Colm Caffrey

VIEWER PERCEPTION OF VISUAL NONVERBAL CUES IN SUBTITLED TV ANIME

The article discusses the results of a questionnaire into viewer perception of subtitled culturally marked visual nonverbal cues in TV anime using the series せんせいのお時間 (Doki-doki School Hours). Two groups of students were used for the study, one Japanese learner group, and one non-Japanese speaker group. A smaller study was also carried out using an eye-tracking monitor to gain insights into non-Japanese speakers’ visual attention to visual nonverbal cues. The anime genre is one of the few forms of animation that is as often, if not more often, subtitled as dubbed. Online surveys have shown a preference among anime fans for subtitling over dubbing (Igarashi, 2007). The abundance of on-screen information adds to this genre’s suitability for study into viewer perception of nonverbal cues and the effectiveness of the translation strategies used to tackle them.

Keywords Japanese subtitling; nonverbal cues; viewer perception; TV anime

Introduction: TV anime

The worldwide popularity of anime and manga, Japanese animation and comics respectively, has led to the current situation where Japan produces around 60% of worldwide animation (Shūji, 2006). According to a report from the Japan External Trade Organization (2005) the market for anime content, which includes video games and character merchandising, was estimated to be worth around 13 billion Euro in 2004. Despite the impact of translation on the genre’s worldwide success, there is a paucity of research into audience perception of anime translation. Covering a large range of genres, the anime market can be divided broadly into three categories: feature-length films, TV shows, and video/DVD releases. The TV anime genre, which has experienced an increase in domestic broadcasting within Japan, often shares links with other media such as videogames and manga, or Japanese comics (Natsume, 2000; Kubo, 2004).

Language transfer in TV anime on DVD

There are essentially two forms of language transfer available to anime distributors in North America and the UK, subtitling and dubbing. Here we will define subtitling as...
the ‘prepared communication using written language acting as an additive and synchronous semiotic channel, as part of a transient and polysemiotic text’ (Gottlieb, 2005b: 16). The subtitling process adds a semiotic channel to the audiovisual (AV) content, that of written text on the screen. The spatiotemporal constraints imposed on the subtitler by the medium mean that they usually have a limited amount of time and space to convey the spoken dialogue in text. The advent of DVD technology has enabled the possibility of up to 32 subtitle tracks per DVD. This allows the distributor to make a DVD with several language versions and also to provide various levels of viewer accessibility. The same DVD can hold subtitles for the deaf and hard of hearing and several foreign language options. A further level of accessibility is the provision of alternative subtitles in the same language, falling into line with Gottlieb’s (1998: 248) ‘personal subtitling’. Here, the viewer can choose between styles and levels of subtitling in the same language, so one set of subtitles may contain more culture-specific references for viewers more familiar with the source culture, while others may make explicit, ignore, or change such references. These heightened levels of accessibility can provide the viewer with more possibilities to experience the audiovisual text, and with a higher level of awareness of translation processes, while the distributor has more space to experiment with their subtitles. Some DVD releases of TV anime have an option to view what the North American distributor ADV films call AD Vidnotes. In this mode, a pop-up explanation appears on screen when culture-specific references are shown.

The form of transfer under investigation in this study is interlingual subtitling, which is the subtitling of speech from one language into text in