 3.4 Organisation of participants into participant groups

These include three learner groups that comprise of JLLs and one Reference group that consists of Japanese speakers. Figure 3.4 shows 48 participants which means that four participants have been excluded. These participants are P45, JP01, JP04 and P39. P45 was a fourth-year student who had not been to Japan on exchange. It was determined that P45 could not be placed in Group 1: pre-exchange due to the considerable difference in years of formal Japanese study. JP01 and JP04 were Japanese speakers who had a poor calibration and an insufficient gaze sample ratio respectively. The data of P39, a first-year JLL, was not discarded based on the selection criteria but due to technological issues that arose during the fixation-data analysis. The eye-tracking software showed an error message each

time it tried to generate fixation-data on the eye-tracking measures for this particular participant. It was decided not to include the empirical data of P39 based on the consultation with Tobii AB Technical Support through email and a remote session.

3.4.4 Profile of participant groups

Table 3.1 shows participants' age, gender, L1, experience with learning Japanese and other FLs, time spent in Japan on exchange, exposure to Japanese television programmes, and extent of prior knowledge of the video stimulus per participant group.

Table 3.1 Background per participant group

Aspect	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Age *(average)*	20 years	21 years*	22 years	22 years
Gender	52.9% female 47.1% male	68.4% female* 26.3% male* 5.3% prefer not to say*	50% female 50% male	60% female 40% male
L1	64.7% English 35.3% other	79% English* 15.8% other* 5.3% unspecified*	83.3% English 16.7% other	Japanese
Experience learning other FLs	Yes	Yes*	83.3% yes 16.7% no	Yes
Academic year	88.2% second year 11.8% first year	Third year	Fourth year	Fourth year
Average length of exchange in Japan	No	3 ½ months*	1 year	-
Exposure to Japanese television media on average per week	76.5% 1-10 hours 17.6% 0 hours 5.9% 11-19 hours	65% 1-10 hours 20% 0 hours 15% 11-19 hours	83.3% 1-10 hours 16.7% 0 hours	60% 1-10 hours 40% 11-19 hours
Prior knowledge of video material	82.4% no 17.6% yes	89.5% no** 10.5% yes**	83.3% no 16.7% yes	Yes
Experience of watching video material	88.2% first time 11.8% sometimes	85% first time 10% sometimes 5% I do not know	83.3% first time 16.7% sometimes	Sometimes

** Response of P04 is missing. The value is calculated over the total number of responses, not participants (i.e. n = 19).*
*** Response of P13 is missing. The value is calculated over the total number of responses, not participants (i.e. n = 19).*

The JLL cohort consists of 43 students enrolled for five different BA degrees. These degrees are Applied Language and Translation Studies, Business Studies International, and three types of Joint Honours; Media Studies and International languages, Irish and International Languages, and Law and International Languages. Participants in the Reference group are five students pursuing a BA degree in Policy Studies. The subjects were 19 to 27 years of age with an overall average of 21 years. One participant preferred not to specify the gender while 28 subjects were female and 18 participants were male.

The majority of participants in each learner group are native English speakers. However, each of these groups also included JLLs who reported that their L1 is a language other than English. Six of these JLLs are in Group 1: pre-exchange and each of them reported a different L1. Their responses included French, Hungarian, Cantonese, German, English-Polish and Vietnamese-French. Three more JLLs are in Group 2: exchange. Again, all three of them reported a different L1. They responded with Chinese, Lithuanian and Italian. Only one JLL in Group 3: post-exchange indicated not to be an English native speaker. Mandarin is the L1 of that particular participant. All of the Japanese speakers in the Reference group reported to be native Japanese speakers.

The JLLs were in four academic years at the time of the empirical data collection. Two were first-years, 15 participants were second-years, 20 third-year students and six subjects from the fourth academic year. Two participants in Group 1: pre-exchange were study abroad students. They were in the second academic year at DCU but had enrolled in the third academic year of their home university. Participants in Group 1: pre-exchange reported never to have been to Japan except for one second-year student who had spent a holiday there for ten days. The participants in Group 2: exchange were in Kyōto and Tōkyō at the time of the experiment sessions and had already spent approximately three-and-a-half months studying at a Japanese university. All the learners in Group 3: post-exchange had completed a full exchange year in Japan.

A large majority of JLLs reported to be familiar with watching Japanese television and to have some experience with text on screen in Japanese television programmes. Only 18.6% of respondents did not watch Japanese television programmes while 72.1% of them watches 1-10 hours of Japanese television programmes on average per week. The remaining 9.3% watches 11-19 hours on average per week. Of those JLLs who responded that they watch Japanese television programmes, 91.4% indicated that they watch them with subtitles or captions. Furthermore, 48.6% of these JLLs reported watching Japanese variety shows. All of the participants in the Reference group reported watching Japanese television programmes, including variety shows.

It appeared that approximately 85.7% of JLLs did not know about the variety show before participating and approximately 86% of JLLs watched it for the first time during the experiment session. Only 14.3% knew about it previously and 11.6% of JLLs sometimes even watch the programme. All participants in the Reference group indicated knowing

about the variety show prior to participating and watching it sometimes. This shows that subjects in the Reference group had more pre-existing knowledge of the video material and that a vast majority of JLLs were unfamiliar with the authentic Japanese variety show before participating in this study. Furthermore, none of the participants knew anything about the study or the experiment at the start of their experiment sessions.

Table 3.2 gives a summary of *Appendix F. Evaluation of collected data* and displays the combined assessments of the three data quality indicators per participant group. As noted previously, *post1* and *post2* refer to the two parts of the post-task questionnaire. The third abbreviation, *pre*, means pre-task questionnaire.

Table 3.2 Data quality per participant group

Aspect	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Calibration	Good	Good	Good	Good
Gaze samples *(average)*	90.3%	92.1%	91.5%	81.4%
Responses to *pre*	Complete	Nearly complete	Complete	Complete
Responses to *post1*	Complete	Complete	Complete	Complete
Responses to *post2*	Complete	Complete	Complete	Complete
Field notes	Yes	Yes	Yes	Yes
Prolonged sound interference during experiment session	No	Yes*	No	No

** P14 and P15 preferred a higher volume for the video stimulus due to the sound of the heater, and P12 and P16 were bothered by the sound of the heater.*

All participants watched the whole video stimulus which means that full recordings with gaze data are generated for each subject. As shown in Table 3.2, the gaze data quality was assessed as satisfactory for each of these recordings. The data set of post-task questionnaire responses is complete. The data set of pre-task questionnaires is the only incomplete data set with one questionnaire missing for P04. Field notes are available for each participant.

Two types of field notes were generated during the empirical data collection. These comprised notes that relate to the experimental task, and notes in which the debriefings and experiment sessions are summarised. Field notes on the experimental task are available for each participant. Field notes regarding the debriefings and experiment sessions were not available for all subjects as these were only generated in cases when some important observation was made or when a participant wished to talk about the video stimulus or experiment session. Field notes unrelated to the experimental task were generated for all participants in Group 3: post-exchange and the Reference group. The ratios were lower for

the other two learner groups with 52.9% for Group 1: pre-exchange and 55% for Group 2: exchange.

3.5 Summary

This chapter explained the conceptual underpinnings of this thesis and presented an exploratory MMR design consisting of three stages to analyse JLLs' multimodal perception and visual attention while watching an excerpt of a Japanese variety show that features telops. These stages focus on a multimodal analysis of the chosen video material, the collection of empirical data through questionnaires, field notes and eye-tracking technology, and empirical data analysis. Two aspects in particular were emphasised in the research design. Leeway to analyse the data in an exploratory fashion in order to allow for observations to pave the way for subsequent analytical steps is one of these aspects. The second aspect related to an overall qualitative approach which allows for the multimodal nature of the video material to be addressed and for the empirical data, otherwise analysed from a quantitative perspective, to be explored from different angles.

This research focuses on JLLs at different stages of their Japanese language study. A total of 43 participants was recruited among students registered at DCU and organised into three groups. The number of years JLLs had formally studied the Japanese language at university, time spent on exchange in Japan, and data quality assessment were established as the three main selection criteria for the formation of coherent participant groups. A fourth group consisting of five native Japanese speakers recruited in Japan formed the Reference group with which JLLs' multimodal perception and visual attention is compared.

The following chapter delves further into the applied aspects of this study's research design. It describes the setting in which the empirical data was collected, and recounts all the steps taken to analyse the collated data.

4. METHODS AND PROCEDURES FOR EMPIRICAL DATA ANALYSIS

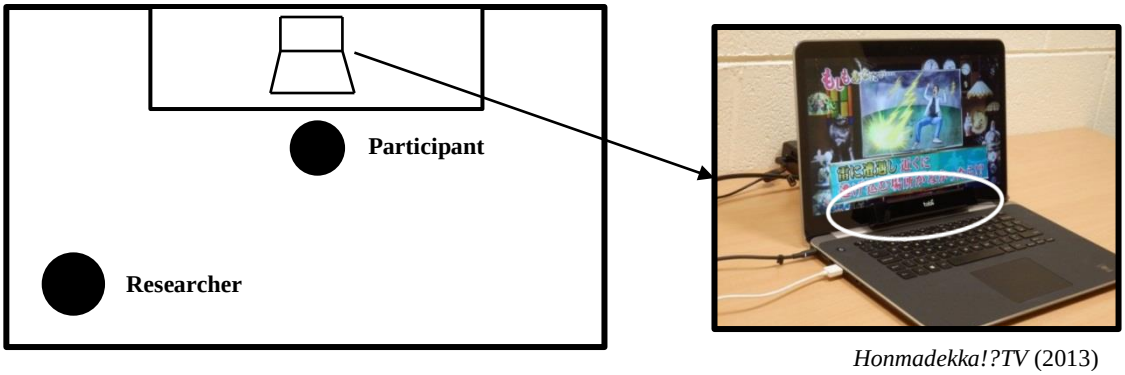
Chapter 4 is the second methodological chapter of this thesis. It demonstrates the implementation of the research design described in the previous chapter, recounts the empirical data collection, and gives a more in-depth explanation of the procedures that were followed for the empirical data analyses. The purpose of this chapter is to illustrate how the operationalised concepts from the previous chapter, multimodal perception and visual attention, have been applied to the collated data.

Section 4.1 starts with a description of the empirical data collection. This includes the actual conditions and experiment settings in which the data was collected, and the materials used to obtain the empirical data. Section 4.2 gives an in-depth account of the steps taken to analyse JLLs' multimodal perception. It explains how this study progressed from its conceptualisation to actual findings on viewer experiences and details the formatting of results in Chapter 6. Section 4.3 does the same for analyses of JLLs' visual attention. It provides an overview of analytical steps and explains the settings in the eye-tracking software for the majority of findings on viewer reception in Chapter 7.

4.1 Empirical data collection

4.1.1 Experiment setting

The experiment sessions were held in two rounds. The first round was conducted in January 2016 and started with the JLLs who were on exchange. Their sessions took place in rental classrooms in Tōkyō and Kyōto. The sessions with Japanese speakers were conducted in a classroom and an office at a Japanese university. The second round of experiment sessions took place from April to May 2016 with JLLs who resided in Ireland. These sessions were held in an interview room at DCU. Although the experiment sessions were conducted at more than one location, the setting has been kept the same as much as possible across sessions. Figure 4.1 shows the experiment set-up.



Honmadekka!?TV (2013)

Figure 4.1 Set-up of main experiment sessions

Further to the eye-tracking system, a spare pre-task questionnaire, a plain language statement, two copies of the informed consent form, a post-task questionnaire, a blank sheet, pens, a cleaning kit for glasses, make-up remover pads, tissues and a token of appreciation were prepared for each experiment session. The token of appreciation was sent by email to Japanese speakers after sessions. During the experimental task participants were seated at a table with the eye-tracking system placed in front of them. As shown on the right in Figure 4.1, the eye-tracker was attached to a laptop directly beneath the screen. The researcher was present in the same room but out of sight during the recording. The ethics forms and questionnaire were filled out at the same or another table depending on the facilities present in the experiment room.

All of the precautionary measures to protect data validity, mentioned previously in *Subsection 3.2.3 Stage 2*, were taken throughout the data collection. However, it appeared that some additional precautions needed to be taken for the experiment sessions in Dublin as another researcher was using the eye-tracking system during the same time period for a different type of experiment set-up. In order to maintain consistency with the experiment set-up for the sessions at DCU the location of the laptop on the table was marked with transparent scotch tape and the positioning parameters of the eye-tracker device on the laptop was saved as an .xconf file with the *X Configuration Tool* (Tobii AB, 2014a, pp. 15-18). This allowed for the experiment set-up to be recreated each time it had been changed. In addition, signs requesting not to disturb or enter the room were put up during experiment sessions at DCU.

Although the conditions in the experiment rooms were controlled for as much as possible, it appeared to be difficult to keep the experiment setting devoid from noise in some cases. In one of the rental rooms in Japan there was a heater on that had the tendency to grow

louder when the temperature in the room changed. Participants who sat through the experiment in this room were asked after the experimental task whether they were bothered by this background noise. A record of their answers was kept in the field notes. Further to this, it proved difficult to control noise outside of experiment rooms. Although it is recommended to choose particular times of the day when corridors are less crowded (Saldanha and O'Brien, 2014, pp. 138-139), it was not always possible to secure such time slots due to high demand on the eye-tracking system by other researchers and availability of experiment locations.

There was one calibration problem in particular that occurred a few times. This problem may have related to glasses with a particular coating. The eye-tracker could not detect a participant's eyes or could not run a satisfactory calibration when students were wearing such glasses. Experiment sessions with these students were therefore run without glasses. None of them appeared to have an issue with this as they confirmed that they were still able to see the video stimulus. In addition, some of them reported not wearing glasses when sitting behind their laptop.

The actual duration of experiment sessions occasionally deviated from the one-hour time slot as the time needed for the post-task questionnaire and debriefing varied between participants. In addition, some of the JLLs had forgot to fill out the pre-task questionnaire. They were invited to fill it out during the experiment session or to send it to me by email afterwards. Participants who were late for their session were asked to fill out both or one part of the post-task questionnaire in another room in order to make sure that the experiment schedule could resume without too much delay. Students who filled out the questionnaire in my absence were given the chance to ask questions in-between sessions. Generally speaking, the experiments in Japan were shorter due to budget constraints which meant that sufficient time could not always be allocated to debriefings. In addition, the time available for the sessions with Japanese speakers was limited. These participants could not be contacted in advance of the experiment sessions as they were recruited on the spot. They filled out the pre-task questionnaire and post-task questionnaire in another room but were also given the chance to ask me questions in-between experiment sessions.

4.1.2 Questionnaires

Two questionnaires were developed for the empirical data collection; one pre-task questionnaire and one post-task questionnaire. These documents were prepared in both

English and Japanese, and checked for their wording by native speaker postdoctoral researchers before finalising their design. Questions focused on participants' background and viewer experience. More specifically, the pre-task questionnaire focused on subjects' linguistic and demographic background while the post-task questionnaire included a comprehension test and questions regarding participants' perceptions of the video stimulus, experience with television programmes, general viewing habits and perceptions of Japanese variety shows.

Three versions of the pre-task questionnaire (*Appendix A. Pre-task questionnaire*) were prepared. These versions did not include any major differences; rather, certain questions were left out or reworded for JLLs on exchange in Japan and Japanese speakers. Considering that JLLs on exchange resided in Japan, it was discerned that this circumstance needed to be accounted for within the questions. In addition, questions relating to Japanese language learning were only asked to JLLs and omitted from questionnaires directed at Japanese speakers.

Two versions of the post-task questionnaire (*Appendix E. Post-task questionnaire*) were developed; one for JLLs and one for Japanese speakers. As mentioned, the main difference between the two are the questions on Japanese language learning. Almost all of these questions were developed for this study specifically. Some of the questions on media exposure to television programmes, general viewing habits and views on telops from the questionnaire used for the ICE project and pilot study were kept. Modified versions of these questions have been incorporated in the post-task questionnaire.

The post-task questionnaire was divided into two parts to break up the large amount of questions. One of the postdoctoral researchers had suggested providing headings in the questionnaire as a support to participants. Sets of questions on the same topic were therefore bundled together and provided with headings. In addition, it was opted to sequence the questions in such a way that the questions proceed from general to specific information and from behavioural to attitudinal questions in order to lower the chance of inconsistencies in participants' responses (Brace, 2004, p. 32). By having subjects think about their behaviour first before giving their opinion it was expected that this would enable them to ease into the different topics without contradicting themselves on their general viewing habits and attitudes (2004, pp. 32-34). Furthermore, in order to prevent bias in the formulation of semantic differential scales, positive and negative statements

were included at both ends of the five-point scales (2004, pp. 65-67). The comprehension test was included at the start of the first part of the questionnaire and filled out directly after the experimental task when the input is still fresh in the minds of participants. A blank sheet was offered to the JLLs when answering these questions in case they wished to write down their thinking process.

Throughout the post-task questionnaire the term "text on screen" was used instead of "telop". It was determined that it would be better not to overcomplicate the questionnaire with definitions and to use a more generic term to refer to the intralingual text. In addition, the terms "caption" and "subtitle" were used in questions on media exposure to television programmes.

4.1.3 Eye-tracking system

A portable Tobii X2-60 eye-tracking system was used for the experiment sessions. This eye-tracking system was chosen as it can be easily transported, is presumably less distracting to participants due to its small size and is more tolerant to head movements than stationary eye-tracker devices. The Tobii X2-60 eye-tracker device was attached to a high-specification *Dell* laptop. This laptop had the *Tobii Studio* software installed on it.

The "Media exposure to Japanese TV programmes"[11] section of the second part of the post-task questionnaire was reviewed for JLLs and native Japanese speakers who had reported watching Japanese television programmes in order to check whether the use of a laptop had created unusual viewing conditions to subjects. The responses showed that 77.8% of JLLs make most use of a laptop when watching Japanese television programmes while 22.2% made most use of another device. Only 8.3% of participants reported making equal use of another device in addition to the laptop and 5.6% of JLLs usually use a TV set. These JLLs are all participants in Group 2: exchange. It appeared, however, that the use of a TV set to watch Japanese television programmes is not exclusive to this particular learner group. Questionnaire responses for the Japanese speakers indicated that all participants make most use of a TV set while 20% of them responded that they make equal use of a laptop.

[11] This section is included as 日本語のテレビ番組の経験について [Regarding experience with Japanese television programmes] in *post2* for Japanese speakers.

This suggests that the portable Tobii X2-60 eye-tracking system was appropriate to the circumstances in which the fieldwork was conducted and the JLLs who are accustomed to watching Japanese television through a laptop. Japanese speakers, on the other hand, generally use a TV set to watch such programmes. The laptop may therefore have been an unusual viewing condition to them.

4.2 Analysis of Japanese language learners' perceptions

4.2.1 Evaluation of post-task questionnaires and field notes

Post-task questionnaires and field notes regarding the debriefings and experiment sessions were examined during the data familiarisation phase. It is important to note that the researcher was involved in the creation of data items to different degrees; open-ended responses were produced entirely by participants for whom the researcher only cleared up certain questions upon request, while the observations from field notes were established through an interactive process and noted down by the researcher directly after each experiment session. Participants were given free rein throughout an experiment session in how much they wished to share. This explains why certain data items are missing or contain insights into thought processes that diverge from original questions asked to participants.

Even though post-task questionnaires were checked for missing responses directly after they were filled out, some missing responses have been detected. However, these instances did not affect the study to a great extent as they appeared to be few. Almost none of the participants made use of the blank sheet that was offered to subjects when answering the comprehension test. Only one participant in Group 2: exchange made use of it for a very brief note. This note consisted of a schematic representation of numbers that was similar to the one used in a number of telops in the video stimulus.

As noted, there was less time available for the debriefing in experiment sessions conducted in Japan. This means that fewer field notes were generated on topics other than the experimental task for participants in Group 2: exchange and the Reference group. It appeared that the field notes gathered for this study related to a variety of topics. These included general comments made by participants on the experiment session, questionnaires, their general interests, their experience with watching AV materials, and their Japanese linguistic skills. Participants also made more specific comments regarding the video

stimulus such as their understanding of the video material or its components, their observations on types of input, and their thoughts on the use of Japanese variety shows in class. There was one participant who reported hearing of impact captions previously in class at DCU. The remaining field notes consisted of my own observations. These related to instances during which I was asked a question on the contents of the video stimulus or saw a student laugh during the experimental task. Reminders of my absence during the experiment session when participants filled out a part or both parts of the post-task questionnaire, and observations on other aspects of an experiment session were also included.

4.2.2 Meaningful study of perception

Upon reviewing the responses to post-task questionnaires and field notes, it was decided that some of the questions were not appropriate to the study of JLLs' perceptions after all. In a similar way as the questions that were deemed to be unsuited for participant selection, it was discerned that some of the questions were not related to the research focus. The (thematic) analysis of questionnaire responses and field notes is geared towards the identification of patterns in participants' multimodal perception. It aims to gain a deeper understanding of the meanings that JLLs attach to sensory information that is not exclusively verbal and the perceived usefulness of authentic Japanese variety shows in the Japanese language classroom.

Questions that asked participants to recall the colours or fonts that were used for telops, and self-assessments of the amount of telops participants had read are not taken up in the remainder of this thesis. It was determined that the former tests participants' memory or recollection of visual-nonverbal input, not their perceptions of this particular type of input. The latter was also considered to be unrelated to JLLs' perceptions. This type of question is important when trying to work out whether participants' ideas of their own gaze behaviour coincides with their actual gaze behaviour. The data familiarisation phase showed that all participants look at telops. Although it is not possible to say with certainty whether participants' gaze behaviour shows signs of reading or just looking, it was expected that this question would not give any further insights into this particular observation or JLLs' perceptions.

4.2.3 Procedure for closed-ended data items

The data set for the analysis of closed-ended data items consists of responses to ten closed-ended questions. Although half of these questions related to perceptions of telops, the other five questions extend the discussion to other types of input and broader topics. The questions are included in the post-task questionnaire and are as follows:

post1 Q13 Did you feel the drawings in the experiment video helped you with understanding its contents?[12]

post1 Q18 What did you think of the difficulty level of the text on screen?[13]

post1 Q19 What helped you with text you did not understand?[14]

post1 Q22a Were there any colours or colour combinations you found difficult to read?[15]

post1 Q20 What did you think of the legibility of the text on screen?[16]

post1 Q25a Did you think the design of the text on screen suited the variety show?[17]

post2 Q19 Would you prefer text on screen to be white and without effects?

post2 Q20 Do you think watching Japanese TV shows helps you with learning Japanese?[18]

post2 Q21 Would you like Japanese variety shows to be used in Japanese language classes?[19]

post2 Q22 Do you think you will watch Japanese variety shows more often after this experiment?[20]

The closed-ended data items were manually counted. Findings from these data items are represented in tables. The wording in these tables have been adjusted slightly as compared to the original questions and answer options due to spacing issues. However, the essence has been kept the same for each aspect and answer option in all of the tables. Hyphens are shown instead of values for some aspects for the Reference group. This does not mean that these aspects are not relevant to Japanese speakers; rather, hyphens indicate that questions on these aspects were not included in the post-task questionnaire for the Reference group.

[12] This question is included as *post1* Q11 for Japanese speakers.
[13] This question is not included for Japanese speakers.
[14] This question is not included for Japanese speakers.
[15] This question is included as *post1* Q18a for Japanese speakers.
[16] This question is included as *post1* Q20 for Japanese speakers.
[17] This question is included as *post1* Q21a for Japanese speakers.
[18] This question is not included for Japanese speakers.
[19] This question is not included for Japanese speakers.
[20] This question is included as *post2* Q20 for Japanese speakers.

4.2.4 Procedure for diverging stacked bar charts

One question included in the post-task questionnaire involved semantic differential scales to collect attitudinal data on telops. This question contained a list of ten statements. Participants were asked to rate their attitudes towards these statements on a five-point scale. Topics of these statements relate to distraction, understanding, confusion, visual clutter, humour, focus of attention, necessity, identification of key points, interpretation of actions and utterances of people appearing on the show, and saliency. The question reads as follows:

post2 Q18 What do you personally feel about the text on screen in Japanese variety shows? They:

Distract me	1	2	3	4	5	Do not distract me
Help me understand the show	1	2	3	4	5	Do not help me understand the show
Confuse me	1	2	3	4	5	Do not confuse me
Clog up the screen	1	2	3	4	5	Do not clog up the screen
Make the show funnier	1	2	3	4	5	Do not add to the comedic nature of the show
Help me retain my attention	1	2	3	4	5	Make me lose my attention
Are indispensible [sic]	1	2	3	4	5	Can be left out from the show
Help me identify important information	1	2	3	4	5	Do not help me identify important information
Change my interpretation of what people on the show do and say	1	2	3	4	5	Do not change my interpretation of what people on the show do and say
Are flashy	1	2	3	4	5	Are plain and simple

The findings from the semantic differential scales in *Subsection 6.1.1 Observations for telop* have been organised into diverging stacked bar charts.[21] These were created in *Excel*. As noted, positive and negative statements were included at both ends of the five-point scales in the original question to prevent bias in its formulation. All of the diverging stacked bar charts show negative statements on the left and positive statements on the right. This means that the scales for some statements have been reversed for the presentation of findings. Take, for example, the five-point scale for the statement on understanding. The

[21] For more information, please see Evergreen (2019).

original question includes “help me understand the show” on the left and “do not help me understand the show” on the right. For such statements number 1 was counted as number 5 and number 2 was counted as number 4, and vice versa.

4.2.5 Procedure for open-ended data items

The data set for the thematic analysis consists of responses to five open-ended questions and all of the field notes that related to the debriefings and experiment sessions, not the experimental task. These include the following open-ended questions from the post-task questionnaire:

post1 Q20	What kind of effects did you notice were used for the text on screen?[22]
post1 Q22b	Please specify.[23]
post1 Q25b	Please explain you answer.[24]
post1 Q26	Did you feel the questionnaire allowed you to say everything you wanted to say? Is there anything you think might be relevant to mention in regard to the experiment video or Japanese text on screen?[25]
post2 Q23	Did you feel the questionnaire allowed you to say everything you wanted to say? Is there anything you think might be relevant to mention in regard to TV programmes or Japanese language learning?[26]

As noted, questions relating to memory were considered to be unsuited for the analysis of JLLs’ perceptions. The same argument can be made for the first question in this list (i.e. *post1* Q20). Nonetheless, of those questions that had tested participants’ recollection, this was the only one that had generated responses that departed from the original question. Responses included comments that went beyond the noticing of effects and formulated perceptions on the functions and interconnectedness of input and their influence on the viewer experience.

Prior to the exploration of the data set a mind map was created. This mind map was based on the researcher’s impression of the responses and field notes after repeated reading of the data items. This mind map represented keywords taken from the data set and paraphrases. Lines were drawn between some of these keywords to represent potential links that the

[22] This question is included as *post1* Q16 for Japanese speakers.
[23] This question is included as *post1* Q18b for Japanese speakers.
[24] This question is included as *post1* Q21b for Japanese speakers.
[25] This question is included as *post1* Q22 for Japanese speakers.
[26] This question is included as *post2* Q21 for Japanese speakers.

researcher thought could exist within the data. “Telop and typography”, “multimodality” and “language learning material” were the main keywords. These were linked to smaller keywords such as “emotions”, “key information”, “fast-paced”, “certain proficiency level perhaps needed”, “understanding of general feel” and “authentic language use”. There appeared to be some contradiction within the data set as both positive and negative aspects were given by participants. The researcher’s impression of the data set was that multimodality, or more specifically telop design, played an important role in JLLs’ understanding of the video material. It was also observed that some aspects of the Japanese variety show may have been a bit challenging to some JLLs.

Data was collated into *Microsoft Word* (Braun and Clarke, 2013, p. 204; Moran, 2017) with data items structured according to participant group and participant number. *Microsoft Word* was chosen over a Computer-Assisted Qualitative Data Analysis Software (CAQDAS) as the data set was smaller and less complex as compared to types of data that are typically associated with the use of a CAQDAS (e.g. interviews). It was opted not to structure the initial coding file according to the open-ended questions and field notes in order to bolster initial coding that cuts across the data set (Braun and Clarke, 2013, p. 227). However, considering the brevity of some responses, the original formulation of questions was kept in mind during the initial coding process in order to identify the context of responses and to work out referents within data items (e.g. “it” referring to “telop design”). Figure 4.2 shows an excerpt of the initial coding file. Abbreviations for the post-task questionnaire have been included with *post1* referring to part one of the post-task questionnaire and *post2* meaning part two of the post-task questionnaire.

Participant group	Participant No.	Data item		Initial code
Group 1: pre-exchange	**P22**	*post1* Q20	3D style text that popped up on screen as it was being said	
		post1 Q22b	The white words that had only one red highlighted word every few sentences.	
		post1 Q25b	The show was somewhat loud and fast-paced. The large & brightly coloured 3D style text was suited as it draws in the viewer's eyes.	
		post1 Q26	There was alot of text, and there was at times not enough screen time for me to try and read certain kanji that I was unsure of.	
		post2 Q23		
		Field notes	P22 tried to concentrate on mouths until P22 noticed the telops popping up. P22 found that green and yellow were easier to read than red and white (on the screens filled with text). P22 found that the use of the box/diagram when explaining the walls facing the outside air was very helpful.	
	P24	*post1* Q20	3d, different colors, big, Bold, white text with colorful Borders.	
		post1 Q22b		
		post1 Q25b	it was quirky and radical, and the guests seemed to have radical comedic reactions to the info. that was presented So a formal/serious text wouldn't have suited.	
		post1 Q26	I find as I'm learnig Japanese, the subtitles/text, contestant body language and speech all coupled together help in the understanding of the show. If one or two would be missing it would be very difficult to understand it	
		post2 Q23	Yeah, Japanese variety shows seem funny and I will definitely look into watching them alot more. They'll help me with my Japanese and will help in understanding people who speak Japanese quickly.	
		Field notes	P24 already knew quite a lot about eye-tracking due to interest in psychology and read a couple of studies on eye-tracking. P24 found the colour blue a bit tricky to read.	

Figure 4.2 Excerpt of initial coding file

The data items either consist of a response to an open-ended question or a set of observations from the field notes. Different notation styles were used for open-ended responses and field note observations. The contents of data items containing questionnaire responses are identical to the original written entries. This means that the spelling, grammar and punctuation have been kept the same in the initial coding file, and that text that had been crossed out in the original response is also omitted from the data set. Sentences were only broken up in cases where the participant appeared to deliberately start another train of thought in a new line irrespective of the space available in the response box. For cases in which this did not apply, questionnaire responses are represented as continuous streams of text. The use of gendered third person pronouns has been avoided for the field notes to further enhance the anonymity of participants.

The initial coding file consists of 288 data items. Of these, 204 data items (70.8%) contained responses and 84 data items (29.2%) were left blank by participants. It appeared that 84 out of those 204 data items (41.2%) with responses contained topics that were unrelated to the research focus. These 84 data items were left black in the initial coding file and were accompanied with a short description of their contents, such as “noticed effects used for telop”, “repetition of closed-ended response”, “note on interests”, “expression of content regarding the questionnaire” and “note on experiment session”. Initial coding was conducted on the remaining 120 data items (58.8%).

4.2.6 Procedure for initial codes

As noted previously, initial coding happened in two rounds. The first round was conducted in a tentative manner to gain a deeper understanding of the contents of the 120 data items. This means that the initial coding was not yet definitive at this stage and that its purpose was to help determine what kind of nuances existed within and between data items, and what kind of coding scheme would be appropriate for the examination of trends in JLLs’ multimodal perception. After the whole data set was provided with provisional labels it was read and revised repeatedly until a satisfactory scheme for initial codes was devised. It was determined that several nuances needed to be kept between initial codes in order to stay close to the subtle differences noticed within the data set and to address the analytic interests as well as possible. In order to unpack JLLs’ perceptions of different types of input in the video stimulus it was decided to apply initial codes to data items that make the following four distinctions between aspects:

1. Telop design as a whole versus individual components of telop design
2. Perceived interrelatedness of types of input versus perceived functions of individual types of input
3. JLLs' viewer experience of the video stimulus versus their viewer experience of individual types of input
4. JLLs' understanding of the video stimulus versus their understanding of individual types of input

It was also determined that colours should be given to initial codes to help the identification of themes based on the patterns that the researcher identified upon finalising the first round of initial coding. Codes relating to the coherence of types of input were yellow, perceived functions of types of input were coded blue, JLLs' viewer experience of combined or individual types of input were green, JLLs' understanding was coded in orange, and pink was used for text that dealt with Japanese television programmes and Japanese Language Acquisition (JLA). It was decided that initial codes for each colour would be given a number to differentiate between aspects. This numbering follows the order of occurrence in the data set.

The second round of initial coding, which consisted of the actual coding of the data set, involved the refinement and application of the coding scheme. The coding scheme consisted of 28 initial codes. These initial codes segmented the data set into groups of data extracts that dealt with similar topics and filtered out text from the data set that was unrelated to the analytic interests. Figure 4.3 shows an excerpt of the initial coding file with initial codes.

Participant group	Participant No.	Data item		Initial code
Group 1: pre-exchange	**P22**	*post1* Q20	3D style text that popped up on screen as it was being said	l. 1: Noticed effect used for telops l. 1: Green 1: appearance of telops linked to audio-verbal and audio-nonverbal input
		post1 Q22b	The white words that had only one red highlighted word every few sentences.	ll. 1-2: Reference to telops that bridge scene change (i.e. screen filled with text)
		post1 Q25b	The show was somewhat loud and fast-paced. The large & brightly coloured 3D style text was suited as it draws in the viewer's eyes.	l. 1: Yellow 1: impression of the Japanese variety show ll. 1-2: Yellow 2: aspects of the telop design that suited or did not suit the Japanese variety show l. 2: Blue 1: perceived effects of the telop design
		post1 Q26	There was alot of text, and there was at times not enough screen time for me to try and read certain kanji that I was unsure of.	ll. 1-2: Green 2: experienced difficulties with telops
		post2 Q23		
		Field notes	P22 tried to concentrate on mouths until P22 noticed the telops popping up. P22 found that green and yellow were easier to read than red and white (on the screens filled with text). P22 found that the use of the box/diagram when explaining the walls facing the outside air was very helpful.	l. 1: Green 3: focus on or preference for certain input ll. 1-2: Green 4: experienced competition between input l. 2: Green 5: legibility of telops l. 3: Reference to telops that bridge scene change (i.e. screen filled with text) ll. 3-5: Orange 1: telop design described as helpful to understanding
	P24	*post1* Q20	3d, different colors, big, Bold, white text with colorful Borders.	l. 1: Noticed effects used for telops
		post1 Q22b		
		post1 Q25b	it was quirky and radical, and the guests seemed to have radical comedic reactions to the info that was presented So a formal/serious text wouldn't have suited.	l. 1: Yellow 3: impression of the telop design ll. 1-3: Yellow 1: impression of the Japanese variety show
		post1 Q26	I find as I'm learnig Japanese, the subtitles/text, contestant body language and speech all coupled together help in the understanding of the show. If one or two would be missing it would be very difficult to understand it	ll. 1-4: Orange 2: combination of input described as helpful to understanding
		post2 Q23	Yeah, Japanese variety shows seem funny and I will definitely look into watching them alot more. They'll help me with my Japanese and will help in understanding people who speak Japanese quickly.	l. 1: Yellow 1: impression of the Japanese variety show ll. 1-2: Repetition of closed-ended response ll. 2-3: Pink 1: expected positive outcomes of watching Japanese television to JLA
		Field notes	P24 already knew quite a lot about eye-tracking due to interest in psychology and read a couple of studies on eye-tracking. P24 found the colour blue a bit tricky to read.	ll. 1-2: Note on interests ll. 2-3: Green 5: legibility of telops

Figure 4.3 Excerpt of initial coding

The data extracts are highlighted in the data items with colours of the corresponding initial codes. Initial codes are represented with their names and locations in the data item through line numbers. As shown in Figure 4.3, initial code names consist of a colour, a number and a description of the common topic of the data extracts assigned to it (e.g. Green 1: appearance of telops linked to audio-verbal and audio-nonverbal input). Text in black represents those segments that were discarded prior to the initial coding process and segments that were left uncoded during the two phases of initial coding.

A total of 174 data extracts of varying lengths were taken from the 120 data items and assigned to the initial codes. The scheme for initial codes and the number of data extracts they include are shown in Table 4.1.

Table 4.1 Scheme for initial codes

Colour	Code No.	Description	No. of extracts*
Yellow	1	Impression of the Japanese variety show	15
	2	Aspects of the telop design that suited or did not suit the Japanese variety show	6
	3	Impression of the telop design	6
	4	Aspects of the Japanese variety show that the telop design resembled	9
Blue	1	Perceived effects of the telop design	8
	2	Perceived functions of colour in the telop design	9
	3	Perceived functions of telop and telop design	6
	4	Perceived functions of font in the telop design	11
	5	Perceived effects of font in the telop design	2
Green	1	Appearance of telops linked to audio-verbal and audio-nonverbal input	9
	2	Experienced difficulties with telops	8
	3	Focus on or preference for certain input	7
	4	Experienced competition between input	8
	5	Legibility of telops	9
	6	Experienced difficulties with the combination of input	3
	7	Perceived resemblances between input	6
	8	Perceived importance of certain input	2
Orange	1	Telop design described as helpful to understanding	12
	2	Combination of input described as helpful to understanding	4
	3	Telop described as helpful to understanding	3
	4	Potential risks to misunderstanding input	5
	5	Self-reported understanding of the Japanese variety show	9
Pink	1	Expected positive outcomes of watching Japanese television to JLA	8
	2	Reservations regarding the use of Japanese television for JLA	4
	3	Aspects of Japanese television that are perceived to be helpful to JLA	5
	4	Perceived better alternatives to Japanese variety shows for JLA	2
	5	Self-proclaimed experience with learning FLs through television	2
	6	Aspects that are perceived to be missing in current Japanese language classes	3

** The total number of extracts amounts to 181 as seven extracts are counted twice considering these were ascribed to two initial codes. Six of these are in blue and one is in orange.*

It appears that green contains the highest number of data extracts (28.7% of data extracts). This is followed by yellow (19.9% of data extracts), blue (19.9% of data extracts), orange (18.2% of data extracts) and pink (13.3% of data extracts). Separate overviews of each initial code with all corresponding data extracts were then prepared. Based on these overviews, a separate document was created in which the essence of each initial code was briefly defined in preparation of the search for (sub)themes. These definitions are included in *Appendix G. Description of initial codes*. The definitions of the initial codes for each colour can be summarised as follows:

Yellow Extracts that illustrate participants' perceptions of the coherence between the image of the video material and the telop design.

Blue Extracts that illustrate participants' perceptions of typographical features and telops.

Green Extracts that illustrate participants' perceptions of their viewer experience of types of input and the multimodal nature of the video stimulus.

Orange Extracts that illustrate participants' perceptions of their understanding of different types of input and the video material.

Pink Extracts that illustrate participants' perceptions of the usefulness of Japanese television programmes to JLA and their attitudes to current Japanese language classes.

This suggests that most of the data extracts relate to participants' perceptions of their viewer experience (green), telops (blue) and overall impression of the video material (yellow). It was expected that further investigation of these topics would provide a good starting point for the search of (sub)themes.

4.2.7 Procedure for (sub)themes

The search for (sub)themes happened in two phases and started with the creation of a list of common issues found within data extracts across the initial codes. As can be seen in Figure 4.4, these common issues comprised recurrent topics (e.g. highlight or makes (main) information stand out) and adjectives (e.g. bright, light-hearted and eye-catching) in particular.

Initial code			No. of extracts	Issues discussed	Candidate (sub)theme
Yellow	1	Impression of the Japanese variety show	15	- Loud / fast-paced (1x) - Fun / funny / light-hearted / wacky (3x) - Formal / serious telop design would not suit (2x) - It is a funny-serious combination (1x) - Guests respond in a radical / comedic way (1x) - Telops seem to be an important component (1x) **6 extracts discarded**	**1.2** Reflects the progression of the Japanese variety show **1.1** Reflects the mood of the Japanese variety show
	2	Aspects of the telop design that suited or did not suit the Japanese variety show	6	- Fast-paced / sensational telop design (1x) - Big / bright / comical / varied telop design (3x) - Fun / exciting telop design (1x) **1 extract discarded**	**1.2** Reflects the progression of the Japanese variety show **1.1** Reflects the mood of the Japanese variety show
	3	Impression of the telop design	6	- Fast-paced / eye-catching (2x) - Fun / comical (2x) - Casual (1x) - Quirky / radical (1x) **1 extract discarded**	**1.2** Reflects the progression of the Japanese variety show **1.1** Reflects the mood of the Japanese variety show
	4	Aspects of the Japanese variety show that the telop design resembled	9	- Speed of telops reflects fast progression (2x) - Telop design reflects eye-catching / attention-seeking nature (1x) - Telop design reflects bold / intense content (1x)	**1.2** Reflects the progression of the Japanese variety show
				- Fun / light-hearted / dynamic typography reflects mood of the show (2x) - Bold / bright / colourful / big typography reflects humour of the show (2x) - Bright typography reflects the set (1x)	**1.1** Reflects the mood of the Japanese variety show
Blue	1	Perceived effects of the telop design	8	- Telop design is eye-catching / impactful (2x) - Telop design makes useful information more memorable (1x)	**1.2** Reflects the progression of the Japanese variety show
				- Telop design helps set the mood (1x) - Telop design helps with the delivery of jokes (1x)	**1.1** Reflects the mood of the Japanese variety show
				- Telop design makes it possible to guess the tone of voice of speech (1x) - Telop design helps emulate the expressions of speakers (1x)	**1.1.2** Contains cues to paralinguistic properties of audio-verbal input
				- Telop design helps emulate the emotions of speakers (1x) - Telop design helps exaggerate the content (1x) **1 extract marked as miscellaneous**	**1.1.1** Contains cues to non-linguistic properties
	2	Perceived functions of colour in the telop design	9	- Colour differences highlight or makes (main) information stand out (5x) - Colour differences are used for emphasis (2x)	**1.2.1** Contains cues to prosodic properties of audio-verbal input
				- Colour differences indicate different speakers (3x) - Colour differences indicate different emotions (1x)	**1.1.1** Contains cues to non-linguistic properties

Figure 4.4 Excerpt of issues discussed per initial code

The common issues are accompanied with a number in parentheses to indicate how often these or similar expressions were encountered in an initial code as an indicator of their prevalence. The colour scheme was further applied to these common issues in order to work out overlap and relations between initial codes. If, for example, an issue on

typographical features or telops in blue also relates to JLA then the colour pink would be used in the column for issues discussed. Different hues of colours were used for any new relationships identified across the initial codes. Descriptions of common issues were kept similar across initial codes to ease the identification of candidate (sub)themes. These descriptions were continuously adjusted throughout this first phase of (sub)theme search to accommodate subsequent data extracts while working through the initial codes.

Data extracts on topics or perceptions that were not patterned were temporarily put in a file with miscellaneous data extracts or placed in a document with discarded data extracts (Braun and Clarke, 2006, p. 90; Moran, 2017). The five initial codes that were excluded altogether did not relate to patterns that the researcher identified and were not internally cohesive as they included data extracts with diverging opinions or viewer experiences. These initial codes were Green 5: legibility of telops, Orange 2: combination of input described as helpful to understanding, Orange 4: potential risks to misunderstanding input, Pink 4: perceived better alternatives to Japanese variety shows for JLA, and Pink 5: Self-proclaimed experience with learning FLs through television. Upon revisiting the data extracts included within these initial codes it also became clear that these not always provided sound proof for the claims that the codes represent. At the end of this stage, 24 extracts were discarded and 49 extracts were marked as miscellaneous. This means that the draft thematic map was constructed based on the remaining 101 data extracts.

It became clear during this first phase of (sub)theme search that blue comprises coherent cohorts of data extracts. It not only links aspects of telop design to their perceived functions; it also describes such aspects as suited to the Japanese variety show of the video material and discusses them in relation to JLA. This became apparent due to the colours in the issues discussed columns of initial codes in yellow, blue and green. The data extracts in Blue 1-5 therefore formed the basis of the draft thematic map. As is shown in Figure 4.4, the candidate (sub)themes in this draft thematic map relate telop design to the mood and progression of the Japanese variety show, non-linguistic input, paralinguistic and prosodic properties of audio-verbal input, context, and authentic Japanese language use. Although this draft thematic map still needed to be considered against the discarded and miscellaneous data extracts, it was determined that the data extracts included in these candidate (sub)themes were likely not to be discarded at a later stage.

A separate document was created in which the data extracts from the draft thematic map were organised according to their corresponding (sub)themes. At this stage the second phase of the (sub)theme search had started. The data extracts for each candidate (sub)theme and the labels given to each of the (sub)themes were checked. Data extracts were moved around between candidate (sub)themes until the (sub)themes were coherent. One of the discarded data extracts and 14 data extracts that were marked as miscellaneous were added to the scope of extracts for the narrative of the thematic map. Two data extracts that were initially included within the scope were discarded in the end. The candidate (sub)themes were then checked against the data set. The overviews of candidate (sub)themes and data extracts were printed and cut into smaller chunks. Chunks that were considered to be more suited elsewhere in the draft thematic map were repositioned before fixing the names of (sub)themes and thematic networks. This resulted in the final thematic map.

The final thematic map was constructed based on data extracts from 84 out of 120 data items. A total of 114 out of 174 extracts were taken from these data items for the creation of the thematic map. *Appendix H. Representation of initial codes in thematic map* shows how these extracts are distributed between the initial codes. This means that ultimately an additional 36 data items were discarded during the formation of (sub)themes and that a total of 60 data extracts were discarded by the end of the search for (sub)themes.

4.2.8 Procedure for thematic map

After reviewing the entire data set against the final thematic map it was decided that two uncoded segments that had been discarded prior to the initial coding process, i.e. segments that had been left black, needed to be included as they related to the (sub)themes. The thematic map therefore represents a total of 114 data extracts (66.7% of data extracts) and two uncoded segments from 84 data items (70% of the data set). JLLs are represented in 91.7% of these data extracts while 8.3% of extracts are from Japanese speakers in the Reference group. A more detailed breakdown of proportions per (sub)theme is included in *Table 6.3 Representation of participant groups in (sub)themes*. Even though a small proportion of extracts in the thematic map is derived from the Reference group, the thematic networks are structured based on a data extract of one of the Japanese speakers as it encapsulates the essence of JLLs' perceptions well:

Reference group, JP05

テロップは色や存在自体もそうですが、動きや動きのスピードが見ようとする意識につながるように感じました。 [I felt that telops are more than just their colours and presence; they also show movement and speed of movement.]

post2 Q21, ll. 1-2

JP05 pointed out that telops are not only defined by their linguistic contents; they also comprise visual-nonverbal types of input or telop design which consists of typography (i.e. colour and font), display rate (i.e. ways of appearing on screen), and movement. It became evident during the thematic analysis that the aspects mentioned by JP05 were described and discussed in participants' responses throughout the data set, and that these appeared to have played an important role in subjects' viewer experience. It was therefore decided to structure the thematic networks in the thematic map around these three components of the telop design (i.e. typography, display rate, and movement). More specifically, the three components were used to create two thematic networks. As shown in *Figure 6.5 Organisation of data extracts and uncoded segments into (sub)themes*, Thematic Network 1 focuses on typography while Thematic Network 2 concentrates on display rate and movement.

The first thematic network includes 67 data extracts and one uncoded segment from 52 data items. The second thematic network comprises 47 data extracts and one uncoded segment from 38 data items. It appeared that some of the data extracts in the thematic networks contained comments on more than one (sub)theme. These were therefore further divided and split into smaller units. This resulted in a total of 71 data extracts for the first thematic network as four extracts were split and 50 data extracts for the second thematic network as three data extracts from the first network overlapped with the second network. The total number of data extracts therefore amounts to 121. The total number of data items amounts to 90 as six items are included in both thematic networks.

Data extracts that best represent the (sub)themes are presented in the narrative in Chapter 6. These are presented according to data item, which means that one data item in the narrative can include more than one data extract. The narrative comprises 49 data extracts and two uncoded segments taken from 38 data items. The remaining 69 data extracts from 58 data items are included in *Appendix I. List of data extracts*. This excludes the beforementioned data extract of JP05. Three suspension points indicate that parts of a data extract have been omitted. As noted, the spelling, grammar and punctuation in the data extracts are identical to those encountered in the original written entries in open-ended data items.

It is important to note that the post-task questionnaire used the term “text on screen” instead of “telop”. This means that the type of telop participants referred to had to be inferred from their responses. During the multimodal analysis it was determined that the great variety of typographical features, display rate and movement are aspects that distinguish telops that directly render dialogue from other types of telop. It is therefore argued here that the thematic map relates to telops that directly render dialogue.

4.3 Analysis of Japanese language learners’ gaze behaviour

4.3.1 Evaluation of gaze data

The *Tobii Studio* project of this study was subdivided into four tests in which participants’ recordings were stored by academic year and order of data collection. One test comprised the data of JLLs on exchange and Japanese speakers, another test consisted of recordings of first-year JLLs, a third test contained gaze data of JLLs in the second academic year, and the fourth test comprised recordings of fourth-year JLLs. The gaze videos were replayed during the data familiarisation phase to check participants’ gaze paths. It appeared that the quality of gaze data varied within and between recordings. Although this variation in gaze data quality was deemed to be in an acceptable range, there were two observations in particular that needed to be taken into account for the generation of dynamic AOIs and the examination of gaze replay videos during the data exploration phase. The first one of these observations is shown in Figure 4.5.

Group 2: exchange, P01

Group 1: pre-exchange, P24

Group 1: pre-exchange, P26

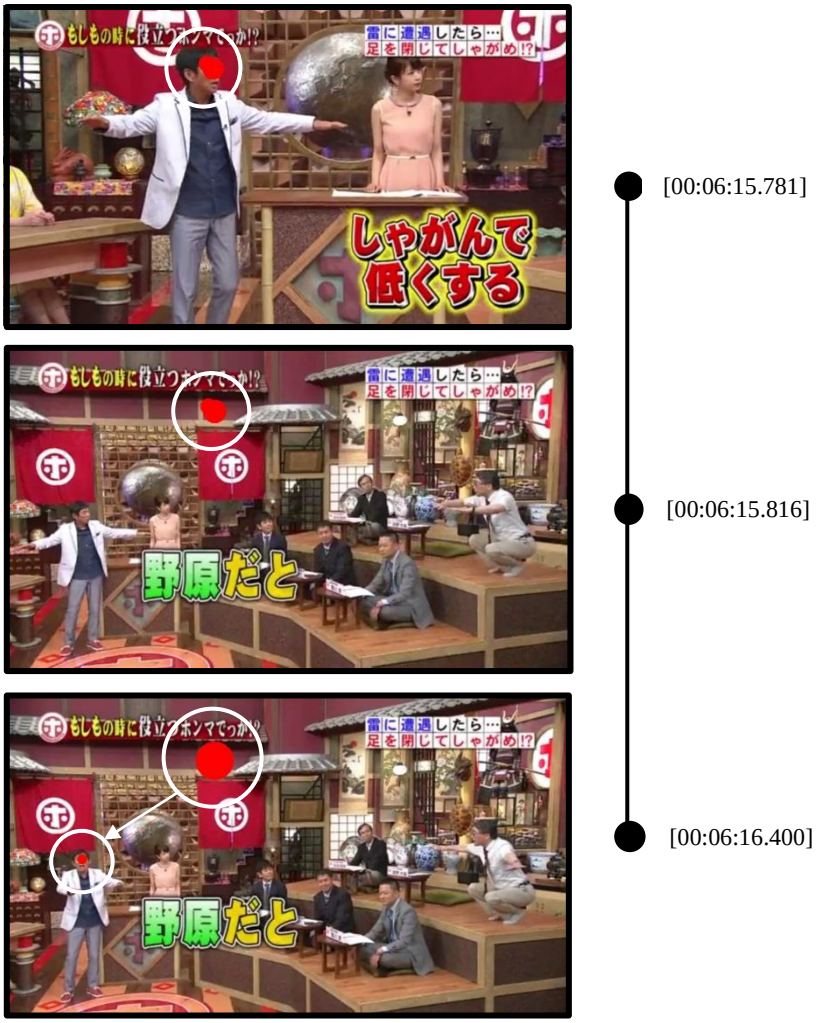
Group 2: exchange, P06

Honmadekka!?TV (2013)

Figure 4.5 Observations for accuracy

It was noticed that the gaze paths of some participants do not always land on particular types of input. As is shown in Figure 4.5, gaze paths of P26 and P06 sometimes hover above telops or next to faces as opposed to P01 and P24 who have their fixations aligned with these types of input. It appeared that this observation was not limited to those participants included in Figure 4.5. The second observation is shown in Figure 4.6.

Group 1: pre-exchange, P41



Honmadekka!?TV (2013)

Figure 4.6 Observations for shot changes

It was also observed that sudden shot changes in the video stimulus caused participants' gaze to land on types of input other than those being fixated on. The gaze path sequence shown in Figure 4.6 illustrates how the visual image changes to a wider view of the set causing for the participants' fixation to land on the wall instead of the speaker's face. The participant quickly corrects this and moves the gaze back to the speaker again, now located at the bottom left of the screen. It was deemed appropriate to keep such gaze path sequences in mind when taking a closer look at participants' gaze behaviour.

Considering that recordings were stored in four tests and that one of these tests contained recordings from two participant groups, additional procedures were taken to facilitate proper data exploration in the Visualisation tab and data analysis in the Statistics tab.

These procedures related to the creation of an independent variable in the eye-tracking software (Tobii AB, 2014b, p. 31).

The independent variable that was created for the purposes of this study was named "Group" and its values were labelled "Group 1: pre-exchange", "Group 2: exchange", "Group 3: post-exchange", "Reference group", and "Discarded". Participants were assigned one of these values in accordance with their participant group. The four participants that have been excluded from this study were given the discarded label. Participant group labels were thereafter created in the Visualisation tab of *Tobii Studio* and assigned the corresponding values of the independent variable. This procedure only needed to be taken once as the created participant group labels were also available in the Statistics tab. This procedure allowed for the recordings in the Visualisation tab to be filtered (2014b, pp. 62-63). This in turn facilitated the separate analyses of Group 2: exchange and the Reference group. However, it appeared that the independent variable does not facilitate the aggregation of gaze data across tests in the Visualisation tab. The independent variable enabled this for AOI group analyses only.

4.3.2 Procedure for heat maps

Heat maps included in this thesis show accumulated data on fixation count and absolute gaze duration. Relative gaze duration heat maps are not presented. When the exposure time to the video stimulus is the same for all participants, as is the case in this study, then the gaze data does not need to be normalised. Relative gaze duration heat maps were checked in the eye-tracking software to confirm this and it appeared that, despite the scale differences, the colour patterns were exactly the same as those encountered in the absolute gaze duration heat maps. This coincides with claims from the established literature (Bojko, 2009, p. 33). Contrary to the other three participant groups, Group 1: pre-exchange has been represented by two heat maps for each fixation measure. Gaze data for this particular participant group could not be integrated into one heat map as recordings for first-year and second-year JLLs were stored in two different tests. It is not possible to generate heat maps across tests in *Tobii Studio*. The two first-year JLLs and 15 second-year JLLs were therefore represented with their own heat maps.

The heat maps have been generated with default settings,[27] which do not set a Scale Max value. Setting this value was considered to keep the scales identical across heat maps, in order to facilitate a better comparison of the gaze data (2009, p. 36). However, it appeared to be difficult to find an upper scale threshold that would accommodate the differences in fixation count and absolute gaze duration between participant groups. It was also difficult to ascertain when an upper scale threshold is not arbitrarily chosen and when a heat map can be seen as an accurate representation of the gaze data. Higher scales tended to wash out hotspots in heat maps of first-year students in Group 1: pre-exchange to the extent that it became hard to distinguish which areas of the screen gathered visual attention. Conversely, lower scales caused heat maps of larger participant groups to get completely covered in splotches that showed less distinction in colour intensity (2009, p. 36). It was therefore decided to leave the Scale Max value deactivated after several attempts in finding an appropriate upper scale threshold and to have the eye-tracking software generate legends for each heat map.

Time intervals need to be set in order to generate heat maps on video segments (Tobii AB, 2014b, p. 74). Table 4.2 gives an overview of timings and durations which were followed to set the time intervals of heat maps on individual video segments and all three video segments together.

Table 4.2 Timings of analysed video segments

Aspect	Segment 2	Segment 3	Segment 4	Segments 2-4
Start	[00:01:08.881]	[00:04:58.769]	[00:08:48.330]	[00:01:08.881]
Duration	[00:03:49.887]	[00:03:49.560]	[00:02:14.185]	[00:09:53.634]
Finish	[00:04:58.768]	[00:08:48.329]	[00:11:02.515]	[00:11:02.515]

The time marker needs to be dragged to the end of a time interval in order to include all the fixation-data. The static image on which the data has been visualised in a heat map therefore shows the first still image of the following video segment rather than the last still image of the video segment on which the data has been generated.

Heat maps that best represent participant's visual attention are presented in Chapter 7. These are ordered according to participant group. Heat maps that are mentioned but not discussed in further detail are included in *Appendix J. Heat maps per analysed video segment*.

[27] The default settings for fixation count are: (1) radius set at 50px and (2) opacity set at 100%, and the default settings for absolute gaze duration are: (1) radius set at 50px and (2) opacity set at 100%.

4.3.3 Procedure for gaze plots

Findings on gaze paths have been supplemented with accumulated gaze plots instead of screen shots from gaze replay videos to illustrate observations on participants' gaze behaviour. It was decided to use gaze plots as gaze points cannot be accumulated in the Replay tab of *Tobii Studio*. The use of screen shots from gaze replay videos would therefore have complicated a proper representation of gaze point sequences. White circles and arrows have been added to the gaze plots in order to make the order of fixations more distinct. The gaze plots have been generated with default settings.[28]

4.3.4 Procedure for areas of interest

As noted, two observations were made during the data familiarisation phase that carried implications for the creation of AOIs. One of these is the slight offset in the gaze paths of some participants. Although it cannot be said with certainty that participants did not mean to look at these areas of the screen, it was determined that this observation should be taken into account; especially since this study involves an eye-tracking system with a 60 Hz sampling frequency. Small margins around the different types of input have been included in the AOIs to compensate for any offset in the gaze data (Holmqvist *et al.*, 2011, pp. 224-225; Bisson *et al.*, 2014, p. 405). Dynamic AOIs were generated in such a way that the size of the margin did not cause AOIs to overlap with other AOIs. The second observation on participants' gaze paths during shot changes was not considered to be problematic for the creation of AOIs. Although shot changes may influence the fixation-data on the different types of input, they are part of the nature of authentic Japanese variety shows and may therefore also be considered as part of natural gaze behaviour. The dynamic AOIs in this study have been created to activate and align with input when they appear on screen. It is therefore expected that the fixation-data for types of input give a good representation of allocated visual attention.

Identical sets of dynamic AOIs were created in the four tests in the *Tobii Studio* project. A set of guidelines was drawn up to ensure that the dynamic AOIs were duplicated in a consistent and uniform manner across tests (Conklin and Pellicer-Sánchez, 2016, p. 11). This set of guidelines comprises information on all AOI properties during their activation state. It provides a record of modifications that were made to the shape and positioning of

[28] The default settings are: (1) fixations set at duration, (2) scale set at 100%, (3) default as border colour, (4) recording colours as fill colour, (5) gaze order enabled, (6) gaze trail enabled, and (7) opacity set at 100%.

AOIs, the exact timings and keyframes in the timeline during which these occur, and details on the vertices of AOIs.

The set of guidelines on the AOI properties was fixed based on the first set of dynamic AOIs created in one of the tests. Each time a new AOI was created the number of vertices was determined based on the keyframe in which a type of input had its most complicated shape. Adjustments to vertices were thereafter entered in the timeline in accordance with the movement and size of the type of input (Tobii AB, 2014b, pp. 78-81). It was observed during the creation of this first set of dynamic AOIs that AOIs tend to transform too early if they are not fixed in place in the keyframe that comes right before the keyframe in which the transformation happens. Such keyframes were therefore added for all transformations of AOIs to ensure that AOIs stay properly aligned with the type of input. These have been called *in-between keyframes* in this thesis. After all of the AOI properties were fixed for the first set, it was duplicated across the other three tests. The same procedure for duplication was taken for each AOI.

The shape of an AOI in which it activates for the first time was copied from the original batch and pasted-in-place in the corresponding keyframe in the timeline of the other tests (Tobii AB, 2014b, p. 82). The progression of the AOI on the timeline was reconstructed from this keyframe onwards. There were a few instances in which a keyframe in one test did not match with the keyframe identified in the other three tests. This difference comprised one keyframe in all of these cases. These slight discrepancies were not considered to be problematic for the gaze data analysis.

4.3.5 Procedure for area of interest groups

After all the dynamic AOIs were copied and recreated in each of the tests the user-defined AOIs were organised into AOI groups that exist across tests (Tobii AB, 2014b, pp. 83-84). First, the AOI groups were created. Second, the dynamic AOIs were added to the AOI groups they are associated with. Whereas the AOI groups only need to be created once, the latter procedure needed to be repeated for each test in order for all dynamic AOIs to be associated with an AOI group and to ensure that they are all included in the AOI group analysis.

4.4 Summary

This chapter recounted the empirical data collection of this study. It detailed the experiment conditions that were recreated at five different locations in Ireland and Japan, the design of the pre-task and post-task questionnaires tailored to the four participant groups, and the type of portable eye-tracking system that was used for each of the experiment sessions.

The remainder of this chapter described each of the procedures of analysis which, taken together, established a transparent and consistent basis for the exploration of JLLs' multimodal perception and visual attention. The procedures for the analysis of multimodal perception comprised explorations of closed-ended data items and a thematic analysis on open-ended data items. The procedures for the analysis of visual attention consisted of explorations of heat maps and gaze replay videos, and the analysis of fixation-data through AOI groups. It was shown that the empirical data analysis was not only guided by the research questions and objectives; it was also based on decisions taken during the data familiarisation and data exploration phases. This resulted in data analyses that are firmly grounded in both the established literature and the collated empirical data of this study.

The following chapter discusses the final necessary step to bridge the gap between the conceptual framework and the actual findings. It takes a closer look at the multimodal analysis of the video material for the formation of units of analysis.

5. MULTIMODAL ANALYSIS OF A JAPANESE VARIETY SHOW

Chapter 5 combines both data analysis and methodology. It describes the multimodal analysis of the video material and presents the methodological choices that followed from these findings in relation to the analysis of fixation-data through AOI groups. The purpose of this chapter is to outline the multimodal nature of the video material, to tailor concepts from the established literature to the video stimulus of this study, and to provide a clear and concise overview of the types of input depicted in the video stimulus to facilitate a proper interpretation of the empirical data.

Section 5.1 explains the contents and structure of the video stimulus. It not only outlines the order of video segments and the scope of video material that was taken up for analysis; it also briefly explains the key points that were discussed in each topic of conversation. Section 5.2 follows with an exploration of the video material. It gives an account of the scope of input that was encountered and determines which of these types of input were thought to contribute to a better understanding of the contents of the video stimulus for JLLs. Section 5.3 describes the multimodal transcription and explains the units of analysis that were detected in the multimodal transcript. Section 5.4 discusses the multimodal nature of the video material. Section 5.5 explains how the AOI groups were defined based on both the multimodal landscape identified in the video stimulus and findings from the exploration phase of the empirical data analysis.

5.1 Video stimulus

5.1.1 Structure of video segments

The video stimulus comprises an excerpt from an episode of *Honmadekka!?TV* that was broadcast and recorded in Japan on 7th of August 2013 (Sasamoto, O'Hagan and Doherty, 2016, p. 6). As noted in *Subsection 3.3.2 Video material*, *Honmadekka!?TV* is aired by Fuji Television Network, Inc. The video material was edited in *Adobe Premiere Pro* and exported in high-definition as an .wmv file with a 1920x1080 resolution. This video format was chosen based on its compatibility with the eye-tracking software. The 1920x1080 resolution coincided with the screen resolution of the laptop and allowed for the video stimulus to be run in full-screen mode. The final edit of the video stimulus was reviewed by an employee of Tobii AB in Tōkyō before commencing the experiment sessions.

During that meeting it was agreed that the design of the video stimulus is appropriate for this study's objectives (Sikkema, 2016). The video stimulus was added as a movie element to the timeline in each of the four tests in the *Tobii Studio* project. The total duration of the video stimulus is 11 minutes and 24 seconds. Its structure is divided into five segments as can be seen in Figure 5.1.

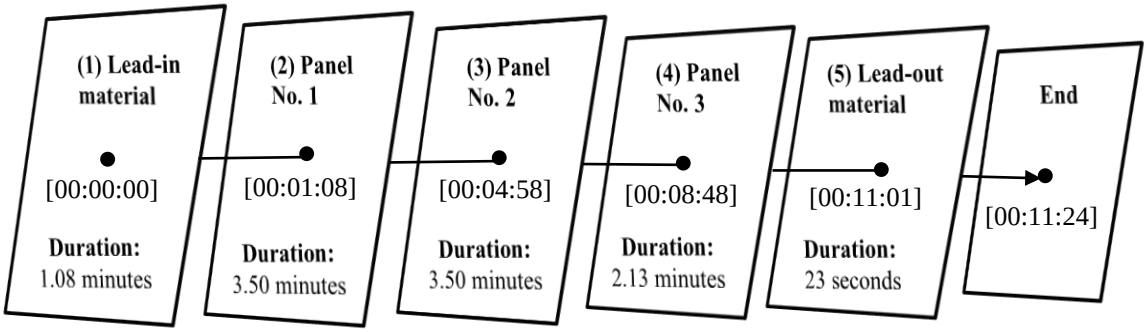


Figure 5.1 Structure of video stimulus

The multimodal analysis and empirical data analysis were conducted on Segments 2-4. Segment 1 and Segment 5 are excluded from data analysis as they were included to serve a different purpose. Segment 1 helped participants ease into the experimental task and become familiar with the programme format and eye-tracking system. Segment 5, on the other hand, helped participants ease out of the authentic AV material. The segments containing the lead-in and lead-out material consisted of the introductory scene to the programme item of Segments 2-4 and the concluding scene that shows the disclaimer and sponsors. This selection of material created natural transitions from non-target to target input and vice versa (Sikkema, 2016).

As shown in Figure 5.1, each of the analysed video segments comprises a panel discussion. Each panel discussion takes up a particular topic. Segment 2 focuses on utility costs and apartment types. Segment 3 discusses advice on how to keep yourself safe when lightning strikes. This advice focuses on cases where someone cannot hide from the lightning. Segment 4 concentrates on a suggested course of action to mitigate a tricky situation with bosses at work. While Segments 2-3 comprise of panel discussions that relate to more general themes and knowledge, Segment 4 involves etiquette and hierarchical work relations that may be rather specific to the Japanese context. Segment 4 was therefore expected to be most difficult to understand for the JLLs. For the purposes of this thesis, it was therefore determined to include only one such topic in the video stimulus and that it should be shorter compared to the other two segments.

5.1.2 Contents of analysed video segments

Segment 2 involves a panel discussion in which utility costs are compared between three types of apartment. These apartment types are a detached house, a room at the corner of the highest floor of an apartment complex and a room at the centre of the middle floor of an apartment complex. The segment starts with the overarching question on how to reduce utility costs. This is followed with brief comments of the hosts and panellists. The expert guest on housing conditions then responds and starts explaining several key points in relation to the overarching question. These explanations are met with intermittent banter and questions from the hosts and panellists.

The points that the expert guest makes throughout Segment 2 can be reduced to five basic claims. First, the room at the centre of the middle floor of an apartment complex has the lowest utility costs. This type of apartment has the least amount of walls facing the outside air which lessens the chance of heat leaving the room. In addition, its residents benefit from overflowing heat from neighbouring apartments above and beside them. Second, the costs of the room at the centre of the middle floor of an apartment complex are two-thirds of the costs for the room at the corner of the highest floor of an apartment complex. Furthermore, the costs for the room at the corner are two-thirds of the costs for a detached house. This makes the utility costs for the room at the centre half the costs for the detached house. Third, the expenses for heating are generally speaking more expensive than those for cooling. The heating costs are higher as it is used for more months in a row and because it needs to make up for a bigger difference in temperature between the room and the outside air. Fifth, further to the utility costs, costs such as rent and housing prices are lower for the room at the centre of the middle floor of an apartment complex than the room at the corner of the highest floor of an apartment complex. This illustrates that the room at the centre is not only cheaper based on its low utility costs; it is also cheaper based on the rent and housing price.

Segment 3 introduces the second panel discussion and continues with advice on dealing with lightning in different hypothetical situations. The segment starts with the overarching question of what to do when lightning strikes while there is no place nearby to hide. The expert guest on lightning starts explaining some key points with regards to the overarching question after several comments from the hosts and celebrity panellists. This segment also comprises banter and questions from hosts and panellists in addition to the expert guests' comments.

The expert guest makes four main points throughout Segment 3. First, he explains that it is best to crouch down or to make yourself as small as possible while putting your feet together when you cannot hide from the lightning. Putting your feet together or making sure that they touch each other prevents the electric current from running through your whole body as it allows the current to leave directly through the feet back into the ground. Second, lightning stops and chooses a high place before it strikes and can travel through the ground. There are no known examples of people being hit by lightning while they were crouched down, but there are examples of people having been hit by lightning while they were standing as they resemble lightning rods. Third, people run the risk of cardiac arrest when the electric current cannot escape the body and travels higher than the feet. Fourth, it is inadvisable to lie down on the ground in order to make yourself as small as possible. In such a case the electric current can travel through the whole body, from head to toe, and can put someone at risk of getting burned.

The final panel discussion, Segment 4, concentrates on a recommendation made by the expert guest on competence for cases in which someone has made a mistake and would like to mitigate the situation in response to his or her boss's anger. This segment starts with the overarching question of how to shift the focus when a boss gets angry. This expert guest starts explaining directly after this overarching question has disappeared from the screen. Banter and questions from hosts and panellists persist throughout Segment 4.

This expert guest makes the last three main points of the analysed video segments. First, he argues that it is best to respond with "Ah, is this how you interpreted that?" to show that the core of the problem could have been caused by miscommunication and the inability of the speaker to give an appropriate explanation. This takes the focus away from the mistake. Second, it is best to have this phrase followed with an apology as it would otherwise imply that your boss is at fault which stirs anger and shows a lack of responsibility taken by the speaker. Third, the described recommendation of the expert guest is also a technique used by frauds to come across as competent.

5.2 Navigating the rugged multimodal terrain

5.2.1 Meaningful selection of input

It became clear during the selection of video material that all three panel discussions included in the analysed video segments make use of manifold types of input for their

multimodal messages. It was argued in *Subsection 3.2.2 Stage 1* that it is important to narrow down the multimodal transcription to those types of input that were expected to contribute to a better understanding of the video material for JLLs and to be visually attended to by participants. Familiarisation with the analysed video segments is a good way to start this process. This not only facilitates a more targeted exploration and multimodal transcription afterwards; it also helps determine the level of precision that is needed in the description of the communicative situation.

All four forms of expression were encountered in the video material; audio-verbal input, audio-nonverbal input, visual-verbal input and visual-nonverbal input. These comprise dialogue, sound, text in the background, telops, telop design, imagery, and kinesic action. The types of input that are addressed in this thesis are those that are thought to be communicative (Bateman, Wildfeuer and Hiippala, 2017, p. 80). These include types of input that make a contribution to the multimodal messages and communicate information that is useful to the viewer. The researcher discerned that not all of the observed varieties in these types of input needed to be taken up for the data exploration phase. Three varieties have been excluded from the data exploration based on the familiarisation with the video material. These are included in the text in the background and imagery.

Two varieties of text were used in the background of the analysed video segments. These are nameplates and katakana characters displayed in the set. It was discerned that the latter did not need to be taken up in the exploration phase as only one was displayed in the set and visible on the screen on occasion. This is the character ホ [ho], which functions as a symbol for the programme title. It was expected that this character would not take up an important role in JLLs' viewer experience. Imagery, on the other hand, comprises three varieties; drawings, physical appearance of people appearing on the show (e.g. clothing and hairstyle), and the set. Only drawings were included in the exploration phase. It was discerned that the physical appearance of people appearing on the show would not add much meaning to their kinesic action for this particular communicative situation. The set in the video material is very busy with many objects hanging on the walls. Each of these objects are unrelated to the topics of the panel discussions. Although such visual-nonverbal input could distract participants, it was not expected that the set would prove to be communicative or useful to the viewer experience of participants.

This means that dialogue, sound, text in background (i.e. nameplates), telops, telop design, imagery (i.e. drawings), and kinesic action were taken up for the exploration phase. As noted in *Subsection 3.2.2 Stage 1*, the taxonomy of Kimura *et al.* (2000) as described by Sakamoto (2009b), and the framework of gestural meanings of Kalantzis *et al.* (2016) were consulted for this phase. An overview of the scope of input and varieties of input that followed from this exploration is taken up in *Appendix K. Scope of input for analysed video segments* and further discussed in the following subsections.

5.2.2 Audio-verbal input

It was determined during the exploration of audio-verbal input that its scope comprises two varieties of dialogue. These are utterances of speakers and narrations. The utterances of speakers consisted of utterances made by two hosts, five panellists, and three expert guests. The narrations were uttered by a male voice.

5.2.3 Audio-nonverbal input

The scope of audio-nonverbal input consists of four varieties of sound. These include noises from the audience, music, sound effects, and a variety named "Other". It was observed that noises from the audience comprise three types of sound. These are phatic expressions, laughing and clapping. The music includes several brief tunes that are repeated throughout the analysed video segments. Sound effects are those sounds that accompany the progression of text units or particular movements of telop. These not only accompany the first text unit of a telop but also text units that appear later on screen as a telop progresses (i.e. add-on text units). Seven sound effects were detected in the video material. These are sharp sound effects, ping sound effects, floating sound effects, shining sound effects, bang sound effects, thunder sound effects, and squeaky sound effects. The variety of other sounds includes sounds that do not fit the other three varieties. These are call bells and a hand that claps on a table. The call bell was used by the expert guests at the start of each analysed video segment. This means that it could be heard three times. The hand that claps on the table, on the other hand, was detected only once.

5.2.4 Visual-verbal input

The scope of visual-verbal forms of expression comprises two types of input: text in background and telops. Although it has been mentioned that only one variety of text in background was selected during the familiarisation phase, nameplates, further exploration

was necessary to gain a deeper understanding of this variety and those varieties included in telops. Examples of nameplates are shown in Figure 5.2.

Honmadekka!?TV (2013)

Figure 5.2 Varieties of text in background

Five nameplates are used in the video material displaying the names and professions of five expert guests. This means that each expert guest has his own nameplate as is shown on the right in Figure 5.2. Only three of these nameplates are clearly visible. These are the nameplates of the three expert guests who take part in the panel discussions of the analysed video segments. One of these nameplates is shown on the left in Figure 5.2 at the bottom of the still image. It translates to "Masashi Yamazaki: commentator on competence". The two nameplates that are less clearly visible throughout the video stimulus show the names and professions of two expert guests who are the commentators of two other panel discussions of the same programme item. However, as these panel discussions are not included in the video stimulus, these two expert guests do not participate in the discussions and have less screen time. Consequently, less screen time is allotted to their nameplates. Considering that the same group of people are present in each of the analysed video segments, all five nameplates are shown throughout the video material. However, their visibility and readability varies between shots.

The taxonomy of Kimura *et al.* (2000) as described by Sakamoto (2009b), included in *Table 2.1 Taxonomy of telops (translated and adapted from Sakamoto, 2009b)*, was consulted for the exploration of telops. More specifically, the typology of function was used to determine the nature of telops in the analysed video segments. Although all three functions (i.e. direct rendering of dialogue, facilitation of comprehension, and bridging of scene change) were detected in the video material, not all of the varieties were encountered.

This thesis includes six varieties of telop. These are utterance of speaker, narration, theme, title, speaker identification, and screen filled with text. Whereas five of these are included in the original taxonomy, the variety of speaker identification was added for the purposes of this study. Figure 5.3 gives examples of the varieties included in telops that directly render dialogue: utterance of speaker and narration.

Honmadekka!?TV (2013)

Figure 5.3 Varieties of telops that directly render dialogue

Narration telops only appear on screen when the narrator speaks. This happens each time a panel discussion is introduced. This means that narrations appear on screen three times. For each of these occasions they introduce the overarching question of the panel discussion. The narration telops included on the left in Figure 5.3 introduce the overarching question of Segment 2. The telop in the top left corner translates to "If you…" and the telop at the bottom of the screen means "Like to lower your utility costs!?". Utterances of speakers generally appear at the bottom of the screen as is shown on the right in Figure 5.3. This telop represents an utterance of the expert guest on housing conditions. His utterance means "It is best to live in a home at the centre of the middle floor of an apartment complex".

Figure 5.4 includes still images that illustrate the three varieties of telops that facilitate comprehension. These include theme, title, and speaker identification.

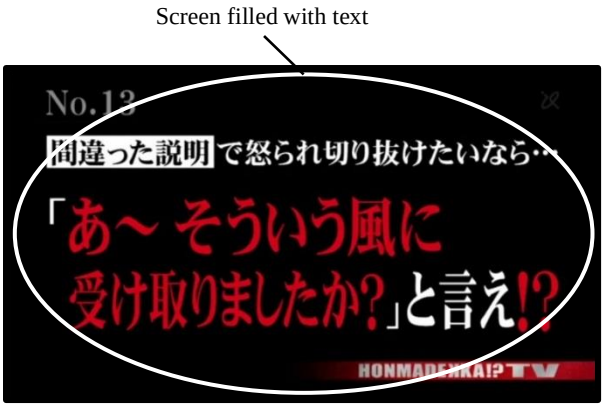
Honmadekka!?TV (2013)

Figure 5.4 Varieties of telops that facilitate comprehension

Themes comprise statements that summarise key points from the panel discussions. A theme is generally shown in the top right corner of the screen and changes its contents each time a key point is made. This means that a theme remains visible on the screen for a certain amount of time until it changes into the next key point. The theme shown on the left in Figure 5.4 means “The utility costs for the centre of the middle floor are two-thirds of the costs for the room in the corner of the highest floor of an apartment complex!?”. The theme on the right translates to “Lightning stops at a certain height for a split second!?”.

Only one title is used throughout the analysed video segments. This title is the name of the programme item and means “Useful Honmadekka!? for unexpected times”. It usually appears in the top left corner of the screen. Only under special circumstances is it placed elsewhere on the screen. This is shown in Figure 5.4 on the left; the title is located in an area of the screen where it does not overlap with the telop that directly renders dialogue. The variety of speaker identification is only used for expert guests and appears with their utterances. The speaker identification on the left in Figure 5.4 means “According to Mr Bōgaki (housing conditions)” and the one on the right translates to “According to Mr Okano (lightning)”.

The variety that is included in telops that bridge scene change is shown in Figure 5.5.



Honmadekka!?TV (2013)

Figure 5.5 Varieties of telops that bridge scene change

This variety is screen filled with text and provides summaries of the most important statements in the panel discussions. The function of this variety is therefore very similar to that of theme and it was observed that their contents are very much alike. The telop included in Figure 5.5 means "If you want to find your way out of the trouble caused by your faulty explanation…Say "Aah, is that how you understood it?"!?".

5.2.5 Visual-nonverbal input

As noted, it was observed that three forms of expression comprise the scope of visual-nonverbal input. These are telop design, imagery, and kinesic action. Whereas the number of varieties included in imagery were already determined during the familiarisation phase, it was necessary to further explore the use of drawings in the video material, and the scope of varieties of input for telop design and kinesic action.

The taxonomy of Kimura *et al.* (2000), as described by Sakamoto (2009b), included in *Table 2.1 Taxonomy of telops (translated and adapted from Sakamoto, 2009b)*, was also consulted for the exploration of the telop design. This time the typology of form was used to work out the range of stylistic effects that are employed for the telops in the analysed video segments. It appeared that the application of this typology was less straightforward as compared to the typology of function. This was due to four factors. First, it was observed that more than one type of form is often used for one telop. This means that types overlap in the video material even though they are presented as separate categories in the original framework. Second, the use of colours, colour gradients, typefaces and fonts have not been incorporated into the typology. Third, the original framework does not discuss the different aspects of telop design. This means that it does not distinguish between the

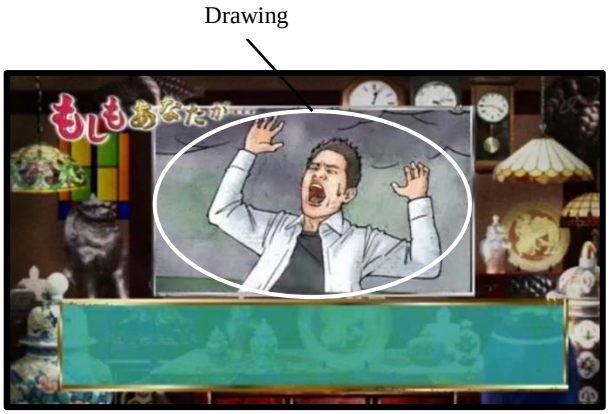
typography, display rate or movement of telop. Fourth, it is argued that most of the types of form in the typology are used for emphasis. However, it proved to be difficult to ascertain whether the stylistic effects are used for the same purposes in the video material and not for more practical reasons such as readability or legibility. This thesis includes six varieties of telop design: expansion, background colour, punctuation, special effect, picture, and a variety named "Other". This list illustrates that five varieties from the typology of form have been included. The scope of two of these have been narrowed down. Figure 5.6 shows examples from the video material of the six varieties of telop design.

Figure 5.6 Varieties of telop design

It was observed that seven types of punctuation are used in the analysed video segments, i.e. exclamation marks, question marks, three suspension points, rightwards arrows, less-than signs, tildes, and the degree symbol. It appeared that the variety of special effect comprises three sets of form types that each relate to one particular aspect of telop design. First, it was observed that some telops had shining or flickering effects that were incorporated into the design through typography. Second, another set of special effects were related to the way in which telops appear on screen. Five ways of appearing on screen were detected for telop in the video material. These include sliding in, scrolling in, slowly appearing, popping up, and shooting in. Third, it appeared that four types of special effect were established through movement; shaking, quivering, moving and floating. Three types of pictures were encountered in the analysed video segments. These are faces of expert guests and the host, images of lightning, and the nameplate of one expert guest. The variety of other includes any telop designs that do not fit in the other five varieties. These involve a graphic representation of an apartment complex, speech bubbles, stick figures, text boxes, and borders. The latter is also known as *ejji* [*edging*] (Shitara, 2011, p. 6).

It is important to note that the combination of the varieties of picture and other shown on the left in Figure 5.6 showcases a rather interesting use of telop design. This particular telop shows an utterance of the host while its contents comprises an utterance made previously by the expert guest. It reads as “Fushigi desu ne” [Amazing isn’t it]. This is the concluding remark of the expert guest to his introduction on the comparison of utility costs for the three types of apartment. The host jokes about this comment as he finds it a wonderful way to start the panel discussion, and may even be surprised that the expert guest states this about his own explanation. After saying “Ii desu yo!” [That’s nice!] he continues with “sono hairikata mo nē” […to start like that]. The second part of the utterance relates to the beforementioned telop. Although this telop and the corresponding utterance of the speaker show a mismatch in verbal content, the telop design manages to visually supplement the necessary information as it shows the moment when the expert guest makes the comment that starts the panel discussion.

Figure 5.7 shows an example of the variety included in imagery. This is the variety of drawing.



Honmadekka!?TV (2013)

Figure 5.7 Varieties of imagery

It was observed that drawings were only used at the start of each panel discussion. More specifically, each time that a topic of a panel discussion is introduced. This means that three drawings were used and that they appeared on screen approximately at the same time as the narration telops.

Figure 5.8 includes examples of the three varieties that are included in the scope of kinesic action, i.e. gesticulation, bodily configuration, and facial expression.

Figure 5.8 Varieties of kinesic action

The names of these varieties were adopted from the framework of gestural meanings as defined by Kalantzis *et al.* (2016, pp. 386-388). Whereas the original framework consists of more types of gestural meanings and includes more comprehensive definitions of the chosen terminology, the varieties have been used in this thesis to distinguish between movements of the hands, body and face respectively. The majority of people appearing on the show are seated in the analysed video segments. Only the two hosts stand throughout the video material. One of the panellists and one of the expert guests stand for a certain timeframe during Segment 3 to enact the hypothetical situations for dealing with lightning. The variety of facial expression includes only those expressions that the researcher considered to be explicit or exaggerated. This happens only three times in the video material. Combinations of the varieties were detected throughout the analysed video segments.

5.3 Multimodal transcription

5.3.1 Meaningful study of multimodal nature

As noted previously, it was expected that the gaze behaviour of participants in this study would be characterised by a search for meaning in the multimodal messages. This means that it was anticipated that, more than low-level factors (e.g. colour, contrast or motion), high-level factors (e.g. users' goals, expectations or previous knowledge) would play an important role in participants' viewer experience and viewer reception (Holsanova, 2014, p. 289). Further to the established literature on eye-tracking of moving images (Redmond and Batty, 2015), this expectation was also based on the pilot experiment sessions, when it was

observed that participants' gaze paths seemed to follow the appearance of text units of telops that directly render dialogue and to be less affected by varieties of telop design that are associated with typography such as the variety of other. Although this could mean that gaze behaviour may have been guided by special effects that relate to the movement of telop or its way of appearing on screen, it was considered that this could be the result of the information it represents and its close tie to the narrative of the communicative situation (Batty, Perkins and Sita, 2015; Kruger, Doherty and Soto-Sanfiel, 2017).

Nonetheless, this does not mean that low-level factors should not be included in the multimodal transcription. On the contrary, research has urged that visual-nonverbal input and interrelationships between forms of expression should be taken up in multimodal analyses (Jewitt, 2014). Furthermore, studies on authentic Japanese variety shows have described the potential of telop design to communicate affective information (Sasamoto and O'Hagan, 2014; Sasamoto, O'Hagan and Doherty, 2016). What this means is that the structure of the multimodal transcript needs to be considered carefully in order for it to elicit these aspects.

It was determined that the multimodal transcript of this study should be structured according to telops that directly render dialogue. Further to their connection to the narrative and dialogue, it was observed during the exploration phase that varieties of telop design appeared to be incorporated into the meaning of such telops in some cases. As discussed in the previous subsection, these cases illustrate that telops sometimes highly resemble dialogue through such visual-nonverbal input. This observation gave rise to the idea that such telops may be more accessible to JLLs or could bring about a different kind of gaze behaviour as compared to more complex telops that are less similar to the corresponding audio-verbal input. A structure according to telops that directly render dialogue could therefore prove to be a good starting point for the identification of interrelationships between types of input and meaningful multimodal units in the analysed video segments.

This means that the multimodal transcription in this study is not only geared towards the creation of a structured overview of the contents of the video material, it is also focused on the identification of meaningful multimodal units and degrees of resemblance between telops and their corresponding dialogue. As discussed in *Subsection 3.2.2 Stage 1*, the

transcription frameworks of Baldry and Thibault (2006), and Taylor (2003), and the typologies of Shitara (2005; 2006) were consulted to this end.

5.3.2 Procedure for multimodal transcript

The multimodal transcript was created in *Microsoft Excel*. It consists of seven columns and 262 rows. The first row contains the headings for each column while the remaining 261 rows include data. An excerpt of the multimodal transcript is shown in Figure 5.9.

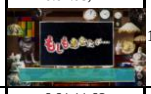
Shot No./ Timecode	Still image	Visual image	1. Content text unit(s)	Dialogue	Kinesic action	Sound
1. 0:01:08;27-0:01:10;25	0:01:08;27 	1. 1 text unit	もしもあなたが…			Music
2. 0:01:10;26-0:01:11;05	0:01:11;02 	1. 1 text unit	もしもあなたが… (moves to upper left corner)			Music, Sharp sound with text unit
3. 0:01:11;06-0:01:20;14	0:01:11;20 	1. 1 text unit, Drawing	もしもあなたが…	**Narrator:** もしもあなたが...		Music
4. 0:11:14;29-0:01:16;25	0:01:16;10 	1. 1 text unit, Drawing	光熱費を (scrolls into screen from left to right)	**Narrator:** 光熱費を...		Music
5. 0:01:16;26-0:01:20;14	0:01:19;05 	1. 2 text units, Drawing	光熱費を安くしたかったら！？ (scrolls into screen from left to right)	**Narrator:** 安くしたかったら。 Panellist (Yashima Norito): お、したい。 Panellist (Kiriko Isono): したい...		Music, Audience (あー)
6. 0:01:20;15-0:01:22;24	0:01:20;29 	2. 2 text units, Panellists		Panellist (Yashima Norito): したい、したい... それはしたい。 Panellist (Kiriko Isono): よねー、したい、したいです！はい。 Panellist (Matsuko Deluxe): した、したい。		Audience (あー), Ping sound with top right telop

Figure 5.9 Excerpt of multimodal transcript

The rows are arranged according to the chronological sequencing of events in the video stimulus and segment the video material into shots. A new shot starts each time a text unit of a telop that directly renders dialogue appears on screen and each time the camera shifts to another person. A more detailed breakdown of the timeline into frames was deemed inappropriate for the purposes of this thesis as it complicated the identification of meaningful multimodal units. Each shot has been numbered in bold in the transcript. This numbering is shown in the column that also displays the timecodes of shots. The timings of timecodes and still images were checked in *Camtasia 7.1*. Of the 261 shots, Segment 2 comprises 107 shots. Segment 3 is slightly shorter and contains 106 shots. As noted previously, Segment 4 is the shortest segment and involves the remaining 48 shots.

A different division of columns was adopted in the transcript as compared to the multimodal transcription frameworks of Baldry and Thibault (2006), and Taylor (2003). Taylor (2003) includes two columns to describe the visual image; one for imagery and one for subtitles. The visual image and contents of telop that directly render dialogue have also been presented in two separate columns in this thesis. However, rather than including one column for "soundtrack" (2003, p. 196), two separate columns have been used for dialogue and sound to better facilitate the analysis of audio-verbal and audio-nonverbal input. The juxtaposition of the contents of telops that directly render dialogue and the corresponding dialogue was also thought to enhance the analysis of degrees of resemblance between the two.

The column for visual image describes the salient types of input visible on screen. These are text units of telops, people appearing on the show, nameplates and drawings. Only those instances during which a nameplate is clearly visible are counted. The number that is mentioned before the number of text units refers to the function of telop, in which direct rendering of dialogue is 1, facilitation of comprehension is 2, and bridging of scene change is 3. Take, for example, shot No. 6 in Figure 5.9. The two text units comprise units of telops that facilitate comprehension as they are preceded with number 2. It is important to note that the column for visual image in the frameworks of Baldry and Thibault (2006), and Taylor (2003) also includes transcriptions of other types of input such as camera position and distance between speakers. Whereas these forms of expression are important for the viewer experience of film, they are not considered to be communicative types of input in the video material of this study. These are therefore not included in the multimodal transcript.

The column for 1. content text unit(s) includes the same numbering for the functions of telop as the one used in the column for visual image. The addition of number 1 in the heading therefore indicates that this column only gives information on the text units of telops that directly render dialogue. The notation style of the contents mimics what is visible in the still image. Ways of appearing on screen and any movements of telop are indicated in brackets. Varieties of telop design are only mentioned in these brackets for those telops that incorporate the telop design to resemble the corresponding dialogue.

The column for dialogue presents all utterances of speakers and narrations in the analysed video segments. Names of speakers are indicated with each of these. The names of

captioned speakers are included in bold. Three suspension points can mean three things. They can indicate a fade-out in volume, time lapse between utterances or the continuation of dialogue into subsequent shot(s). A full stop indicates the end of an utterance of a speaker or narration.

The columns for kinesic action and sound present all of the varieties from the scope of input that are associated with these types of input. Varieties overlapped in both of these columns. Instances in the analysed video segments during which, for example, hands and other body parts were used to enact hypothetical situations were transcribed as a combination of gesticulation and bodily configuration. The same was done in the column for sound where combinations of, for example, jingles and ping sound effects were transcribed as a combination of music and sound effect. The types of noises from the audience are shown in brackets.

Further to the multimodal transcript another overview was made of the still images. All of the still images were transferred from the transcript to *Microsoft PowerPoint* and overlaid with a grid. This technique is used in the ICE project to work out the composition of an authentic Japanese variety show. Such an overview serves another purpose in this thesis; this condensed version of the multimodal transcript helped ease the counting process of occurrences of types of input.

5.3.3 Procedure for multimodal units

Although this arrangement of columns and rows gives a structured overview of the co-occurrence of input in each shot it does not show how these shots interrelate. Furthermore, even though text units have been identified in each of these shots, information on the number of telops that directly render dialogue is still missing. To this end, an eighth column was added in the multimodal transcript. As shown in Figure 5.10, this column was named “Telop No.” and added in front of the column in which the shot numbers and timecodes are included.

Telop No.	Shot No./ Timecode	Still image	Visual image	1. Content text unit(s)	Dialogue	Kinesic action	Sound
1	**1.** 0:01:08;27-0:01:10;25	0:01:08;27 	1. 1 text unit	もしもあなたが…			Music
	2. 0:01:10;26-0:01:11;05	0:01:11;02 	1. 1 text unit	もしもあなたが… (moves to upper left corner)			Music, Sharp sound with text unit
	3. 0:01:11;06-0:01:20;14	0:01:11;20 	1. 1 text unit, Drawing	もしもあなたが…	**Narrator:** もしもあなたが…		Music
2	**4.** 0:11:14;29-0:01:16;25	0:01:16;10 	1. 1 text unit, Drawing	光熱費を (scrolls into screen from left to right)	**Narrator:** 光熱費を…		Music
	5. 0:01:16;26-0:01:20;14	0:01:19;05 	1. 2 text units, Drawing	光熱費を安くしたかったら！？ (scrolls into screen from left to right)	**Narrator:** 安くしたかったら。 Panellist (Yashima Norito): お、したい。 Panellist (Kiriko Isono): したい…		Music, Audience (あー)
	6. 0:01:20;15-0:01:22;24	0:01:20;29 	2. 2 text units, Panellists		Panellist (Yashima Norito): したい、したい…それはしたい。 Panellist (Kiriko Isono): よねー、したい、したいです！はい。 Panellist (Matsuko Deluxe): した、したい。		Audience (あー), Ping sound with top right telop

Figure 5.10 Excerpt of multimodal units

It is argued here that text units do not exist on their own in a meaningful way and that they combine to form telops. It was observed that utterances of speakers and narrations do not always appear on screen at once; in most cases parts of an utterance or narration appear on screen as text units. These text units appear in succession and form sequences. Such sequences and text units that represent whole utterances or narrations are defined as a telop in this thesis. This means that telops span different time frames. The data in the columns for 1. content text unit(s) and dialogue were compared to work out which dialogue occurs with which text unit. Based on this comparison it was determined which text units form telops on their own and which text units needed to be grouped to form one telop.

As shown in Figure 5.10, sets of shots that feature the text units of one telop were grouped and given a number. This numbering follows the order of occurrence in the multimodal transcript. It forms the overarching structure of the transcript and visually represents meaningful multimodal units in the analysed video segments; both horizontally, in relation to co-occurring input, and vertically, regarding their temporal and conceptual sequencing. Sets of shots that do not feature text units only have shot numbers. These comprise a total of 38 shots. A total of 66 telops that directly render dialogue were identified over the remaining 223 shots. Of these, 6 telops represent narrations (9.1% of telops) while the remaining 60 telops comprise utterances of speakers (90.9% of telops).

5.3.4 Procedure for types of resemblance

The identification of degrees of resemblance between telops that directly render dialogue and corresponding utterances of speakers and narrations happened in two rounds of colour coding. The first round was conducted in a tentative manner to gain a deeper understanding of the kinds of relationship that existed between the two types of input in the 223 shots. It became clear during the exploration phase that telops include words or expressions that are identical, similar or different from those used in the dialogue. It was decided that these sections needed to be distinguished with different colours in the columns for 1. content text unit(s) and dialogue of the multimodal transcript. The use of colours in this first round is shown in Figure 5.11.

Telop No.	Shot No./ Timecode	Still image	Visual image	1. Content text unit(s)	Dialogue	Kinesic action	Sound
1	1. 0:01:08;27-0:01:10;25	0:01:08;27 	1. 1 text unit	もしもあなたが…			Music
	2. 0:01:10;26-0:01:11;05	0:01:11;02 	1. 1 text unit	もしもあなたが… (moves to upper left corner)			Music, Sharp sound with text unit
	3. 0:01:11;06-0:01:20;14	0:01:11;20 	1. 1 text unit, Drawing	もしもあなたが…	**Narrator:** もしもあなたが…		Music
2	4. 0:11:14;29-0:01:16;25	0:01:16;10 	1. 1 text unit, Drawing	光熱費を (scrolls into screen from left to right)	**Narrator:** 光熱費を…		Music
	5. 0:01:16;26-0:01:20;14	0:01:19;05 	1. 2 text units, Drawing	光熱費を安くしたかったら！？ (scrolls into screen from left to right)	**Narrator:** 安くしたかったら。 Panellist (Yashima Norito): お、したい。 Panellist (Kiriko Isono): したい…		Music, Audience (あー)
	6. 0:01:20;15-0:01:22;24	0:01:20;29 	2. 2 text units, Panellists		Panellist (Yashima Norito): したい、したい… それはしたい。 Panellist (Kiriko Isono): よねー、したい、したいです！はい。 Panellist (Matsuko Deluxe): した、したい。		Audience (あー), Ping sound with top right telop

Figure 5.11 Excerpt of colour coding

The colour green was applied to those sections of telops and dialogue that were identical. Orange was used for synonyms and expressions that were considered to be similar. The third colour, blue, was given to text that caused for discrepancies between the contents of the two columns. Such discrepancies were the result of additions or omissions of information in one of the two columns. All remaining text was left black in the transcript.

Upon completing this first round of colour coding, it was discerned that a framework needed to be devised that focused on the identification of degrees of resemblance for the type of input (i.e. telops that directly render dialogue) rather than the individual varieties

(i.e. utterances of speaker, and narrations). This was determined based on the expectation that the function of such telops (i.e. direct rendering of dialogue) would take precedence in participants' gaze paths in their search for meaning. To this end, two guiding principles were adopted for the definition of a joint framework with which degrees of resemblance could be identified in the video material. First, the verbal content of telops should function as the starting point. Second, the still images should be taken into account as a high degree of resemblance can also be established through telop design; even when the verbal content is not identical. The observations from the first round of colour coding were then revisited with these guiding principles in mind. The typologies of utterance of speaker and narration telop as defined by Shitara (2005; 2006), included in *Subsection 2.2.2 Overview of telops*, were used as a point of reference during this revision of the multimodal transcript.

Five types of resemblance were identified for the purposes of this study. The scheme for these types of resemblance is shown in Table 5.1.

Table 5.1 Scheme for types of resemblance

Colour	Type No.	Description	Definition	No. of telops
Green	1	Identical	Very high degree of resemblance due to timing and wording	23
Pink	2	Other	High degree of resemblance due to telop design	5
Blue	3	Omission	Medium degree of resemblance due to omissions of words/phrases	15
Brown	4	Recast	Low degree of resemblance due to synonyms/alternative phrases	16
Purple	5	Incongruous	Very low degree of resemblance due to discrepancies in contents	7

A number, a description and a definition have been assigned to each of these types of resemblance. The name of a type of resemblance consists of the number and the description (e.g. Type 1: identical). The types of resemblance have also been allocated a colour in preparation of the gaze data analysis. This colour coding scheme was determined based on the range of default colours available in the eye-tracking software and therefore does not carry any added meaning.

This list of types of resemblance illustrates that the scheme draws from Shitara's typology of narration telop (2005) in particular. However, the definitions divert from the original framework as they have been shaped by the telops in the analysed video segments and the beforementioned guiding principles. Brief definitions have been included in Table 5.1. Telops that are included in Type 1: identical appear at the same time as an utterance of the speaker or narration and use the exact same wording. Type 2: other comprises telops that highly resemble the dialogue due to visual-nonverbal input (i.e. telop design). Telops in

Type 3: omission have omitted words or expressions from the dialogue. Such telops present the key points of the dialogue in the same wording without altering the contents. Type 4: recast includes telops that also keep the core of the dialogue intact, but use synonyms or similar expressions to represent the information contained in the utterances. Telops that constitute Type 5: incongruous show discrepancies in the contents of telops and dialogue through the addition or omission of information.

The second round of colour coding involved the assignment of the colours of types of resemblance to the corresponding telops. This is shown in the first column of Figure 5.11. The included excerpt demonstrates that Telop No. 1-2 were grouped into Type 1: identical as they have been coloured green. This procedure was repeated for all of the 66 telops that directly render dialogue. Table 5.1 includes an overview of the number of telop that were grouped into the types of resemblance. It appeared that 34.8% of telops is assigned to Type 1: identical. These ratios are 7.6% for Type 2: other, 22.7% for Type 3: omission, 24.2% for Type 4: recast, and 10.6% for Type 5: incongruous. These proportions suggest that the analysed video segments contain more telops with lower degrees of resemblance (i.e. Types 3-5).

5.4 Multimodal landscape

5.4.1 Spatial dimension per analysed video segment

Table 5.2 gives an overview of the prevalence of types of input in the analysed video segments. It shows the rates of occurrence for each variety of input in the shots of the multimodal transcript. These percentages are calculated over the total number of shots per analysed video segment and all analysed video segments combined.

Table 5.2 Rates of occurrence for types of input

Type of input	Variety of input	Segment 2	Segment 3	Segment 4	Segments 2-4
Dialogue	Utterance of speaker	87.9%	85.8%	83.3%	86.2%
	Narration	2.8%	3.8%	6.3%	3.8%
Sound	Noise from audience	29.0%	25.5%	43.8%	30.3%
	Music	8.4%	10.4%	20.8%	11.5%
	Sound effect	39.3%	17.0%	29.2%	28.4%
	Other	0.9%	2.8%	2.1%	1.9%
Text in background	Nameplate	15.0%	3.8%	16.7%	10.7%
Telop	Utterance of speaker	79.4%	59.4%	62.5%	68.2%
	Narration	4.7%	5.7%	10.4%	6.1%
	Theme	91.6%	88.7%	85.4%	89.3%
	Title	91.6%	83.0%	87.5%	87.4%
	Speaker identification	17.8%	17.0%	8.3%	15.7%
	Screen filled with text	3.7%	4.7%	2.1%	3.8%
Imagery	Drawing	2.8%	3.8%	6.3%	3.8%
Kinesic action	Gesticulation	43.0%	42.5%	35.4%	41.4%
	Bodily configuration	1.0%	13.2%	2.1%	6.1%
	Facial expression	1.0%	-	4.2%	1.1%

The proportions demonstrate that utterances of speakers in both written and spoken form, themes, titles and gesticulation were observed most frequently and consistently throughout the video material. Of these, themes and titles show the highest percentages. They are absent in those instances during which narrations and screens filled with text were presented on screen. Two additional observations were made regarding the differences in percentage between the varieties of dialogue and their written counterparts in telops. First, it appears that not all utterances of speakers have been visually represented by telops. As can be seen in Table 5.2, the proportions for telops that show utterances of speakers are lower. Second, it appears that the opposite applies to narrations as narration telops have more screen time as compared to the time it takes for the narrator to say the narrations.

When taking a closer look at the multimodal structure for each analysed video segment it can be argued that bodily configuration is much more frequent in Segment 3 as opposed to the other two segments. This fits with the topic of the panel discussion as people appearing on the show enact hypothetical situations in that particular segment. This enactment happens through a combination of gesticulation and bodily configuration. These include standing, crouching, arm movements to indicate lying on your belly, pointing and gestures to show the direction and movement of electric current. Kinesic action in the other two segments is restricted to movements with the hands as the majority of people are seated throughout Segment 2 and Segment 4. Kinesic action for these segments includes holding up fingers to indicate numbers, apologetic gestures, and hand movements that specify floors of an apartment complex or the walls of a room. This also explains why nameplates show a low rate of occurrence for Segment 3. Nameplates are placed in front of the seats of

expert guests; less sitting therefore means less screen time for the nameplate of the commentator for Segment 3. As observed during the data exploration phase, the variety of other comprises the lowest percentage of the varieties included in sound. Table 5.3 further elaborates on the rates of occurrence for types of input and shows the prevalence of types of resemblance.

Table 5.3 Rates of occurrence for types of resemblance

Type of resemblance	Segment 2	Segment 3	Segment 4	Segments 2-4
Type 1: identical	17.8%	20.3%	51.4%	24.7%
Type 2: other	20%	4.3%	14.3%	13.4%
Type 3: omission	25.6%	33.3%	-	23.7%
Type 4: recast	14.4%	26.1%	31.4%	21.6%
Type 5: incongruous	22.2%	15.9%	2.9%	16.5%

It appears that telops in Type 1: identical are visible on screen the most. This is followed with Type 3: omission, Type 4: recast, Type 5: incongruous, and Type 2: other. Not only the first three of these types show percentages that are similar, the same can be observed for the least frequent types of resemblance. This appears to be related to the number of telops included in the types of resemblance. It was shown in Table 5.1 that Type 2: other and Type 5: incongruous comprise the lowest numbers of telop. When taking a closer look at the rates of occurrence per analysed video segment it appears that certain types of resemblance have more screen time in particular video segments.

It is important to take the number of text units of telops and text in background into account for the interpretation of these percentages as the display rate and informative nature of such types of input can influence the role they play in JLLs' viewer experience and visual attention. A table that shows the number of text units gives an accurate description of saliency of visual-verbal types of input. This is demonstrated in the following subsection.

5.4.2 Temporal dimension for analysed video segments

Table 5.4 gives an overview of saliency of visual-verbal types of input in the analysed video segments. It shows the counts for each variety of input and the number of text units these comprise for all video segments combined. The number of add-on text units have been included to demonstrate how many of these appear after first units of text in background and telops have become visible on screen. It appears that utterances of speakers and narrations are the only visual-verbal types of input that involve add-on text units.

Table 5.4 Display rate for visual-verbal types of input

Type of input	Variety of input	Count	No. of text units	No. of add-on text units
Text in background	Nameplate	5	5	-
Telop	Utterance of speaker	60	102	42
	Narration	6	10	4
	Theme	13	13	-
	Title	1	1	-
	Speaker identification	3	3	-
	Screen filled with text	10	10	-

As can be seen in Table 5.4, the visual-verbal types of input that occurred most frequently in the analysed video segments, theme and title, comprise a relatively low number of text units. In addition, it was observed that each time a theme appeared on screen it was accompanied with a sound effect. Although such sound effects may direct viewers' visual attention to such telops, it can be argued that they remain static for relatively long periods of time and do not necessarily display information that is new to the user. The same applies to nameplates, speaker identifications and screens filled with text. Of these, nameplates are not accompanied with any music or sound effects. It was observed that utterances of speakers and narrations are the only dynamic visual-verbal type of input. Both varieties combined feature a total of 112 text units, of which 46 are added after the initial unit has appeared on screen. They are paced and time-bound due to their close tie to the dialogue and a large portion of the text units are accompanied with music or sound effects. Although telops that directly render dialogue reiterate spoken utterances of speakers and narrations, it can be argued that they include information that is pertinent and new at the moment they appear on screen. Table 5.5 gives further details on the saliency of telops per type of resemblance.

Table 5.5 Display rate for types of resemblance

Type of resemblance	Count	No. of text units	No. of add-on text units
Type 1: identical	23	31	8
Type 2: other	5	13	8
Type 3: omission	15	28	13
Type 4: recast	16	24	8
Type 5: incongruous	7	16	9

It illustrates that Type 3: omission has the highest number of add-on text units. The other four types of resemblance, on the other hand, appear to have a similar number of add-on text units.

5.5 Area of interest groups

5.5.1 Meaningful extraction of gaze data

The choice of AOI groups is based on the findings from the multimodal analysis and the exploration of the gaze data. It was determined that two phases of AOI group analyses would be appropriate for this study. First, it was decided that those areas of the screen that gather the most visual attention needed to be investigated further with the help of fixation-data. The observations from heat maps and gaze replay videos were consulted prior to the creation of these AOI groups. The AOI group analyses in this first phase focused on the generation of fixation-data on areas of the screen that show faces and telops that directly render dialogue. Second, it was discerned that the gaze data could give insights into differences in JLLs' visual attention on visual-verbal input that showcases different degrees of resemblance to the corresponding dialogue. This suspicion was based on observations from gaze replay videos on JLLs' gaze behaviour for narration telops. The second phase of AOI group analyses therefore focused on the telops that directly render dialogue. The different types of resemblance established during the multimodal analysis were thus also consulted prior to the definition and generation of AOI groups.

This means that the scheme for dynamic AOI groups had to be devised in such a way that it allows for AOI groups to generate fixation-data on faces, telops that directly render dialogue and the five types of resemblance of these telops. It was determined that the creation of dynamic AOIs should start with the types of resemblance and work its way towards telops that directly render dialogue and faces. It was also decided that an AOI group for the whole screen, similar to that of O'Hagan and Sasamoto (2016), needed to be included. This AOI group is included in order to work out how the viewing time allocated to each AOI group relates to the total viewing time.

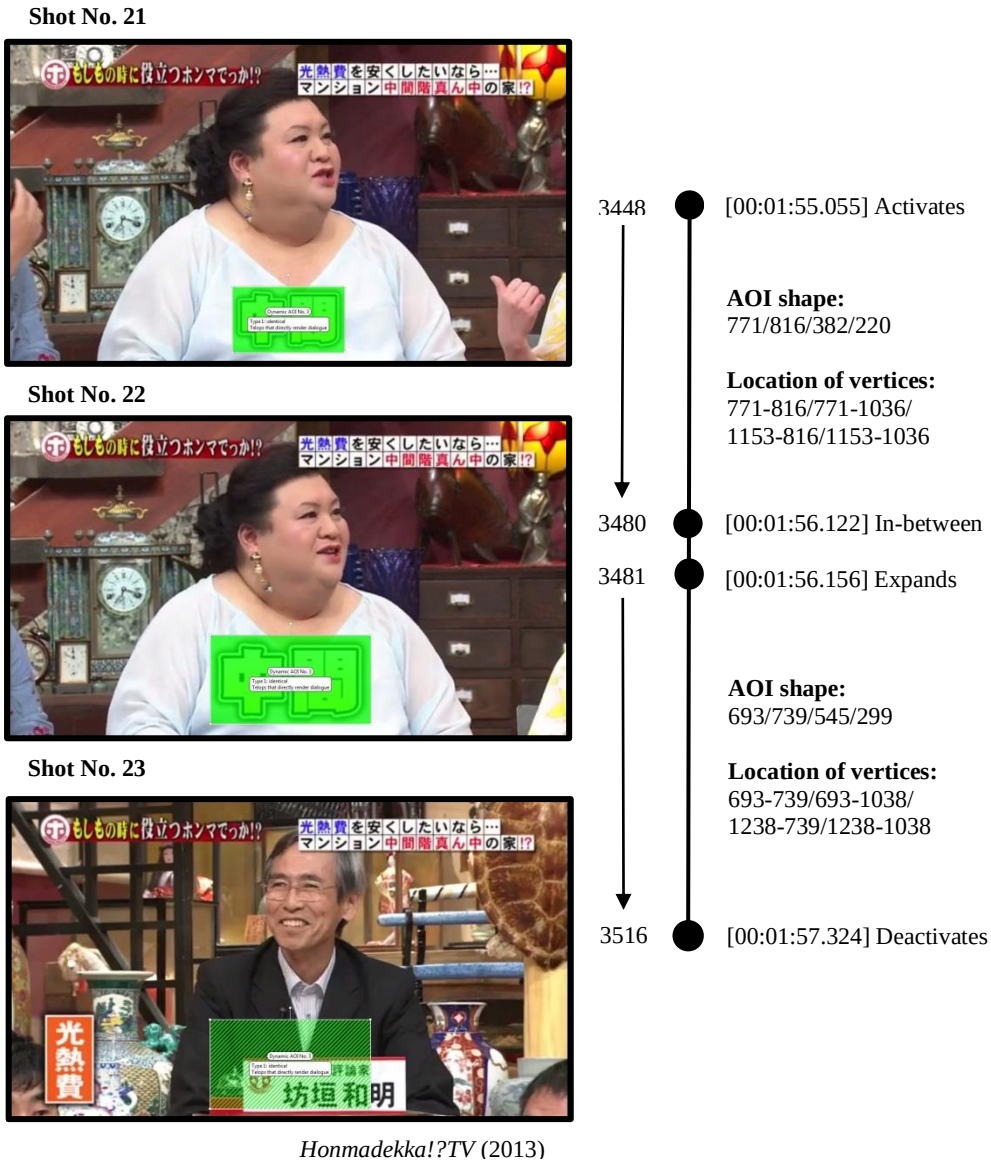
5.5.2 Areas of interest

As noted previously in *Subsection 5.3.3 Procedure for multimodal units*, 66 telops that directly render dialogue were counted in the multimodal transcript. However, rather than creating 66 dynamic AOIs per test in which AOIs are aligned with a single telop, it was decided to reduce the total number of dynamic AOIs to minimise visual clutter and complexity. This was achieved by assigning multiple telops to a single dynamic AOI where possible. This was deemed appropriate to this study's objectives as fixation-data is extracted on groups of telop that directly render dialogue, not each telop individually.

The still images of the multimodal transcript in which telops that directly render dialogue are featured were transferred to a separate document. The still images were then grouped based on the type of resemblance of the featured telops and were subsequently arranged according to the telop's location on the screen and movement in order to work out how many AOIs needed to be generated. Telops that were placed on the screen in a similar way and were assigned the same type of resemblance were included in one dynamic AOI. Telops that showcased a series of special effects such as moving or expansion were assigned their own dynamic AOI. The number of dynamic AOIs assigned to telops that directly render dialogue more than halved through this process to 31 AOIs per test which amounts to 124 AOIs across the four tests. The wide range of special effects used in the video stimulus for the telops made it necessary to set four criteria relating to movement to make sure that the activation of AOIs was aligned with the telops in a consistent manner. Dynamic AOIs activate from the first keyframe in which a telop:

1. Pops up onto the screen
2. Has become fully visible after slowly appearing
3. Remains static after it was shot into the screen
4. Starts scrolling or sliding into the screen from left to right

It was explained that telops appear on screen in five different ways. This list of criteria covers all of those special effects. A dynamic AOI deactivates as soon as a telop disappears. All of the telops disappear only in one way. Figure 5.12 demonstrates how such sequencing looks like and shows the alignment of AOI No. 3 with Telop No. 5. All of the AOI properties from the guidelines for this particular AOI are also included in Figure 5.12.



Honmadekka!?TV (2013)

Figure 5.12 Example of area of interest sequence

Telop No. 5 was given its own dynamic AOI as it involves an expansion of the word chūkan [Middle]. Telop No. 5 contains an utterance of a panellist that expands the second time it is uttered. Figure 5.12 illustrates that the AOI activates as soon as the telop pops up onto screen. An in-between keyframe then keeps the AOI in place until the expansion happens. The AOI deactivates directly after the telop has disappeared off screen.

It is important to note that the AOI group for telops that directly render dialogue consists of all the dynamic AOIs that were assigned to the five types of resemblance. This means that the 124 dynamic AOIs that were created for the extraction of fixation-data on telops have been given two labels; one for the type of resemblance they are affiliated with and one for the type of input they represent.

It was observed in the multimodal transcript that faces of hosts, panellists and expert guests were mostly featured in the upper region of the screen and at the bottom right. This meant that two dynamic AOIs for each test would suffice for the extraction of gaze data on faces. The width of the AOI for the upper region matched with the space needed on the screen to show faces of three people sitting next to each other. It only changed shape in cases where it would overlap with other AOIs. The final AOI was created on the whole screen in each of the four tests. This particular AOI activates as soon as Segment 2 starts and remains activated until Segment 4 ends.

The outcome of the arrangement and generation of AOIs can be found in *Appendix L. Scheme for area of interest groups*. A total number of 136 AOIs was created; 124 AOIs for the telops that directly render dialogue combined with eight AOIs aligned with faces and four AOIs that covered the whole screen.

5.5.3 Grouping of areas of interest

As shown in *Appendix L. Scheme for area of interest groups*, eight AOI groups were created for the fixation-data analysis. AOI groups have been differentiated by colour. The corresponding dynamic AOIs have been given the same colour although different hues were used for each test to distinguish between them during the playback of the video stimulus and the data analysis. The colours that were used to indicate the types of resemblance in the multimodal transcript coincide with those used for the AOI groups of the types of resemblance; green, pink, blue, brown and purple. As noted previously, the colours for the types of resemblance were chosen based on availability of default colours in the eye-tracking software. The colours for the AOI groups on telops that directly render dialogue (i.e. red), faces (i.e. grey) and the whole screen (i.e. orange) were chosen for the same reason.

Figure 5.13 shows how the four identical sets of dynamic AOIs are linked to the AOI groups. Dynamic AOIs were placed in intervals on the timeline of each test in which the types of input appeared and aligned with the exact location of the stimuli on the screen. As shown in Figure 5.13, this means that a large portion of the video stimulus has a combination of AOIs activated.

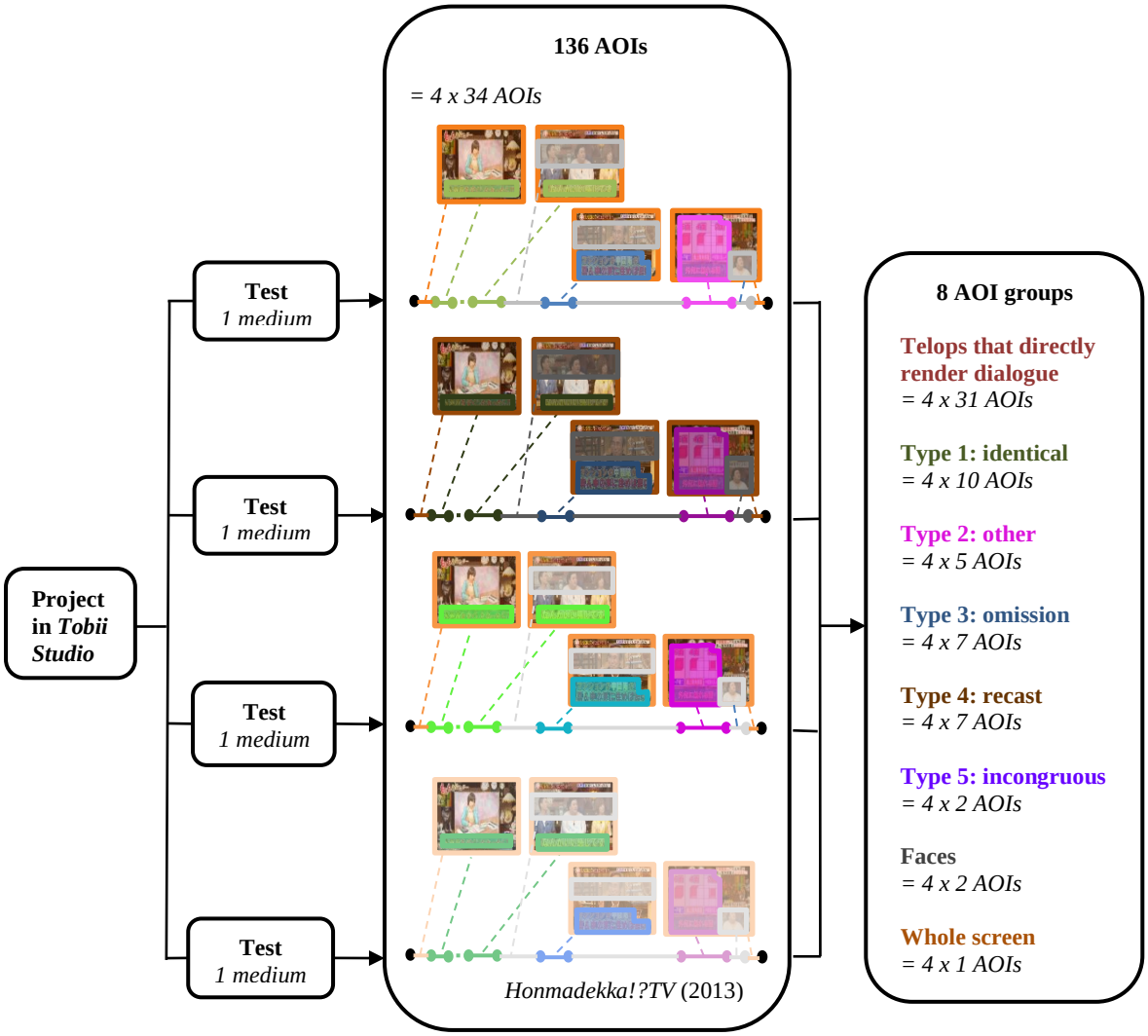


Figure 5.13 Organisation of areas of interest into area of interest groups

It was explained in the previous subsection that the AOI group for faces contains dynamic AOIs that align with almost the complete upper region of the screen. More specifically, it covers the area of the screen where three faces can appear. When relating the placement of these four AOIs back to the multimodal transcript, it can be argued that this AOI group also includes areas of the screen where gesticulation appears. However, based on the findings from the data exploration phase, it became clear that little visual attention appeared to be directed towards gesticulation. It is therefore expected that the AOI group for faces gives an appropriate representation of the visual attention paid to faces of hosts, panellists and expert guests.

5.5.4 Procedure for bar charts

The bar charts in this thesis have been created and formatted in *Microsoft Excel* with the .txt files that were exported from *Tobii Studio*. These bar charts and files show AOI

group summary data per participant group. The fixation-data exported from the eye-tracking software is shown in *Appendix M. Fixation-data exports*. The bar charts are based on the fixation-data included in the sum columns of the appended tables. The sum columns show the total number of fixation counts and the total duration of these fixations in seconds per AOI group. The bar charts on types of input include a bar that displays fixation-data on other types of input. This bar represents all of the remaining visual types of input that were not included in the eight AOI groups; text in background (i.e. nameplates), telops that facilitate comprehension (i.e. titles, themes, and speaker identifications), telops that bridge scene change (i.e. screens filled with text), imagery (i.e. drawings), and kinesic action (i.e. gesticulation, and bodily configuration). As can be seen in *Appendix M. Fixation-data exports*, there was no fixation-data exported on an AOI group named "Other types of input". The summary data of other types of input was calculated from the .txt files with the help of the AOI group for the whole screen. The data from the sum columns of faces and telops that directly render dialogue were subtracted from the sum column of the whole screen to get this fixation-data. Proportions were thereafter calculated and added to the bar charts.

Normalisation of the gaze data of this study was considered to calculate the fixation density for each AOI group due to the differences in number and screen times of telops (Conklin and Pellicer-Sánchez, 2016, p. 9). However, since a comparison is made between participant groups who looked at the same video stimulus with AOIs kept constant across the analysis it was deemed inappropriate for the purposes of this study to scale the gaze data (Holmqvist *et al.*, 2011, p. 226).

5.6 Summary

This chapter illustrated that the focus of the multimodal and empirical data analyses is not on the entire video stimulus; it was decided during Stage 1 of the research design that only three of the five video segments are taken up for analysis. Audio-verbal, audio-nonverbal, visual-verbal and visual-nonverbal types of input within these video segments were explored further at the start of this chapter. It was discerned that not all of the encountered types of input could be considered as potential contributors to a better understanding of the video material for JLLs or as types of input that were expected to gather much visual attention from participants. Seven types of input that together comprise 23 varieties of input were selected across the four forms of expression and theorised as helpful to JLLs' understanding of the analysed video segments' contents.

The selected types and varieties of input were brought together in a multimodal transcript. The transcript was divided into 66 multimodal units that coincided with the number of telops that directly render dialogue encountered in the analysed video segments. These units were thereafter assigned a type of resemblance which indicated the extent to which the contents of the telop resembles the contents of the corresponding dialogue. The scope of visual-nonverbal input was consulted during this procedure. The addition of types of resemblance to the multimodal transcript followed from findings of the exploration phase of the empirical data analysis in an attempt to further explore the relationship between visual attention and audio-verbal input.

This chapter also reported on the multimodal landscape of the analysed video segments. It demonstrated that those varieties of input that occurred most frequently in the video material in a spatial sense were not necessarily the most informative varieties or the most attention-seeking varieties in a temporal sense. Whereas telops that directly render dialogue were prevalent in shots and consisted of a high number of add-on text units, telops that facilitate comprehension and telops that bridge scene change were fewer in number and showed a tendency to repeat information from one another in addition to the information contained in utterances of speakers despite their high rates of occurrence within shots. These observations combined with the findings from the exploration phase of the empirical data analysis informed the choice of 136 AOIs and eight AOI groups discussed at the end of this chapter.

The following chapter focuses on the findings from the empirical data analyses of JLLs' multimodal perception. It not only describes the data; it also explains the implications these results carry for answering this study's research questions and the established literature.

6. JAPANESE LANGUAGE LEARNERS' PERCEPTIONS OF A JAPANESE VARIETY SHOW

Chapter 6 is the first empirical chapter of this thesis. It builds further on the account given in the previous chapter of the scope of input and multimodal landscape identified in the analysed video segments and presents the findings on JLLs' multimodal perception of the video material. The purpose of this chapter is to demonstrate participants' viewer experience of the video material and to give insights into their thoughts on the use of Japanese variety shows in the language classroom.

Section 6.1 starts with a description of findings from closed-ended data items and semantic differential scales regarding particular types of input. Section 6.2 elaborates on one of these types of input in particular, telop design, and presents the outcome of the thematic analysis on open-ended data items. This not only includes an explanation of the thematic map; it also reports on the thematic networks and shows JLLs' perceptions of different aspects of the telop design in relation to other types of input with the help of representative data extracts. Section 6.3 presents results from responses to the comprehension test and closed-ended data items on JLLs' overall viewer experience of the video stimulus. Section 6.4 then summarises all of these findings and interprets the empirical data in connection to this study's research questions and the established literature.

6.1 Viewer experience of types of input

6.1.1 Observations for telop

The diverging stacked bar charts that summarise participants' perceptions of telops consist of colour gradients. These colour gradients include red, grey and green colours. Red hues indicate negative viewer experiences, grey signifies neutral attitudes, while green hues demonstrate positive perceptions. The more saturated the colour the more negative or positive the experience has been for participants. Figure 6.1 shows the results on JLLs' perceptions of telops for Group 1: pre-exchange.

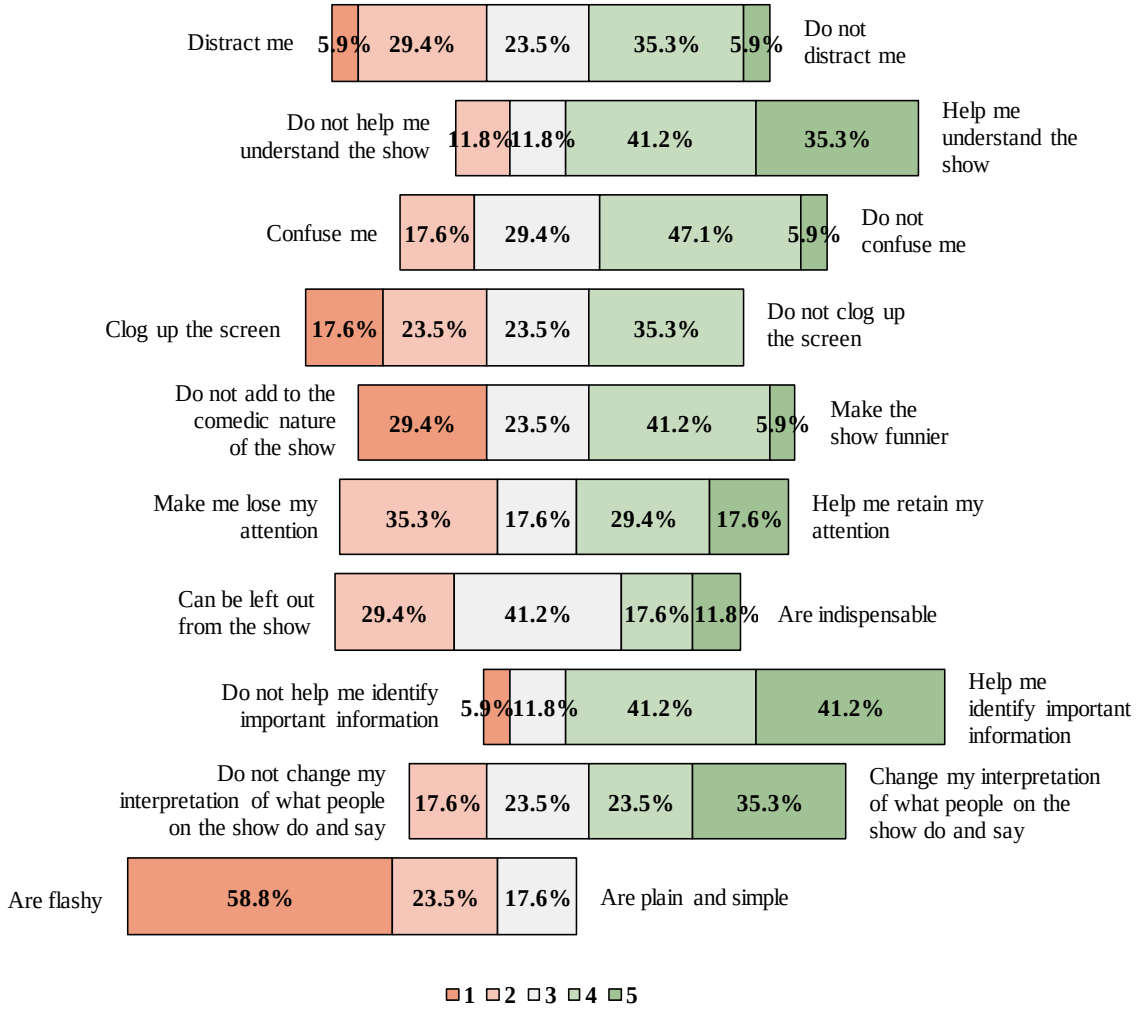


Figure 6.1 Perceptions of telops for Group 1: pre-exchange

The diverging stacked bar charts in Figure 6.1 demonstrate that JLLs in Group 1: pre-exchange had a predominantly positive attitude towards four statements regarding telop. According to the respondents, telops help with understanding the video material (76.5%), are not confusing (53%), help with the identification of important information (82.4%), and change the interpretation of actions and utterances of people appearing on the show (58.8%). The bar chart on confusion shows that a high percentage of responses is neutral. The charts show that opinions are divided regarding five other statements. These involve distraction (35.3% versus 41.2%), visual clutter (41.1% versus 35.3%), humour (29.4% versus 47.1%), focus of attention (35.3% versus 47%), and necessity (29.4% versus 29.4%). The latter in particular appears to have the highest percentage of neutral responses. The proportions for the statement regarding the saliency of telops, on the other hand, illustrate that a large majority of JLLs (82.3%) find telops flashy. Figure 6.2 gives details on the diverging stacked bar charts of JLLs in Group 2: exchange.

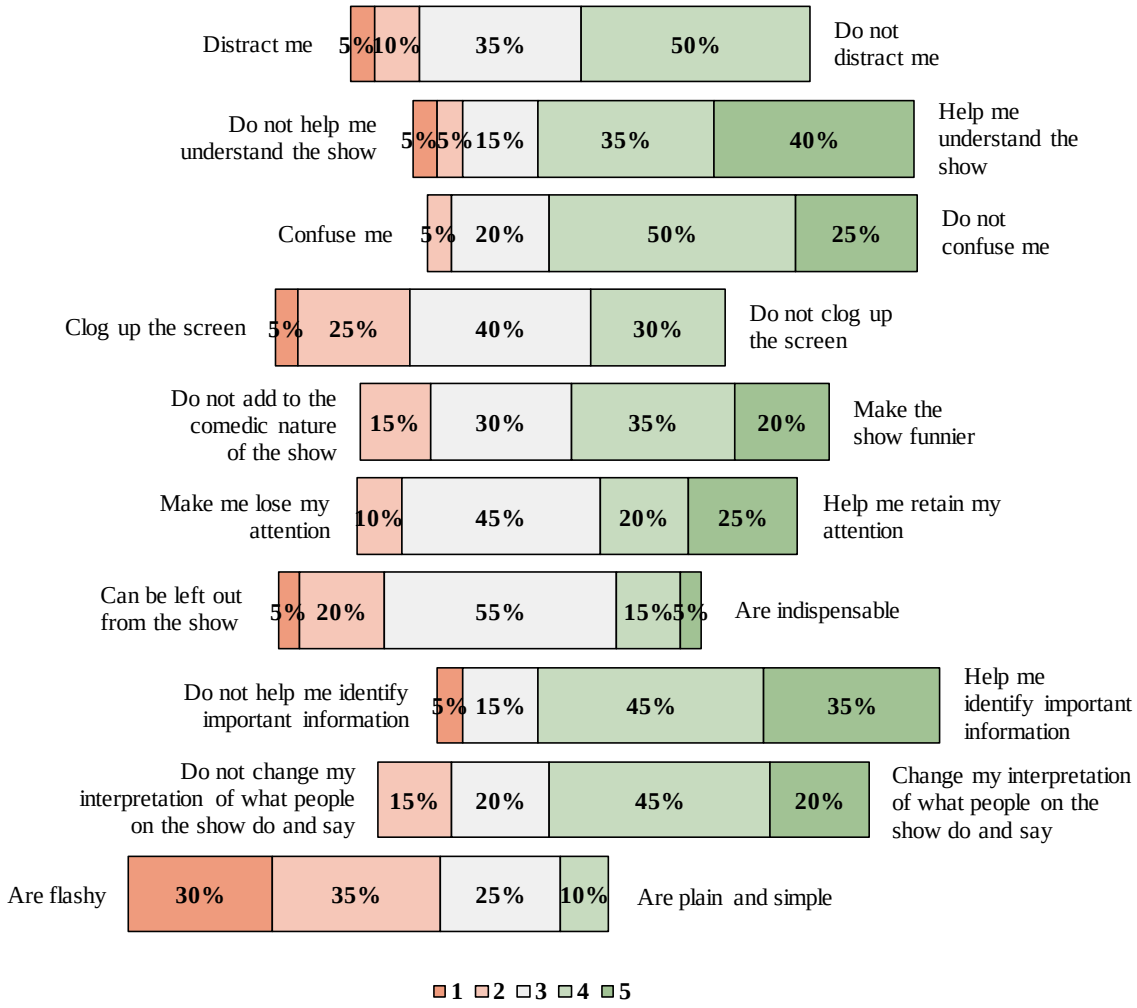


Figure 6.2 Perceptions of telops for Group 2: exchange

The percentages demonstrate that JLLs in this particular learner group have a predominantly positive attitude towards seven out of ten statements. JLLs not only feel that telops do not distract them (50%), they also think that telops help them understand the video material (75%), do not confuse them (75%), make the Japanese variety show funnier (55%), help them retain their attention (45%), help them identify important information (80%), and change their interpretation of actions and utterances of people appearing on the show (65%). Nonetheless, three of these statements (i.e. distraction, humour, and focus of attention) also show high percentages of neutral responses. Opinions are divided for only two statements. These relate to visual clutter (30% versus 30%) and necessity (25% versus 20%). These statements also show large proportions for neutral responses. Again, participants agree that telops are flashy (65%). Figure 6.3 gives an overview of JLLs' perceptions in Group 3: post-exchange.

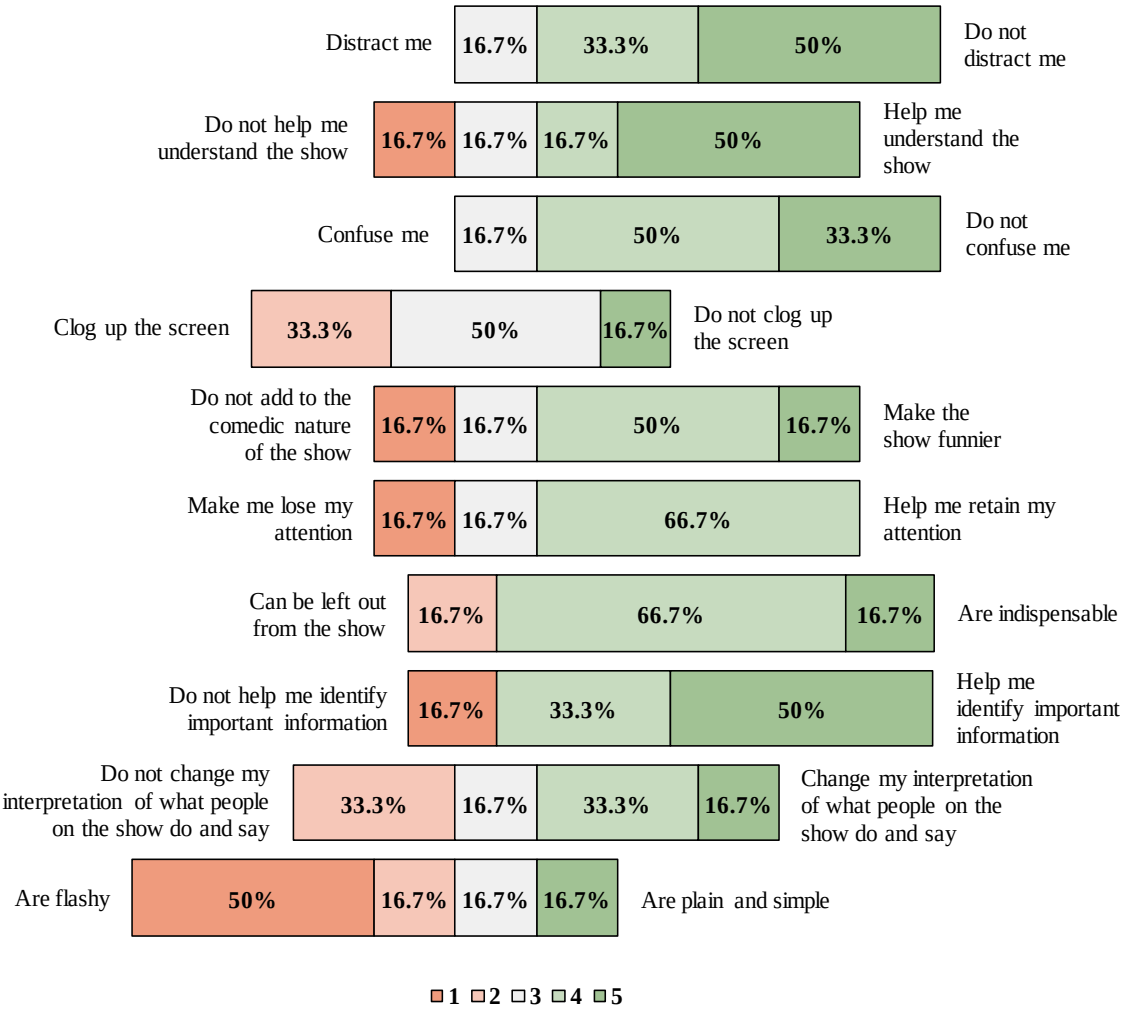


Figure 6.3 Perceptions of telops for Group 3: post-exchange

It demonstrates that participants have a predominantly positive attitude towards telops as seven out of ten statements show a majority of responses at the positive end of the scales. Participants feel that telops are not distracting (83.3%), help with understanding the video material (66.7%), are not confusing (83.3%), make the show funnier (66.7%), help with the retention of attention (66.7%), are indispensable (83.4%), and help with the identification of important information (83.3%). It was observed that none of the JLLs in Group 3: post-exchange feel that telops distract or confuse them. Attitudes are divided for the statements on visual clutter (33.3% versus 16.7%), and interpretation of actions and utterances of people appearing on the show (33.3% versus 50%). The former even illustrates that half of the participants selected a neutral response. Respondents in Group 3: post-exchange also think that telops are flashy (66.7%). The diverging stacked bar charts in Figure 6.4 give an overview of the results on Japanese speakers' perceptions of telops in the Reference group.

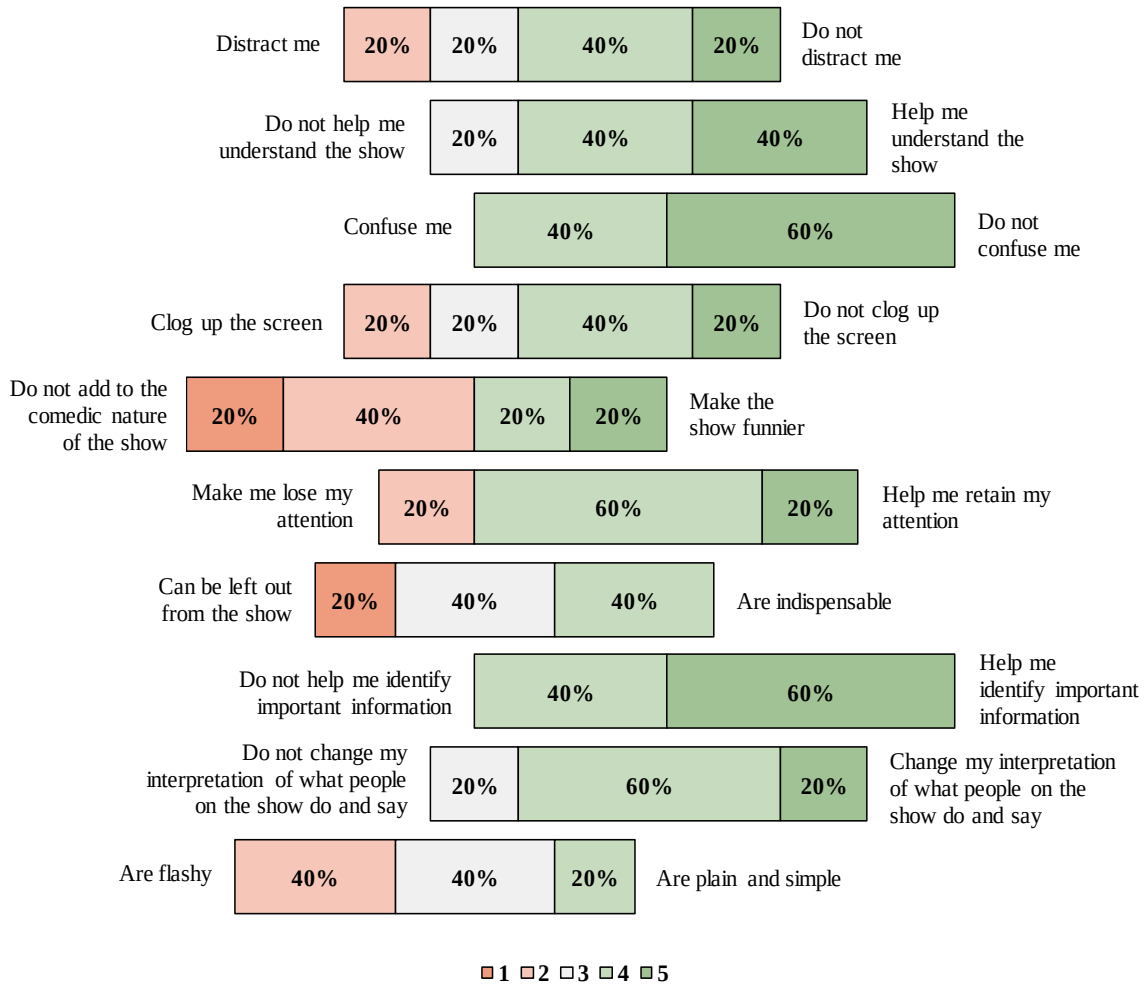


Figure 6.4 Perceptions of telops for Reference group

It demonstrates that participants had a predominantly positive attitude towards seven statements. Japanese speakers feel that telops do not distract them (60%), help them understand the video material (80%), do not confuse them (100%), do not clog up the screen (60%), help them retain their attention (80%), help them identify important information (100%), and change their interpretation of actions and utterances of people appearing on the show (80%). It appears that none of the respondents think that telops do not help with understanding the Japanese variety show, are confusing, do not help with the identification of important information, and do not change their interpretation of what people appearing on the show do and say. Attitudes are divided on three statements. These relate to humour (60% versus 40%), necessity (20% versus 40%) and saliency (40% versus 20%).

Overall, these numbers suggest that all participant groups have positive attitudes towards telops. However, it appears that participants in Group 1: pre-exchange in particular

experience more difficulties with telops as compared to the other three participant groups based on the results of the diverging stacked bar charts. JLLs in Group 1: pre-exchange feel more distracted by telops, feel that telops clog up the screen, have more difficulty retaining their attention, and do not necessarily think that telops are indispensable or add humour to the video material. There appears to be an upward trend in participants' perceptions from Group 1: pre-exchange to the Reference group regarding these matters.

6.1.2 Observations for telop design

Table 6.1 gives an overview of JLLs' attitudes towards the difficulty level of telops and participants' views on three different aspects of the telop design in the video stimulus. These include the readability of colours and colour combinations, the legibility of the text, and the suitability of the telop design to the Japanese variety show. Table 6.1 also shows participants' attitudes towards a telop design that includes white text and no effects.

Table 6.1 Perceptions of telop design

Aspect	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Comprehension of text	52.9% a few lines 41.2% half of it 5.9% most of it	55% a few lines 35% half of it 5% most of it 5% everything	50% most of it 33.3% half of it 16.7% a few lines	-
Difficult colours or colour combinations	76.5% no 23.5% yes	75% no 25% yes	No	No
Extent of legible text	52.9% majority 47.1% everything	69% majority 40% everything	66.7% everything 33.3% majority	60% majority 40% everything
Suitability to Japanese variety show	94.1% suitable 5.9% unsuitable	90% suitable 10% unsuitable	Suitable	80% suitable 20% unsuitable
Preference to plain design	82.4% no 17.6% yes	90% no 10% yes	No	No

The responses demonstrate that none of the JLLs felt they were not able to understand any of the telops. It appears that Group 3: post-exchange is the only learner group in which a large percentage of JLLs responded to have understood most of the telops. The majority of JLLs in Groups 1-2, on the other hand, indicated to have understood a few lines to half of the telops that appear on screen.

It was observed during the multimodal transcription that a variety of colours, colour gradients, colour combinations, typefaces and fonts were used for telops in the video material. A majority of subjects in each participant group indicated not to have experienced any difficulties with the colours or colour combinations of telops. In addition, all participant groups responded that the design of at least the majority of telops was legible.

A vast majority of subjects in each participant group thought that the design of telop suits the variety show. Only 8.3% of respondents thought that the design was inappropriate. It appeared that a large majority of participants did not prefer a telop design with fewer colours or effects. Only 10.4% preferred such a design of telop. Overall, these numbers suggest that the telop design was considered to be appropriate to the video material by all participant groups.

6.1.3 Observations for other types of input

Table 6.2 shows participants' attitudes towards the use of drawings in the video stimulus, and gives an overview of the types of input that JLLs considered as helpful for their understanding of difficult telops.

Table 6.2 Perceptions of other types of input

Aspect	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Helpfulness of drawings	94.1% helpful 5.9% unhelpful	75% helpful 25% I do not know	Helpful	Helpful
Support for comprehension	82.4% non-verbal clues 82.4% known kanji 64.7% context 52.9% dialogue 11.8% colours/fonts 5.9% other	80% non-verbal clues 80% known kanji 65% context 55% dialogue 35% colours/fonts 15% radicals 5% phonetic elements 5% other	100% non-verbal clues 100% context 83.3% known kanji 83.3% dialogue 50% colours/fonts 16.7% radicals 16.7% phonetic elements	-

The responses illustrate that participants have a predominantly positive attitude towards the use of drawings. A large majority, 86.0% of JLLs, experienced them as helpful to a better understanding of the Japanese variety show. Only one participant in Group 1: pre-exchange indicated that such visual-nonverbal input did not help with understanding the programme contents.

JLLs indicated that non-verbal clues (e.g. imagery, body language, and facial expression) in particular help with their understanding of difficult telops. This type of input is followed by those kanji that were known to participants, context (e.g. hiragana and katakana), dialogue, and colours and fonts. The radicals and phonetic elements mentioned in the table refer to those used in kanji that were unknown to participants. These were only selected by JLLs in Groups 2-3. Two JLLs had noted down other types of input in the other section of the response box. The participant in Group 1: pre-exchange indicated that "the images/drawings provided helped with understanding". The JLL in Group 2: exchange had included "the narrator read as the words appeared in some instances". This further suggests

the importance of non-verbal input for JLLs and the usefulness of the co-occurrence of narrations and narration telops.

6.2 Viewer experience of multimodal landscape

6.2.1 Overview of (sub)themes

The thematic map is shown in Figure 6.5. It includes the thematic networks, the total number of data extracts and uncoded segments included in the (sub)themes, and the proportions that these extracts represent from the total number of data extracts.

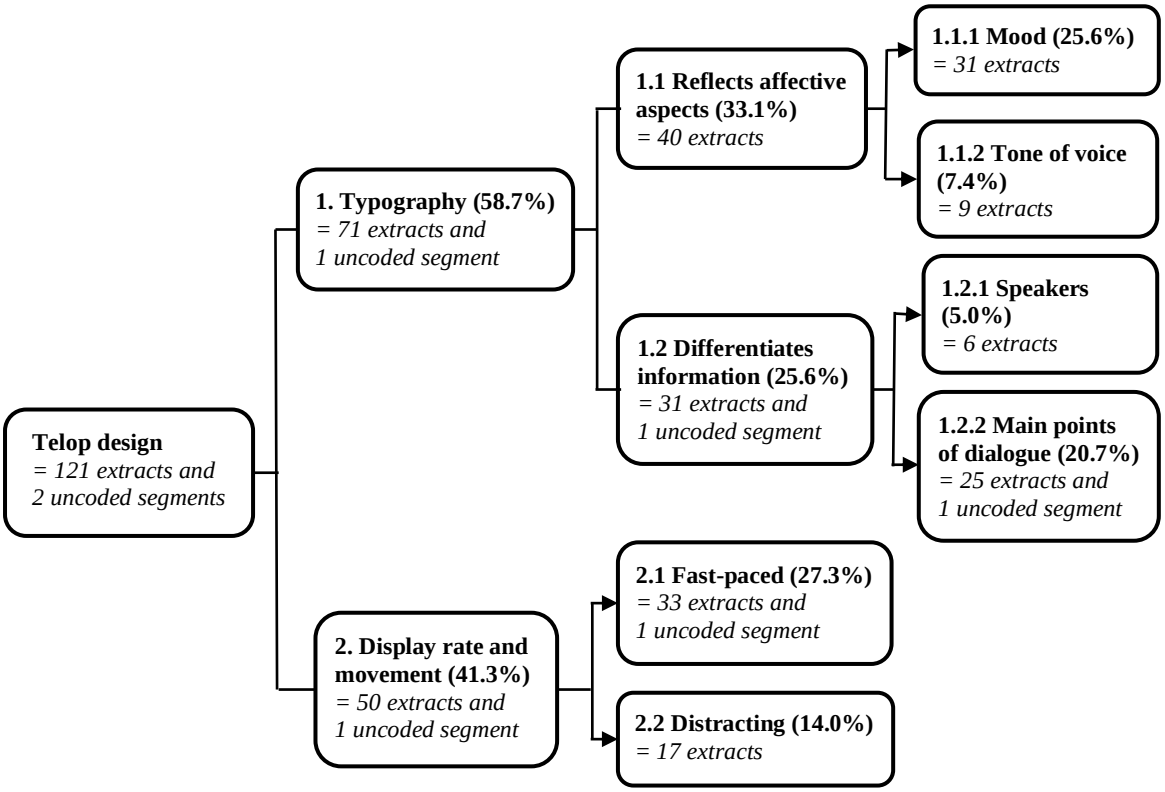


Figure 6.5 Organisation of data extracts and uncoded segments into (sub)themes

The name of a (sub)theme consists of a number and a description (e.g. Theme 1: typography, and Subtheme 1.1: reflects affective aspects). Figure 6.5 shows that all (sub)themes relate to the design of telops. As noted previously, these telops are those that directly render dialogue.

The thematic map comprises two thematic networks. The first network includes the highest number of data extracts and takes a closer look at typography. It consists of one theme, typography, and two layers of subthemes. These comprise a total of six subthemes. The subthemes of Thematic Network 1 follow two strands that structure part of the overall

narrative and focus on JLLs' perceptions of helpful aspects of the telop design. According to JLLs, typography reflects affective aspects and differentiates information. The second layer of subthemes then further outlines the two affective aspects and two types of information these perceptions are concerned with. The second thematic network delves deeper into JLLs' perceptions of the display rate and movement of telops. These are included in one theme as they both appeared to pose challenges to JLLs. Thematic Network 2 also follows two strands but it comprises one layer of subthemes. It contains only two subthemes which illustrates that this second network is less complex than the first one. According to JLLs, telops that directly render dialogue are also fast-paced and distracting.

Table 6.3 shows the extent to which participant groups are represented in the (sub)themes.

Table 6.3 Representation of participant groups in (sub)themes

Theme	Subtheme	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Typography	Reflects affective aspects	35%	42.5%	12.5%	10%
	Mood	38.7%	35.5%	16.1%	9.7%
	Tone of voice	22.2%	66.7%	-	11.1%
	Differentiates information	25.8%	32.3%	35.5%	6.5%
	Speakers	16.7%	66.7%	16.7%	-
	Main points of dialogue	28%	24%	40%	8%
Display rate and movement	Fast-paced	39.4%	39.4%	18.2%	3%
	Distracting	23.5%	47.1%	11.8%	17.6%

The percentages are calculated from the total number of data extracts per (sub)theme. Overall, 32.2% of data extracts in the thematic map is derived from data items of Group 1: pre-exchange. These ratios are 39.7% for Group 2: exchange, 19.8% for Group 3: post-exchange, and 8.3% for the Reference group. These percentages are more or less the same for each of the thematic networks.

It appears that the second thematic network in particular shows a big difference in representation of participant groups. Fewer responses of participants in Group 3: post-exchange are included as compared to JLLs in Group 1: pre-exchange and Group 2: exchange. The percentages appear to be more evenly distributed in the first thematic network. However, it appears that the second layer of (sub)themes in the first thematic network also showcases predominance of particular participant groups. Mood comprises more data extracts of participants in Groups 1-2. Tone of voice and speakers are more often mentioned in data extracts of JLLs in Group 2: exchange. Main points of dialogue, on the other hand, is more often referred to by subjects in Group 3: post-exchange.

It is important to take the number of subjects per participant group into account for the interpretation of these percentages as fewer participants means that fewer data items were available in the data set. A table that shows the proportions of data extracts calculated over the total number of data extracts included for each participant group gives a more accurate description of prevalence of (sub)themes. This is demonstrated in the following subsection.

6.2.2 Prevalence of (sub)themes

Table 6.4 shows how prevalent each (sub)theme is per participant group.

Table 6.4 Rates of occurrence for (sub)themes

Theme	Subtheme	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Typography	Reflects affective aspects	35.9%	35.4%	20.8%	40%
	Mood	30.8%	22.9%	20.8%	30%
	Tone of voice	5.1%	12.5%	-	10%
	Differentiates information	20.5%	20.8%	45.8%	20%
	Speakers	2.6%	8.3%	4.2%	-
	Main points of dialogue	17.9%	12.5%	41.7%	20%
Display rate and movement	Fast-paced	33.3%	27.1%	25%	10%
	Distracting	10.3%	16.7%	8.3%	30%

The percentages are calculated from the total number of data extracts per participant group. The thematic map includes a total of 39 data extracts from data items that belong to JLLs in Group 1: pre-exchange. These are 48 data extracts for Group 2: exchange, 24 data extracts for Group 3: post-exchange, and 10 data extracts for the Reference group.

It appears that different (sub)themes are prevalent in the data items for each participant group. Reflects affective aspects is more prevalent among JLLs in Groups 1-2. It is suggested that participants in Group 1: pre-exchange speak more about the mood while subjects in Group 2: exchange have more data items on the tone of voice. Differentiates information is prevalent among JLLs in Group 3: post-exchange in particular. The majority of these data extracts are in main points of dialogue. It is suggested that both subthemes of the second thematic network are more prevalent among JLLs in Groups 1-2. Whereas JLLs in Group 1: pre-exchange mention fast-paced more often, distracting is referred to by participants in Group 2: exchange to a greater extent.

6.2.3 Thematic Network 1

The first subtheme in the second layer of Thematic Network 1 (i.e. Subtheme 1.1.1: mood) includes data extracts in which JLLs relate the typographical features of telop to the overall

mood of the video material. The features that are mentioned most often are colour and font. JLLs perceive these to be coherent with their impression of the authentic Japanese variety show. The data extracts support this observation in three ways. First, they include responses in which participants briefly describe their impression of the Japanese variety show. According to them, the video material appears to be “funny”, “quirky and radical”, “funny/serious”, “tabloid-esque” [sic], and not suited to a formal or serious telop design. Second, the extracts contain responses in which JLLs describe their impression of the telop design. The adjectives they use for this depiction are similar to those used for the Japanese variety show and include “fun and exciting”, “comical”, “casual”, and “informal”. Third, a number of data extracts explicitly state that particular aspects of the video material are reflected in the telop design. Such responses link, for example, the variety in brightness and boldness of the design to the funny and light-hearted nature of the Japanese variety show:

Group 1: pre-exchange, P25

The font used was bright and bold which reflected the comedy aspects of the scene well.

post1 Q25b, ll. 1-2

Group 1: pre-exchange, P34

It seemed fun and dynamic and interesting, which seemed to be the mood that the show was trying to create

post1 Q25b, ll. 1-2

Group 2: exchange, P15

The variety show seemed like a funny + light-hearted talkshow, thus the varied and comical text design was fitting

post1 Q25b, ll. 1-2

It appears that the use of fonts and colours not only matches the mood and dynamic nature of the video stimulus; it also reflects affective aspects such as emotion and humour:

Group 3: post-exchange, P28

Font reflected that the text was meant to be funny but it was still legible so I could get the information.

post1 Q25b, ll. 1-2

Group 3: post-exchange, P43

Also it was used well to express emotions and comedy ...

post1 Q25b, l. 3

The data extracts suggest that this perceived coherence between the telop design and the overall impression of the authentic Japanese variety show enables JLLs to pick up on the affective context regardless of their understanding of the verbal content. As noted in *Subsection 3.3.2 Video material*, the chosen excerpt includes an authentic talk-show that combines jokes and humour with interesting facts on various topics. This means that JLLs understand the format of the video material. This also suggests that the perceived

coherence between telop design and the overall mood of the video material gives JLLs the feeling that they understand the general idea of the Japanese variety show. This is evidenced by their data extracts:

Group 1: pre-exchange, P27

(even though it does help Japanese learners to catch the context).

post1 Q26, ll. 4-5

Group 1: pre-exchange, P41

Watching those kind of realistic movies can be really usefull to learn Japanese. They are not so hard to understand what is say in general, funny and is a good way to understand the culture.

post2 Q23, ll. 1-3

Group 2: exchange, P08

It helped set the mood and helped with the delivery of some of the jokes. It also made it easier to understand which is one of the aims of the variety show

post1 Q25b, ll. 1-3

Group 3: post-exchange, P46

It was fun and interesting, and it helped explain the concepts.

post1 Q25b, l. 1

The second subtheme in the second layer of Thematic Network 1 (i.e. Subtheme 1.1.2: tone of voice) is closely related to this. Although this subtheme is much smaller in size (9 versus 31 data extracts), it extends the claim that participants can understand the general feel of the video stimulus to that of the tone of voice of utterances made by speakers. Respondents relate the colours, fonts and movements to "emotions in words", and exaggeration of the content. The data extracts suggest that participants think that the telop design reflects the expressions or feelings of the speakers, and feel this helps them determine affective aspects of utterances of speakers:

Group 1: pre-exchange, P41

you can almost guess if what is said is funny, surprising, if someone gets angry..

post1 Q25b, ll. 1-2

Group 2: exchange, P05

helped enulate the expressions/feelings of the speakers

post1 Q25b, l. 1

Group 2: exchange, P13

The style used represented how things were said
eg) Bold font for things said loudly or in a shocked voice.

post1 Q25b, ll. 1-2

Group 2: exchange, P14

The colours, fonts and movements illustrated the emotion behind the text. It felt a bit like when a sitcom laugh track indicates a joke.

post1 Q20, ll. 1-2

Reference group, JP02

気持ちや言い方が強かったりすると字体が変わっていた。 [The font changed when emotions and expressions were strong.]

post1 Q16, l. 1

As was observed with Subtheme 1.1.1: mood, colours and fonts give indications of the emotional state or attitude of speakers, irrespective of the verbal content of their utterances. This suggests that coherence is not only perceived between typographical features and the overall mood of the video material; it is also observed between telop design and the mood of speakers appearing on the show. According to the participants, the use of different colours and fonts signal nuances in the tone of voice which help them identify the context of an utterance. Although this does not guarantee that participants fully understand the panel discussions in the authentic Japanese variety show, such layers of coherent types of input at least help them figure out the gist of the dialogue. When taking a closer look at the multimodal transcript it appears that a clear link between the telop design and the speakers' tones of voice can be detected for a number of shots. These include, for example, a telop that expands when the speaker reiterates an utterance with emphasis and surprise, and a telop that quivers when the host utters something in an accusatory tone of voice. This demonstrates that JLLs pick up on this particular use of telop design.

The third subtheme in the second layer of Thematic Network 1 (i.e. Subtheme 1.2.1: speakers) is part of the second strand in the narrative and comprises a relatively small number of data extracts in which JLLs explain that colours and fonts differentiate speakers appearing on the Japanese variety show:

Group 1: pre-exchange, P32

P32 explained that colours helped identify who is/was speaking.

Field notes, ll. 2-3

Group 2: exchange, P01

Variety of colours reminded me of variety of topics & speakers.

post1 Q25b, ll. 1-2

It was observed during the exploration phase of the multimodal analysis that there appeared to be no relationship between the varieties of telop design and speakers. Nonetheless, the data extract of P28 of Group 3: post-exchange includes a comment that has also been discussed in *Subsection 5.2.4 Visual-verbal input* for the exploration of visual-verbal input; the difference in telop between hosts and expert guests:

Group 3: post-exchange, P28

I feel like the font is changed depending on who's speaking, masculine/feninine, whether the Japanese is carse or polite; the font used for the experts was different to the host sometimes.

post1 Q26, ll. 1-3

It has been argued that there is one variety of input that is exclusive to expert guests: speaker identification. These are telops that facilitate comprehension and appear with corresponding telops that directly render dialogue. The telops that accompany speaker identifications display condensed versions of utterances, have a different font, are written in red and black, and appear with a background colour. This was shown on the right in *Figure 5.4 Varieties of telops that facilitate comprehension*. This is the only format in telop design that can be linked to a particular group of speakers (i.e. expert guests). Although it is difficult to ascertain whether the data item refers to such speaker identifications, it appears that subtle differences in telop design are noticed by JLLs and used for their understanding of the video material.

The fourth subtheme in the second layer of Thematic Network 1 (i.e. Subtheme 1.2.2: main points of dialogue) is much larger in size and consists of data extracts in which participants argue that typographical features differentiate important from less important points in the dialogue. Again, colours and fonts are mentioned most often. According to the participants, such features are used to draw viewers' attention to those points of the dialogue that are perceived to be most important. This observation is supported by the data extracts in two ways. First, the extracts comprise responses in which the participants state that typography is used to "catch your eye", "accentuate punchlines" or to show "emphasis". This suggests that typographical features are thought to stress certain parts of the utterances. Second, the data extracts include comments in which participants explain that colours and fonts are used to "highlight important info" or "make them stand out". In other words, it appears that typography is perceived to be useful for the structuring and compartmentalisation of strands of text, and points out which parts of a telop needs to be visually attended to as mentioned by various participants:

Group 1: pre-exchange, P36

different highlighted words.

post1 Q20, l. 2

Group 2: exchange, P13

Size: bigger to make important points stand out or if something was said with enphasis.
Colour: … and to make key parts of a sentence stand out

post1 Q20, ll. 1-4

Group 3: post-exchange, P42

The most imoportant part or key word (sentences) were taken out of the conversation (explanation by the expert guestes), which makes me understand the key point quickly, and it also could catch my eyes relatively easier, as well

post1 Q25b, ll. 1-4

Group 3: post-exchange, P43

P43 thought that telops were maybe used because Japanese words often have the same pronunciation. In this way it would be easier to distinguish between the words. P43 felt that typography helped with dividing up the sentences in chunks.

Field notes, ll. 2-6

Reference group, JP06

大きくて色もついていて、重要なところだけをピックアップしていてよかった。 [It was great that only important parts were picked up due to their larger size or the addition of colours.]

post1 Q21b, ll. 1-2

When taking a closer look at the multimodal transcript it is difficult to establish a clear link between the use of colours and the importance of text units to understanding the panel discussions. Key words are sometimes shown in a brighter colour as compared to the colours that are used for other text units, but such choice of colours is not applied in a consistent manner. Nonetheless, it is observed that the use of two or more colours for one telop is not unusual in the analysed video segments.

It is important to note that data extracts in which participants argue that typography helps them to quickly identify key points are mostly from JLLs in Group 3: post-exchange. Data extracts in which typographical features are described to highlight and emphasise words are included in all three learner groups. Although it is difficult to determine whether this means that JLLs in Groups 1-2 did not understand the contents of the highlighted words, this could suggest that such use of typography may carry more implications to understanding for JLLs in Group 3: post-exchange.

6.2.4 Thematic Network 2

The subtheme in the first strand of Thematic Network 2 (i.e. Subtheme 2.1: fast-paced) involves data extracts in which JLLs relate the display rate of telops to the progression of the video material. According to participants, these aspects are coherent. The data extracts support this observation in three ways. First, it includes a data extract in which an impression of the authentic Japanese variety show is given. According to P22 of Group 1: pre-exchange, the video material is “loud and fast-paced”. Second, a number of data extracts explain how this impression is reflected in the telop design, which is described in the following terms:

Group 1: pre-exchange, P29

The text was quick and snappy, like the progression of the show.

post1 Q25b, l. 1

Group 2: exchange, P02

it was very fast-paced and sensational. The big, bright font added to the show's pace and exaggerated certain moments of the show.

post1 Q25b, ll. 1-3

Group 2: exchange, P16

It was bold and intense similar to the content

post1 Q25b, l. 1

Third, JLLs in Groups 1-2 link the appearance of telops to the dialogue. Although not explicitly stated, such responses appear to focus on both utterances of speakers and narrations. This could suggest that these participants are well aware that telops are visual representations of the audio-verbal input. Whether participants feel that the rate at which people speak is also fast-paced is hard to determine based on the data extracts:

Group 1: pre-exchange, P27

The words appear one after the other one following the spoken dialogue

post1 Q20, ll. 1-2

Even though it was perceived by participants that the display rate of telops coheres with the dialogue and the overall pace of the authentic Japanese variety show, JLLs experienced a mismatch between their reading speed and the screen time of telops:

Group 1: pre-exchange, P22

There was alot of text, and there was at times not enough screen time for me to try and read certain kanji that I was unsure of.

post1 Q26, ll. 1-2

Group 1: pre-exchange, P36

I was slow to read the text and often times speed was the greatest hindrance to understanding

post1 Q26, ll. 1-2

This mismatch is coupled with a loss of focus or an inability to finish (re)reading telops before they disappear from the screen. One of these data extracts involves P44 of Group 3: post-exchange:

Group 3: post-exchange, P44

may have had a bit of difficulty with the reading speed during the sponsors and the questions (other parts of the video were not mentioned).

Field notes, ll. 2-4

It is important to note that P44 expresses this difficulty in relation to telops that are not included in the analysed video segments and to telops that facilitate comprehension. Even though this comment comprises a reference to types of telop other than those that directly render dialogue, it is included in the thematic map to illustrate that participants in Group 3:

post-exchange did not report any problems with the display rate of telops that directly render dialogue.

Of those comments that discuss the implications of the telop design to JLA, a vast majority report a positive attitude towards the use of authentic materials as it exposes learners to "real" Japanese language use. Whereas P08 of Group 2: exchange raises the concern that Japanese variety shows may not be suitable to learners at beginner level due to the fast pace, other respondents agree that its use provides learners with the opportunity to gain familiarity with "fast-paced" speech, "dialects" and "slang". Moreover, they think that the video material will help them get "accustomed" to the sound of the language, as attested by the following citations:

Group 1: pre-exchange, P24

They'll help me with my Japanese and will help in understanding people who speak Japanese quickly.

post2 Q23, ll. 2-3

Group 2: exchange, P09

it will make the student be accustomed to the sound of it, and also learn the 'slang' of the language.

post2 Q23, ll. 2-3

Group 3: post-exchange, P43

I think it is a good way to get people used to the kind of fast-paced, more casual style and accented versions of Japanese that you will not learn in a classroom. While at the same time, on screen text helps you to keep up with what's going on if you miss some of what they are saying.

post2 Q23, ll. 1-5

These demonstrate that participants feel that a full understanding of the Japanese variety show does not need to be the only objective of watching authentic AV materials and that authentic materials facilitate learning outcomes that can also be achieved by JLLs with beginner proficiency levels such as familiarity with hearing "natural" Japanese language use. P28 of Group 3: post-exchange argues that Japanese variety shows provide examples of how Japanese speakers actually talk to each other which would help learners acquire more natural language skills in addition to an increased awareness of how natural Japanese language use sounds:

Group 3: post-exchange, P28

Japanese variety shows exposed me to 'real' Japanese unlike what I learned in class and helped me to converse more easily as I got an idea of how Japanese people actually speak to each other.

post2 Q23, ll. 1-3

The subtheme in the second strand of Thematic Network 2 (i.e. Subtheme 2.2: distracting) includes data extracts that relate the ways in which telops appear on screen to the overall attention-seeking nature of the video material. Again, JLLs perceive that such a coherence

exists between these two aspects. Nonetheless, this type of coherence appears to pose challenges to JLLs. The data extracts support this in three ways. First, several data extracts describe the Japanese variety show as "flamboyant" and impactful. Second, it includes a data extract in which P37 of Group 1: pre-exchange states that the telop design is "eye-catching" and "calling for attention":

Group 1: pre-exchange, P37

Very eye-catching, sped-up and calling for attention, much like the show itself.

post1 Q25b, ll. 1-2

Third, a number of responses argue that the way in which telops pop up onto screen is "distracting":

Group 2: exchange, P01

I think the amount of text on screen distracted me from paying attention to the dialogue, as I felt the need to read the text as it appeared.

post1 Q26, ll. 1-3

The eye-catching nature of telops has also been discussed in the previous subsection in relation to the use of typography. Highlighted words are considered to be helpful as it makes key points of the dialogue stand out. However, the eye-catching movements of telops when appearing on screen distract participants. As noted, telops appear on screen in five different ways. Respondents reported that such movements distract them from other types of input such as the people appearing on the show. Some participants also feel that they are obliged to look at the telops as soon as they appear on screen. The experienced difficulties with the distracting nature of telops is most often related to the amount of text and the busyness of the screen:

Group 2: exchange, P18

Sometimes there was too much to read, which lost some of my concentration

post1 Q26, ll. 1-2

Group 3: post-exchange, P43

The only trouble I had, and this might be because I am not a native speaker, but at times there was too much text on screen for me to read and it became confusing.

post1 Q26, ll. 1-3

Group 3: post-exchange, P44

P44 sometimes felt there was too much text and did not always know where to look at.

Field notes, ll. 8-9

The majority of these extracts are from JLLs in Groups 1-2 but when taking a closer look at the contents, it appears that participants in Groups 1-2 and Group 3: post-exchange

describe the experienced difficulties in different ways. Subjects in Groups 1-2 mention distraction in combination with a loss of concentration or a trade-off in attention while respondents in Group 3: post-exchange talk about confusion regarding “where to look at” on the screen. This suggests that learners in Groups 1-2 may be affected more in their understanding of the programme contents by the ways in which telop appear on screen than JLLs in Group 3: post-exchange.

6.3 Viewer experience of video stimulus

6.3.1 Comprehension of analysed video segments

Answers to the comprehension test were reviewed to follow up on these observations. None of the participants answered all questions incorrectly. Respondents with many correct answers are represented in all participant groups. Nonetheless, questionnaires with responses that are partly correct and questionnaires with few correctly answered questions were detected more frequently for JLLs in Groups 1-2. JLLs in Group 3: post-exchange in particular did well on the comprehension test. They responded with longer and more complete answers in the response boxes as compared to JLLs in Groups 1-2. They also left fewer questions blank. In some cases JLLs in Group 3: post-exchange responded with more precise answers to open-ended questions than the Japanese speakers in the Reference group.

Further to the comprehension test, self-reports were also reviewed. These self-reports concern the extent to which participants felt they understood the video stimulus. They also describe how certain subjects were of their answers to the comprehension test. Approximately 48.8% of JLLs felt that they understood 0-40% of the video stimulus directly after watching it. These are all JLLs in Groups 1-2. Participants who thought they understood 60-80% after completing the experimental task are represented in all learner groups and include 44.2% of JLLs. JLLs who felt they understood 80-100% of the video stimulus are all in Groups 2-3 and comprise 4.7% of respondents. A majority of 48.6% of participants in Groups 1-2 did not feel so certain about the answers they had provided. None of the participants in Group 1: pre-exchange indicated to be very certain about their answers to the comprehension test. This is followed with 32.4% of JLLs in Groups 1-2 who responded to be quite certain about their answers. All participants in Group 3: post-exchange and the Reference group were quite to very certain of their answers. It appears that some of these confidence levels had changed after participants filled out the

comprehension test. It was observed that Group 2: exchange comprises the highest number of JLLs who felt they understood less the second time they answered this question. Group 1: pre-exchange, on the other hand, included the highest number of participants who felt they understood more after they filled out the comprehension test.

These findings suggest that JLLs were not overwhelmed by the video stimulus and that they were actively engaged in their search for meaning. However, the answers to the comprehension test and self-reports demonstrate that these aspects vary between participant groups. JLLs in Group 3: post-exchange in particular understand the analysed video segments well and feel confident about their understanding of the video material.

6.3.2 Observations for further use

Table 6.5 gives an overview of participants' attitudes towards the use of Japanese television programmes for Japanese language learning and individual use:

Table 6.5 Perceptions of Japanese television programmes

Aspect	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Japanese television programmes help JLA	Yes	85% yes 15% I do not know	Yes	-
Japanese variety shows in language classes	82.4% yes 17.6% no	80% yes 20% no	Yes	-
Intention of watching Japanese variety shows	58.8% yes 41.2% maybe	50% maybe 30% yes 20% no	66.7% yes 16.7% maybe 16.7% no	80% yes 20% maybe

The responses demonstrate that JLLs share a predominantly positive attitude towards the use of Japanese television programmes for language learning purposes. A vast majority, 93.0%, think that watching Japanese television programmes will help them with their Japanese language studies. All of the participants in Group 1: pre-exchange and Group 3: post-exchange support this claim. None of the JLLs think that the use of such AV materials would be unhelpful. In addition, a large majority of 83.7% of JLLs would like Japanese variety shows to be used in the Japanese language classroom. Only 16.3% of them did not wish to see such authentic materials be used in class. JLLs showed less enthusiasm when asked whether they intended to watch Japanese variety shows more often after participating in this study. Only 46.5% of them intend to watch Japanese variety shows more often while 41.9% is not sure whether they will. Overall, these numbers suggest that participants in Group 3: post-exchange have the most affirmative attitude towards the use of Japanese television programmes.

6.4 Summary and discussion

The results demonstrate that JLLs think that Japanese television programmes can be helpful for learning the Japanese language. The findings also suggest that many JLLs would like to see Japanese variety shows be used in the Japanese language classroom. Participants' perceptions of the scope of input explain why such authentic AV materials are perceived to be useful by each learner group. Table 6.6 provides a joint display of findings from questionnaires and field notes per participant group for the video stimulus. It includes a summary of the results on closed-ended and open-ended data items from the diverging stacked bar charts, tables, and thematic map:

Table 6.6 Joint display of findings on perceptions

Type of input	Type of data	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Telops	Diverging stacked bar chart	4 positive views 5 divided views 1 negative view	7 positive views 2 divided views 1 negative view	7 positive views 2 divided views 1 negative view	7 positive views 3 divided views
Telop design	Table	Majority understood a few lines of text Appropriate for video material	Majority understood a few lines of text Appropriate for video material	Majority understood most of the text Appropriate for video material	- Appropriate for video material
	Thematic map*	33.3% fast-paced 30.8% mood	27.1% fast-paced	41.7% main points of dialogue	No prevalent subthemes
Other	Table	3 types of input are helpful	3 types of input are helpful	3 types of input are helpful	1 type of input is helpful**

** Telop design that is specific to telops that directly render dialogue.*
*** Result based on responses to one closed-ended data item instead of two as opposed to the other three participant groups.*

As shown in Table 6.6, JLLs in Groups 2-3 and Japanese speakers in the Reference group have a predominantly positive attitude towards telops. Their views are divided on only two to three statements respectively. Such division was observed for statements on visual clutter, humour, necessity, interpretation of actions and utterances of people appearing on the show, and saliency; albeit in different combinations for each of these participant groups. JLLs in Group 1: pre-exchange, on the other hand, feel positive regarding four out of ten statements on telops. Subjects in this particular participant group are divided on five statements that deal with distraction, visual clutter, humour, focus of attention, and necessity. This means that more distraction and loss of attention was experienced as a result of telops by participants in Group 1: pre-exchange as opposed to the other three participant groups. Nonetheless, all four participant groups think that telops are not confusing, and help them understand the video material and identify important information.

It was discussed in *Subsection 2.2.3 Japanese speakers' perceptions of telops* that Japanese users' attitudes towards telops tend to differ according to generation, and that younger users often feel that telops are necessary and helpful to their understanding when watching television as opposed to adults (Shitara, 2008; Sakamoto, 2009b; Shitara, 2009). As noted, the observations from diverging stacked bar charts show that participants in this study also feel that telops are helpful to their understanding of the video material. However, the results also suggest that native Japanese speakers feel divided on the necessity of telops. Furthermore, Group 3: post-exchange was the only learner group in which a vast majority of JLLs felt that telops are indispensable. Responses were more evenly distributed between negative, neutral and positive attitudes for JLLs in Groups 1-2. Even though participants were not asked to explain their views on this statement it can be argued that the results relate to the concept of semiotic redundancy in the case of Japanese speakers considering that they can understand the programme contents without difficulty. For learners, it could be that JLLs in Groups 1-2 perceived telops to be less useful as they experienced more trouble with understanding the linguistic contents as compared to JLLs in Group 3: post-exchange.

It has also been maintained in the established literature that telops are used to frame humour in the programme contents of a Japanese variety show (O'Hagan, 2010). The observations from diverging stacked bar charts illustrate that not all four participant groups in this study think that telops make the video material funnier. Opinions were divided on this statement for Group 1: pre exchange and the Reference group. While it is difficult to ascertain whether these results are related to JLLs' lack of understanding of the jokes, these findings demonstrate that a production team's intentions for the use of telops are not necessarily reflected in users' perceptions.

The findings from diverging stacked bar charts are closely related to the results from the thematic map. It was observed that participants perceive the design of telops that directly render dialogue to be suitable to the video stimulus and that this perceived suitability affects their viewer experience in mostly two ways. First, JLLs argue that the variation in typographical features helps them with the identification of certain aspects of the video material. According to them, typography reflects the mood of the Japanese variety show and the tone of voice of utterances, and helps them differentiate speakers appearing on the show and main points of the dialogue. Second, the display rate and movement of telops are described as fast-paced and distracting which poses challenges to JLLs. Table 6.6 includes

the names of subthemes that are most prevalent for each participant group based on the rates of occurrence and contents of data extracts. The subthemes that are not mentioned (i.e. Subtheme 1.1.2: tone of voice, Subtheme 1.2.1: speakers, and Subtheme 2.2: distracting) show the highest rates of occurrence for Group 2: exchange. As is shown in Table 6.6, challenges from the second thematic network are discussed more often by JLLs in Groups 1-2. Nonetheless, when taking a closer look at their feedback, it appears that they associate such challenges with potential positive learning outcomes.

The perceptions of JLLs in this study demonstrate that telops support learners' understanding of the programme contents in a way that would normally be associated with SDH. Contrary to subtitles that are created for a hearing audience, SDH makes use of colours and written commentary to clarify speakers, phonological information and audio-nonverbal input (De Linde and Kay, 1999, pp. 12-15; Neves, 2008, p. 178; Díaz Cintas, 2013, p. 280; O'Hagan, 2013). The results from the thematic map show that the telop design provides language learners with similar clues. Telops are perceived to be indicative of speakers, and to be signalling aspects of the dialogue such as tone of voice, phrasing and emphasis. These observations suggest that telops are able to visually represent elements of the dialogue that are narrative-informative, interactional, and structuring (Remael, 2004, p. 115; Díaz Cintas and Remael, 2014, p. 49). That is to say, they include references to the narrative through linguistic contents, reflect interactions between people appearing on the show via typographical features, and structure the dialogue by means of typography. The latter was evidenced by data extracts in Subtheme 1.2.2: main points of dialogue. These data extracts not only showed that colours and fonts differentiated information for participants; P43 of Group 3: post-exchange indicated that such features also help with speech segmentation.

The thematic map demonstrates trends in JLLs' perceptions that were not expected based on the established literature and the multimodal analysis. It was argued that the design of telop in Japanese variety shows is largely based on the intuitions of the production team (Sasamoto and Doherty, 2016; Sasamoto, O'Hagan and Doherty, 2016, p. 2). It was also determined during the exploration of the video material that the video stimulus does not include a consistent telop design. Nonetheless, the observed trends appear to suggest that JLLs are able to link the telop design, or typography in particular, to other types of input in the video stimulus in a way that makes sense to them. It appeared that this was in large part related to the coherence that participants felt existed between different types of input.

Table 6.6 also includes results from tables on types of input that helped participants with telops they did not understand. JLLs indicated that such other types of input include imagery, kinesic action, and dialogue. Japanese speakers in this study find drawings helpful. This suggests that, further to typography, other non-verbal types of input were experienced as helpful to understanding the programme contents.

Although the data sets suggest that the video stimulus was not overwhelming to JLLs in Groups 1-2, especially since the general feel of the video material appears to be accessible to them through non-verbal input, findings from tables and the comprehension test suggest that a full understanding was mostly achieved by JLLs in Group 3: post-exchange. Results demonstrate that these participants understood the video stimulus well and felt more confident about their understanding of the programme contents as opposed to the other two learner groups.

6.5 Conclusion

This chapter demonstrated that the JLLs had a predominantly positive attitude towards the use of Japanese variety shows in the language classroom. This was illustrated in both the closed-ended data items on the overall video stimulus and open-ended data items included in the thematic map. JLLs in each of the three participant groups generally felt that they were able to pick up on aspects of the video material and that positive outcomes were to be expected from repeated exposure to such authentic AV materials.

It appeared that telops or telop design in particular played an important role in JLLs' viewer experience of the video material. Participants felt they were able to understand affective aspects and to distinguish between certain types of information in the Japanese variety show through typographical features such as colour and font. It was argued in this chapter that such visual-nonverbal input is one of the most accessible types of input to JLLs at any stage of their Japanese language study and that it can be tapped into regardless of a proper understanding of the contents of a Japanese variety show. It was also discussed that, although JLLs' perceptions may have diverged slightly from the findings of the multimodal analysis, the perceived interrelationship of typography with other types of input could potentially have lowered JLLs' feeling of being overwhelmed. Two other aspects of the telop design on the other hand, display rate and movement, appeared to pose challenges to JLLs in Groups 1-2 in particular. Even though telops provide clues to the

affective context and mood of the Japanese variety show, they were also experienced as fast-paced and distracting.

This chapter also discussed the implications of these empirical findings in relation to the established literature. These related to the parallels that were found between SDH and telops that directly render dialogue in particular. The similarities not only appeared to exist in the design of such types of intralingual text; they also applied to the roles that these texts play in users' understanding of the contents of a television programme.

The following chapter interrogates the findings on JLL's multimodal perception and takes a closer look at the eye-tracking data. It describes the data on JLLs' visual attention and discusses the implications of the results for this study's research questions and the established literature.

7. JAPANESE LANGUAGE LEARNERS' GAZE BEHAVIOUR WHILE WATCHING A JAPANESE VARIETY SHOW

Chapter 7 is the second empirical chapter of this thesis. It provides evidence on JLLs' visual attention while watching the video material. The purpose of this chapter is to demonstrate JLLs' viewer reception of the multimodal landscape of the analysed video segments and to illustrate which types of input gathered the most fixation counts and were fixated on the longest by participants. This chapter also aims to further examine JLLs' perceptions based on the results from the collated gaze data.

Section 7.1 starts with a description of findings on visual attention in a spatial sense. This entails results on the areas of the screen in which JLLs' visual attention was clustered. This section presents the observations from heat maps for Segments 2-4 of the video stimulus and each of the analysed video segments separately. Section 7.2 further elaborates on these results and explains JLLs' gaze behaviour in the gaze replay videos for the types of input encountered in those areas of the screen on which JLLs' visual attention was concentrated. Section 7.3 presents the fixation-data. Section 7.4 then gives a summary of the eye-tracking findings and discusses the implications that these results carry for this study's research questions and the established literature.

7.1 Viewer reception of spatial dimension

7.1.1 Observations for analysed video segments

Heat maps that summarise participants' gaze behaviour use colour gradients to display the concentration of visual attention on screen. Such colour gradients range from green to red. Green represents the lower end of the scale, yellow and orange signifies a medium range of concentration, and red is used for those areas of the screen with the highest intensity of visual attention. Areas of the screen that are not overlaid with a colour have not been visually attended to by participants. It was discussed in *Subsection 4.3.2 Procedure for heat maps* that data on first-year and second-year JLLs could not be integrated into one heat map due to the way in which recordings have been stored in the *Tobii Studio* project. Results for Group 1: pre-exchange have therefore been represented by two heat maps; a heat map for first-years on the left and a heat map for second-year JLLs on the right.

Figure 7.1 shows fixation count heat maps per participant group that were generated on all analysed video segments combined.

Honmadekka!?TV (2013)

Figure 7.1 Fixation count aggregated onto screen

As is shown in Figure 7.1, visual attention is primarily concentrated in two horizontal green-yellow lines. A red hotspot is situated at the centre of the top horizontal line in each of these heat maps. When dividing the screen horizontally into three regions, the top green-yellow line is positioned in-between the top and middle regions. The second line covers the bottom region. If the screen were to be divided into two vertically, the red hotspot would be situated on the dividing line. It appears that the top green-yellow line is less distinct than the bottom line. The colour intensity of the bottom green-yellow line in particular varies between participant groups.

The separate overview of still images that are overlaid with grids was checked in order to contextualise these hotspots. Figure 7.2 includes two examples of still images from the overview. The hotspots were recreated on these still images to work out how the fixation count heat maps can be put into context. It is important to note that different types of input pass through the areas of the screen that have been identified as hotspots of visual attention due to the dynamic nature of the video stimulus. However, it appears that two types of input in particular are featured within the hotspots. These are faces of people appearing on the show, and telops that directly render dialogue. This is displayed in Figure 7.2:

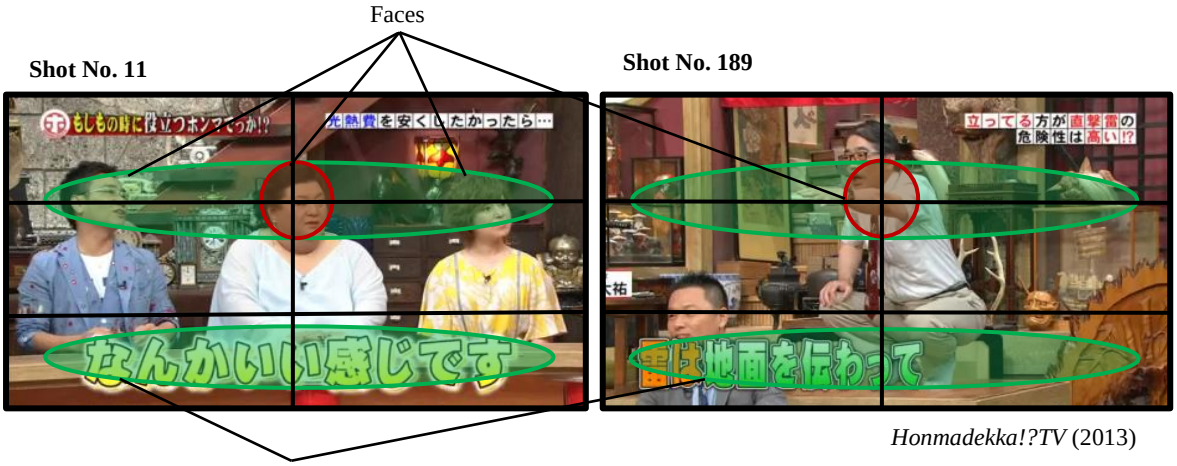


Figure 7.2 Contextualisation of fixation count aggregated onto screen

It was observed that faces appear most often in the area of the screen that was marked by the top green-yellow line. Faces of people appearing on the show were featured there in 85.4% of shots in the multimodal transcript. Of these shots, 68.6% had faces visible in the area that was overlaid with the red hotspot in the fixation count heat maps of Figure 7.1. Both varieties of telops that directly render dialogue (i.e. utterance of speaker, and narration) were positioned on the top line during 6.9% of shots. When taking a closer look at the ratios for the bottom green-yellow line an opposite trend was detected. Utterances of speakers appeared most frequently in that area of the screen and were observed for 70.9% of shots in the multimodal transcript. Faces of people appearing on the show were shown there for 4.6% of shots. It is important to note that 18.4% of shots did not feature any of the varieties that were included in the scope of input in the bottom region of the screen. This demonstrates how prevalent telops that directly render dialogue are for this region of the screen.

Heat maps on absolute gaze duration were also examined to further explore the concentration of visual attention on screen. Figure 7.3 shows absolute gaze duration heat maps per participant group that were generated on all analysed video segments combined:

Honmadekka!?TV (2013)

Figure 7.3 Absolute gaze duration aggregated onto screen

The hotspots are distributed in a similar fashion to those featured in the fixation count heat maps shown in Figure 7.1. Again, the colour red is superimposed on areas of the screen where faces appear. The green-yellow lines appear on approximately the same regions as the fixation count heat maps and overlap with areas of the screen where faces and telops that directly render dialogue are shown. This indicates that participants not only fixated most often on these particular types of input; they also spent most time processing faces of people appearing on the show and telops that directly render dialogue.

7.1.2 Observations per analysed video segment

Further to the heat maps on all analysed video segments combined, fixation count and absolute gaze duration heat maps were generated per analysed video segment to check whether different multimodal landscapes affect the concentration of visual attention on the screen. It appeared that the same observations for all analysed video segments combined were made for each analysed video segment. This is shown in *Appendix J. Heat maps per analysed video segment*. However, there are several observations that are specific to particular multimodal landscapes in the video stimulus. Figure 7.4 demonstrates one of these observations and takes a closer look at Segment 4. It includes the fixation count heat map of Group 2: exchange and a still image that is overlaid with a grid from the separate overview of still images. Hotspots have been recreated in the still image to facilitate a better contextualisation of the hotspots.

Figure 7.4 Contextualisation of fixation count aggregated onto screen for Segment 4

It was observed that fixation count and absolute gaze duration heat maps for Segment 4 show an extra green-yellow hotspot for the area of the screen where Telop No. 62 appears. This happens during Shot No. 245-249 of the multimodal transcript. This telop is shown on the right in Figure 7.4. The still image illustrates that this particular telop extends into the middle region of the screen leaving a distinct mark on the heat map. Further to this telop, this particular green-yellow hotspot overlaps with faces for 18.8% of shots that are included in Segment 4. There were no other types of input observed in this particular area of the screen in the remaining shots of this video segment. This finding strengthens the assumptions made earlier that faces and telops that directly render dialogue are main points of visual attention for participants in this study.

Another observation was made for the fixation count and absolute gaze duration heat maps that were generated on Segment 3. Of these, the absolute gaze duration heat maps have been included in *Appendix J. Heat maps per analysed video segment*. The fixation count heat maps are included in Figure 7.5.

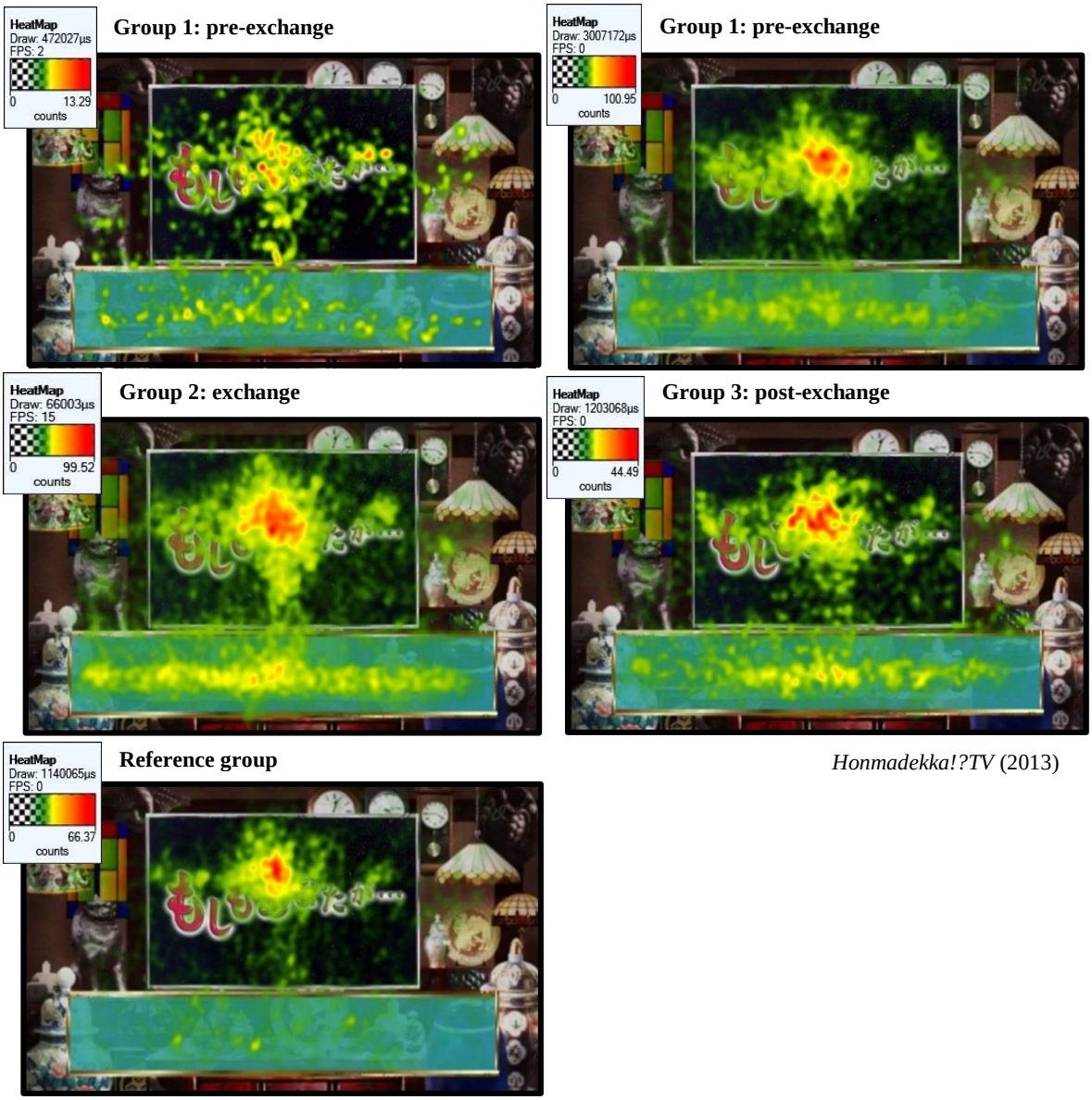


Figure 7.5 Fixation count aggregated onto screen for Segment 3

Similar to the trend observed in Figure 7.4, there appears to be an additional green-yellow line in the same area of the screen for each of these heat maps. This time it runs down the centre connecting the top and bottom green-yellow lines. The separate overview of still images was once again consulted to contextualise this observation. Example still images overlaid with grids and reconstructed hotspots are included in Figure 7.6.

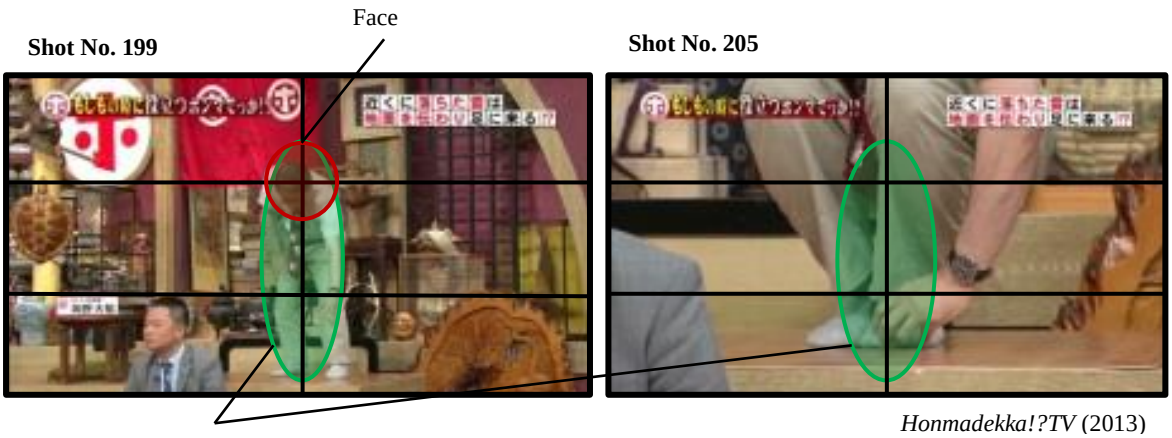


Figure 7.6 Contextualisation of fixation count aggregated onto screen for Segment 3

It appears that the hotspot overlaps with kinesic action. More specifically, the hotspot is overlaid on the area of the screen where gesticulation and bodily configuration are shown for the enactment of hypothetical situations for dealing with lightning. As noted previously, bodily configuration is used to a great extent within Segment 3 as opposed to the other two panel discussions. Of those shots allocated to this particular video segment, approximately 42.5% displays a combination of bodily configuration and gesticulation in the centre of the screen in accordance with the vertical green-yellow line that was observed in the heat maps of Figure 7.5. Movements include standing, crouching, arm movements to indicate lying on your belly, pointing and gestures to show the direction and movement of electric current. The latter is shown in Figure 7.6. This suggests that kinesic action, a type of input that did not appear as a hotspot in the heat maps for all analysed video segments combined, gathers much visual attention in the video segment in which kinesic action is used extensively.

This demonstrates that gesticulation and bodily configuration comprise two more varieties of input that gather visual attention from participants although to a lesser extent as compared to faces of people appearing on the show and telops that directly render dialogue. The heat maps per analysed video segment also suggest that the multimodal landscape of analysed video segments affects the types of input that participants visually attend to.

7.2 Viewer reception of temporal dimension

7.2.1 Meaningful study of gaze paths

It was noted previously that further exploration of the gaze data is needed to cross-check the findings from heat maps. It was argued that heat maps alone cannot provide an accurate

account of the concentration of visual attention considering that they visualise gaze data on static images even though the eye-movements have been recorded on a dynamic video stimulus. It was therefore discerned that the gaze videos needed to be replayed in another round of data exploration to rule out the possibility that types of input other than those identified through the heat maps and multimodal transcript have been fixated on by participants. As noted, the exploration phase of the multimodal transcript resulted in the selection of seven types of input and 23 varieties of input across four forms of expression (*Appendix K. Scope of input for analysed video segments*). However, not all of these can be focused on simultaneously during the replay of gaze videos. Target varieties were therefore chosen prior to the second round of gaze data exploration. These were selected from the scope of visual forms of expression considering that gaze paths are overlaid on visual input.

Seventeen varieties in the scope of input are communicated through visual forms of expression. It was determined during the multimodal analysis that four of these occurred most frequently and consistently throughout the video material. These are utterances of speakers, themes, titles, and gesticulation. Furthermore, utterances of speakers and narrations were identified as the most salient visual-verbal types of input. The heat maps showed that only five varieties of visual input gather visual attention from participants: faces, utterances of speakers, narrations, gesticulation, and bodily configuration. This suggests that faces are fixated on even when people appearing on the show do not make explicit or exaggerated facial expressions. It appeared that other prevalent varieties of input such as themes and titles do not gather visual attention from participants as the top left and right corners of the static images were not overlaid with colour gradients. In addition, the hotspots do not follow the screen lay-out of shots that feature telops that bridge scene change. This suggests that those visual-verbal types of input that were identified as few in number and static were not visually attended to by participants. However, considering that nameplates generally appear on screen in the region where the bottom green-yellow line is shown in the heat maps it cannot be ruled out that they are not fixated on by participants until a closer look is taken at the gaze replay videos.

The exploration of gaze replay videos is focused on the identification of gaze behaviour for those varieties of input that were identified as main focal points. For the purposes of this study, faces of people appearing on the show, telops that directly render dialogue, and kinesic action were selected as these main focal points. In addition, the possibility of digressions in gaze behaviour to other visual-verbal types of input is kept in mind during

this exploration. This not only helps ascertain whether the main focal points are the only varieties of input that participants fixate on; it also helps determine how participants visually attend to such types of input and whether the presence of aural input can be inferred from participants' gaze behaviour based on observations for telops that directly render dialogue.

7.2.2 Observations for main focal points

Gaze plots that display participants' gaze paths use different sizes of dots, numbers and trails to illustrate the duration and order of fixations. It was discussed in *Subsection 4.3.3 Procedure for gaze plots* that gaze plots included in this thesis have been enhanced with white circles and arrows to make such gaze point sequences more distinct. This thesis recognises that tendencies in gaze behaviour vary within and between recordings. Gaze plots in this section therefore represent eye-movements that the researcher considered to be recurrent within the data set regardless of the participant group or analysed video segment. Figure 7.7 demonstrates four accumulated gaze plots that showcase participants' gaze behaviour during instances when faces and telops that directly render dialogue are visible on screen. These examples include faces of people appearing on the show and faces in drawings.

Honmadekka!?TV (2013)

Figure 7.7 Gaze behaviour for faces and telops that directly render dialogue

The gaze paths of P42, P43 and JP06 start with a fixation that lands on one of the faces that are visible on screen. The first fixation of the gaze path of P41 is on the telop that directly renders dialogue. The white arrows in the accumulated gaze plots illustrate that gaze paths of JLLs and Japanese speakers go back and forth between faces and telops that directly render dialogue. As shown in Figure 7.7, this was observed for both utterances of speakers and narration telops. In addition, such continuous jumps in the gaze paths were observed for all kinds of faces that make their appearance in the analysed video segments. These include faces of speakers, faces of other people appearing on the show, and those faces that are depicted in drawings. This means that visual attention directed to faces is not exclusive to those of speakers that correspond to the telops that directly render dialogue. None of the other types of input that are visible on the screen (i.e. set, and telops that facilitate comprehension) are fixated on in these examples.

Another observation was made in relation to narration telops. These are the telops that introduce the overarching questions at the start of each panel discussion as shown in the accumulated gaze plots of P43 and JP06. It appeared that participants' gaze paths follow

the appearance of text units in accordance with the progression of the corresponding narrations. This means that participants matched their reading speed with the speed at which the narrator utters the narrations. This type of gaze behaviour has not been observed for any of the other visual-verbal types of input. Figure 7.8 displays four more accumulated gaze plots in which participants' gaze paths tend to focus on faces of people appearing on the show when telops that directly render dialogue are not featured on screen.

Honmadekka!?TV (2013)

Figure 7.8 Gaze behaviour for faces and kinesic action

The first fixation for each of these depicted gaze paths is on the set. The white arrows illustrate that participants' gaze paths thereafter go directly to the face that is visible on screen or starts a search for faces in long shots. Moreover, the gaze path of P14 goes straight for the speaker and his kinesic action after the shot change. This suggests that faces are visually attended to even in shots that feature faces in areas of the screen other than where the red hotspot had appeared in the fixation count and absolute gaze duration heat maps. Again, none of the other types of input are fixated on in these examples.

It is important to note that the gaze path of P29 demonstrates the same kind of gaze behaviour as was observed in Figure 7.7. Although not visible in the accumulated gaze plot

included in Figure 7.8, directly after the depicted still image a telop that directly renders dialogue pops up onto screen. As soon as it appears on screen the gaze path of P29 initiates the continuous shift between the face of the expert guest and the telop. This observation strengthens the assumption made earlier that this back-and-forth movement of gaze paths is less established with visual-verbal types of input other than those that directly render dialogue.

7.2.3 Observations for other types of visual input

Nonetheless, this does not mean that gaze paths are never directed towards other types of visual input on the screen. Figure 7.9 demonstrates four accumulated gaze plots in which participants' gaze behaviour is shown for types of input other than those that comprise the main focal points.

Honmadekka!?TV (2013)

Figure 7.9 Gaze behaviour for telops that facilitate comprehension and nameplates

The gaze paths of P38 and P09 start with the fixations visible on the title and theme. The first fixations for the gaze paths of P36 and P03, on the other hand, are on the set. The white arrows show that their gaze paths move between faces and telops that directly render dialogue while other types of telop and nameplates are also visually attended to. The gaze

paths of P38 and P09 illustrate that participants intermittently glance at the title, themes and speaker identifications. Such gaze behaviour has also been observed for nameplates and katakana characters displayed in the set. The theme in the accumulated gaze plot of P09 is accompanied with a sound effect; this may have instigated P09 to look at this particular telop that facilitates comprehension. The gaze plots of P36 and P03, on the other hand, demonstrate that participants at times fixate more and longer on such visual-verbal types of input which may suggest a more conscious attempt at understanding the contents. It is important to note that the gesticulation visible in the accumulated gaze plot of P36 was not visually attended to at all. This was not unusual gaze behaviour.

This suggests that gesticulation alone does not gather much visual attention from participants as opposed to the combination of gesticulation and bodily configuration as observed in Figure 7.8. In addition, it appears that faces of people appearing on the show and telops that directly render dialogue are indeed visually attended to the most by all participants. However, some visual attention was paid to types of input other than the identified main focal points.

7.3 Distribution of visual attention

7.3.1 Fixation count for analysed video segments

Figure 7.10 shows the total number of fixations per participant group on faces, telops that directly render dialogue, and other types of input for all analysed video segments combined.

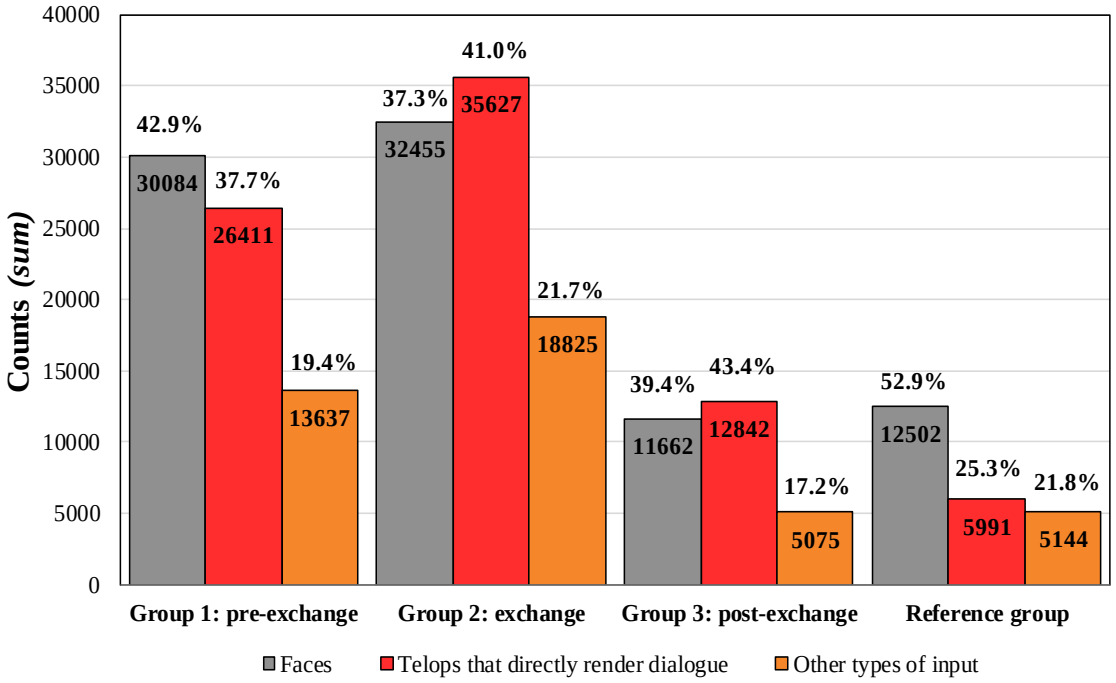


Figure 7.10 Fixation count on types of input

The counts and ratios shown in Figure 7.10 illustrate that faces and telops that directly render dialogue are indeed visually attended to the most with a combined proportion of approximately 80% for each participant group. The ratios are in respective group order 80.6%, 78.3%, 82.8%, and 78.2%. This means that for each participant group approximately 20% of fixations accumulated over the duration of the analysed video segments landed on text in background (i.e. nameplates), telops that facilitate comprehension (i.e. themes, titles, and speaker identifications), telops that bridge scene change (i.e. screens filled with text), imagery (i.e. drawings), and kinesic action (i.e. gesticulation, and bodily configuration).

When taking a closer look at the proportions of fixation counts directed at faces and telops that directly render dialogue it appears that the ratios are rather similar for the three learner groups. Overall, the proportions of fixation counts that landed on telops that directly render dialogue are approximately 40% for each of these participant groups. It appears that the proportions of visual attention directed at faces is also approximately 40% for each learner group. Nonetheless, it is suggested in Figure 7.10 that participants in Group 1: pre-exchange fixate more often on faces as opposed to subjects in Groups 2-3. The latter show a tendency to look more at telops that directly render dialogue. Furthermore, JLLs in Group 3: post-exchange directed their visual attention most often at telops that directly render dialogue as opposed to the other two learner groups.

The Japanese speakers in the Reference group have a stronger tendency to look at faces. For this particular participant group, approximately twice as many fixations were counted in the area of the screen where faces appear than the region of the screen where telops that directly render dialogue are showed. Figure 7.10 also demonstrates that the proportion for telops that directly render dialogue is similar to the ratio for other types of input. This suggests that, whereas the visual attention of JLLs was more evenly distributed between faces and telops that directly render dialogue, such an even distribution was detected between telops that directly render dialogue and other types of input for the Reference group. However, for JLLs this even distribution is between the types of input that they visually attended to the most. For participants in the Reference group this even distribution was between the types of input that they fixated on the least.

The proportions in Figure 7.10 thus suggest an upward trend in participants' visual attention on telops that directly render dialogue from Group 1: pre-exchange to Group 3: post-exchange. This could mean that the more linguistic skills a JLL acquires in Japanese the more they rely on such input. Nonetheless, the participant group with the highest proficiency level in this study, i.e. Japanese speakers, relied much less on telops as compared to the JLLs in Group 3: post-exchange. Moreover, they exhibited viewing behaviour that is similar to that of participants in Group 1: pre-exchange as both participant groups fixated on faces of people appearing on the show the most. This thesis proposes several reasons for their focus on faces. It could be that JLLs in Group 1: pre-exchange found telops too difficult and visually searched for other clues in the video material to support their understanding of the programme contents, which translated into fewer attempts at reading telops. It can be argued that participants in the Reference group, on the other hand, did not need telops for their understanding of the analysed video segments due to the semiotically redundant nature of such intralingual text. Furthermore, participants in the Reference group may have needed fewer fixations on telops to grasp their linguistic contents considering they are native speakers of the Japanese language.

Figure 7.11 gives an overview of the total number of fixations per participant group that were counted for each type of resemblance for all analysed video segments combined.

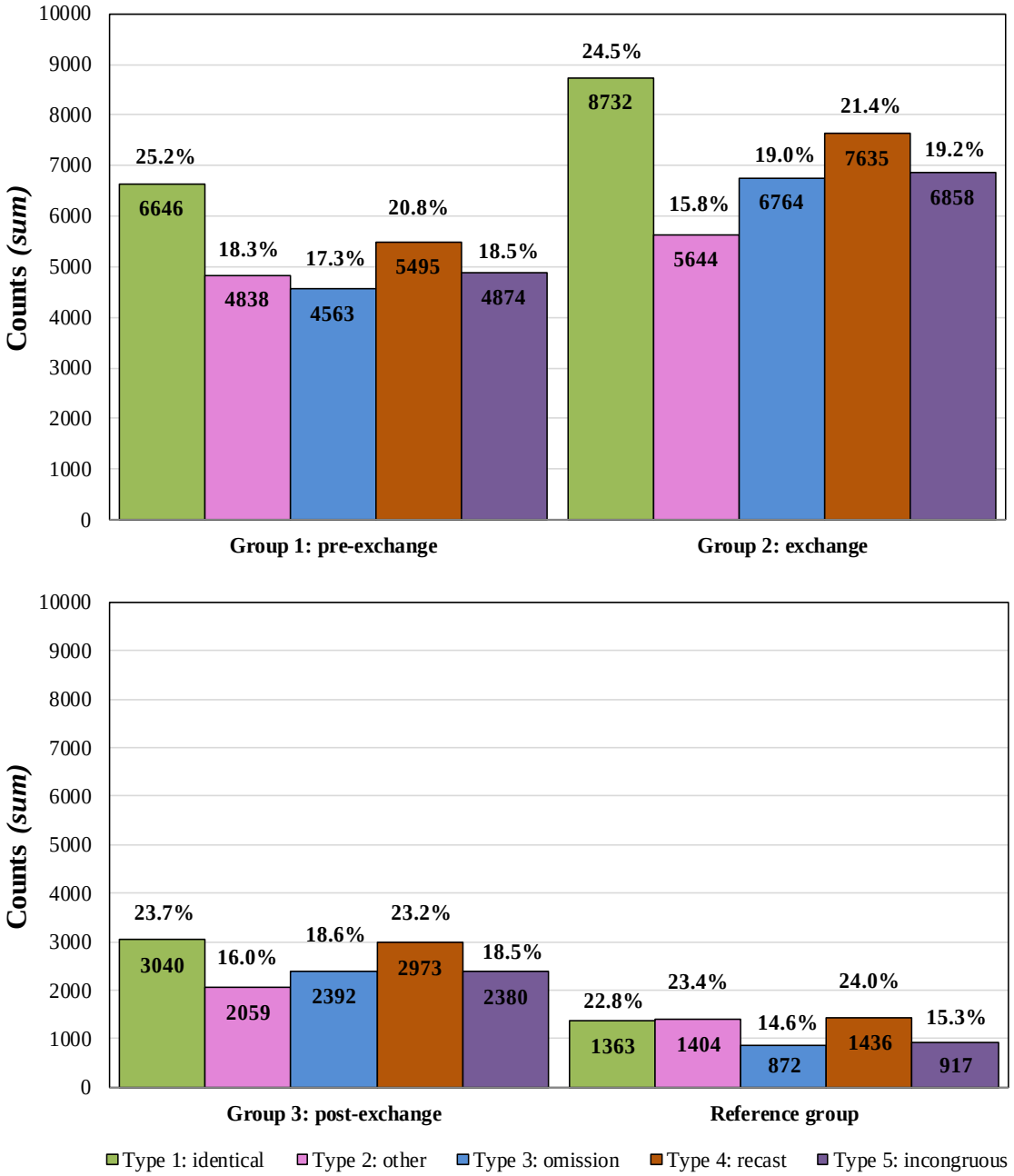


Figure 7.11 Fixation count on types of resemblance

As shown in Figure 7.11, less distinct trends can be detected when sorting the data on telops that directly render dialogue according to types of resemblance. It appears that subjects in all four participant groups fixated most on telops in Type 1: identical, and Type 4: recast. Only the Reference group showed a tendency to focus more on telops in Type 2: other instead of telops in Type 1: identical. As noted in *Table 5.1 Scheme for types of resemblance*, these types are far removed from each other in their degree of resemblance to the corresponding dialogue. Participants in Group 1: pre-exchange and the Reference group fixated the least on telops in Type 3: omission. The type of resemblance visually

attended to the least by JLLs in Groups 2-3 is Type 2: other. Overall, it appears that the ratios are relatively evenly distributed between types of resemblance.

When adding the proportions of telops that show high levels of resemblance to the dialogue (Types 1-2) and those telops that showcase lower levels of resemblance to the utterances of speakers and narrations (Types 3-5) it appears that more fixations are counted for the latter. The proportions for these three types combined in respective group order are 56.6%, 59.6%, 60.3%, and 53.9%. This illustrates that telops with lower degrees of resemblance to the dialogue not only occur more frequently in the video material; they also gather more visual attention from all participant groups.

7.3.2 Total fixation duration for analysed video segments

Figure 7.12 demonstrates the total duration of fixations in seconds per participant group on faces, telops that directly render dialogue, and other types of input for all analysed video segments combined.

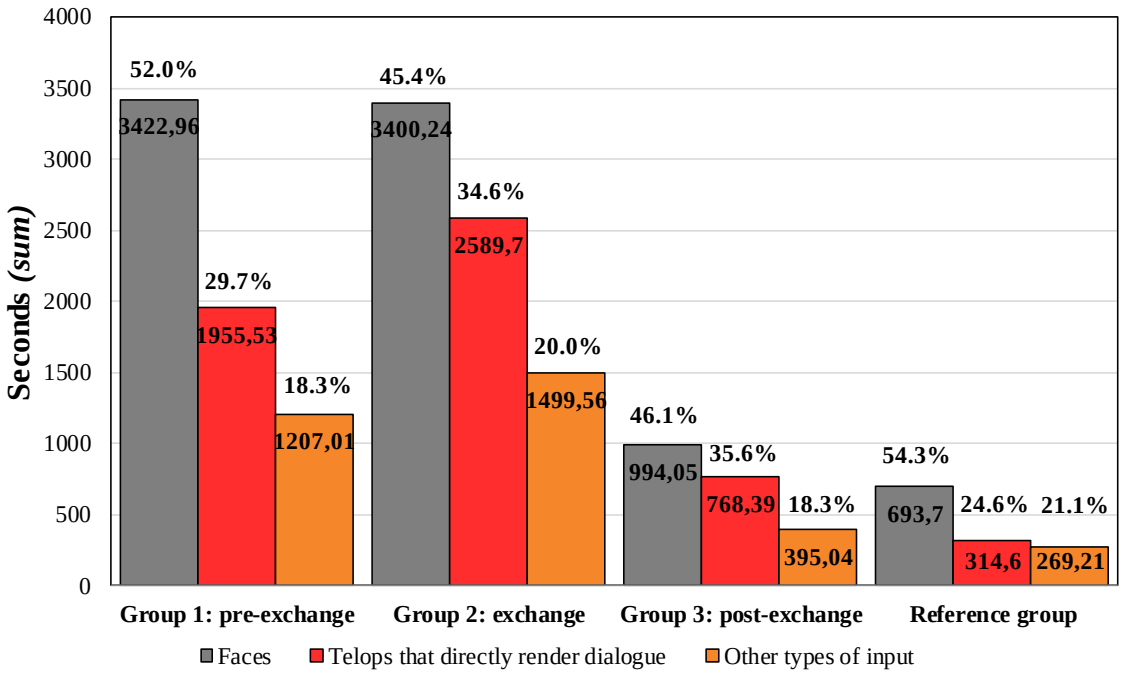


Figure 7.12 Total fixation duration on types of input

The counts and ratios illustrate that faces of people appearing on the show and telops that directly render dialogue are fixated on the longest. Again, the proportions for faces and telops that directly render dialogue combined amount to approximately 80% for each participant group. The ratios are in respective group order 81.7%, 80%, 81.7%, and 78.9%.

This not only means that for each participant group approximately 20% of fixation counts is directed at other types of input; approximately 20% of the total viewing time for the analysed video segments was spent looking at text in background (i.e. nameplates), telops that facilitate comprehension (i.e. themes, titles, and speaker identifications), telops that bridge scene change (i.e. screens filled with text), imagery (i.e. drawings), and kinesic action (i.e. gesticulation, and bodily configuration).

The ratios in Figure 7.12 show an upward trend in the time that subjects spend fixating on telops that directly render dialogue from Group 1: pre-exchange to Group 3: post-exchange, which strengthens the assumption that JLLs make more attempts at understanding their linguistic contents when they have a better command of the Japanese language. As opposed to the findings on fixation count in Figure 7.10, it appears that JLLs in the three learner groups fixate the longest on faces of people appearing on the show with approximately 50% of the overall time spent looking at the analysed video segments allocated to this particular type of input. In addition, it appears that visual attention is less evenly distributed between faces and telops as compared to Figure 7.10. The gap between the proportions for these two types of input narrows when moving from Group 1: pre-exchange to Group 3: post-exchange with the latter showing a more even distribution between faces and telops that directly render dialogue. JLLs in Group 1: pre-exchange show the strongest tendency to focus on faces. It is important to note that faces, categorised as a visual-nonverbal form of expression in the scope of input, can be considered as the source of audio-verbal stimuli in the video material. Further to telops, it could be that visual information from faces supports JLLs' understanding of the dialogue.

Japanese speakers also fixated longest on faces of people appearing on the show. Figure 7.12 shows that the time spent looking at faces is approximately twice as long as the viewing time for telops that directly render dialogue. As noted, native Japanese speakers can grasp the information from telops through audio-verbal stimuli. This could explain why they fixate longer on nonverbal input such as faces, and to a lesser extent other types of input, as compared to learners. While it was observed that the proportions for faces of people appearing on the show and telops that directly render dialogue are less evenly distributed for JLLs in the learner groups, it appears that the proportions for telops that directly render dialogue and other types of input are more evenly distributed for Japanese speakers in the Reference group. Again, this even distribution is between the types of input

that Japanese speakers spent little time fixating on. This coincides with the observations on the distribution of fixation counts.

Figure 7.13 gives an overview of the total duration of fixations in seconds per participant group on each type of resemblance for all analysed video segments combined.

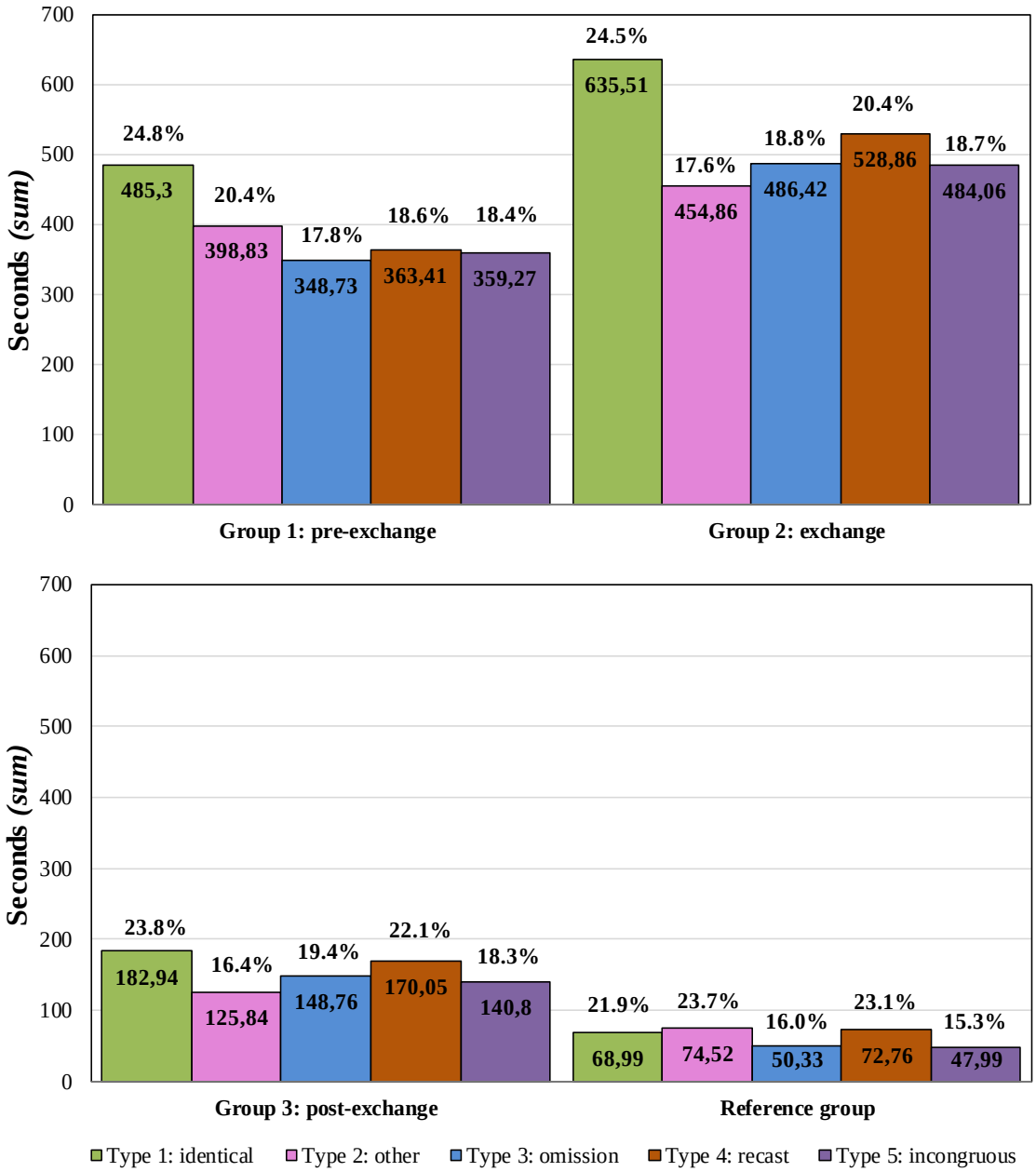


Figure 7.13 Total fixation duration on types of resemblance

Again, less distinct trends can be detected when sorting the data on telops that directly render dialogue according to types of resemblance. It appears that proportions of time spent looking at telops that directly render dialogue are evenly distributed between the five types of resemblance. The two types of resemblance that are fixated on the longest differ

per participant group. Group 1: pre-exchange shows the longest fixation durations for telops in Types 1-2, Groups 2-3 have the highest proportions for telops in Type 1: identical and Type 4: recast, and the Reference group has the longest fixation durations for telops in Type 2: other and Type 4: recast. It was observed that variation also exists between participant groups for the types of resemblance that show the shortest fixation durations. This is Type 3: omission for Group 1: pre-exchange, Type 2: other for Groups 2-3, and Type 5: incongruous for the Reference group.

It appears that telops in types of resemblance that are less similar to the dialogue (Types 3-5) are fixated on longer as opposed to telops that have higher degrees of resemblance to the corresponding audio-verbal input (Types 1-2). Proportions of the total viewing time spent looking at telops with lower degrees of resemblance are in respective group order 54.8%, 57.9%, 59.8%, and 54.4%. These proportions demonstrate that telops in Types 3-5 not only gather the most visual attention when it comes to fixation counts; it is also evidenced in the total fixation durations.

7.4 Summary and discussion

Table 7.1 includes a joint display of gaze data findings per participant group for all analysed video segments combined. It includes a summary of the results from heat maps, gaze plots, and bar charts on the scope of visual types of input.

Table 7.1 Joint display of findings on gaze behaviour

Type of input	Type of data	Group 1: pre-exchange	Group 2: exchange	Group 3: post-exchange	Reference group
Faces	Heat map	Hotspot	Hotspot	Hotspot	Hotspot
	Gaze plot	Main focal point	Main focal point	Main focal point	Main focal point
	Bar chart	42.9% counts 52.0% duration	37.3% counts 45.4% duration	39.4% counts 46.1% duration	52.9% counts 54.3% duration
Telops that directly render dialogue	Heat map	Hotspot	Hotspot	Hotspot	Hotspot
	Gaze plot	Main focal point	Main focal point	Main focal point	Main focal point
	Bar chart	37.7% counts 29.7% duration Majority to lower degrees of resemblance	41.0% counts 34.6% duration Majority to lower degrees of resemblance	43.4% counts 35.6% duration Majority to lower degrees of resemblance	25.3% counts 24.6% duration Majority to lower degrees of resemblance
Other	Heat map	1 hotspot*	1 hotspot*	1 hotspot*	1 hotspot*
	Gaze plot	1 focal point	1 focal point	1 focal point	1 focal point
	Bar chart	19.4% counts 18.3% duration	21.7% counts 20.0% duration	17.2% counts 18.3% duration	21.8% counts 21.1% duration

** Only visible in fixation count and absolute gaze duration heat maps for Segment 3 of the video stimulus.*

As shown in Table 7.1, all three types of data suggest that faces of people appearing on the show and telops that directly render dialogue are the types of input in the video material

that gather the most visual attention from each participant group. Such types of input not only appear in areas of the screen that are overlaid with red and green-yellow hotspots in the heat maps; participants' visual attention is often drawn to faces and this particular type of telop in the gaze replay videos. Participants' gaze paths constantly switch between telops that directly render dialogue and faces, and show a tendency to fixate on faces when utterances of speakers are not displayed on the screen. It was shown in the previous chapter that JLLs found the telop design helpful for identifying speakers appearing on the show. Data extracts in which this was explained are included in Subtheme 1.2.1: speakers. These open-ended responses provide a possible explanation for the continuous shifts in participants' gaze paths between faces of people appearing on the show and telops that directly render dialogue as it could indicate a consistent effort from participants in linking speakers to corresponding utterances.

Another possible explanation for such gaze behaviour can be found in the established literature. Stickler and Shi (2015, p. 72) also identified a regular alternation between social and content types of input in gaze paths of their participants. Their study focused on Chinese language learners who participated in a reading task and an interactive task during online Chinese tutorials. The continuous shift in visual attention between social and content types of input was observed in relation to the interactive task. For their data analysis they defined content areas of the screen as those regions that include language teaching content while social areas referred to those regions that displayed information on other participants (Stickler and Shi, 2015, p. 67; Stickler and Shi, 2017, p. 169). The participants clarified in subsequent stimulated recall interviews that they felt the need to see the names of other participants and to know who is speaking (Stickler and Shi, 2015, p. 72; Stickler and Shi, 2017, p. 169). While the current study does not focus on a synchronous communicative situation it could be that the focus on faces and the frequent jumps in gaze paths between faces of people appearing on the show and telops that directly render dialogue indicate a similar need to see speakers of corresponding utterances.

Furthermore, the established literature on speech perception provides another possible explanation for participants' focus on faces in AV contexts. The effects of facial clues on audiovisual speech perception has been a topic of particular discussion ever since McGurk and MacDonald introduced the McGurk effect in 1976 (MacDonald, 2018). They published a paper in that year that focused on a multisensory illusion that occurs when users are presented with a voice and face that articulate incongruent consonants (Tiippana,

2014, p. 1; MacDonald, 2018). Under such circumstances users perceive consonants other than those being articulated, which has come to be known as the McGurk effect (Tiippana, 2014, p. 1). Their study showed that speech perception is based on both aural information and facial clues (MacDonald, 2018, p. 16). Recent research on audiovisual speech perception has maintained this claim and argue that facial clues from the eyes and mouth play an important role in speech segmentation (Lusk and Mitchel, 2016) and human development (Irwin, Brancazio and Volpe, 2017). Considering that the eye-tracker device used for the experiment sessions had a relatively low sampling frequency, it was not possible to distinguish between visual attention directed towards the eyes and mouths of people appearing on the show. Nonetheless, JLLs' focus on faces could indicate that they integrate facial clues into their speech perception.

As shown in Table 7.1, the fixation-data suggests that subjects in each participant group direct approximately 80% of fixations and time spent looking at the analysed video segments towards the main focal points. JLLs in Group 1: pre-exchange and Japanese speakers in the Reference group show higher proportions of fixation counts and total fixation duration for faces. JLLs in Groups 2-3 look more often at telops that directly render dialogue, but fixate longer on faces. The empirical analysis of the gaze data also shows that less clear-cut results could be generated on types of resemblance. Even though JLLs in Groups 1-2 appear to fixate more often and longer on telops in Type 1: identical, proportions for other types of resemblance and participant groups are more evenly distributed. Nonetheless, as illustrated in Table 7.1, the results demonstrate that all four participant groups look more often and longer at telops in Types 3-5. These include telops that directly render dialogue with lower degrees of resemblance to corresponding utterances of speakers and narrations.

The results from fixation-data on types of input correspond to observations of O'Hagan and Sasamoto (2016). Their study focused on visual attention of native Japanese speakers while watching an excerpt of an authentic Japanese variety show. They used three AOIs to generate fixation-data on the whole screen, and two regions in which telops and faces are featured (2016, p. 47). Their findings demonstrated that Japanese speakers' gaze paths mostly ignored telops that directly render dialogue and that the area of the screen in which faces are featured gathered the most visual attention, at least in the case of fixation counts (2016, p. 47). The observation that faces of people appearing on the show are a main focal point for native Japanese speakers was made for the current study as well. The results on

total fixation duration in Table 7.1, however, differ from the findings of Stickler and Shi (2015, p. 68). Their study showed that percentages of the total fixation duration was highest for the content region followed by social and technical areas of the screen. As shown in Table 7.1, the total fixation duration for all participant groups in this study was highest for faces of people appearing on the show.

Table 7.1 also demonstrates that subjects in each participant group direct approximately 20% of fixations and time spent looking at the analysed video segments to the eight varieties included in the other types of input. None of these are overlaid with red or green-yellow hotspots in the heatmaps on all analysed video segments combined. Visual attentional shifts to such varieties of input are minimal in the gaze replay videos. Nonetheless, the area of the screen where kinesic action appears is overlaid with a green-yellow hotspot in the heat maps for Segment 3. Furthermore, participants' gaze paths showed a tendency to fixate on the combination of gesticulation and bodily configuration.

One of the participants explained during the debriefing why little visual attention was directed towards telops other than those that directly render dialogue. This explanation is included in one of the data extracts that was discarded during the thematic analysis. P44 did not make a conscious effort to look at telops that facilitate comprehension (i.e. titles and themes) as they did not change much:

Group 3: post-exchange, P44

P44 did not really look at the telops in the top left and right corners as P44 said they did not change (much).

Field notes, ll. 5-7

This could mean that the contents and the dynamic nature of types of input, as observed in *Subsection 5.4.2 Temporal dimension for analysed video segments*, play an important role in the distribution of participants' visual attention. It was argued that telops that facilitate comprehension or bridge scene change are less informative of the narrative as compared to telops that directly render dialogue considering that they repeat information. Furthermore, they are less salient as they do not include add-on text units and are accompanied with less special effects. The dynamic nature of kinesic action and its supportive role to the narrative could explain why much visual attention was directed towards the combination of gesticulation and bodily configuration.

7.5 Conclusion

This chapter illustrated that JLLs' visual attention was directed at faces of people appearing on the show and telops that directly render dialogue the most. This observation was recurrent for each of the eye-tracking data analyses. The extent to which JLLs visually attended to these types of input varied per participant group. It appeared that participants in Group 1: pre-exchange in particular had a higher count of fixations on faces of people appearing on the show than telops that directly render dialogue. It was also suggested that JLLs in each of the three participant groups fixated longest on faces of people appearing on the show when taking a closer look at the total duration of these fixations. Although clear trends were identified for the types of input, the same was not achieved for the types of resemblance that further categorise the telops that directly render dialogue.

It was discussed that those visual-verbal types of input that were identified as less prevalent in the multimodal analysis gathered less visual attention from participants. It was suggested that the lay-out of the video stimulus may indeed be an important factor in participants' viewer experience. It was also argued that JLLs' perceptions were mirrored in their gaze behaviour as telops that directly render dialogue and nonverbal clues such as faces and, to a lesser extent, bodily configuration gathered the most visual attention. However, the extent to which typography influences JLLs' visual attention could not be determined.

This chapter also linked the empirical findings to the established literature. These were made in relation to social and content regions of the screen on the one hand, and low-level and high-level factors in the video material on the other hand. It was not only suggested that social and content areas of the screen gathered the most visual attention; it was also discussed that high-level factors may have been more influential in participants' gaze behaviour than low-level factors.

The following chapter concludes this thesis. It brings together everything that has been discussed so far, reflects on the problem space and this study, and closes with directions for further research.

8. CONCLUSIONS

This thesis presented an interdisciplinary study of JLLs' perceptions of an authentic Japanese variety show. It was discussed in Chapter 1 and Chapter 2 that this research project not only developed from areas of research that focus on the use of authentic materials for FL learning purposes, FL learners' online learning behaviour while using learning materials, and users' viewer experience and viewer reception of moving images. It also evolved from studies that had already been completed as part of the overarching ICE project, and the established literature on telops and Japanese variety shows.

This thesis situated itself at the crossroads of these fields of study by taking a closer look at the appropriateness of authentic Japanese variety shows to JLLs at different stages of their Japanese language study. This research focus followed from my personal interest as a learner of the Japanese language, and studies on materials development that claim authentic materials can cater for FL learners at any stage when used appropriately (Danan, 2004, pp. 74-75; Gilmore, 2007; Gilmore, 2011; Tomlinson, 2012). This study had two objectives. It aimed to provide insights into the appropriateness of Japanese variety shows to JLT, on the one hand, and exploratory mixed-methods approaches to perception studies on the other. It was suggested in Chapter 1 and Chapter 2 that multimodality is a recurring concept in the research context of this study and that the application of an eye-tracking method would make for a valuable contribution to the established literature. This study therefore adopted a research design that combines multimodal analysis with an experimental task. It was described in Chapter 3 and Chapter 4 that this research design consisted of three main stages. These stages not only involved a protocol for empirical data collection; they also comprised procedures for complementary analyses of the video material, detailed in Chapter 5, and the empirical data, detailed in Chapter 6 and Chapter 7.

This chapter brings the findings from this research project together. It first answers the three research questions based on the observations from Chapter 6 and Chapter 7. It then discusses the implications of this study for the didactic use of authentic AV materials that feature intralingual text, the analysis of FL learners' visual attention while watching such materials, the application of an eye-tracking method to Applied Linguistics research, and the use of multimodal analysis to further explore the relationship between moving images, viewer experience, and viewer reception.

8.1 Research questions answered

8.1.1 Question 1

The first research question of this study focused on JLLs' perceptions of the video stimulus used for the experimental task. As described in *Subsection 3.1.2 Methodological approach of this study*, data was gathered on subjects' multimodal perception to gain a better understanding of JLLs' thoughts on the video material and its multimodal landscape. This concept was addressed during the first stage of the empirical data analysis. Findings from Chapter 6 were interpreted in order to answer the following question:

RQ1 What are Japanese language learners' perceptions and attitudes towards authentic Japanese variety shows?

Findings from closed-ended and open-ended data items of questionnaires, a comprehension test, and field notes demonstrated that JLLs felt predominantly positive towards the video material and that a vast majority of JLLs think that Japanese television programmes can be used for language learning purposes. This positive attitude was observed for all three learner groups which suggests that language proficiency does not play a role in the perceived usefulness of authentic Japanese variety shows to learning the Japanese language.

According to JLLs, the video material included a communicative situation that is funny and light-hearted, on the one hand, and fast-paced and flamboyant on the other. They anticipated positive learning outcomes from watching authentic Japanese variety shows despite the challenges JLLs in Groups 1-2 experienced as a result of the fast-paced and flamboyant nature of the video stimulus. A number of data extracts in each learner group demonstrated that JLLs expect that exposure to natural and "real" language use, as included in the video stimulus, could help them grow accustomed to the sound of the language and gain familiarity with fast-paced and accented versions of Japanese, and cultural aspects of Japan. In other words, the majority of JLLs who participated in this study found authentic Japanese variety shows fun to watch and helpful to raising their Japanese language awareness.

8.1.2 Question 2

The second research question focused on the second central concept of this study: visual attention. The second stage of the empirical data analysis dealt with this particular concept

in order to cross-check findings on JLLs' perceptions with actual gaze behaviour. Findings from Chapter 7 were interpreted in order to answer the following question:

RQ2 To what extent does eye-tracking data allow us to interrogate and further examine those perceptions?

The gaze data of this study facilitated further analysis of participants' multimodal perception of visual types of input. These included types of input that were reported as both helpful and unhelpful to understanding the video material. According to JLLs, telops, imagery and kinesic action supported a better understanding of the programme contents. These perceptions corresponded to observations from heat maps, gaze replay videos and fixation-data. Telops and kinesic action in particular were visually attended to by JLLs in each learner group. More specifically, JLLs' visual attention was mostly directed towards telops that directly render dialogue and faces of people appearing on the show. It was shown that JLLs in Group 1: pre-exchange had a stronger tendency to fixate on faces as compared to the other two learner groups. This, as well, seemed to correspond to JLLs' perceptions as learners from this particular participant group experienced more difficulties with the fast-paced and flamboyant nature of the video material.

According to JLLs, one more type of input was helpful to understanding the video material. This fourth type of input comprised the only audio-verbal form of expression included in the analysed video segments: dialogue. Even though participants' focus on faces could have indicated that they direct their visual attention towards the source of the dialogue, it proved to be difficult to further interrogate this perception through gaze data analyses. Observations from gaze replay videos and fixation-data on types of resemblance did not generate clear-cut results regarding JLLs' viewer reception of spoken utterances and narrations. Furthermore, it was demonstrated in the thematic map that telop design played an important role in JLLs' perceptions of the multimodal landscape of the video material. However, considering that telops and telop design overlap in the video stimulus, it was not possible to determine participants' gaze behaviour for these types of input separately.

8.1.3 Question 3

The third research question of this study followed from the empirical data findings and focused on the interpretation of QUAL and quan results in order to formulate suggestions

as to how authentic Japanese variety shows can complement Japanese language learning. The question that guided the interpretation of the empirical data was formulated as follows:

RQ3 What principles can be derived for the use of authentic Japanese variety shows in the Japanese language classroom?

The empirical data findings demonstrated that the video material allowed for JLLs in all three learner groups to perceive the programme contents in a way that makes sense to them. The multimodal landscape of the authentic Japanese variety show comprised a number of types of input that were accessible to all learners. This was evidenced by JLLs' perceptions of the telop design. The format and affective aspects of the television programme were understood by each learner group and none of the participants reported to be overwhelmed by the programme contents. This suggests that authentic Japanese variety shows can be used by JLLs at any stage of their Japanese language study.

Nonetheless, it was also observed that the multimodal landscape included types of input that were mostly picked up by more advanced learners. This was, for example, illustrated by JLLs' understanding of the (linguistic) programme contents of the video material. Furthermore, participants in Groups 1-2 reported difficulties with the fast-paced and flamboyant nature of the video stimulus, which showed up as a stronger tendency to fixate on faces, rather than telops that directly render dialogue, in the gaze data of JLLs in Group 1: pre-exchange. This suggests that JLLs in Group 3: post-exchange may benefit the most from exposure to authentic Japanese variety shows as compared to the other two learner groups.

Nevertheless, this thesis maintains that authentic Japanese variety shows can supplement Japanese language learning at any stage. This thesis derived three principles with which such materials can be used in the Japanese language classroom across academic years:

1. Careful selection of video material
2. Variation in learning goals
3. Variation in learning tasks

This study demonstrated that the choice of video material had an impact on JLLs' understanding of the programme contents. Participants who took part in the pilot

experiment sessions watched an excerpt of a talk-show in which the communicative situation made little use of visual-nonverbal input. Answers to comprehension questions suggested that the video material was too difficult for JLLs. A different excerpt was therefore used for the video stimulus of the main experiment sessions. It included more visual-nonverbal types of input (i.e. drawings and different types of kinesic action) which resulted in a better understanding of the programme contents by JLLs in each learner group. JLLs who participated in the main study also indicated that telop design, imagery and kinesic action helped them understand the video stimulus. This suggests that the presence of visual-nonverbal input is important when integrating authentic Japanese variety shows into the Japanese language classroom; at least in the case of JLLs with lower or beginner proficiency levels in Japanese. Furthermore, participants described the video stimulus as funny and light-hearted. They also seemed to be actively engaged in trying to understand the programme contents. This suggests that the leisure-oriented nature of the video material is not problematic for the integration of authentic Japanese variety shows into the Japanese language classroom.

The potential positive learning outcomes that JLLs in this study anticipated from watching authentic Japanese variety shows did not require a full understanding of the video material. They mostly related to improvements in linguistic skills that are difficult to trace such as raising Japanese language awareness. This thesis recognises that traceable gains in linguistic skills are not a prerequisite for use of authentic Japanese variety shows in class, and suggests that learning goals, which do not focus on such traceable gains in linguistic skills should also be considered when integrating authentic Japanese variety shows into the language classroom.

This study demonstrated that the formulation of comprehension questions affected the extent to which participants could showcase their understanding of the programme contents. Comprehension questions of the pilot study were relatively broad and contained mostly questions that tested memory. The comprehension test of the main study, on the other hand, referred to specific instances in the video segments. This format seemed to work well for JLLs in Group 3: post-exchange. However, questionnaires with partly correct or incorrect answers and blank responses were more frequent for JLLs in Groups 1-2. This suggests that different learning tasks should be used for learners with lower or beginner proficiency levels in Japanese.

8.2 Problem space revisited

8.2.1 Contributions of this study

The impact of this study is twofold. Firstly, this thesis provided teachers of the Japanese language insights into the appropriateness of authentic Japanese variety shows to JLLs at different stages of their Japanese language study. It generated findings that support the claim from research on materials development that authentic materials can cater for FL learners at any stage. This thesis showed that this has in large part to do with the multimodal nature of FL learning. It evidenced that the highly contextualised nature of authentic Japanese variety shows enables JLLs to avail of both linguistic and non-linguistic input for their understanding of such materials, and to pick up on information that fits their abilities. This thesis therefore suggests that authentic Japanese variety shows can be used by JLLs at any stage provided that they are used in combination with learning goals and tasks that are suited to the capabilities of language learners.

Secondly, this thesis provided researchers insights into the appropriateness of mixed-methods approaches to perception studies. It demonstrated that the mixed-methods approach of this study was appropriate for the chosen research topic. It was argued in Chapter 1 and Chapter 2 that the viewer experience and viewer reception of authentic Japanese variety shows had not yet been researched in relation to JLLs. It was therefore opted to base this study on an exploratory MMR design that is led by the research questions. This thesis demonstrated that such a research design not only allows for analytical steps to be informed by observations from the data sets; it also facilitates analyses of the multimodal nature of the video material and empirical data from different angles. Furthermore, it was shown that the overall qualitative approach enhanced the otherwise quantitative use of eye-tracking technology. The four analytical phases of data familiarisation, data exploration, data analysis, and data interpretation aided the extraction of data sets from the collated empirical data, the identification of potential trends in and between data sets, the preparation of data sets for analysis, and the triangulation of findings. The mixing of methods in each stage of the research design reinforced the robustness of this study in five ways:

1. The multimodal transcript contextualised observations from the empirical data
2. The multimodal transcript informed the proper placement of dynamic AOIs
3. The empirical data contextualised the multimodal landscape
4. The empirical data provided further insights into other empirical data sets

5. Data transformation further refined observations from data sets

The multimodal transcript contextualised findings from the empirical data sets on two occasions. First, it allowed for observations on participants' perceptions of types of input to be compared with the multimodal landscape. This showed that some of the links that JLLs thought existed between the telop design and other types of input were not identified in the multimodal transcript (e.g. the relationship between typography and the importance of information contained in telops). Second, the still images in the multimodal transcript helped with the identification of types of input in the hotspots of heat maps.

Furthermore, the multimodal transcript included necessary information for the creation and duplication of identical sets of dynamic AOIs across the four tests. The still images informed the correct placement and number of AOIs for the extraction of fixation-data on faces and telops that directly render dialogue. The information included on the telop design also helped establish proper alignment of AOIs with telops that directly render dialogue. Keyframes were determined based on these observations in order to make sure that dynamic AOIs (de)activate in accordance with the screen time of such telops.

Conversely, it can also be argued that the empirical data contextualised the findings from the multimodal transcript. It was argued that the scope of input comprised seven types and 23 varieties of input which were expected to be useful to JLLs' understanding of the programme contents. However, it was not possible to determine which of these participants would find helpful or would visually attend to without the empirical data. The data visualisation tools were useful for this in three ways. First, heat maps showed where visual attention was concentrated on the screen. Second, hotspots in the heat maps narrowed down the scope of visual types of input to those that gathered the most visual attention. This demonstrated that visual types of input with high rates of occurrence were not necessarily associated with high levels of visual attention. Third, gaze replay videos further contextualised the findings from the multimodal transcript and provided conclusive evidence for the definition of AOI groups.

The analyses of the empirical data also demonstrated how data sets can complement one another with insights into multimodal perception and visual attention. This was evidenced on three occasions. First, the thematic analysis provided insights into participants' perceptions of the telop design. Gaze behaviour for telop design could not be determined

within the scope of this study which made the observations from open-ended data items very useful. These open-ended responses and observations from field notes provided evidence against the initial assumption that mostly high-level factors such as users' goals, expectations or previous knowledge (Holsanova, 2014, p. 289) would play an important role in participants' viewer experience and viewer reception. Second, gaze replay videos supplemented heat maps with observations on participants' gaze behaviour for types of input identified in hotspots on static images. Third, the fixation-data provided numerical results on the distribution of visual attention per participant group which could not have been achieved with heat maps alone.

Data in this study has been transformed on two occasions. Data transformation happens when qualitative data is converted into quantitative data or vice versa (Johnson, 2018). For the purposes of this study, data from the multimodal transcript was transformed into rates of occurrence and counts. These included number of text units and number of add-on text units for visual-verbal types of input. Such numerical data gave insights into the spatial and temporal dimensions of the video material, and helped summarise the findings from the multimodal transcription. The second time data was transformed in this study happened during the thematic analysis. The prevalence of (sub)themes was calculated over the textual data included in the thematic map. While this was useful for determining to what extent (sub)themes are mentioned for each participant group, only the contents of data extracts could give details on how (sub)themes are discussed by participants. The narrative showed that high prevalence of (sub)themes was not necessarily accompanied with experienced difficulties from participants. This demonstrated that counts or the "size" of (sub)themes alone do not determine how a thematic map should be interpreted (Braun and Clarke, 2006, pp. 82-83).

Further to the insights gained on the appropriateness of this study's mixed-methods approach, this research made four additional contributions that carry implications for methodological choices in perception studies. First, this thesis presented how the confines of a single telop can be identified within a multimodal transcript. While clear definitions of telops exist in the established literature it proved to be difficult to uncover how researchers determine the number of telops in their studies. The exploration of the multimodal nature of the video material showed that telops comprise many facets (e.g. text units, linguistic contents, and stylistic effects). Depending on the research objectives, it can be argued that each of these facets can be used for the definition of a telop. This thesis opted to seek out

multimodal units within the shots of the multimodal transcript and to combine text units to form telops.

Second, this study built further on established taxonomies and typologies of telops, and frameworks for multimodal transcription. This thesis detailed the procedures that were used to tailor these frameworks to the chosen video material. Considering that differences exist in the composition of Japanese television programmes and that the appearance of new usages and designs is characteristic of the development of telops, it is expected that these procedures will be useful to further research on Japanese television. Furthermore, it is anticipated that the multimodal transcription framework used in this thesis will be helpful to further research that focuses on the interrelationships between types of input in AV texts or the multimodal nature of FL learning.

Third, observations from the gaze data of this study suggest that the multimodal landscape of a video stimulus may indeed influence participants' gaze behaviour. It was shown that rates of occurrence of types of input and counts of text units for visual-verbal types of input can give insights into these matters. This thesis therefore agrees with the recommendations made by Spinner, Gass and Behney (2013) and argues that further studies should report on the lay-out of a test and consider the potential influence it may have on gaze data findings.

Fourth, this study took an alternative approach to the use of AOIs. Generally speaking, studies that employ eye-tracking technology often make use of static AOIs. It was demonstrated in this thesis that the use of dynamic AOIs for a video stimulus containing linear dynamic stimuli ensured that the extraction of fixation-data happened only on those types of input that are represented by the AOIs.

8.2.2 Limitations of this study

Despite the careful consideration given to procedures for experiment conduct and empirical data analysis, several limitations of this study were determined during self-reflexive phases of this research. These limitations concern two stages of the research design in particular; the protocol for empirical data collection (i.e. Stage 2), and the selection of participants and empirical data analysis (i.e. Stage 3). The insights gained into these matters constitute some additional considerations for further research on the analysis of FL learners' visual attention while watching authentic AV materials that feature

intralingual text, and the application of an eye-tracking method to examine FL learners' visual attention.

Three limitations were identified in relation to Stage 2 of this study's research design. These involve the script for the experiment sessions and the comprehension test included in the post-task questionnaire. First, it has proved to be difficult to explain the calibration in the exact same way for each participant; especially when subjects ask questions during the calibration process. As noted previously, the script was included in the protocol to ensure that participants were exposed to the same amount of information during each experiment session. Although the information included in the script was explained to all subjects, this means that they were not addressed with the exact same words.

Second, a question regarding any eye conditions was not included in the protocol for empirical data collection. It has been argued that this study focuses on a general impression of participants' gaze behaviour and that some leeway in the level of precision and accuracy in the gaze data quality was therefore justified. Generally speaking, eye conditions were only discussed with participants in cases where the calibration was unsatisfactory at the first attempt. This means that eye conditions were not discussed with all JLLs and Japanese speakers. A full record of eye conditions was therefore not available in this study and could not be linked to the quality of the gaze data.

Third, the answers to one of the questions in the comprehension test (i.e. *post1* Q11)[29] had not been formulated in such a way that only one answer can be considered as correct. This particular question asks for the reason behind a panellist's facial expression as shown in Figure 1 in the comprehension test (*Appendix E. Post-task questionnaire*). The correct answer is the third box in which it is explained that facial expressions reveal someone's true feelings despite what (s)he is saying. However, in hindsight it appeared that one of the other answers could also be considered as correct although it provides an incomplete answer as compared to the third box. This is the first box in which it is argued that the panellist makes fun of the expert guest. This means that this particular question may have caused confusion among participants when filling out the post-task questionnaire.

Four more limitations were detected for Stage 3 in the research design of this study. These not only relate to the selection of participants; they also involve three of the four analytical

[29] This question is included as *post1* Q10 for Japanese speakers.

phases of the empirical data analysis. First, missing responses or information were not included in the criteria for participant selection. It has been argued that the participant groups in this thesis are thought to comprise cohorts of JLLs at roughly the same stage of their Japanese language study due to individual differences between learners. However, for three JLLs a less clear picture could be uncovered of their experience with learning Japanese and their linguistic abilities in the Japanese language. Two participants in Group 1: pre-exchange were study abroad students, and one participant in Group 2: exchange had not filled out the pre-task questionnaire. This made it difficult to ascertain whether the enrolment of these study abroad students in the third academic year of their home university had carried serious implications for the coherence of Group 1: pre-exchange. Further to this, the missing responses of the JLL in Group 2: exchange has made the inclusion of this particular participant debatable as it cannot be maintained with certainty that this subject spent the same amount of time on exchange in Japan as compared to the other JLLs in the same participant group.

Second, while reviewing the questionnaire responses it appeared that some of the questions may have been misinterpreted by JLLs. This suspicion is based on responses to two questions in particular. One of these questions is *post2* Q7.[30] It asks participants how many hours on average per week they watch Japanese television programmes. This particular question instructs respondents to skip several questions on Japanese television programmes (i.e. *post2* Q8-12)[31] in cases where subjects do not watch any Japanese television. It was observed that several JLLs who had responded with "I do not watch Japanese TV programmes" did not continue answering from *post2* Q13;[32] rather, they also filled out the questions they were instructed to skip. Although such responses may be the result of a misunderstanding, these answers could also indicate that JLLs wanted to demonstrate that they watch Japanese television occasionally instead of weekly or that they wished to explain what they would do if they were to watch Japanese television programmes. Nonetheless, responses to *post2* Q8-12 from participants who had indicated not to watch Japanese television programmes were excluded from the data analysis.

The second question is *post2* Q18, which includes the semantic differential scales. For some participants a discrepancy was observed between responses to several statements, and answers to other questions included in the questionnaire and observations from field

[30] This question is included as *post2* Q1 for Japanese speakers.
[31] These questions are included as *post2* Q2-6 for Japanese speakers.
[32] This question is included as *post2* Q7 for Japanese speakers.

notes. As noted previously, the particular question that includes semantic differential scales was designed in such a way that both ends of the scales feature opposite statements. These positive and negative statements appeared at both ends of the scales in order to establish a neutral and unbiased formulation of the question. At times it appeared that some of the statements were treated as Likert scales and that the numbers were thought to represent the extent to which someone can (dis)agree (i.e. strongly agree, agree, neutral, disagree, and strongly disagree) with a statement without noticing the opposite statement mentioned at the other end of the scale. While potential misinterpretations cannot be ruled out entirely when conducting questionnaires, misinterpretations of semantic differential scales can perhaps be mitigated through the inclusion of an example scale in the question.

Third, two limitations were observed in relation to the field notes that were collated for the purposes of this study. For some participants the field notes include comments on times when loud noises could be heard outside of the experiment room. However, it was not possible to link these back to the timeline of the recordings as the starting times of recordings have not been included in the field notes. It was therefore opted to consult the user cam instead to try work out, where possible, whether such noises accompany visual attentional shifts in participants' gaze paths. In addition, the debriefings during which the field notes have been generated were not (semi-)structured. This means that questions were not asked on a pre-established set of topics or scenes in the video stimulus and that field notes therefore dealt with diverging concepts on different scenes in the video material.

Fourth, two limitations were detected in the empirical analysis of the gaze data. The first one of these relates to the gaze replay videos. Considering that gaze behaviour and tendencies in visual attention vary within and between portions of recordings, the results presented in the accumulated gaze plots in this thesis comprise recurrent movements in gaze paths encountered throughout the analysed video segments and across participant groups. These form an impressionistic account and a systematic analysis between participant groups in a CAQDAS could not be conducted within the scope of this thesis due to time constraints.

The second limitation is concerned with the size of the user-defined dynamic AOIs. Whereas the size of AOIs in this study was appropriate for the extraction of fixation-data on types of input, it was discerned that the size may have played a part in the lack of clear-cut findings on types of resemblance. More specifically, AOIs may have been too big for

the analysis of types of resemblance. Considering that telops stretch over a certain time frame and can consist of more than one text unit, types of resemblance in this study were determined based on the overall degree of resemblance observed for all text units of a telop combined. However, it is possible that individual text units show different degrees of resemblance to corresponding utterances as opposed to the one assigned to the telop such text units are affiliated with. This means that the definition of types of resemblance and the choice of user-defined AOIs may not have been appropriate for the study of degrees of resemblance between telops that directly render dialogue and corresponding utterances of speakers and narrations.

Furthermore, it was very time-consuming to create the set of guidelines on the AOI properties, and to generate the 136 dynamic AOIs and eight AOI groups. It also appeared that, even though the AOI groups defined in this study were appropriate for the generation of AOI group summary data per participant group, the design of AOI group analyses complicated the extraction of fixation-data on individual dynamic AOIs and raw data per participant. The use of dynamic AOIs and AOI groups therefore needs to be considered carefully for any further studies that require extraction of such fixation-data.

Whereas these limitations can be considered as areas for improvement when replicating this study, the limitations that were detected for Stage 3 of the research design also suggest several directions for further research. Such directions relate to the use of methods that can offset the identified weaknesses of questionnaires, field notes, and gaze replay videos. This brings us back to those methods that were recommended in the established literature on SCMC but could not be applied in the scope of this thesis due to time and financial constraints: stimulated recall interviews, and interpretative analysis of gaze replay videos.

8.3 Directions for further research

The limitations of questionnaires and field notes can be minimised when using stimulated recall interviews, which allow for further exploration of observations made before, during or after an experimental task. They can be structured according to questions that are prepared in advance of the empirical data collection when the research focus is on target scenes or stimuli in a test. They can also be based on questions that arise during an experimental task. Such stimulated recall interviews are most effectively done in an experiment set-up that utilises two screens; one screen to record the eye-movements of participants and a secondary screen to monitor participants' eye-movements in real time in

preparation of the stimulated recall interviews (O'Rourke *et al.*, 2016, p. 295). Alternatively, researchers can opt to watch a participant's gaze replay video with the participant and ask questions that come to mind while replaying the recording (Smith, 2012, p. 61; Stickler and Shi, 2015, p. 61; Stickler and Shi, 2017, p. 167). This method therefore reduces room for misinterpretations of questions and allows for observations to be linked directly to the relevant scenes or stimuli in a test as it is based on the interaction between prompts in the gaze replay video, the participant, and the researcher. Furthermore, the various applications of stimulated recall interviews demonstrate that such a method can be incorporated into many types of MMR design and can elicit more in-depth thought processes as compared to questionnaire responses and field notes.

This thesis suggests three directions for further research on JLLs' viewer experience and viewer reception of authentic Japanese variety shows that apply to the use of stimulated recall interviews. First, a deeper understanding is needed of the role of telop design in learners' understanding of such video materials. Participants in this study described the usefulness of low-level factors such as colour and font in open-ended data items. However, it proved to be difficult to further explore this finding in the empirical analysis of the gaze data as telops and telop design are not visually separated in the video stimulus. Distinction between visual attention directed towards telops, on the one hand, and telop design on the other is therefore not possible based on eye-tracking data alone. Stimulated recall interviews during which participants are asked questions regarding the three components (i.e. typography, display rate, and movement) and six varieties (i.e. expansion, background colour, punctuation, special effect, picture, and other) of telop design can provide insights into these matters. They can also shed light on the parallels that were found between SDH and telops that directly render dialogue, and the roles of high-level and low-level factors in JLLs' viewer experience.

Second, a deeper understanding needs to be gained of the relationship between JLLs' visual attentional shifts, telops that directly render dialogue, and the audio track. Even though this thesis did not generate clear-cut results on types of resemblance, observations on narration telops from gaze replay videos suggest that there might be a link between the degrees of resemblance of telops that directly render dialogue to corresponding utterances of speakers and narrations, and participants' visual attention. Stimulated recall interviews that give participants the opportunity to reflect on their gaze behaviour in relation to such

similarities can expand our knowledge on speech-monitoring practices and the appropriateness of telops for the Japanese language classroom.

Third, ways in which different scopes of input or multimodal landscapes affect JLLs' viewer experience and viewer reception need further investigation. It was argued in this thesis that the three analysed video segments in the video stimulus differ in their multimodal structure. This was discussed in relation to the high rates of occurrence of bodily configuration in Segment 3 in particular. Observations from fixation count and absolute gaze duration heat maps suggest that this difference in multimodal landscapes affects the concentration of visual attention on the screen. However, it was not possible to further analyse this finding within the scope of this thesis. Stimulated recall interviews that include questions regarding JLLs' perceptions of different multimodal landscapes can give further insights into the appropriateness of particular combinations of types of input to learners at different stages of their Japanese language study.

The established literature proposes several experiment set-ups that facilitate the incorporation of stimulated recall interviews into a research project. The Tobii X2-60 eye-tracking device that was used for the purposes of this study can be set up in two different ways that allow for a researcher to monitor a participant's gaze path on a secondary screen while recording (Tobii AB, 2014a, p. 22). Furthermore, *Tobii Studio* includes a feature called Retrospective Think Aloud (RTA) with which stimulated recall interviews can be recorded (Tobii AB, 2014b, pp. 40-41). Alternatively, audio and screen capture software such as *Camtasia* (Smith, 2012, p. 60) or mounted cameras (Stickler and Shi, 2015, p. 60; Stickler and Shi, 2017, p. 166) can also be used to record stimulated recall interviews.

The latter two of these directions for further research can also be investigated from another perspective through the use of interpretative analysis of gaze replay videos. O'Rourke (2008, p. 246; 2012, p. 319) argues that such interpretational analysis of gaze paths can give valuable insights into FL learner's online learning behaviour that cannot be derived from other data sets. While O'Rourke makes this claim in relation to SCMC, it can be argued that interpretative analysis of gaze paths can, for example, give indications of vision-sound matching practices which either confirm or debunk the observations on JLLs' gaze behaviour for narration telops. It can also expand our knowledge on the ways in which JLLs integrate forms of expression in authentic Japanese variety shows into their viewer reception. O'Rourke (2012, p. 318) argues that interpretive analysis needs to be

rigorous and consistent in its definition and application of operational definitions (i.e. codes) in order to minimise the subjective nature of this method. However, it appears that this method is faced with many challenges as there are no established operational definitions on gaze behaviour available to researchers (2012, pp. 335-336). Further research on methodology is therefore recommended for this particular field of study.

This thesis suggests the use of CAQDAS in order to facilitate a more systematic analysis of gaze paths that can offset the beforementioned limitation of gaze replay video analysis (Flewitt *et al.*, 2014, p. 49). While nowadays a vast array of software packages exist with which qualitative data analysis can be conducted, the choice of CAQDAS depends on several factors such as the research question and the type of data that needs to be analysed (Bassett, 2011). The established literature on SCMC recommends the use of *ELAN*[33] (O'Rourke *et al.*, 2016, p. 295). Another option is *MAXQDA*.[34] This is a software package that is designed for qualitative and mixed-methods research specifically and includes a feature with which video files can be coded.[35] It is important to consider the appropriateness of a CAQDAS in relation to the types of analysis a particular software package allows and the objectives of a study; the choice of software package is therefore not limited to these suggestions. Considering the complexity of interview data, this thesis also recommends the use of CAQDAS for the analysis of data obtained through the beforementioned stimulated recall interviews.

Whereas these suggested pathways provide methods for improving the current study, it should be noted that it is important to examine several issues when building on this research. First, application of multimodal transcription and eye-tracking analysis should be carefully planned as they both comprise labour-intensive and time-consuming procedures. Moreover, the latter involves a steep learning curve and possibly limited access to the costly hardware and software (Stickler, Smith and Shi, 2016, p. 182). It has therefore been suggested that the necessity of eye-tracking technology to answering research questions should be carefully considered before commencing a research project (2016, p. 180). Second, the inclusion of a Reference group comprising native Japanese speakers to study JLL's perceptions should be reviewed. Whereas the Japanese speakers in this study functioned as a point of reference, exclusion of such a participant group when focussing on

[33] *ELAN* is available at: https://tla.mpi.nl/tools/tla-tools/elan/.
[34] *MAXQDA* is available at: https://www.maxqda.com/.
[35] For more information, please see MAXQDA (2019).

learners of the Japanese language could be considered for further studies on the research topic.

While these directions for further research are related to the limitations of this study and directed specifically towards research on JLLs' viewer experience and viewer reception of authentic Japanese variety shows, this thesis suggests some additional directions which can help put the findings of this study into context. Now that it has been established that authentic Japanese variety shows provide a low-anxiety setting with programme contents that is useful to learners at different stages of their Japanese language study it is imperative that a closer look is taken at how such materials can be integrated into the Japanese language classroom. This involves the development of appropriate learning goals and tasks (Gilmore, 2007; Tomlinson, 2012), and effective learner strategies (Danan, 2004).

The findings of this study showed that more advanced JLLs in Group 3: post-exchange developed a good understanding of the video material and that they performed well in answering the comprehension questions. Learners in Groups 1-2, on the other hand, were able to pick up on fewer aspects of the analysed video segments as they experienced some difficulties with the programme contents. It was therefore suggested that learning goals, which do not focus on traceable gains in linguistic skills should be considered as learning outcomes when integrating authentic Japanese variety shows into language classes, and that different learning tasks should be developed for JLLs at beginner level. These suggestions were not only grounded in feedback from participants; they were also based on observations with regard to the types of input that learners understood during experiment sessions. This thesis proposes that, based on these findings, learning goals for JLLs at beginner level should be more focused on raising awareness of the Japanese language and culture while learning goals for more advanced JLLs should be centred on the characteristics and contents of authentic Japanese language use. Furthermore, this thesis argues that tasks regarding aspects that JLLs in Groups 1-2 were able to pick up on form good alternatives to the type of questions included in the comprehension test. This implies awareness-raising exercises on the format and affective aspects of authentic Japanese variety shows, and the interactions between people appearing on them.

It was discussed in the literature review that research has maintained that a strong element of manipulation underlies the use of telops and that such text is employed to guide users' interpretation processes (O'Hagan, 2010; Sasamoto, 2014). Furthermore, it was argued that

telops are seen as a safety net by younger Japanese users. Telops not only help such users with their understanding of the programme contents; they also facilitate viewing habits of users who do several activities at the same time (Shitara, 2008; Shitara, 2009). These claims suggest that telop designs influence how viewers watch Japanese television. Whereas previous studies on intralingual text have tested FL learners' performance with different text designs (Montero Perez, Peters and Desmet, 2013: Montero Perez, Peters and Desmet, 2015), this thesis proposes development of learner strategies without affecting the authenticity of the materials. It is suggested here that effective learner strategies should be achieved through longitudinal research in which JLLs get ample opportunities to develop such strategies by means of exposure to authentic Japanese television programmes. To this end, viewing habits of native Japanese speakers could inform strategies of more advanced JLLs considering that they have developed effective viewing habits that allow them to do other activities while watching Japanese television.

Nonetheless, this thesis recognises that access to authentic AV materials in a FL is not limited to the language classroom; especially now that the number of streaming services with in-built subtitling features is on the rise. Before commencing this research project the established literature had already called for research on autonomous language learning and longitudinal studies on the pedagogical use of intralingual text (Danan, 2004; Vanderplank, 2015; Vanderplank, 2016, p. 243). These pathways for further research are still relevant to this day.

It is hoped that this thesis has sparked interest in areas of research that deal with the use of authentic AV materials with intralingual text for FL learning, FL learners' online learning behaviour while using such materials, and the role of multimodality in viewer experience and viewer reception of moving images. It is also hoped that this thesis has inspired researchers to investigate these areas of research in relation to authentic Japanese variety shows whether through replication of this study, exploration of the suggested directions for further research or reconsideration of the effects of intralingual text on learners albeit in the context of telop. Such studies on the multimodal nature of FL learning make for valuable contributions to FL learning practices and better equip FL learners and teachers in the technology-rich and globalised world we live in today.

REFERENCES

@DIME (2014) *"Mita koto ga aru" "kakasazu miru" "manzokudo ga takai" baraeti bangumi rankingu* [*"Have seen it before", "watch without missing an episode", "very satisfied" variety show ranking*], 20 November. Copyright of Shogakukan Inc., 2018. Available at: http://dime.jp/genre/166985/ (Accessed: 9 December 2014).

Arai, K. (2014) 'Jimaku to iu media no kanōsei: moji ga ayadoru atarashii hyōgen hōhō [The possibilities of the subtitle medium: new methods of expression with decorated characters]', *BLOGOS* (copyright of LINE Corporation), 26 November. Available at: http://blogos.com/article/99784/ (Accessed: 10 December 2014).

Araújo, V.L.S. (2008) 'The educational use of subtitled films in EFL teaching', in Díaz Cintas, J. (ed.) *The didactics of Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 227-238.

Attride-Stirling, J. (2001) 'Thematic networks: an analytic tool for qualitative research', *Qualitative Research*, 1(3), pp. 385-405. doi: 10.1177/146879410100100307.

Baldry, A. and Thibault, P.J. (2006) *Multimodal transcription and text analysis: a multimedia toolkit and coursebook with associated on-line course*. London: Equinox Publishing Limited.

Bartoll, E. (2004) 'Parameters for the classification of subtitles', in Orero, P. (ed.) *Topics in Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 53-60.

Bassett, R. (2011) 'Visual conceptualization opportunities with qualitative data analysis software', in Margolis, E. and Pauwels, L. (eds.) *The SAGE handbook of visual research methods*. London: SAGE, pp. 530-545.

Bateman, J., Wildfeuer, J. and Hiippala, T. (2017) *Multimodality. Foundations, research and analysis: a problem-oriented introduction*. Berlin/Boston: De Gruyter Mouton.

Batty, C., Perkins, C. and Sita, J. (2015) 'How we came to eye tracking animation: a cross-disciplinary approach to researching the moving image', *Refractory: A Journal of Entertainment Media*, 25, 6 February. Available at: https://refractory-journal.com/batty-perkins-sita/ (Accessed: 3 December 2018).

Bazeley, P. (2016) *Integration in mixed methods research* [Master class webinar], *Mixed Methods webinar series*. International Institute for Qualitative Methodology. 8 November.

Bisson, M., Van Heuven, W.J.B., Conklin, K. and Tunney R.J. (2014) 'Processing of native and foreign language subtitles in films: an eye tracking study', *Applied Psycholinguistics*, 35, pp. 399-418. doi: 10.1017/S0142716412000434.

Blignaut, P. and Wium, D. (2014) 'Eye-tracking data quality as affected by ethnicity and experimental design', *Behavior Research Methods*, 46(1), pp. 67-80. doi: 10.3758/s13428-013-0343-0.

Bojko, A. (2009) ‘Informative or misleading? Heatmaps deconstructed’ (from the Proceedings of the 13th HCI international conference, San Diego, 19-24 July 2009), in Jacko, J.A. (ed.) *Human-computer interaction. New trends: Part 1*. Berlin/Heidelberg: Springer-Verlag, pp. 30-39. doi: 10.1007/978-3-642-02574-7_4.

Brace, I. (2004) *Questionnaire design: how to plan, structure and write survey material for effective market research*. 3rd edn. London: Kogan Page Limited.

Braun, V. and Clarke, V. (2006) ‘Using thematic analysis in psychology’, *Qualitative Research in Psychology*, 3(2), pp. 77-101. doi: 10.1191/1478088706qp063oa.

Braun, V. and Clarke, V. (2013) *Successful qualitative research: a practical guide for beginners*. Los Angeles: SAGE.

Charles, T.J. and Trenkic, D. (2015) ‘Speech segmentation in a second language: the role of bi-modal input’, in Gambier, Y., Caimi, A. and Mariotti, C. (eds.) *Subtitles and language learning: principles, strategies and practical experiences*. Bern: Peter Lang, pp. 173-198.

Conklin, K. and Pellicer-Sánchez, A. (2016) ‘Using eye-tracking in Applied Linguistics and second language research’, *Second Language Research*, 32(3), pp. 1-15. doi: 10.1177/0267658316637401.

Creswell, J.W. and Plano Clark, V.L. (2007) *Designing and conducting mixed methods research*. Thousand Oaks: SAGE Publications.

Danan, M. (2004) ‘Captioning and subtitling: undervalued language learning strategies’, *Meta*, 49(1), pp. 67-77. doi: 10.7202/009021ar.

De Bruycker, W. and d'Ydewalle, G. (2003) ‘Reading native and foreign language television subtitles in children and adults’, in Hyönä, J., Radach, R. and Deubel, H. (eds.) *The mind's eye: cognitive and applied aspects of eye movement research*. Amsterdam/London: Elsevier, pp. 671-684.

DeKeyser, R. and Juffs, A. (2005) ‘Cognitive considerations in L2 learning’, in Hinkel, E. (ed.) *Handbook of research in second language teaching and learning*. New York/London: Routledge Taylor & Francis Group, pp. 437-454.

Delabastita, D. (1989) ‘Translation and mass-communication: film and T.V. translation as evidence of cultural dynamics’, *Babel*, 35(4), pp. 193-218. doi: 10.1075/babel.35.4.02del.

De Linde, Z. and Kay, N. (1999) *The semiotics of subtitling*. Manchester: St Jerome Publishing.

DENTSU INC. (2015) *Terebi o katsuyōshita arata na puromōshon fōmatto no kaihatsu ni chōsen. Hayaoshi kuizu rendō dorama “Dassen keiji” ga 12-gatsu 13-nichi hōsō!* [*The attempt to take on the development of a new production format using the TV set. The drama series “Dassen Police” which is operated through a push-button quiz will be broadcast on 13 December!*], 10 December. Copyright of DENTSU INC., 2013. Available at: http://dentsu-ho.com/articles/3458 (Accessed: 11 December 2015).

Díaz Cintas, J. (ed.) (2008) *The didactics of Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Díaz Cintas, J. (2013) 'Subtitling: theory, practice and research', in Millán, C. and Bartrina, F. (eds.) *The Routledge handbook of Translation Studies*. London/New York: Routledge Taylor & Francis Group, pp. 273-287.

Díaz Cintas, J. and Fernández Cruz, M. (2008) 'Using subtitled video materials for foreign language instruction', in Díaz Cintas, J. (ed.) *The didactics of Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 201-214.

Díaz Cintas, J. and Remael, A. (2014) *Audiovisual Translation: subtitling*. Oxon/New York: Routledge Taylor & Francis Group.

Dublin City University (2016) *Research on Japanese impact captions which shake, pulse and capture your attention on TV screen*, 9 June. Copyright of Dublin City University. Available at: https://www.dcu.ie/research/news/2016/jun/research-japanese-impact-captions-which-shake-pulse-and-capture-your (Accessed: 24 April 2019).

Duchowski, A.T. (2003) *Eye tracking methodology: theory and practice*. New York: Springer.

Dwyer, T. (2015) 'From subtitles to SMS: eye tracking, texting and Sherlock', *Refractory: A Journal of Entertainment Media*, 25, 7 February. Available at: https://refractory-journal.com/dwyer/ (Accessed: 12 February 2015).

Dyer, A.G. and Pink, S. (2015) 'Movement, attention and movies: the possibilities and limitations of eye tracking?', *Refractory: A Journal of Entertainment Media*, 25, 6 February. Available at: https://refractory-journal.com/dyer-pink/ (Accessed: 30 November 2018).

Eskey, D.E. (2005) 'Reading in a second language', in Hinkel, E. (ed.) *Handbook of research in second language teaching and learning*. New York/London: Routledge Taylor & Francis Group, pp. 563-580.

Evergreen, S. (2019) *How to make a diverging stacked bar chart in Excel*. Available at: https://stephanieevergreen.com/diverging-stacked-bars/ (Accessed: 13 October 2018).

Flewitt, R., Hampel, R., Hauck, M. and Lancaster, L. (2014) 'What are multimodal data and transcription?', in Jewitt, C. (ed.) *The Routledge handbook of multimodal analysis*. 2nd edn. Oxon/New York: Routledge Handbooks, pp. 44-59.

Frith, H. and Gleeson, K. (2004) 'Clothing and embodiment: men managing body image and appearance', *Psychology of Men & Masculinity*, 5(1), pp. 40-48. doi: 10.1037/1524-9220.5.1.40.

Fukunaga, S., Katayama, T. and Shoyama, S. (2014a) 'Kotonaru fonto to haichi no teroppu o mochi'ita washoku no ryōri kōtei gamen no inshō hyōka [The impressions of a scene of cooking Japanese food on the screen with captions written using different types of font and arrangement]', *Journal of Human and Living Environment*, 21(1), pp. 25-34. doi: 10.24538/jhesj.21.1_25.

Fukunaga, S., Katayama, T. and Shoyama, S. (2014b) 'Kotonaru mojishoku no teroppu o mochi'ita washoku no ryōri bangumi no inshō hyōka [Assessment of viewers' impressions of a Japanese cooking program using tickers in different colors]', *Journal of Human and Living Environment*, 21(2), pp. 85-91. doi: 10.24538/jhesj.21.2_85.

Gambier, Y. (2013) 'The position of Audiovisual Translation studies', in Millán, C. and Bartrina, F. (eds.) *The Routledge handbook of Translation Studies*. London/New York: Routledge Taylor & Francis Group, pp. 45-59.

Gambier, Y. (2015) 'Subtitles and Language Learning (SLL): theoretical background', in Gambier, Y., Caimi, A. and Mariotti, C. (eds.) *Subtitles and language learning: principles, strategies and practical experiences*. Bern: Peter Lang, pp. 63-82.

Gilmore, A. (2007) 'Authentic materials and authenticity in foreign language learning', *Language Teaching*, 40(2), pp. 97-118. doi: 10.1017/S0261444807004144.

Gilmore, A. (2011) '"I prefer not text": developing Japanese learners' communicative competence with authentic materials', *Language Learning*, 61(3), pp. 786-819. doi: 10.1111/j.1467-9922.2011.00634.x.

Godfroid, A., Winke, P. and Gass, S.M. (eds.) (2013) 'Eye-movement recordings in second language research', *Studies in Second Language Research: Special Issue*, 35(2). Available at: http://journals.cambridge.org/action/displayIssue?decade=2010&jid=SLA&volumeId=35&issueId=02&iid=8937225 (Accessed: 24 November 2016).

Goodman, K. (1988) 'The reading process', in Carrell, P.L., Devine, J. and Eskey, D.E. (eds.) *Interactive approaches to second language reading*. Cambridge/New York/Melbourne: Cambridge University Press, pp. 11-21.

Gottlieb, H. (2004) 'Language-political implications of subtitling', in Orero, P. (ed.) *Topics in Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 83-100.

Gyōretsu no Dekiru Hōritsu Sōdanjo (2014) Nippon Television Network Corporation, 22 June.

Harada Senmonka (2015) 'Baraeti bangumi wa heikin "3.75-byō" de katto ga kawaru! (Harada Senmonka no TV no mikata) [Variety shows cut scenes every "3.75 seconds" on average! (The view of specialist Harada on TV)]', *NEWS Posuto Sebun* (copyright of Shogakukan Inc., 2018), 26 April. Available at: http://www.news-postseven.com/archives/20150426_319262.html (Accessed: 1 June 2015).

Holsanova, J. (2014) 'Reception of multimodality: applying eye-tracking methodology in multimodal research', in Jewitt, C. (ed.) *The Routledge handbook of multimodal analysis*. 2nd edn. Oxon/New York: Routledge Handbooks, pp. 287-298.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. and Van de Weijer, J. (2011) *Eye tracking: a comprehensive guide to methods and measures*. Oxford/New York: Oxford University Press.

Honmadekka!?TV (2013) Fuji Television Network, Inc., 7 August.

Irwin, J., Brancazio, L. and Volpe, N. (2017) 'The development of gaze to a speaking face', *The Journal of the Acoustical Society of America*, 141(5), pp. 3145-3150. doi: 10.1121/1.4982727.

Jakobson, R. (1959) 'On linguistic aspects of translation', in Venuti, L. (ed.) *The Translation Studies reader*. London/New York: Routledge Taylor & Francis Group, pp. 232-239.

Jewitt, C. (ed.) (2014) *The Routledge handbook of multimodal analysis*. 2nd edn. Oxon/New York: Routledge Handbooks.

Johnson, R. (2018) *Understanding and conducting mixed methods research* [Summer school], *11th Qualitative Research Summer School*. Dublin City University: School of Nursing and Human Sciences. 7-8 June.

Just, M.A. and Carpenter, P.A. (1980) 'A theory of reading: from eye fixations to comprehension', *Psychological Review*, 87(4), pp. 329-354. doi: 10.1037/0033-295X.87.4.329.

Kaindl, K. (2013) 'Multimodality and translation', in Millán, C. and Bartrina, F. (eds.) *The Routledge handbook of Translation Studies*. London/New York: Routledge Taylor & Francis Group, pp. 257-270.

Kalantzis, M., Cope, B., Chan, E. and Dalley-Trim, L. (2016) *Literacies*. Port Melbourne: Cambridge University Press.

Kami'iyu, B. (2014) 'Chijōha eiga hōsō ni aho mitai na teroppu ga tsuku mittsu no riyū [Three reasons for adding idiotic telops to terrestrial movie broadcastings]', *BLOGOS* (copyright of LINE Corporation), 16 November. Available at: http://blogos.com/article/98884/ (Accessed: 10 December 2014).

Kanai, M. (2011) 'Baraeti teimeiki ni kōchō na "Honmadekka!?TV", ninki no riyū [The favourable conditions of "Honmadekka!?TV" during a period in which variety shows floundered, reasons for its popularity]', *NIKKEI STYLE* (managed by Nikkei Inc. and Nikkei Business Publications, Inc.), 6 June. Available at: https://style.nikkei.com/article/DGXNASFK3102V_R30C11A5000000?channel=DF280120166614&style=1 (Accessed: 21 November 2017).

Kasper, G. and Roever, C. (2005) 'Pragmatics in second language learning', in Hinkel, E. (ed.) *Handbook of research in second language teaching and learning*. New York/London: Routledge Taylor & Francis Group, pp. 317-334.

Kawabata, M. (2006) 'Terebi nyūsu bangumi ni okeru keishikiteki gorakuka no genjō to sono mondai: jimaku, teroppu o chūshin toshite [Well-guided or misled?: open captions and the tabloidization of the TV news programs in Japan]', *Mejiro Journal of Social and Natural Sciences*, 2, pp. 209-219.

Koga, T. (2013) 'Chōsa kenkyū nōto: jōhō, baraetī bangumi no enshutsu yōso teiryō bunseki no kokoromi: puraimutaimu bangumi no bunseki chōsa kara [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), pp. 68-81. Available at:

https://www.nhk.or.jp/bunken/summary/research/report/2013_05/20130505.pdf (Accessed: 26 March 2014).

Koolstra, C.M. and Beentjes, J.W.J. (1999) 'Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home', *Educational Technology Research and Development*, 47(1), pp. 51-60. doi: 10.1007/BF02299476.

Koolstra, C.M., Peeters, A.L. and Spinhof, M. (2002) 'The pros and cons of dubbing and subtitling', *European Journal of Communication*, 17(3), pp. 325-354. Available at: https://doi-org.dcu.idm.oclc.org/10.1177/0267323102017003694 (Accessed: 6 June 2019).

Kothari, B., Pandey, A. and Chudgar, A.R. (2004) 'Reading out of the "idiot box": same-language subtitling on television in India', *Information Technologies and International Development*, 2(1), pp. 23-44. Available at: https://itidjournal.org/index.php/itid/article/view/191/61 (Accessed: 16 March 2017).

Kothari, B., Takeda, J., Joshi, A. and Pandey, A. (2002) 'Same language subtitling: a butterfly for literacy?', *International Journal of Lifelong Education*, 21(1), pp. 55-66. doi: 10.1080/02601370110099515.

Krashen, S. (2008) 'Language education: past, present and future', *RELC Journal*, 39(2), pp. 178-187. doi: 10.1177/0033688208092183.

Kruger, J., Doherty, S. and Soto-Sanfiel, M.T. (2017) 'Original language subtitles: their effects on the native and foreign viewer', *Comunicar*, 50(XXV), pp. 23-32. doi: 10.3916/C50-2017-02.

Kruger, J., Szarkowska, A. and Krejtz, I. (2015) 'Subtitles on the moving image: an overview of eye tracking studies', *Refractory: A Journal of Entertainment Media*, 25, 7 February. Available at: https://refractory-journal.com/kruger-szarkowska-krejtz/ (Accessed: 30 November 2018).

Kuppens, A.H. (2010) 'Incidental foreign language acquisition from media exposure', *Learning, Media and Technology*, 35(1), pp. 65-85. doi: 10.1080/17439880903561876.

Leow, R.P., Grey, S., Marijuan, S. and Moorman, C. (2014) 'Concurrent data elicitation procedures, processes, and the early stages of L2 learning: a critical overview', *Second Language Research*, 30(2), pp. 111-127. doi: 10.1177/0267658313511979.

Library of Congress (2012) *Japanese*. Available at: https://www.loc.gov/catdir/cpso/romanization/japanese.pdf (Accessed: 19 October 2019).

livedoor NEWS (2014) *Nittere "Kin'yō rōdo SHOW!" no teroppu ni shichōsha kara hihan ga sattō* [*Telops in Nippon TV's "Kin'yō rōdo SHOW!" under heavy criticism from viewers*], 16 November. Copyright of LINE Corporation. Available at: http://news.livedoor.com/article/detail/9472947/ (Accessed: 9 December 2014).

livedoor NEWS (2015) *Fuji ga Teretō ni rekishiteki zanpai dōjikantai & dōnaiyō kaiketsu de… gamen gochagocha sugi de mizurai* [*Fuji's historical defeat against TV Tōkyo: showdown of the same contents aired at the same time…so messy on the screen that it is hard to watch*], 9 March. Copyright of LINE Corporation. Available at: http://news.livedoor.com/article/detail/9869342/ (Accessed: 2 April 2015).

livedoor NEWS (2016) "*Hiruobi!" de hōsō toraburu Suga Yoshihide Kanbō Chōkan no kaiken de jimaku sūpā ni misu* [*Trouble during the broadcast of "Hiruobi!": a mistake in the captions of an interview with the Chief Cabinet Secretary Yoshihide Suga*], 25 April. Copyright of LINE Corporation. Available at: http://news.livedoor.com/article/detail/11453992/ (Accessed: 26 April 2016).

Lusk, L.G. and Mitchel, A.D. (2016) 'Differential gaze patterns on eyes and mouth during audiovisual speech segmentation', *Frontiers in Psychology*, 7(52). doi: 10.3389/fpsyg.2016.00052.

MacDonald, J. (2018) 'Hearing lips and seeing voices: the origins and development of the 'McGurk effect' and reflections on audio–visual speech perception over the last 40 years', *Multisensory Research*, 31, pp. 7-18. doi: 10.1163/22134808-00002548.

Maree, C. (2013) *"Onē kotoba" ron* [*"Onē-language" studies*]. Tōkyo: Seidosha.

Maree, C. (2014) '"Henshinshitai desu" teroppu o tōshite gamen ni kakikomareru yokubō to aidentiti ['I want to be made-over': identity and desire written onto the screen in telop]', *Kotoba to Shakai*, 16, pp. 57-85.

Maree, C. (2015) 'The perils of paisley and weird manwomen. Queer crossings into primetime J-TV via telops', in Curran, B., Sato-Rossberg, N. and Tanabe, K. (eds.) *Multiple translation communities in contemporary Japan*. New York/London: Routledge Taylor & Francis Group, pp. 124-147.

Masuji, H. (2013) 'Terebi bangumi no moji jōhō ni okeru mojishu no sentaku: bangumi no janru to goyōronteki yōso ni chūmoku shite [The choice of character type as encountered in character news of television programmes: focusing on pragmatic elements and programme genre]', *Waseda Nihongo Kenkyū*, 22, pp. 24-38.

Matsukawa, R., Miyata, Y. and Ueda, S. (2009) 'Terebi nyūsu shichō e no jōhō jūfukusei no kōka: gankyū undō sokutei to kakunin tesuto ni yoru bunseki [Information redundancy effect on watching TV news: analysis of eye tracking data and examination of the contents]', *Library and Information Science*, 62, pp. 193-205.

MAXQDA (2019) *F. 3: code video files*. Copyright of MAXQDA, 1995-2019 (product of VERBI GmbH). Available at: https://www.maxqda.com/max12-tutorial/e-importing-av-files/j-audiovisual-files (Accessed: 27 October 2016).

McLoughlin, L.I. and Lertola, J. (2014) 'Audiovisual Translation in Second Language Acquisition. Integrating subtitling in the foreign-language curriculum', *The Interpreter and Translator Trainer*, 8(1), pp. 70-83. doi: 10.1080/1750399X.2014.908558.

McMeekin, A. (2015) 'Mango Premiere – Japanese videos for English speakers', *CALICO Journal*, 32(1), pp. 211-220. doi: 10.1558/calico.v32i1.25645.

Mishan, F. (2005) *Designing authenticity into language learning materials*. Bristol: Intellect Books.

Mitterer, H. and McQueen, J.M. (2009) 'Foreign subtitles help but native-language subtitles harm foreign speech perception', *PLoS ONE*, 4(11), pp. 1-5. doi: 10.1371/journal.pone.0007785.

Montero Perez, M., Peters, E. and Desmet, P. (2013) 'Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension', *ReCALL*, 26(1), pp. 21-43. doi: 10.1017/S0958344013000256.

Montero Perez, M., Peters, E. and Desmet, P. (2015) 'Enhancing vocabulary learning through captioned video: an eye-tracking study', *The Modern Language Journal*, 99(2), pp. 308-328. doi: 10.1111/modl.12215.

Moran, C. (2017) *Thematic analysis* [Summer school], *10th Qualitative Research Summer School*. Dublin City University: School of Nursing and Human Sciences. 27-28 April.

Moreno, R. (2006) 'Learning in high-tech and multimedia environments', *Current Directions in Psychological Science*, 15(2), pp. 63-67. doi: 10.1111/j.0963-7214.2006.00408.x.

Moreno, R. and Mayer, R. (2007) 'Interactive multimodal learning environments. Special issue on interactive learning environments: contemporary issues and trends', *Educational Psychology Review*, 19(3), pp. 309-326. doi: 10.1007/s10648-007-9047-2.

Neves, J. (2008) 'Training in subtitling for the d/Deaf and the hard-of hearing', in Díaz Cintas, J. (ed.) *The didactics of Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 171-190.

Nornes, A.M. (1999) 'For an abusive subtitling', *Film Quarterly*, 52(3), pp. 17-34. doi: 10.2307/1213822.

Nornes, A.M. (2007) *Cinema babel: translating global cinema*. Minneapolis/London: University of Minnesota Press.

O'Hagan, M. (2010) 'Japanese TV entertainment: framing humour with open caption telop', in Chiaro, D. (ed.) *Translation, humour and the media: Vol. 2 Translation and humour*. London/New York: Continuum International Publishing Group, pp. 70-88.

O'Hagan, M. (2013) 'Subtitling as a languaging test bed in the 21st century: from fansubs to impact captioning', *Subtitling: a collective approach*. Nottingham University, Nottingham, 12-13 July.

O'Hagan, M. and Sasamoto, R. (2016) 'Crazy Japanese subtitles? Shedding light on the impact of impact captions with a focus on research methodology', in Hansen-Schirra, S. and Grucza, S. (eds.) *Eyetracking and Applied Linguistics*. Berlin: Language Science Press, pp. 31-58.

Olk, B. and Kappas, A. (2011) 'Eye tracking as a tool for visual research', in Margolis, E. and Pauwels, L. (eds.) *The SAGE handbook of visual research methods*. London: SAGE, pp. 433-451.

Olsen, A. (2012) *The Tobii I-VT fixation filter: algorithm description*. Tobii AB, pp. 1-21. Available at: https://www.tobiipro.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/tobii-pro-i-vt-fixation-filter.pdf (Accessed: 6 December 2017).

O'Rourke, B. (2008) 'The other C in CMC: what alternative data sources can tell us about text-based synchronous computer mediated communication and language learning', *Computer Assisted Language Learning*, 21(3), pp. 227-251. doi: 10.1080/09588220802090253.

O'Rourke, B. (2012) 'Using eye-tracking to investigate gaze behaviour in synchronous computer-mediated communication for language learning', in Dooly, M. and O'Dowd, R. (eds.) *Researching online foreign language interaction and exchange: theories, methods and challenges: Vol. 3 Telecollaboration in education*. Bern: Peter Lang, pp. 305-341.

O'Rourke, B., Prendergast, C., Shi, L., Smith, B. and Stickler, U. (2016) 'Eyetracking in CALL – present and future', in Gimeno, A., Levy, M., Blin, F. and Barr, D. (eds.) *WorldCALL: sustainability and Computer-Assisted Language Learning*. London/New York: Bloomsbury Academic, pp. 285-298.

O'Rourke, B., Michel, M., Satar, M. and Balaman, U. (2017) *Research methods in CMC* [CMC SIG workshop]. University of Southampton. 23 August.

Ortega, L. (2011) 'Second Language Acquisition', in Simpson, J. (ed.) *The Routledge handbook of Applied Linguistics*. London/New York: Routledge Taylor & Francis Group, pp. 171-184.

Paivio, A. (2008) *Mental representations: a dual coding approach*. Oxford/New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195066661.003.0004.

Park, J.S. (2009) 'Regimenting languages on Korean television: subtitles and institutional authority', *Text & Talk*, 29(5), pp. 547-570. doi: 10.1515/TEXT.2009.029.

Pérez-González, L. (2012) 'Co-creational subtitling in the digital media: transformative and authorial practices', *International Journal of Cultural Studies*, 16(1), pp. 3-21. doi: 10.1177/1367877912459145.

Rayner, K. (1998) 'Eye movements in reading and information processing: 20 years of research', *Psychological Bulletin*, 124(3), pp. 372-422. doi: 10.1037//0033-2909.124.3.372.

Rayner, K. (2009) 'Eye movements and attention in reading, scene perception, and visual search', *The Quarterly Journal of Experimental Psychology*, 62(8), pp. 1457-1506. doi: 10.1080/17470210902816461.

Redmond, S. and Batty, C. (eds.) (2015) 'Eye-tracking the moving image', *Refractory: A Journal of Entertainment Media*, 25, 6 February. Available at: https://refractory-journal.com/volume-25-2015/ (Accessed: 30 November 2018).

Remael, A. (2004) 'A place for film dialogue analysis in subtitling courses', in Orero, P. (ed.) *Topics in Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 103-126.

Remael, A. (2010) 'Audiovisual Translation', in Gambier, Y. and Van Doorslaer, L. (eds.) *Handbook of Translation Studies: Vol. 1*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 12-17.

Riazi, A.M. and Candlin, C.N. (2014) ‘Mixed-methods research in language teaching and learning: opportunities, issues and challenges’, *Language Teaching*, 47(2), pp. 135-173. doi: 10.1017/S0261444813000505.

Roberts, L. and Siyanova-Chanturia, A. (2013) ‘Using eye-tracking to investigate topics in L2 acquisition and L2 processing’, *Studies in Second Language Acquisition*, 35(2), pp. 213-235. doi: 10.1017/S0272263112000861.

Robinson, P. (2003) ‘Attention and memory during SLA’, in Doughty, C.J. and Long, M.H. (eds.) *The handbook of Second Language Acquisition*. Malden: Blackwell Publishing, pp. 631-678.

Robinson, J., Stadler, J. and Rassell, A. (2015) ‘Sound and sight: an exploratory look at Saving Private Ryan through the eye tracking lens’, *Refractory: A Journal of Entertainment Media*, 25. Available at: https://refractory-journal.com/robinson-stadler-rassell/ (Accessed: 3 December 2018).

Rost, M. (2005) ‘L2 listening’, in Hinkel, E. (ed.) *Handbook of research in second language teaching and learning*. New York/London: Routledge Taylor & Francis Group, pp. 503-528.

Sakamoto, M. (2009a) ‘Hanransuru jimaku bangumi no kōzai [The good and bad points of the plethora of subtitled programmes]’, *Subete o utagae!! MAMO’s Site* (copyright of Mamoru Sakamoto, 2002-2015). Available at: http://www.maroon.dti.ne.jp/mamos/tv/jimaku.html (Accessed: 13 December 2014).

Sakamoto, M. (2009b) ‘Teroppu seitaigaku (terebi jimaku no kenkyū) [The ecology of telop (research on television subtitles)]’, *Subete o utagae!! MAMO’s Site* (copyright of Mamoru Sakamoto, 2002-2015). Available at: http://www.maroon.dti.ne.jp/mamos/lecture/jimaku99.html (Accessed: 9 December 2014).

Sakellariou, P. (2012) ‘The semiotic status of interlingual subtitling’, *Meta*, 57(3), pp. 677-693. doi: 10.7202/1017086ar.

Saldanha, G. and O'Brien, S. (2014) *Research methodologies in Translation Studies*. London/New York: Routledge Taylor & Francis Group.

Sasamoto, R. (2014) ‘Impact caption as a highlighting device: attempts at viewer manipulation on TV’, *Discourse, Context and Media*, pp. 1-10. doi: 10.1016/j.dcm.2014.03.003i.

Sasamoto, R. and Doherty, S. (2016) ‘Towards the optimal use of impact captions on TV programmes’, in O’Hagan, M. and Zhang, Q. (eds.) *Conflict and communication: a changing Asia in a globalising world*. Hauppauge: Nova Science Publishers, pp. 210-247.

Sasamoto, R. and O’Hagan, M. (2014) ‘Relevance, style and multimodality: typographical features as stylistic devices’, *Interpreting for relevance: discourse and translation conference 7*. Warsaw, 29-30 September.

Sasamoto, R., O’Hagan, M. and Doherty, S. (2016) ‘Telop, affect, and media design: a multimodal analysis of a Japanese TV program’, *Television & New Media*, pp. 1-14. doi: 10.1177/1527476416677099.

Shiota, E. (2003) ‘Kanrensei riron to teroppu rikai [Relevance Theory and understanding telops]’, *The Graduate School Review of the English Language and Literature*, 31, pp. 63-91.

Shitara, K. (2005) ‘Narēshon no moji teroppu [The terop of narration]’, *Mukogawa Women’s University Annual Report of Research Institute for Linguistic Cultural Studies*, 17, pp. 13-28.

Shitara, K. (2006) ‘Terebi no tōkukōnā o yomu: dōitsu no hatsuwa o tomonawanai moji teroppu no jittai [There are differences between ‘the telop’ and the actual audio on Japanese talk shows]’, *Mukogawa Women’s University Annual Report of Research Institute for Linguistic Cultural Studies*, 18, pp. 37-61.

Shitara, K. (2008) ‘Terebi shichō taido to moji teroppu: gakusei to seijin no taihi [The way how to watch TV with the telop: the distinctions between students and adults]’, *Mukogawa Women’s University Annual Report of Research Institute for Linguistic Cultural Studies*, 20, pp. 29-54.

Shitara, K. (2009) ‘Moji teroppu to shichōsha ishiki [Telop and viewer awareness]’, pp. 1-6. Available at: http://www.hituzi.co.jp/kenkyukai/happyo-rejume/shitarakaoru05Dec09.pdf (Accessed: 15 December 2014).

Shitara, K. (2011) ‘NHK baraeti bangumi ni miru moji teroppu no hensen: terebi ni okeru hyōki jittai to kinō no bunka [The change of TELOPs on NHK variety shows: the actual conditions and two functions of writing information]’, *The Bulletin of Mukogawa Women’s University. Humanities and Social Science*, 59, pp. 1-9. doi: 10.14993/00000052.

Shitara, K. (2012) ‘Eizō media to kodomotachi: moji teroppu to chūgakusei no gengo hyōgenryoku [Image-based media and children: telops and the verbal expressive skills of junior high school students]’, *Nihongo Nihonbungaku Ronsō*, 7, pp. 123-138. doi: 10.14993/00000541.

Sikkema, E.C. (2015) *Minutes of eye-tracking training by Acuity ETS 10 December 2015*, Dublin.

Sikkema, E.C. (2016) *Minutes of meeting with Tobii AB 8 January 2016*, Tōkyo.

Smith, B. (2012) ‘Eye tracking as a measure of noticing: a study of explicit recasts in SCMC’, *Language Learning & Technology*, 16(3), pp. 53-81. doi: 10125/44300.

Sperber, D. and Wilson, D. (1995) *Relevance: communication and cognition*. 2nd edn. Oxford/Cambridge: Blackwell Publishing.

Spinner, P., Gass, S.M. and Behney, J. (2013) ‘Ecological validity in eye-tracking: an empirical study’, *Studies in Second Language Acquisition*, 35(2), pp. 389-415. doi: 10.1017/S0272263112000927.

Stickler, U. and Shi, L. (2015) ‘Eye movements of online Chinese learners’, *CALICO Journal*, 32(1), pp. 52-81. doi: 10.1558/calico.v32i1.25964.

Stickler, U. and Shi, L. (2017) ‘Eyetracking methodology in SCMC: a tool for empowering learning and teaching’, *ReCALL*, 29(2), pp. 160-177. doi: 10.1017/S0958344017000040.

Stickler, U., Smith, B. and Shi, L. (2016) 'Using eye-tracking technology to explore online learner interactions', in Caws, C. and Hamel, M. (eds.) *Language-learner computer interactions: theory, methodology and CALL applications*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 163-186.

Sullivan, B. (2016) *Eye tracking data quality, validation, and fixation classification* [Webinar], *Tobii Pro webinar: tools and tips to assess eye tracking data quality*. Tobii AB. 11 October.

Suto, H. (2008) 'Teroppu ga shichōsha no kaishaku ni ataeru eikyō [Influences of telops on the receivers' interpretation]', *Informatics*, 1(2), pp. 13-20.

Tama-chan (2014) '*Hannin wa Yasu! Kinō no "kajō teroppu jiken" de Nittere ni hinan sattō* [*The culprit is Yasu! Nippon TV under heavy criticism over yesterday's "incident with redundant telops"*]', *Byōkan SUNDAY* (managed by MEDIANO Co.,Ltd.), 15 November. Available at: http://www.yukawanet.com/archives/4776598.html (Accessed: 9 December 2014).

Taylor, C.J. (2003) 'Multimodal transcription in the analysis, translation and subtitling of Italian films', *The Translator*, 9(2), pp. 191-205. doi: 10.1080/13556509.2003.10799153.

The Japan Foundation (2012) *JF Standard for Japanese-language education 2010*. 2nd edn. Available at: https://jfstandard.jp/pdf/jfs2010_all_en.pdf (Accessed: 4 June 2016).

THE TETSUWAN DASH (2014), Nippon Television Network Corporation, 22 June.

Tiippana, K. (2014) 'What is the McGurk effect?', *Frontiers in Psychology*, 5(725), pp. 1-3. doi: 10.3389/fpsyg.2014.00725.

Tobii AB (2013a) *Webinar – Tobii Studio and dynamic stimuli*. Available at: https://www.youtube.com/watch?reload=9&v=vLnCevz6QgA (Accessed: 24 April 2015).

Tobii AB (2013b) *Tobii free webinar: introduction to eye tracking research*. Available at: https://www.youtube.com/watch?reload=9&v=VLzSxygcBF8&feature=youtu.be (Accessed: 3 October 2014).

Tobii AB (2014a) *User's manual: Tobii X2-60 eye-tracker*. Version 1.0.3. Available at: https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-x2-60-eye-tracker-user-manual.pdf/?v=1.0.3 (Accessed: 14 December 2015).

Tobii AB (2014b) *User manual: Tobii Studio version 3.3.0*. Version 3.3.0. Available at: https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf (Accessed: 3 November 2016).

Tobii AB (2018) *Eye tracker accuracy and precision*. Available at: https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/what-affects-the-accuracy-and-precision-of-an-eye-tracker/ (Accessed: 28 November 2016).

Tomlinson, B. (2012) 'Materials development for language learning and teaching', *Language Teaching*, 45(2), pp. 143-179. doi: 10.1017/S0261444811000528.

Vanderplank, R. (1988) ‘The value of teletext sub-titles in language learning’, *ELT Journal*, 42(4), pp. 272-281. doi: 10.1093/elt/42.4.272.

Vanderplank, R. (2010) ‘Déjà vu? A decade of research on language laboratories, television and video in language learning’, *Language Teaching*, 43(1), pp. 1-37. doi: 10.1017/S0261444809990267.

Vanderplank, R. (2015) ‘Thirty years of research into captions/same language subtitles and second/foreign language learning: distinguishing between ‘effects of’ subtitles and ‘effects with’ subtitles for future research’, in Gambier, Y., Caimi, A. and Mariotti, C. (eds.) *Subtitles and language learning: principles, strategies and practical experiences*. Bern: Peter Lang, pp. 19-40.

Vanderplank, R. (2016) *Captioned media in foreign language learning and teaching: subtitles for the deaf and hard-of-hearing as tools for language learning*. Basingstoke: Palgrave Macmillan.

Van Leeuwen, T. (2011) ‘Multimodality’, in Simpson, J. (ed.) *The Routledge handbook of Applied Linguistics*. London/New York: Routledge Taylor & Francis Group, pp. 668-682.

Williams, H. and Thorne, D. (2000) ‘The value of teletext subtitling as a medium for language learning’, *System*, 28, pp. 217-228. doi: 10.1016/S0346-251X(00)00008-7.

Winke, P.M, Gass, S. and Sydorenko, T. (2013) ‘Factors influencing the use of captions by foreign language learners: an eye-tracking study’, *The Modern Language Journal*, 97(1), pp. 254-275. doi: 10.1111/j.1540-4781.2012.01432.x.

Winke, P.M., Godfroid, A. and Gass, S.M. (2013) ‘Introduction to the special issue. Eye-movement recordings in second language research’, *Studies in Second Language Acquisition*, 35(2), pp. 205-212. doi: 10.1017/S027226311200085X.

Yoshizawa, H. (2015) ‘Dorama ni mo teroppu ga!? TV gyōkai no kurēmu taisaku ga funpan mono! [Even drama series have telops!? The measures taken by the TV industry to counteract complaints are downright absurd]’, *TOCANA* (copyright of CYZO Inc.), 1 June. Available at: http://tocana.jp/2015/06/post_6484_entry.html (Accessed: 1 June 2015).

Zabalbeascoa, P. (2008) ‘The nature of the audiovisual text and its parameters’, in Díaz Cintas, J. (ed.) *The didactics of Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 21-37.

Zarei, A.A. and Rashvand, Z. (2011) ‘The effect of interlingual and intralingual, verbatim and nonverbatim subtitles on L2 vocabulary comprehension and production’, *Journal of Language Teaching and Research*, 2(3), pp. 618-625. doi: 10.4304/jltr.2.3.618-625.

APPENDICES

Appendix A. Pre-task questionnaire

Appendix A.1 English version

Appendix A.1.1 Japanese language learners at Dublin City University

General information

Please answer the following questions and tick the box(es) that apply to you the most.

Q1. What is the name of your studies at Dublin City University?

Q2. Which academic year are you enrolled in?

❒ BA degree - 1st year	❒ MA degree - 1st year
❒ BA degree - 2nd year	❒ MA degree - 2nd year
❒ BA degree - 4th year	❒ Other *(specify)*:

Q3. Have you been to Japan?

❒ Yes ❒ No *(please go to Q5)*

Q4. For how long and for what purpose were you in Japan?

Q5. Did you study the Japanese language before entering university?

❒ Yes ❒ No *(please go to Q7)*

Q6. For what purpose or how did you study the Japanese language before entering university? *(more than one answer possible)*

❒ During my transition year	❒ I did a summer course in Japanese
❒ For my leaving cert	❒ I went to Japanese language school
❒ I did self-study	❒ Other *(specify)*:
❒ Through Japanese relatives	

Q7. Is English your mother tongue?

❒ Yes ❒ No *(specify)*:

Q8. Do you have experience with learning other foreign languages besides Japanese?

❒ Yes ❒ No *(please go to Q10)*

Q9. What other foreign languages would these be? *(more than one answer possible)*

❒ French	❒ Italian
❒ German	❒ Chinese
❒ Spanish	❒ Other (*specify*):

Q10. How would you rank a. your reading comprehension and b. your listening comprehension in Japanese?

	Q10a. (reading comprehension)	**Q10b. (listening comprehension)**
Excellent	❒	❒
Very good	❒	❒
Good	❒	❒
Reasonable	❒	❒
Poor	❒	❒

Q11. Have you done a Japanese language proficiency test (e.g. J-CAT, JLPT) before?

❒ Yes ❒ No *(please go to Q13)*

Q12. What was the name of that Japanese language proficiency test and what was your score? (*more than one answer possible if you have done multiple tests*)

Q13. Do you use the Japanese language outside of class?

❒ Yes ❒ No

Q14. How do you practice kanji outside of class? *(more than one answer possible)*

❒ Flashcards	❒ Kanji apps and/or games
❒ Repetitive writing	❒ Memorisation programmes (e.g. *Anki*)
❒ I read texts in Japanese	❒ I do not study kanji outside of class
(specify text types):	❒ Other *(specify)*:

Q15. How would you rank a. your reading comprehension and
b. your listening comprehension in your mother tongue?

	Q15a. (reading comprehension)	**Q15b. (listening comprehension)**
Excellent	❒	❒
Very good	❒	❒
Good	❒	❒
Reasonable	❒	❒
Poor	❒	❒

Q16. Do you wear contacts or glasses?

❒ Yes ❒ No

Q17. Age

❒ 18 years old	❒ 21 years old
❒ 19 years old	❒ 22 years old
❒ 20 years old	❒ Other *(specify)*:

Q18. Gender

❒ Female ❒ Male ❒ Prefer not to say

That was the final question of the questionnaire.

Thank you for your cooperation!

Appendix A.1.2 Japanese language learners on exchange in Japan

General information

Please answer the following questions and tick the box(es) that apply to you the most.

Q1. What is the name of your studies at Dublin City University?

Q2. What is the name of the Japanese university you are currently studying at?

Q3. Have you been to Japan before going on your exchange year?

❒ Yes ❒ No *(please go to Q5)*

Q4. For how long and for what purpose were you in Japan before going on exchange?

Q5. Did you study the Japanese language before entering university?

❒ Yes ❒ No *(please go to Q7)*

Q6. For what purpose or how did you study the Japanese language before entering university? *(more than one answer possible)*

❒ During my transition year	❒ I did a summer course in Japanese
❒ For my leaving cert	❒ I went to Japanese language school
❒ I did self-study	❒ Other *(specify)*:
❒ Through Japanese relatives	

Q7. Is English your mother tongue?

❒ Yes ❒ No *(specify)*:

Q8. Do you have experience with learning other foreign languages besides Japanese?

❒ Yes ❒ No *(please go to Q10)*

Q9. What other foreign languages would these be? *(more than one answer possible)*

❒ French	❒ Italian
❒ German	❒ Chinese
❒ Spanish	❒ Other *(specify)*:

Q10. How would you rank a. your reading comprehension and
b. your listening comprehension in Japanese?

	Q10a. (reading comprehension)	**Q10b. (listening comprehension)**
Excellent	❒	❒
Very good	❒	❒
Good	❒	❒
Reasonable	❒	❒
Poor	❒	❒

Q11. Have you done a Japanese language proficiency test (e.g. J-CAT, JLPT) before?

❒ Yes ❒ No *(please go to Q13)*

Q12. What was the name of that Japanese language proficiency test and what was your score? *(more than one answer possible if you have done multiple tests)*

Q13. How do you practice kanji outside of class? *(more than one answer possible)*

❒ Flashcards	❒ Kanji apps and/or games
❒ Repetitive writing	❒ Memorisation programmes (e.g. *Anki*)
❒ I read texts in Japanese *(specify text types):*	❒ I do not practice kanji outside of class
	❒ Other *(specify)*:

Q14. How would you rank a. your reading comprehension and
b. your listening comprehension in your mother tongue?

	Q14a. (reading comprehension)	**Q14b. (listening comprehension)**
Excellent	❒	❒
Very good	❒	❒
Good	❒	❒
Reasonable	❒	❒
Poor	❒	❒

Q15. Do you wear contacts or glasses?

❒ Yes ❒ No

Q16. Age

❒ 18 years old	❒ 21 years old
❒ 19 years old	❒ 22 years old
❒ 20 years old	❒ Other *(specify)*:

Q17. Gender

❒ Female ❒ Male ❒ Prefer not to say

That was the final question of the questionnaire.

Thank you for your cooperation!

Appendix A.2 Japanese version

個人情報

質問に答え、最も当てはまる箇所に X をつけてください。

質問1. 大学名は何ですか。

質問2. 所属（学部・学科名）は何ですか。

質問3. 何年生ですか。

❒ 学士課程 - 1年	❒ 修士課程 - 1年
❒ 学士課程 - 2年	❒ 修士課程 - 2年
❒ 学士課程 - 3年	❒ その他：
❒ 学士課程 - 4年	

質問4. 母国語は日本語ですか。

❒ はい　　❒ いいえ *（母国語は）*：

質問5. 外国語を勉強したことがありますか。

❒ はい　　❒ いいえ *（質問7に進んでください）*

質問6. 勉強したことがある外国語はどれですか? *（複数の回答が可能）*

❒ 英語	❒ スペイン語
❒ ドイツ語	❒ 中国語
❒ フランス語	❒ その他：

質問7a. 自分の母語読解力をどう評価しますか。

b. 自分の母語聴解力をどう評価しますか。

	質問7a.（読解力）	**質問7b.（聴解力）**
完璧	❒	❒
とても良い	❒	❒
良い	❒	❒
まあまあ	❒	❒
良くない	❒	❒

質問8. めがね、または、コンタクト･レンズを着用していますか。

❒ はい　　❒ いいえ

質問9. 年齢

❒ 18歳	❒ 21歳
❒ 19歳	❒ 22歳
❒ 20歳	❒ その他：

質問10. 性別

❒ 女性　　❒ 男性　　❒ 無回答

これで以上です。

ご協力ありがとうございます！

Appendix B. Plain language statement

Appendix B.1 English version

I. Introduction to the Research Study

Reception of Japanese variety shows: A comparative study of multimodal media comprehension between native speakers and language learners of Japanese

Principal investigator: Eline Sikkema (eline.sikkema2@mail.dcu.ie)
School of Applied Language and Intercultural Studies (SALIS)

Supervisors: Dr Ryoko Sasamoto (ryoko.sasamoto@dcu.ie)
Dr Minako O'Hagan (minako.ohagan@dcu.ie)
School of Applied Language and Intercultural Studies (SALIS)

II. Details of what involvement in the Research Study will require

For this Research Study, participants will be asked to attend one session that should last no more than 1 hour. Before starting the experiment, the participant will read the Plain Language Statement and sign the Informed Consent Form. During the experiment the participant will look at a short video via an eye-tracker and fill in a questionnaire afterwards regarding some general information about the participant, the video and its contents.

III. Potential risks to participants from involvement in the Research Study (if greater than that encountered in everyday life)

There are no risks to participants from involvement in the Research Study.

IV. Benefits (direct or indirect) to participants from involvement in the Research Study

- The participants' involvement will contribute to the research field as little work has been done on this topic to date.
- The results of this Research Study may in particular prove to be beneficial to language learners of Japanese.

V. Advice as to arrangements to be made to protect confidentiality of data, including that confidentiality of information provided is subject to legal limitations

The anonymity of the participants will be protected at all times. Participants will be given an identifier such as "Participant A" and no link will ever be made to their real identity. The data collated will be used only by Eline Sikkema, Dr Ryoko Sasamoto and Dr Minako O'Hagan and will not be given to anybody else.

VI. Advice as to whether or not data is to be destroyed after a minimum period

The data will be stored in a secure location only at DCU. The data will be destroyed within five years of its acquisition.

VII. Statement that involvement in the Research Study is voluntary

Involvement in this study is voluntary. You may withdraw from the Research Study at any point. There will be no penalty for withdrawing before all stages of the

Research Study have been completed. Involvement/non-involvement in this study will not affect your relationship with DCU and/or the university you are currently affiliated with in any way.

VIII. Any other relevant information
If participants have concerns about this study and wish to contact an independent person, please contact:

The Secretary, Dublin City University Research Ethics Committee, c/o Research and Innovation Support, Dublin City University, Dublin 9. Tel 01-7008000

Appendix B.2 Japanese version

I. **本研究プロジェクト名**
日本バラエティ番組の受容：日本語母語話者と日本語学習者による視聴内容の理解の比較研究

研究責任者: エリーネ・シッケマ (eline.sikkema2@mail.dcu.ie)
School of Applied Language and Intercultural Studies (SALIS)

指導教官: 笹本涼子博士 (ryoko.sasamoto@dcu.ie)
オヘイガン統子博士 (minako.ohagan@dcu.ie)
School of Applied Language and Intercultural Studies (SALIS)

II. **参加者の方に求めらること**
本研究プロジェクトの参加者の方に求められる内容は以下の通りです：
実験は約 1 時間ほどで終了します。実験を始める前に、参加者は「実験概要」を読み、「告知に基づく合意」にサインします。実験中、眼球運動の測定のため、参加者はビデオを短時間見ることになります。ビデオを見終えた後、参加者ご本人、そしてビデオおよびその内容に関するアンケートに記入していただきます。

III. **参加者の方への日常の範囲を超えたリスクの有無**
本実験では、日常の範囲を超えたリスクはありません。

IV. **参加者の方への利点（直接的・間接的）**
– 本実験の参加によって新たな研究分野へ貢献することになります。
– 本実験の結果は、特に日本語学習者にとって有益であると想定されます。

V. **法的制限内におけるデータ守秘・保存等の説明**
参加者の個人情報は常に守秘されます。各個人からのデータはすべて匿名扱いとし、「参加者 A」のように表記され、個人の特定に使われることはありません。回答結果はプロジェクトメンバー（エリーネ・シッケマ、笹本涼子博士、オヘイガン統子博士）のみによって使用され、その他の人に提供されることはありません。

VI. **データ保管期間**
本実験結果はダブリン・シティ大学内で厳重に保管され、5 年以内に破棄されます。

VII. **任意参加**
本実験への参加は強制ではなく、個人の意思でいずれの時点でも任意に参加を取り止めることができます。参加取り止めに対するなんらの罰則もありません。また、本実験への参

加の有無により、参加者と参加者の所属大学との関係がなんらの影響も受けることはありません。

VIII. その他

不明な点がある場合には下記までご連絡ください:

The Secretary, Dublin City University Research Ethics Committee, c/o Research and Innovation Support, Dublin City University, Dublin 9. Tel 01-7008000

Appendix C. Informed consent form

Appendix C.1 English version

I. Research Study Title

Reception of Japanese variety shows: A comparative study of multimodal media comprehension between native speakers and language learners of Japanese

Principal investigator: Eline Sikkema (eline.sikkema2@mail.dcu.ie)
School of Applied Language and Intercultural Studies (SALIS)

Supervisors: Dr Ryoko Sasamoto (ryoko.sasamoto@dcu.ie)
Dr Minako O'Hagan (minako.ohagan@dcu.ie)
School of Applied Language and Intercultural Studies (SALIS)

II. Clarification of the purpose of the research

The purpose of the research is to gain insights into the viewing behaviour of native speakers and language learners of Japanese while watching a Japanese variety show.

III. Confirmation of particular requirements as highlighted in the Plain Language Statement

I will be asked to attend one session that should last no more than 1 hour. Before starting the experiment, I will first read the Plain Language Statement and sign the Informed Consent Form. During the experiment I will look at a short video via an eye-tracker and fill in a questionnaire afterwards regarding some general information about myself, the video and its contents.

Participant – please complete the following (Circle Yes or No for each question)

I have read the Plain Language Statement (or had it read to me)	*Yes/No*
I understand the information provided	*Yes/No*
I have had an opportunity to ask questions and discuss this study	*Yes/No*
I have received satisfactory answers to all my questions	*Yes/No*

IV. Confirmation that involvement in the Research Study is voluntary

Involvement in this Research Study is voluntary. I may withdraw from the Research Study at any point. There will be no penalty for withdrawing before all stages of the Research Study have been completed. Involvement/non-involvement in this study will not affect my relationship with DCU and/or the university I am currently affiliated with in any way.

V. Advice as to arrangements to be made to protect confidentiality of data, including that confidentiality of information provided is subject to legal limitations

My anonymity will be protected at all times. I will be given an identifier such as "Participant A" and no link will ever be made to my real identity. The data collated will be used only by Eline Sikkema, Dr Ryoko Sasamoto and Dr Minako O'Hagan and will not be given to anybody else.

VI. Any other relevant information

If participants have concerns about this study and wish to contact an independent person, please contact:

The Secretary, Dublin City University Research Ethics Committee, c/o Research and Innovation Support, Dublin City University, Dublin 9. Tel 01-7008000

VII. Signature:

I have read and understood the information in this form. My questions and concerns have been answered by the researchers, and I have a copy of this consent form. Therefore, I consent to take part in this research project

Participants Signature: _____

Name in Block Capitals: _____

Witness: _____

Date: _____

Appendix C.2 Japanese version

I. 本研究プロジェクト名

日本バラエティ番組の受容：日本語母語話者と日本語学習者による視聴内容の理解の比較研究

研究責任者: エリーネ・シッケマ (eline.sikkema2@mail.dcu.ie)
School of Applied Language and Intercultural Studies (SALIS)

指導教官: 笹本涼子博士 (ryoko.sasamoto@dcu.ie)
オヘイガン統子博士 (minako.ohagan@dcu.ie)
School of Applied Language and Intercultural Studies (SALIS)

II. 本研究プロジェクトの目的

本研究プロジェクトの目的は日本語母語話者と日本語学習者による日本バラエティ番組の視聴の傾向を分析することです。

III. 「実験概要」で示された参加者の方の役割

約 1 時間ほどのセッションに参加します。実験を始める前に、「実験概要」を読み、「告知に基づく合意」にサインします。実験中、眼球運動の測定のため、ビデオを短時間見ます。ビデオを見終えた後、私自身、そしてビデオおよびその内容に関するアンケートに記入します。

参加者の方へ - 次の質問に答えてください（適切なほうに丸を付けてください）

「実験概要」を読みました	*はい / いいえ*
内容を理解しました	*はい / いいえ*
実験内容について質問したり話し合う機会がありました	*はい / いいえ*
質問に対して満足な説明がなされました	*はい / いいえ*

IV. 任意参加

本実験への参加は強制ではなく、自分の意思でいずれの時点でも任意に参加を取り止めることができます。参加取り止めに対するなんらの罰則もありません。また、本実験への参加の有無により、私と私の所属大学との関係がなんらの影響も受けることはありません。

V. 法的制限内におけるデータ守秘・保存等の説明

私の個人情報は常に守秘されます。各個人からのデータはすべて匿名扱いとし、自分のデータは「参加者 A」のように表記され、私個人の特定に使われることはありません。回答結果はプロジェクトメンバー（エリーネ・シッケマ、笹本涼子博士、オヘイガン統子博士）のみによって使用され、その他の人に提供されることはありません。

VI. その他

不明な点がある場合には下記までご連絡ください:

The Secretary, Dublin City University Research Ethics Committee, c/o Research and Innovation Support, Dublin City University, Dublin 9. Tel 01-7008000

VII. サイン:

本書を読み、内容を理解しました。質問や疑問に対してすべて満足のいく説明が行われ、本書のコピーも受け取りました。したがって、この実験への参加に合意します。

参加者　サイン: _____

参加者　氏名（活字体）: _____

立会人: _____

年月日: _____

Appendix D. Script for main experiment sessions

Appendix D.1 English version

An eye-tracker is a tool that tracks and records eye-movements through the use of infrared light. It does this while a participant watches something or does an activity. In this way, an eye-tracker can show a researcher afterwards where a participant has looked at. However, before an eye-tracker can do this accurately it first needs to get used to a participant's eyes. This happens through the so-called calibration process. If we were to skip this calibration process we might run the risk of collecting data that has a low accuracy rate, risking the validity of the results.

An eye-tracker's calibration process requires for a participant to follow an orange ball that is displayed on the screen. The participant does this by only moving his/her eyes which means that the participant should not move the head too much. Please take a seat behind the eye-tracker and look at the screen. When I initiate the calibration process you will see an orange ball on the screen with a black dot at its centre. This dot functions as an aid to the participant for fixating the eyes on the orange ball. During the calibration process the orange ball will not stay in one place, but will move around the screen so that the eye-tracker can check whether it can trace your eyes over the whole surface of the screen. The job of the participant is to follow this ball by moving his/her eyes. So if the ball goes left, you will try and follow it by focusing your gaze on the black dot at its centre. This will take approximately 20-30 seconds. We will start the actual experiment after this has been completed successfully.

Appendix D.2 Japanese version

アイ・トラッカーという機械は赤外線を通して眼球運動を自動的に計る技術で、このように収集してきたデータを記録します。参加者はある活動をしたり、または、あるビデオを観ったりしながら、アイ・トラッカーが眼球運動を自動的に計ります。このように、研究者はアイ・トラッカーに参加者が何を見たのか、または、何を見つめたのかを示していただきます。しかし、眼球運動を自動的に計る前に、アイ・トラッカーが参加者の目に慣れなければなりません。これはキャリブレーションのおかげにできます。キャリブレーションせずに実験を行ったら、収集したデータの正確さも妥当性も微妙になってしまいます。

キャリブレーションする時、参加者は画面上に表示されたオレンジボールを目で追います。具体的には、参加者は頭をあまり動かないよう、目だけで追うのです。アイ・トラッカーの前にお掛けになり、画面をごらんください。キャリブレーションを開始すると、黒い点を真ん中にあるオレンジボールが画面上に出てきます。この黒い点は参加者が目でオレンジボールを追うのを支援するために表示されています。キャリブレーション中に、アイ・トラッカーは参加者が画面のどこからにも観ても眼球運動を自動的に計れるかどうか確認するため、オレンジボールが一か所にとどまらず画面上に移動します。キャリブレーション中、参加者の役割はこのオレンジボールを目で追うということです。例えば、オレンジボールが左に移動したら、目で黒い点を見ながら追います。キャリブレーションは 20-30 秒ぐらいかかります。首尾よく達成したら、本実験を始めます。

Appendix E. Post-task questionnaire

Appendix E.1 English version

Appendix E.1.1 Part 1

Contents of the experiment video

Please answer the following questions and tick the box(es) that apply to you the most.

Q1. How well did you feel you understood the experiment video?

❒ I understood 0%–20%	❒ I understood 60%–80%
❒ I understood 20%–40%	❒ I understood 80%–100%
❒ I understood 40%–60%	❒ I do not know

Q2-6 are about the first scenario of 「もしもの時に役立つホンマでっか！？」.

Q2. What does the expert guest recommend if you want to lower your heating costs?

❒ To live in a room that is at the centre of the middle floor of an apartment block
❒ To live in a detached house
❒ To live at the corner of the highest floor of an apartment block

Q3. How does he explain this?

❒ The costs for a detached house are 2/3 of the costs for the room at the corner of the highest floor of an apartment block
❒ The costs for a room at the centre of the middle floor of an apartment block are 1/2 of the costs for a detached house
❒ The costs for the room at the corner of the highest floor of an apartment block are 2/3 of the costs for a room at the centre of the middle floor of an apartment block

Q4. How come the costs are cheaper if you live in the recommended accommodation?

Q5. Generally speaking, are the costs cheaper for airconditioning (cooling) or heating?

❒ Airconditioning (cooling) ❒ Heating

Q6. What is the explanation behind this?

❒ People commonly use the heater for twice as many months as the airconditioner (cooling)
❒ People commonly use the airconditioner (cooling) longer than the heater
❒ The difference between the temperature outside and inside the room is higher when using the airconditioner (cooling)

Q7-9 are about the second scenario of 「もしもの時に役立つホンマでっか！？」.

Q7. What are the <u>two</u> most important things someone needs to do when lightning strikes while there are no places to hide?

1.
2.

Q8. Why is that?

❒ There are only known instances of people getting struck by lightning when they were standing
❒ Lightning chooses high places before striking, so people who are standing function as lightning rods
❒ People who stand run the risk of cardiac arrest when struck by lightning
❒ All of the above

Q9. What does the expert guest say about laying down on your belly when lightning strikes?

Q10-11 are about the third scenario of 「もしもの時に役立つホンマでっか！？」.

Q10. Why is it best to say 「あ～そういう風に受け取られましたか？」 when trying to get out of the situation that is described in the third scenario?

Q11. Why does Yashima Norito pull the facial expression as in figure 1?

- ❒ He makes fun of the expert guest
- ❒ He is surpised by what the expert guest is saying
- ❒ He illustrates that someone's facial expression reveals someone's true feelings behind what s/he is saying
- ❒ He pulls this facial expression because he finds it hard to believe that the expert guest has been involved in fraud

Figure 1

About the experiment video

Please answer the following questions and tick the box(es) that apply to you the most.

Q12. How well do you feel you understand the experiment video now that you have answered the questions about 「もしもの時に役立つホンマでっか！？」?

- ❒ I understand 0%–20%
- ❒ I understand 20%–40%
- ❒ I understand 40%–60%
- ❒ I understand 60%–80%
- ❒ I understand 80%–100%
- ❒ I do not know

Q13. Did you feel the drawings in the experiment video helped you with understanding its contents?

❒ Yes ❒ No ❒ I do not know

Q14. How certain are you of your answers on the questions regarding 「もしもの時に役立つホンマでっか！？」?

- ❒ Very certain
- ❒ Quite certain
- ❒ Not so certain
- ❒ I guessed most of my answers
- ❒ I do not know

Q15. Did you know the Japanese variety show used in the experiment video?

❒ Yes ❒ No

Q16. Have you watched this variety show before?

- ❒ I watch it regularly
- ❒ I watch it sometimes
- ❒ This was my first time to watch it
- ❒ I do not know

Text on screen in the experiment video

Please answer the following questions and tick the box(es) that apply to you the most.

Q17. Did you read the text on screen during the experiment video?

- ❒ I read everything
- ❒ I read the majority of the text on screen
- ❒ I read less than half of the text on screen
- ❒ I only read the text in the upper left and right corners
- ❒ I only read the text at the bottom of the screen
- ❒ I did not read the text on screen

Q18. What did you think of the difficulty level of the text on screen?

- ❒ I understood everything
- ❒ I understood most of it
- ❒ I understood half of it
- ❒ I understood only a few lines of the text on screen
- ❒ I did not understand the text on screen

Q19. What helped you with text you did not understand? *(more than one answer possible)*

- ❒ Context (e.g. hiragana and/or katakana)
- ❒ Radicals of kanji unknown to me
- ❒ Phonetic elements of kanji unknown to me
- ❒ Colours and/or font(s) of the text
- ❒ Kanji known to me
- ❒ The spoken dialogue
- ❒ Non-verbal clues (e.g. imagery, body language, facial expressions)
- ❒ Other (*specify*):

Q20. What kind of effects did you notice were used for the text on screen?

Q21. Which colours were used for the the text on screen?

Q22a. Were there any colours or colour combinations you found difficult to read?

❒ Yes ❒ No *(please go to Q23)*

b. Please specify.

Q23. Were different fonts used for the text on screen?

❒ Yes ❒ No ❒ I cannot remember

Q24. What did you think of the legibility of the text on screen?

- ❒ The design of the text was legible
- ❒ The design of the majority of the text was legible
- ❒ The design of the majority of the text was not legible
- ❒ The design of the text was not legible
- ❒ Other (*specify*):

Q25a. Did you think the design of the text on screen suited the variety show?

❒ Yes ❒ No

b. Please explain you answer.

Q26. Did you feel the questionnaire allowed you to say everything you wanted to say? Is there anything you think might be relevant to mention in regard to the experiment video or Japanese text on screen?

That was the final question of the first part of the questionnaire.

Thank you for your cooperation!

Appendix E.1.2 Part 2

Media exposure to English TV programmes

Please answer the following questions and tick the box(es) that apply to you the most.

Q1. How many hours on average per week do you watch English TV programmes?

- ❒ 01-10 hours
- ❒ 11-19 hours
- ❒ 20-29 hours
- ❒ More than 30 hours
- ❒ I do not watch English TV programmes *(please go to Q7)*

Q2. On what devices do you watch English TV programmes? *(more than one answer possible)*

❒ TV	❒ Smartphone
❒ PC	❒ Tablet
❒ Laptop	❒ Other *(specify)*:

Q3. Please give each device a mark between 0 and 10 based on how much you use it for watching English TV programmes. *(0 and 10 represent the extremes on the scale with 0 indicating you do not use it and 10 indicating you make heavy use of it)*

Device	Mark
TV	
PC	
Laptop	
Smartphone	
Tablet	
Other	

Q4. What genre(s) of English TV programme do you watch? *(more than one answer possible)*

❒ Comedy	❒ Variety shows (entertainment)
❒ Sport	❒ Soaps
❒ News	❒ Other *(specify)*:

Q5. Do you watch English TV programmes with any type of (same language or foreign language) subtitles or captions?

❒ Yes ❒ No *(please go to Q7)*

Q6. What type of subtitles or captions do these programmes have? *(more than one answer possible)*

❒ Those that I can turn on and off myself
❒ Those that I cannot turn on or off and are part of the programme
❒ Those in the same language as the dialogue
❒ Those in a different language from the dialogue
❒ Other *(specify)*:

Media exposure to Japanese TV programmes

Please answer the following questions and tick the box(es) that apply to you the most.

Q7. How many hours on average per week do you watch Japanese TV programmes?

❒ 01-10 hours	❒ More than 30 hours
❒ 11-19 hours	❒ I do not watch Japanese TV programmes *(please go to Q13)*
❒ 20-29 hours	

Q8. On what devices do you watch Japanese TV programmes? *(more than one answer possible)*

❒ TV	❒ Smartphone
❒ PC	❒ Tablet
❒ Laptop	❒ Other *(specify)*:

Q9. Please give each device a mark between 0 and 10 based on how much you use it for watching Japanese TV programmes. *(0 and 10 represent the extremes on the scale with 0 indicating you do not use it and 10 indicating you make heavy use of it)*

Device	Mark
TV	
PC	
Laptop	
Smartphone	
Tablet	
Other	

Q10. What genre(s) of Japanese TV programme do you watch? *(more than one answer possible)*

❒ Drama	❒ Variety shows (entertainment)
❒ Anime (cartoons)	❒ Travel
❒ News	❒ Other *(specify)*:

Q11. Do you watch Japanese TV programmes with any type of (same language or foreign language) subtitles or captions?

❒ Yes ❒ No *(please go to Q13)*

Q12. What type of subtitles or captions do these programmes have? *(more than one answer possible)*

- ❒ Those that I can turn on and off myself
- ❒ Those that I cannot turn on or off and are part of the programme
- ❒ Those in the same language as the dialogue
- ❒ Those in a different language from the dialogue
- ❒ Other *(specify)*:

General viewing habits

Please answer the following questions and tick the box(es) that apply to you the most.

Q13. Do you have a TV set where you currently live?

❒ Yes ❒ No

Q14. Where do you normally watch TV programmes? *(more than one answer possible)*

- ❒ At home
- ❒ At a friend's house
- ❒ On public transport
- ❒ Other *(specify)*:

Q15. When do you watch TV programmes? *(more than one answer possible)*

- ❒ When there is a TV programme I want to watch
- ❒ When I want to have a TV programme on in the background
- ❒ When I do not have anything else to do
- ❒ Other *(specify)*:

Q16. What do you base your choice of TV programme on? *(more than one answer possible)*

- ❒ Recommendation
- ❒ Plot/story seems interesting
- ❒ TV guide
- ❒ Grabs my attention in the first few minutes
- ❒ People appearing on the programme
- ❒ Other *(specify):*

Q17. Do you prefer watching TV programmes with or without subtitles or captions?

- ❒ I prefer watching TV programmes with subtitles or captions with the sound on
- ❒ I prefer watching TV programmes with subtitles or captions with the sound off
- ❒ I prefer to have subtitles or captions with Japanese TV programmes, but I rather not have them with English TV programmes
- ❒ I prefer to have subtitles or captions with English TV programmes, but I rather not have them with Japanese TV programmes
- ❒ I prefer watching TV programmes without subtitles or captions
- ❒ Other *(specify)*:

Japanese variety shows

Please answer the following questions and tick the box(es) that apply to you the most.

Q18 represents several scales with opposing views on text on screen in Japanese variety shows. Please indicate which number on these scales best describes what you personally feel about the text on screen in Japanese variety shows.

Q18. What do you personally feel about the text on screen in Japanese variety shows? They:

Distract me	1	2	3	4	5	Do not distract me
Help me understand the show	1	2	3	4	5	Do not help me understand the show
Confuse me	1	2	3	4	5	Do not confuse me
Clog up the screen	1	2	3	4	5	Do not clog up the screen
Make the show funnier	1	2	3	4	5	Do not add to the comedic nature of the show
Help me retain my attention	1	2	3	4	5	Make me lose my attention
Are indispensible	1	2	3	4	5	Can be left out from the show
Help me identify important information	1	2	3	4	5	Do not help me identify important information
Change my interpretation of what people on the show do and say	1	2	3	4	5	Do not change my interpretation of what people on the show do and say
Are flashy	1	2	3	4	5	Are plain and simple

Q19. Would you prefer text on screen to be white and without effects?

❒ Yes ❒ No

Q20. Do you think watching Japanese TV shows helps you with learning Japanese?

❒ Yes ❒ No ❒ I do not know

Q21. Would you like Japanese variety shows to be used in Japanese language classes?

❒ Yes ❒ No

Q22. Do you think you will watch Japanese variety shows more often after this experiment?

❒ Yes ❒ No ❒ Maybe

Q23. Did you feel the questionnaire allowed you to say everything you wanted to say? Is there anything you think might be relevant to mention in regard to TV programmes or Japanese language learning?

That was the final question of the second part of the questionnaire.

Thank you for your cooperation!

Appendix E.2 Japanese version

Appendix E.2.1 Part 1

今見た番組の内容について

質問に答え、最も当てはまる箇所に X をつけてください。

質問1-5は「もしもの時に役立つホンマでっか！？」の最初の事例に関連しています。

質問1. 光熱費を安くしたかったら、どうしたら良いですか。

❒ マンションの中間階の真ん中に住めば良いです
❒ 一戸建てに住めば良いです
❒ マンションの最上階角部屋に住めば良いです

質問2. その理由はどのように説明されましたか。

❒ 一戸建ての光熱費はマンション最上階角部屋の光熱費の2/3
❒ マンション中間階真ん中の光熱費は一戸建ての光熱費の約半分
❒ マンション最上階角部屋の光熱費はマンション中間階真ん中の光熱費の2/3

質問3. なぜその条件の住宅だと、光熱費が最も安くなるのですか。

質問4. 一般的に、冷房と暖房どちらの費用が安いですか。

❒ 冷房 ❒ 暖房

質問5. その理由はどのように説明されましたか。

❒ 一般的に、暖房を使う期間は冷房を使う期間の二倍です
❒ 一般的に、冷房を使う期間は暖房より長いです
❒ 外の温度と室内の温度の温度差は冷房の方が大きいです

質問6-8は「もしもの時に役立つホンマでっか！？」の二つ目の事例に関連しています。

質問6. 雷に遭遇し近くに逃げ込む場所がなかったら、どのような2つの行動が一番大事ですか。

1.
2.

質問7. どうしてですか。

❒ 立っている時に落雷にあった事例しかありません
❒ 雷は高い所を選んで落ちるので、人は立っているだけで避雷針になります
❒ 人は立っている場合に落雷にあったら、心配停止の危険性があります
❒ 以上の全て

質問8. カミナリ評論家は腹這いの場合に落雷される可能性に関して、何と言いましたか。

質問9-10は「もしもの時に役立つホンマでっか！？」の三つ目の事例に関連しています。

質問9. どうして三つ目の事例で紹介された状況を切り抜けたかったら、「あ～そういう風に受け取られましたか？」と言えば良いですか。

質問10. どうして八嶋智人は図1にあるような表情を見せましたか。

❒ 専門家をからかっているから
❒ 専門家が言ったことに驚いたから
❒ 顔の表情と言葉の内容が違うから
❒ 専門家が詐欺にあったということが信じ難いから

図1

今見た番組について

質問に答え、最も当てはまる箇所に X をつけてください。

質問11. 今見た番組で使われていたイラストは内容も理解するのに役立ったと思いますか。

❒ はい　❒ いいえ　❒ 分かりません

質問12. 質問の解答に確信がありますか。

❒ 十分にある　❒ 大半の質問を当て推量で答えました
❒ それなりにある　❒ 分かりません
❒ あまりない

質問13. 今見たテレビ番組について知っていましたか。

❒ はい　❒ いいえ

質問14. 今見たテレビ番組は見たことがありますか。

❒ よく見ます
❒ 時々見ます
❒ 今回初めて見ました
❒ 分かりません

今見た番組のテロップについて

質問に答え、最も当てはまる箇所に X をつけてください。

質問15. 今見たテレビ番組のテロップを読みましたか。最も当てはまる箇所に X をつけてください。

❒ 全部読みました
❒ 大半を読みました
❒ 半分未満だけ読みました
❒ 画面上部のものだけ読みました
❒ 画面下部のものだけ読みました
❒ 何も読みませんでした

質問16. 今見たテレビ番組のテロップで、どのような特殊効果に気がつきましたか。

質問17. テロップに使われていた色を思い出せるだけ書いてください。

質問18a. 読みにくい色や色の組み合わせがありましたか。

❒ はい　❒ いいえ *（質問19に進んでください）*

b. 具体的には何でしたか。

質問19. 今見た番組のテロップでは、色々なフォントが使われていましたか。

❒ はい　❒ いいえ　❒ 覚えていません

質問20. 今見た番組のテロップの読み易すさはどう思いましたか。

❒ テロップの可読性は良かったです
❒ テロップ大半の可読性は良かったです
❒ テロップ大半の可読性は悪かったです
❒ テロップの可読性は悪かったです
❒ その他：

質問21a. 今見たテレビ番組のテロップがこのバラエティ番組に適当だと思いましたか。

❒ はい　　❒ いいえ

b. そう思う理由を説明してください。

質問22. このアンケートで言いたいことを全部言い切れたと思いますか。何か重要だと思うことがあったら、以下にご自由に述べてください。

これで以上です。

ご協力ありがとうございます！

Appendix E.2.2 Part 2

日本語のテレビ番組の経験について

質問に答え、最も当てはまる箇所に X をつけてください。

質問1. 週に何時間ぐらい日本語のテレビ番組を見ますか。

❒ 1～10時間
❒ 11～19時間
❒ 20～29時間
❒ 30時間以上
❒ 全然見ません *（質問7に進んでください）*

質問2. 日本語のテレビ番組は何で見ますか。 *（複数の回答が可能）*

❒ テレビ
❒ コンピューター
❒ パソコン
❒ スマートフォン
❒ タブレット
❒ その他：

質問3. 日本語のテレビ番組を見るとき、どの媒体でどれだけ見ていますか。0～10の数字を以下につけてください。 *（0と10は両極端の表記で、0は「使いません」10は「一番使っています」という意味です）*

媒体	頻度
テレビ	
コンピューター	
パソコン	
スマートフォン	
タブレット	
その他	

質問4. 日本語のテレビ番組のジャンルで、何が好きですか。 *（複数の回答が可能）*

❒ ドラマ
❒ アニメ
❒ ニュース
❒ バラエティ
❒ 旅行
❒ その他：

質問5. 字幕、または、キャプション付きの日本語のテレビ番組を見ますか。

❒ はい　　❒ いいえ *（質問7に進んでください）*

質問6. そのテレビ番組の字幕・キャプションは何の種類ですか。 *（複数の回答が可能）*

❒ 自分でオン・オフができる種類
❒ 自分でオン・オフができないテレビ番組の一部である種類
❒ 番組の言葉の音声と同じ種類
❒ 番組の言葉の音声のと違う種類
❒ その他：

英語のテレビ番組の経験について

質問に答え、最も当てはまる箇所に X をつけてください。

質問7. 週に何時間ぐらい英語のテレビ番組を見ますか。

❒ 1～10時間
❒ 11～19時間
❒ 20～29時間
❒ 30時間以上
❒ 全然見ません *（質問13に進んでください）*

質問8. 英語のテレビ番組は何で見ますか。 *（複数の回答が可能）*

❒ テレビ
❒ コンピューター
❒ パソコン
❒ スマートフォン
❒ タブレット
❒ その他：

質問9. 英語のテレビ番組を見るとき、どの媒体でどれだけ見ていますか。0～10の数字を以下につけてください。 *（0と10は両極端の表記で、0は「使いません」10は「一番使っています」という意味です）*

媒体	頻度
テレビ	
コンピューター	
パソコン	
スマートフォン	
タブレット	
その他	

質問10. 英語のテレビ番組のジャンルで、何が好きですか。 *（複数の回答が可能）*

❒ コメディー
❒ スポーツ
❒ ニュース
❒ バラエティ
❒ 昼ドラ
❒ その他：

質問11. 字幕、または、キャプション付きの英語のテレビ番組を見ますか。

❒ はい
❒ いいえ *（質問13に進んでください）*

質問12. その番組の字幕・キャプションは何の種類ですか。 *（複数の回答が可能）*

❒ 自分でオン・オフができる種類
❒ 自分でオン・オフができない番組の一部である種類
❒ 番組の言葉の音声と同じ種類
❒ 番組の言葉の音声のと違う種類
❒ その他：

番組視聴の習慣

質問に答え、最も当てはまる箇所に X をつけてください。

質問13. 今住んでいる家にテレビがありますか。

❒ はい
❒ いいえ

質問14. 普段、どこでテレビ番組を見ますか。 *（複数の回答が可能）*

❒ 家で
❒ 友人の家で
❒ 公共交通機関に乗っているとき
❒ その他：

質問15. 普段、いつテレビ番組を見ますか。 *（複数の回答が可能）*

❒ 見たい番組があるとき、テレビ番組を見ます。
❒ 他のことをしながら、テレビ番組を見ます。
❒ 他にすることがないとき、テレビ番組を見ます。
❒ その他：

質問16. テレビ番組を選ぶ基準は何ですか。 *(複数の回答が可能)*

❒ 誰かにすすめられて	❒ 2～3分での印象
❒ 話が面白い	❒ 出演者
❒ 番組案内ガイド	❒ その他：

質問17. 字幕・キャプション付きのテレビ番組と字幕・キャプション無しのテレビ番組では、どちらが好みですか。

❒ （音声を消音せず）字幕・キャプション付きのテレビ番組
❒ （音声を消音しながら）字幕・キャプション付きのテレビ番組
❒ 日本語のテレビ番組は字幕・キャプション付きで、英語のテレビ番組は字幕・キャプション無し
❒ 英語のテレビ番組は字幕・キャプション付きで、日本語のテレビ番組は字幕・キャプション無し
❒ 字幕・キャプション無しのテレビ番組
❒ その他：

日本語のバラエティ番組について

質問に答え、最も当てはまる箇所に X をつけてください。

質問18はバラエティ番組のテロップに関する意見のスケールです。バラエティ番組のテロップに関する自身の考え方に最も近い数字に丸をつけてください。

質問18. バラエティ番組のテロップに関しては、どう思いますか。テロップは：

気が散ります	1	2	3	4	5	気が散りません
番組をより分かりやすくします	1	2	3	4	5	番組をよりわかりやすくしません
紛らわしいです	1	2	3	4	5	紛らわしくないです
画面を塞ぎます	1	2	3	4	5	画面を塞ぎません
番組をより面白くします	1	2	3	4	5	番組をより面白くしません
番組に集中できるようにします	1	2	3	4	5	番組に集中できるようにしません
不可欠です	1	2	3	4	5	不可欠ではありません
番組の重要なところを指摘します	1	2	3	4	5	番組の重要なところを指摘しません
番組内での発言や行動の解釈を変えます	1	2	3	4	5	番組内での発言や行動の解釈を変えません
派手です	1	2	3	4	5	地味です

質問19. 特殊効果のない白いテロップの方が良いと思いますか。

❒ はい　　❒ いいえ

質問20. 今見たテレビ番組を見たきっかけに、これから日本のバラティ番組を見ようと思いますか。

❒ はい　　❒ いいえ　　❒ たぶん見ます

質問21. このアンケートで言いたいことを全部言い切れたと思いますか。何か重要だと思うことがあったら、以下にご自由に述べてください。

これで以上です。

ご協力ありがとうございました！

Appendix F. Evaluation of collected data

Participant No.	Quality of recording		Questionnaires			Field notes	Potential influences on gaze data quality
	Calibration	Gaze samples	*pre*	*post1*	*post2*		
P01	Good	94%	Yes	Yes	Yes	Yes	-
P02	Good*	97%	Yes	Yes	Yes	Yes	Astigmatism Loud sound outside of room (1x)
P03	Good	92%	Yes	Yes	Yes	Yes	Loud sound outside of room (1x)
P04	Good	93%	No	Yes	Yes	Yes	Took glasses off before calibration
P05	Good	83%	Yes	Yes	Yes	Yes	Felt ill on day of experiment session
P06	Good	75%	Yes	Yes	Yes	Yes	Loud sound outside of room (1x)
P07	-	-	-	-	-	-	-
P08	Good	96%	Yes	Yes	Yes	Yes	Loud sound outside of room (1x)
P09	Good	92%	Yes	Yes	Yes	Yes	Loud laugh outside of room (1x)
P10	Good	97%	Yes	Yes	Yes	Yes	-
P11	Good	96%	Yes	Yes	Yes	Yes	-
P12	Good	93%	Yes	Yes	Yes	Yes	Facial piercing Sound of heater (was a little bothered by it)
P13	Good	93%	Yes	Yes	Yes	Yes	Sound of heater (was not bothered by it)
P14	Good	99%	Yes	Yes	Yes	Yes	Partially colour blind Preferred higher volume for video
P15	Good	93%	Yes	Yes	Yes	Yes	Facial piercings Preferred higher volume for video Sound of heater (was not bothered by it)
P16	Good	91%	Yes	Yes	Yes	Yes	Sound of heater (may have been bothered by it)
P17	Good	97%	Yes	Yes	Yes	Yes	Loud sound outside of room (1x) Sound of heater (was not bothered by it)
P18	Good	92%	Yes	Yes	Yes	Yes	Make-up Loud sound outside of room (1x) Sound of heater (was not bothered by it)
P19	Good*	92%	Yes	Yes	Yes	Yes	Sound of heater (was not bothered by it)
P20	Good	81%	Yes	Yes	Yes	Yes	Sound of heater (was not bothered by it)
P21	Good	96%	Yes	Yes	Yes	Yes	Earrings Sound of heater (was not bothered by it)
P22	Good	90%	Yes	Yes	Yes	Yes	-
P23	Good	97%	Yes	Yes	Yes	Yes	-
P24	Good	91%	Yes	Yes	Yes	Yes	-
P25	Good	91%	Yes	Yes	Yes	Yes	-
P26	Good**	97%	Yes	Yes	Yes	Yes	Eye condition Colour blind
P27	Good	96%	Yes	Yes	Yes	Yes	-
P28	Good	93%	Yes	Yes	Yes	Yes	Facial piercing
P29	Good	97%	Yes	Yes	Yes	Yes	Loud sound outside of room (1x)
P30	Good	86%	Yes	Yes	Yes	Yes	Make-up
P31	Good**	61%	Yes	Yes	Yes	Yes	-
P32	Good***	96%	Yes	Yes	Yes	Yes	Noise outside of room (1x)
P33	Good	94%	Yes	Yes	Yes	Yes	-
P34	Good	92%	Yes	Yes	Yes	Yes	-
P35	Good	87%	Yes	Yes	Yes	Yes	-
P36	Good	90%	Yes	Yes	Yes	Yes	-
P37	Good	91%	Yes	Yes	Yes	Yes	-
P38	Good	95%	Yes	Yes	Yes	Yes	-
P39	Good	96%	Yes	Yes	Yes	Yes	-
P40	Good	95%	Yes	Yes	Yes	Yes	-

** Could not calibrate with glasses*
*** Recalibrated several times*
**** Different visions of eyes showed in calibration*

(continued)

Participant No.	Quality of recording		Questionnaires			Field notes	Potential influences on gaze data quality
	Calibration	Gaze samples	*pre*	*post1*	*post2*		
P41	Good	86%	Yes	Yes	Yes	Yes	Make-up
P42	Good*	92%	Yes	Yes	Yes	Yes	-
P43	Good	91%	Yes	Yes	Yes	Yes	-
P44	Good	81%	Yes	Yes	Yes	Yes	-
P45	Good	95%	Yes	Yes	Yes	Yes	Make-up Phone beeped (1x)
P46	Good****	95%	Yes	Yes	Yes	Yes	-
JP01	Poor	44%	Yes	Yes	Yes	Yes	Perhaps seated uncomfortably
JP02	Good	74%	Yes	Yes	Yes	Yes	Perhaps seated uncomfortably
JP03	Good	89%	Yes	Yes	Yes	Yes	Perhaps seated uncomfortably
JP04	Good	53%	Yes	Yes	Yes	Yes	Office phone rang (1x)
JP05	Good	84%	Yes	Yes	Yes	Yes	-
JP06	Good	88%	Yes	Yes	Yes	Yes	-
JP07	Good	72%	Yes	Yes	Yes	Yes	-

** Could not calibrate with glasses*

***** Dots in tracker box were flickering a bit*

Appendix G. Description of initial codes

Appendix G.1 Yellow

Yellow 1: impression of the Japanese variety show

Extracts give indications of the image that participants have of the Japanese variety show in the video stimulus. Adjectives that are recurrent in the extracts are loud, fast-paced, funny and suited to an informal telop design. Other similar adjectives that were mentioned by participants, although less frequently in comparison to those referred to previously, are light-hearted, funny/serious or tabloidesque, wacky and flamboyant.

Yellow 2: aspects of the telop design that suited or did not suit the Japanese variety show

Extracts describe aspects of the telop design that participants thought were either appropriate or inappropriate for their use in the Japanese variety show of the video stimulus. The majority of these extracts mention suitable characteristics. A recurrent adjective is varied, but other adjectives mentioned by the participants include large and brightly coloured, sensational, comical, fun, exciting and fast-paced.

Yellow 3: impression of the telop design

Extracts give indications of the image participants have of the telop design that was used in the Japanese variety show of the video stimulus. Adjectives that were mentioned by the participants were similar to those referred to in Yellow 1 and included quirky, radical, comical, fast-paced, eye-catching, fun and casual.

Yellow 4: aspects of the Japanese variety show that the telop design resembled

Extracts mention aspects of the telop design that participants perceived to bear resemblances to aspects of the Japanese variety show of the video stimulus. Links between the speed of telops and the pace of the Japanese variety show, and the varied typographical features of telops and the humour of the Japanese variety show were made in particular. Links between the telop design and the mood, set or overall contents of the Japanese variety show were also made.

Appendix G.2 Blue

Blue 1: perceived effects of the telop design

Extracts show examples of effects perceived by the participants of the telop design that was used in the Japanese variety show of the video stimulus. This cohort of extracts differs from the other initial codes in blue as distinctions between typographical features are not made. Participants mentioned that the telop design was impactful, drew in their eyes, kept them interested, helped them guess or emulate speakers' expressions or feelings, exaggerate the content, set the mood and help with the delivery of the jokes.

Blue 2: perceived functions of colour in the telop design

Extracts show participants' ideas on the reasons behind the use of colours in telops that were used in the Japanese variety show of the video stimulus. Colour differences are linked to differences in speakers and emotions, ascribed to emphasis or thought to make important information stand out to viewers.

Blue 3: perceived functions of telop and telop design

Extracts show participants' ideas on the reasons behind the use of telops or telop design as a whole in the Japanese variety show of the video stimulus rather than describing typographical features separately. The reasons expressed by participants were similar to those mentioned in Blue 2 and Blue 4. However, there was one extract in which a participant had indicated that telop may be used to disambiguate Japanese words that have the same pronunciation.

Blue 4: perceived functions of font in the telop design

Extracts show participants' ideas on the reasons behind the use of fonts in telops that were used in the Japanese variety show of the video stimulus. Font differences are linked to differences in speakers, register, gender and emotions, ascribed to emphasis/exaggeration or thought to make important information stand out to viewers.

Blue 5: perceived effects of font in the telop design

Extracts represent examples of effects perceived by the participants of the font of telops in the Japanese variety show of the video stimulus. Participants mentioned that it adds to the pace of the Japanese variety show, exaggerates certain moments or catches your eyes.

Appendix G.3 Green

Green 1: appearance of telops linked to audio-verbal and audio-nonverbal input

Extracts show that participants link the appearance of telops to the corresponding speech in the Japanese variety show of the video stimulus. Some extracts are more specific and indicate that the appearance of telop is linked to funny, dramatic or relevant speech. Some participants mention that telops appeared with sound effects.

Green 2: experienced difficulties with telops

Extracts show that participants struggled with the amount of text in telops and the speed of telops in the Japanese variety show of the video stimulus in particular.

Green 3: focus on or preference for certain input

Extracts illustrate that some participants felt they focused on particular input in the Japanese variety show of the video stimulus. There was one participant in Group 1: pre-exchange who expressed that too much emphasis is put on telops and that telops were not necessary in the Japanese variety show. This particular participant had a preference to non-verbal input.

Green 4: experienced competition between input

Extracts show different opinions on whether there is a trade-off in concentration as a result of the amount or combination of certain input in the Japanese variety show of the video stimulus. Whereas some participants expressed that there was a lot happening in general, some other participants commented on the distracting nature of telops in particular. Only one participant, organised into Group 1: pre-exchange, indicated that telops were not distracting.

Green 5: legibility of telops

Extracts show different opinions on the legibility of telops in the Japanese variety show of the video stimulus. It appeared that telops with dark colours or those that were blue or yellow were most difficult to read for participants.

Green 6: experienced difficulties with the combination of input

Extracts illustrate that participants struggled with overlapping voices or types of input in the Japanese variety show of the video stimulus.

Green 7: perceived resemblances between input

Extracts mention aspects of the telop design that participants perceived to bear resemblances to input in the Japanese variety show of the video stimulus. The variety of colours, fonts and movements were linked to the emotions or meaning of words and the variety of speakers and topics. This initial code therefore appears to be related to the initial codes in blue.

Green 8: perceived importance of certain input

Extracts describe participants' ideas on the importance of particular input in the Japanese variety show of the video stimulus. These include the importance of audio-nonverbal input (i.e. music and noise from audience) for the general feel, and the display rate and movements of telops.

Appendix G.4 Orange

Orange 1: telop design described as helpful to understanding

Extracts include statements in which the helpful nature of the telop design is expressed and the ways in which aspects of the telop design helped with understanding the Japanese variety show of the video stimulus is described. The ways in which telop design was helpful appeared to relate to the colours and fonts, and schematic representations of the dialogue incorporated within the telop design.

Orange 2: combination of input described as helpful to understanding

Extracts comprise participants' responses in which input other than telop or the combination of input were described as helpful to understanding the Japanese variety show of the video stimulus. Whereas one participant indicated to find the intonation helpful, other participants described the usefulness of kinesic action (i.e. gesticulation and bodily configuration), visual input or the combination of telops, kinesic action and dialogue.

Orange 3: telop described as helpful to understanding

Extracts comprise participants' responses in which the usefulness of telops to understanding the Japanese variety show of the video stimulus is emphasised. Interesting is the comment from a participant in Group 1: pre-exchange who explained that it helped grasp the context and another comment from a participant in Group 3: post-exchange who indicated that telops help the viewer keep up with the dialogue.

Orange 4: potential risks to misunderstanding input

Extracts showcase instances during which it became clear that there was room for a participant to misunderstand input in the Japanese variety show of the video stimulus. Most cases relate to (mis)readings of kanji. However, one participant in Group 1: pre-exchange misinterpreted one of the visual-nonverbal types of input.

Orange 5: self-reported understanding of the Japanese variety show

Extracts illustrate whether participants felt they understood the Japanese variety show of the video stimulus.

Appendix G.5 Pink

Pink 1: expected positive outcomes of watching Japanese television to JLA

Extracts describe expected positive results to JLA from watching Japanese television. These relate to acquiring more natural language skills and familiarity with fast-paced and casual speech (i.e. slang, accented dialogue and dialects) in particular. There was also one participant who pointed out the benefit of gaining acquaintance with the sound of the Japanese language, and two participants who linked Japanese television to vocabulary knowledge.

Pink 2: reservations regarding the use of Japanese television for JLA

Extracts give explanations on why some participants have reservations regarding the use of Japanese television for JLA. One participant thinks that it is not suitable for learners at beginner level, another participant thinks variety shows might be too eccentric for class, while other participants dislike Japanese television in general or Japanese variety shows in particular.

Pink 3: aspects of Japanese television that are perceived to be helpful to JLA

Extracts illustrate why participants think Japanese television may be helpful to JLA. These helpful aspects of Japanese television relate to the relevant, realistic or native setting, repeated exposure to vocabulary and the combination of input.

Pink 4: perceived better alternatives to Japanese variety shows for JLA

Extracts show participants' ideas on what they perceived to be more useful to JLA than watching Japanese variety shows or Japanese television in general. These ideas comprise watching other, more serious, genres of Japanese television programme and practice of conversation skills with native Japanese speakers.

Pink 5: self-proclaimed experience with learning FLs through television

Extracts give anecdotes in which participants describe their experience with television for the acquisition of FLs.

Pink 6: aspects that are perceived to be missing in current Japanese language classes

Extracts describe aspects which participants thought were missing in current Japanese language classes. These relate to real or authentic Japanese language use in particular.

Appendix H. Representation of initial codes in thematic map

Colour	Code No.	Description	No. of extracts*
Yellow	1	Impression of the Japanese variety show	8
	2	Aspects of the telop design that suited or did not suit the Japanese variety show	5
	3	Impression of the telop design	6
	4	Aspects of the Japanese variety show that the telop design resembled	9
Blue	1	Perceived effects of the telop design	8
	2	Perceived functions of colour in the telop design	9
	3	Perceived functions of telop and telop design	6
	4	Perceived functions of font in the telop design	11
	5	Perceived effects of font in the telop design	2
Green	1	Appearance of telops linked to audio-verbal and audio-nonverbal input	7
	2	Experienced difficulties with telops	7
	3	Focus on or preference for certain input	1
	4	Experienced competition between input	6
	5	Legibility of telops	0
	6	Experienced difficulties with the combination of input	1
	7	Perceived resemblances between input	5
	8	Perceived importance of certain input	1
Orange	1	Telop design described as helpful to understanding	9
	2	Combination of input described as helpful to understanding	0
	3	Telop described as helpful to understanding	3
	4	Potential risks to misunderstanding input	0
	5	Self-reported understanding of the Japanese variety show	3
Pink	1	Expected positive outcomes of watching Japanese television to JLA	6
	2	Reservations regarding the use of Japanese television for JLA	2
	3	Aspects of Japanese television that are perceived to be helpful to JLA	3
	4	Perceived better alternatives to Japanese variety shows for JLA	0
	5	Self-proclaimed experience with learning FLs through television	0
	6	Aspects that are perceived to be missing in current Japanese language classes	3

** The total number of extracts amounts to 121 as seven extracts are counted twice considering these were ascribed to two initial codes. Six of these are in blue and one is in orange.*

Appendix I. List of data extracts

Appendix I.1 Subtheme 1.1.1: mood

Group 1: pre-exchange, P24

it was quirky and radical, and the guests seemed to have radical comedic reactions to the info. that was presented
So a formal/serious text wouldn't have suited.

post1 Q25b, ll. 1-3

Group 1: pre-exchange, P24

Japanese variety shows seem funny

post2 Q23, l. 1

Group 1: pre-exchange, P30

the text seemed fun and exciting.

post1 Q25b, l. 1

Group 1: pre-exchange, P32

The text seemed to be designed in a comical style

post1 Q25b, l. 1

Group 1: pre-exchange, P35

I think the variety show was meant to be fun and lighthearted and the text and the colours of the text showed this as well by being different colours etc.

post1 Q25b, ll. 1-3

Group 1: pre-exchange, P36

The whole thing was funny/serious, meaning they get funny people to talk about serious things so it was very tabloid-esque

post1 Q25b, ll. 1-2

Group 1: pre-exchange, P40

The set seemed to be bright in colour and this was reflected in the text.

post1 Q25b, ll. 1-2

Group 2: exchange, P01

… and appeared casual in nature.

post1 Q25b, l. 1

Group 2: exchange, P08

… and emotions

post1 Q20, ll. 1-2

Group 2: exchange, P10

It was colourful and big which goes well with the humour of the show.

post1 Q25b, ll. 1-2

Group 2: exchange, P12

since it is a variety show, I would not expect to see a very 'serious' font on the screen.

post1 Q25b, ll. 1-2

Group 2: exchange, P17

As the show is fun and wacky so it suits this particular type of show to use different fonts and colours.

post1 Q25b, ll. 1-2

Group 2: exchange, P19

The strength / thickness of the radicals seemed informal, equivalent to the less formal fonts often used in English.

post1 Q25b, ll. 1-2

Reference group, JP05

特に違和感も感じず、内容を理解したと思うため。 [Because I understood the contents without it feeling out of place.]

post1 Q21b, l. 1

Reference group, JP07

理解しやすい [It is easy to understand]

post1 Q21b, l. 1

Appendix I.2 Subtheme 1.1.2: tone of voice

Group 1: pre-exchange, P36

colour based on emotions in words

post1 Q20, l. 1

Group 2: exchange, P06

Colours. and sizes when trying to exaggerate

post1 Q20, l. 1

Group 2: exchange, P06

It helps it to exargerrate the content and allow the audience to understand what was spoken more easily.

post1 Q25b, ll. 1-2

Appendix I.3 Subtheme 1.2.1: speakers

Group 2: exchange, P02

There was different colours and fonts used for different speakers.

post1 Q20, l. 1

Group 2: exchange, P08

different fonts and colours for different people …

post1 Q20, l. 1

Group 2: exchange, P13

Colour: Represents different people by changing colour

post1 Q20, l. 3

Appendix I.4 Subtheme 1.2.2: main points of dialogue

Group 1: pre-exchange, P22

The large & brightly coloured 3D style text was suited as it draws in the viewer's eyes.

post1 Q25b, ll. 1-2

Group 1: pre-exchange, P30

different colours highlighting the main words on the screen.

post1 Q20, l. 1

Group 1: pre-exchange, P32

used to accentuate punchlines and moments of surprise

post1 Q25b, ll. 1-2

Group 1: pre-exchange, P34

Different sizes of text and boxes around some key points

post1 Q20, l. 1

Group 1: pre-exchange, P38

It was eye-catching and kept you interested.

post1 Q25b, l. 1

Group 2: exchange, P10

Certain words were in a different colour or bold in order to stand out.

post1 Q20, l. 1

Group 2: exchange, P18

the text would emphasise the words, whether it was something shocking or funny that was said. Made it easier to remember too.

post1 Q25b, ll. 1-3

Group 3: post-exchange, P23

Different colours and fonts were used to emphasise different things.

post1 Q20, l. 1

Group 3: post-exchange, P28

words were outlined in a darker colour to make them stand out

post1 Q20, l. 2

Group 3: post-exchange, P43

Large size letters for emphasis.

post1 Q20, l. 1

Group 3: post-exchange, P43

… and highlight important info.

post1 Q25b, ll. 3-4

Group 3: post-exchange, P44

font changes to catch your eye.

post1 Q20, l. 3

Group 3: post-exchange, P44

P44 definitely felt that the colours helped.

Field notes, l. 8

Reference group, JP06

色をつけて強調していた。 [It was emphasised through colour.]

post1 Q16, l. 1

Appendix I.5 Subtheme 2.1: fast-paced

Group 1: pre-exchange, P22

text that popped up on screen as it was being said

post1 Q20, l. 1

Group 1: pre-exchange, P22

The show was somewhat loud and fast-paced.

post1 Q25b, l. 1

Group 1: pre-exchange, P29

The text appeared in time with, or before the people spoke

post1 Q20, l. 1

Group 1: pre-exchange, P30

The text vanished fast off the screen which match the fast pace of the show.

post1 Q25b, ll. 1-2

Group 1: pre-exchange, P32

Attempting to read the text while not understanding it was quite tiring. I began to lose focus about halfway through the video. The speed at which it appeared and disappeared added to this I think.

post1 Q26, ll. 1-3

Group 1: pre-exchange, P35

I think showing the language both written and orally in a native setting would help to acquire a more natural Japanese.

post2 Q23, ll. 1-3

Group 2: exchange, P01

It was very fast paced

post1 Q25b, l. 1

Group 2: exchange, P03

appeared suddenly → something dramatic was said

post1 Q20, l. 1

Group 2: exchange, P03

Usually when something funny or dramatic was said, it was shown on the screen in writing.

post1 Q25b, ll. 1-2

Group 2: exchange, P08

Because they are very fast paced, the tv shows aren't really suitable for beginners and I think they'll only be helpful if the speaker already has a fair grasp of the language

post2 Q23, ll. 1-3

Group 2: exchange, P12

They would appear suddenly especially when some relevant information was said

post1 Q20, ll. 1-2

Group 2: exchange, P14

it went off screen before I could finish reading often.

post1 Q26, ll. 1-2

Group 2: exchange, P16

Text appearing as it was read out

post1 Q20, l. 1

Group 2: exchange, P21

I feel that not enough real, everyday Japanese is used in classes. If Japanese TV was used as a learning tool, people would probably have better listening skills and understand dialects more.

post2 Q23, ll. 1-3

Appendix I.6 Subtheme 2.2: distracting

Group 1: pre-exchange, P22

P22 tried to concentrate on mouths until P22 noticed the telops popping up.

Field notes, ll. 1-2

Group 1: pre-exchange, P37

Variety shows might be a bit too eccentric for classroom purposes

post2 Q23, l. 1

Group 2: exchange, P08

I think that the text on screen is unusual in Western TV and to me it's quite unusual. It is distracting from the actual people/actors

post1 Q26, ll. 1-2

Group 2: exchange, P10

When I first started watching the show, I found that there was so much happening on screen that it was hard to understand what they were saying, i.e people shouting and laughing with multiple sentences written accross the screen.

post1 Q26, ll. 1-4

Group 2: exchange, P17

I suppose at times it can seem to be a bit too much as there are so many colours combinations being used. As it can make it harder to concentrate as you are focusing on reading as well as trying to understand what is being said.

post1 Q22b, ll. 1-4

Group 2: exchange, P18

making it hard not to look at it

post1 Q20, l. 1

Group 2: exchange, P20

this flamboyant show.

post1 Q25b, l. 2

Reference group, JP02

印象に残りやすいから。 [because it easily leaves an impression.]

post1 Q21b, ll. 1-2

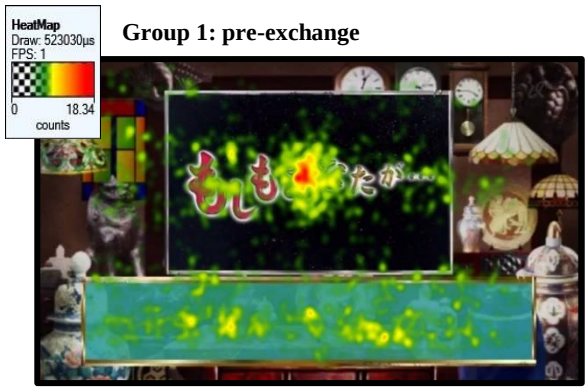
Reference group, JP07

インパクトがある [It has impact]

post1 Q21b, l. 2

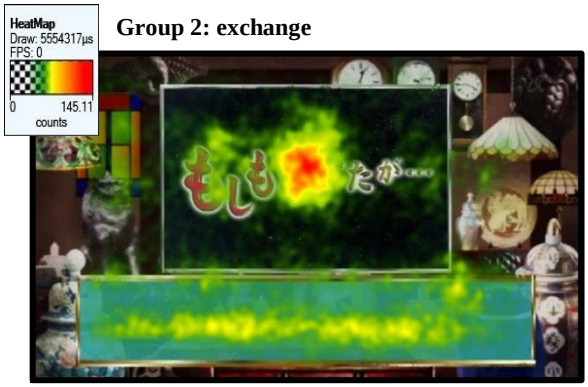
Appendix J. Heat maps per analysed video segment

Appendix J.1 Fixation count aggregated onto screen for Segment 2

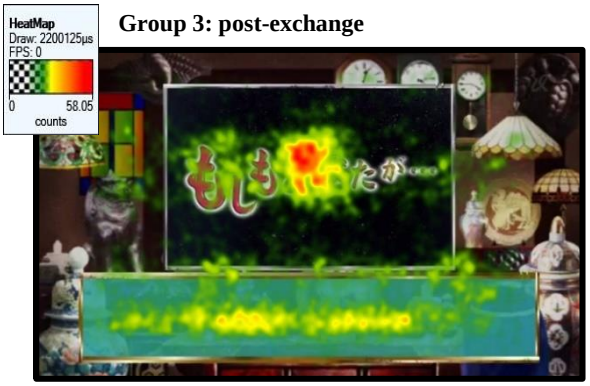


Group 1: pre-exchange

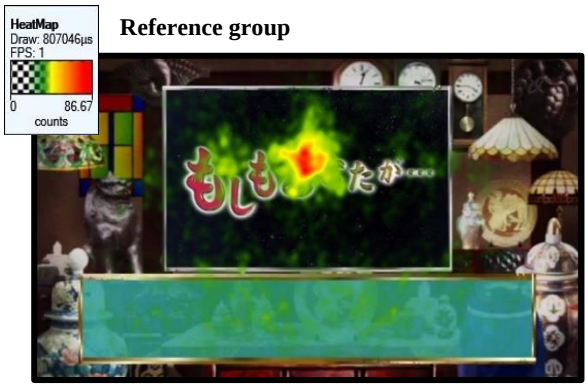
Group 1: pre-exchange



Group 2: exchange



Group 3: post-exchange

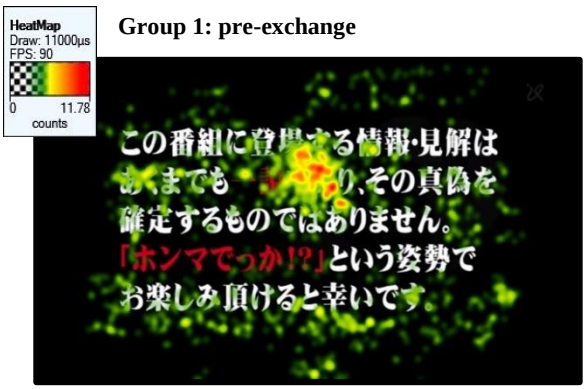


Reference group

Honmadekka!?TV (2013)

Appendix J.2 Fixation count aggregated onto screen for Segment 4

Group 1: pre-exchange



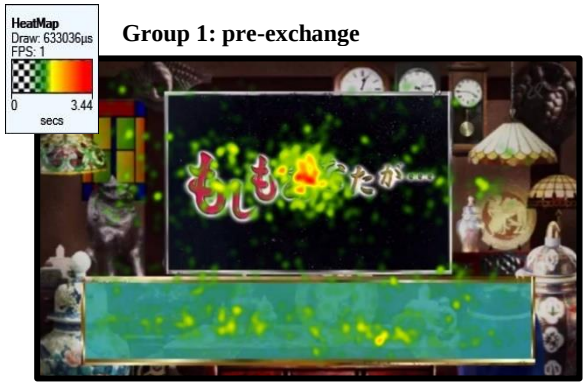
Group 1: pre-exchange

Group 3: post-exchange

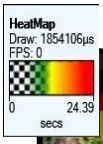
Reference group

Honmadekka!?TV (2013)

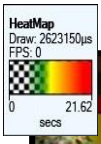
Appendix J.3 Absolute gaze duration aggregated onto screen for Segment 2



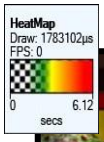
Group 1: pre-exchange



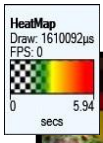
Group 1: pre-exchange



Group 2: exchange



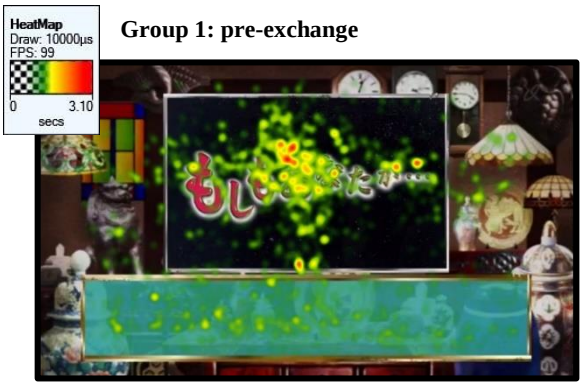
Group 3: post-exchange



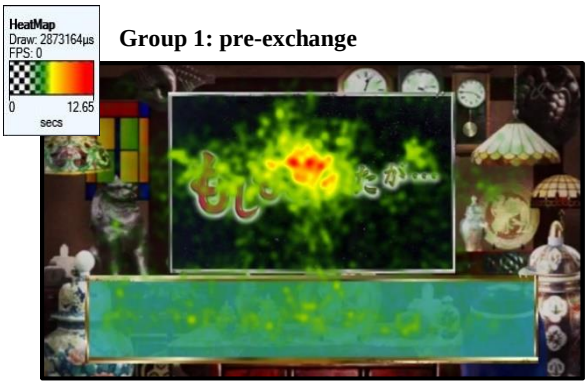
Reference group

Honmadekka!?TV (2013)

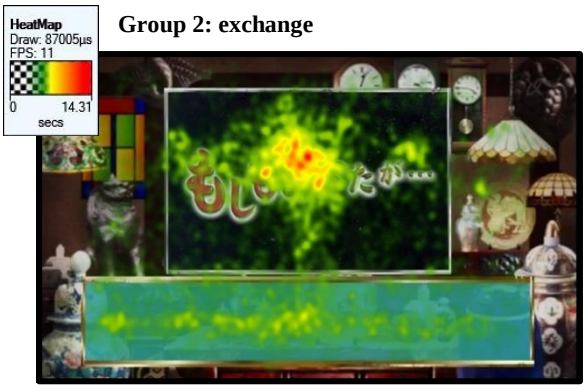
Appendix J.4 Absolute gaze duration aggregated onto screen for Segment 3



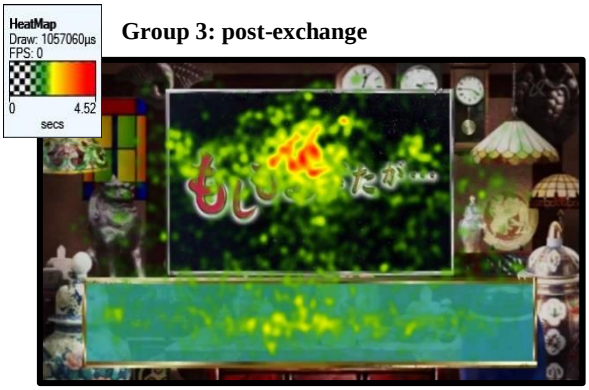
Group 1: pre-exchange



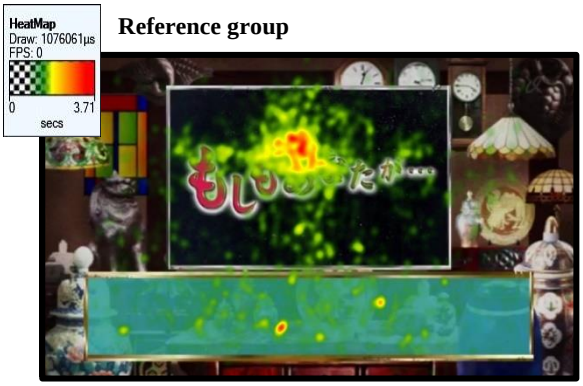
Group 1: pre-exchange



Group 2: exchange



Group 3: post-exchange



Reference group

Honmadekka!?TV (2013)

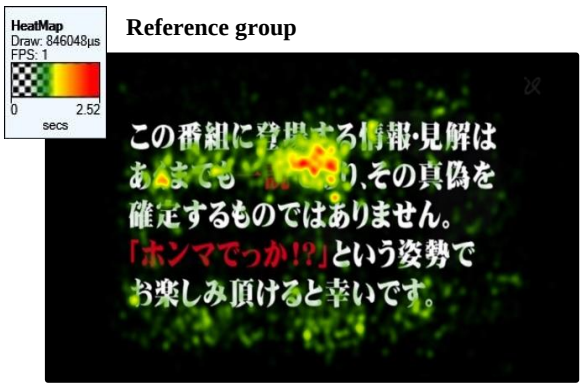
Appendix J.5 Absolute gaze duration aggregated onto screen for Segment 4

Group 1: pre-exchange

Group 1: pre-exchange

Group 2: exchange

Group 3: post-exchange



Reference group

Honmadekka!?TV (2013)

Appendix K. Scope of input for analysed video segments

Form of expression	Type of input	Variety of input
Audio-verbal	**1.** Dialogue	**1.1** Utterance of speaker **1.2** Narration
Audio-nonverbal	**2.** Sound	**2.1** Noise from audience **2.2** Music **2.3** Sound effect **2.4** Other
Visual-verbal	**3.** Text in background	**3.1** Nameplate
	4. Telop	**4.1** Utterance of speaker **4.2** Narration **4.3** Theme **4.4** Title **4.5** Speaker identification **4.6** Screen filled with text
Visual-nonverbal	**5.** Telop design	**5.1** Expansion **5.2** Background colour **5.3** Punctuation **5.4** Special effect **5.5** Picture **5.6** Other
	6. Imagery	**6.1** Drawing
	7. Kinesic action	**7.1** Gesticulation **7.2** Bodily configuration **7.3** Facial expression

Appendix L. Scheme for area of interest groups

Colour	AOI group	AOI No.	Telop No.
Red	Telops that directly render dialogue	1-31	1-66
Green	Type 1: identical	1	1
		2	2, 3, 28, 42, 46, 54, 55, 57, 59, 60
		3	5
		4	8, 12, 13, 29
		5	9
		6	21, 37
		7	27
		8	53
		9	56
		10	64
Pink	Type 2: other	11	10, 38
		12	11
		13	19
		14	38
		15	62
Blue	Type 3: omission	16	4, 6, 39
		17	17, 18, 20, 26, 41
		18	22
		19	30
		20	31, 34
		21	32, 35
		22	47
Brown	Type 4: recast	23	7, 16, 65
		24	14, 15, 33, 50, 58
		25	36
		26	43
		27	44, 45, 61, 66
		28	49
		29	51
Purple	Type 5: incongruous	30	23, 40, 48, 52, 63
		31	24, 25
Grey	Faces	32	*39-50
		33	*6-18, 20-30, 32-38, 51-63, 65-72, 74-107, 114-132, 134-161, 163-180, 182-192, 194-206, 208-213, 219-224, 226-261
Orange	Whole screen	34	*1-261

** Shot No. instead of telop No.*

Appendix M. Fixation-data exports

Appendix M.1 Fixation count

Telops that directly render dialogue

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	1553.59	2628.00	785.00	26411.00	1634.00	463.66
Group 2: exchange	20	1781.35	2740.00	279.00	35627.00	1807.50	703.11
Group 3: post-exchange	6	2140.33	2650.00	1396.00	12842.00	2332.50	592.00
Reference group	5	1198.20	1639.00	567.00	5991.00	1166.00	410.68

Type 1: identical

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	390.94	671.00	140.00	6646.00	414.00	137.70
Group 2: exchange	20	436.60	733.00	44.00	8732.00	424.50	153.21
Group 3: post-exchange	6	506.67	687.00	335.00	3040.00	518.50	147.73
Reference group	5	272.60	346.00	156.00	1363.00	265.00	78.96

Type 2: other

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	284.59	492.00	182.00	4838.00	277.00	73.64
Group 2: exchange	20	282.20	411.00	148.00	5644.00	280.50	91.95
Group 3: post-exchange	6	343.17	460.00	214.00	2059.00	357.50	95.79
Reference group	5	280.80	372.00	156.00	1404.00	322.00	86.39

Type 3: omission

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	268.41	448.00	140.00	4563.00	287.00	87.82
Group 2: exchange	20	338.20	601.00	29.00	6764.00	346.00	157.44
Group 3: post-exchange	6	398.67	560.00	226.00	2392.00	411.50	135.03
Reference group	5	174.40	238.00	63.00	872.00	190.00	72.46

Type 4: recast

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	323.24	570.00	125.00	5495.00	300.00	119.95
Group 2: exchange	20	381.75	667.00	41.00	7635.00	386.00	190.70
Group 3: post-exchange	6	495.50	658.00	353.00	2973.00	479.00	121.51
Reference group	5	287.20	394.00	110.00	1436.00	313.00	111.55

Type 5: incongruous

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	286.71	447.00	141.00	4874.00	280.00	91.91
Group 2: exchange	20	342.90	568.00	17.00	6858.00	336.00	154.53
Group 3: post-exchange	6	396.67	551.00	234.00	2380.00	397.00	132.35
Reference group	5	183.40	293.00	82.00	917.00	182.00	77.65

Faces

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	1769.65	2673.00	935.00	30084.00	1813.00	473.94
Group 2: exchange	20	1622.75	2352.00	585.00	32455.00	1654.50	522.51
Group 3: post-exchange	6	1943.67	2961.00	1005.00	11662.00	1917.50	854.24
Reference group	5	2500.40	3737.00	1382.00	12502.00	2315.00	860.74

Whole screen

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	4125.41	6683.00	2991.00	70132.00	3971.00	943.15
Group 2: exchange	20	4345.35	5918.00	2627.00	86907.00	4398.50	986.09
Group 3: post-exchange	6	4929.83	6404.00	2832.00	29579.00	5572.50	1537.96
Reference group	5	4727.40	5942.00	3003.00	23637.00	4623.00	1226.56

Appendix M.2 Total fixation duration

Telops that directly render dialogue

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	115.03	182.65	22.83	1955.53	116.61	35.84
Group 2: exchange	20	129.48	237.57	7.98	2589.70	126.87	54.91
Group 3: post-exchange	6	128.06	172.18	100.19	768.39	120.76	24.74
Reference group	5	62.92	101.54	12.76	314.60	74.35	34.98

Type 1: identical

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	28.55	47.46	3.82	485.30	32.28	9.96
Group 2: exchange	20	31.78	53.13	1.20	635.51	33.11	12.63
Group 3: post-exchange	6	30.49	43.48	22.24	182.94	29.31	7.00
Reference group	5	13.80	23.15	3.75	68.99	13.39	7.20

Type 2: other

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	23.46	34.32	5.90	398.83	23.30	6.93
Group 2: exchange	20	22.74	36.22	4.09	454.86	24.12	7.82
Group 3: post-exchange	6	20.97	28.42	13.24	125.84	21.52	5.04
Reference group	5	14.90	26.09	3.42	74.52	14.63	8.27

Type 3: omission

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	20.51	30.88	4.42	348.73	20.64	6.08
Group 2: exchange	20	24.32	56.36	0.93	486.42	22.25	12.21
Group 3: post-exchange	6	24.79	32.96	19.22	148.76	23.33	5.00
Reference group	5	10.07	16.27	1.48	50.33	12.18	6.36

Type 4: recast

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	21.38	34.88	4.39	363.41	22.91	7.25
Group 2: exchange	20	26.44	50.62	1.26	528.86	26.98	13.54
Group 3: post-exchange	6	28.34	35.89	20.84	170.05	28.21	5.20
Reference group	5	14.55	22.53	2.25	72.76	17.00	8.72

Type 5: incongruous

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	21.13	41.86	4.31	359.27	22.24	9.41
Group 2: exchange	20	24.20	42.72	0.50	484.06	23.47	11.39
Group 3: post-exchange	6	23.47	31.43	17.51	140.80	22.14	5.39
Reference group	5	9.60	14.16	1.85	47.99	12.98	5.65

Faces

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	201.35	270.59	42.48	3422.96	219.86	61.58
Group 2: exchange	20	170.01	290.66	60.05	3400.24	152.22	62.96
Group 3: post-exchange	6	165.68	193.20	134.21	994.05	166.73	22.54
Reference group	5	138.74	181.46	35.31	693.70	169.28	61.31

Whole screen

Participant group	N	Mean	Maximum	Minimum	Sum	Median	Standard deviation
Group 1: pre-exchange	17	387.38	458.03	95.51	6585.50	419.48	91.45
Group 2: exchange	20	374.47	480.84	127.41	7489.50	402.13	91.49
Group 3: post-exchange	6	359.58	439.69	286.46	2157.48	360.33	50.36
Reference group	5	255.50	348.08	73.87	1277.51	276.82	106.52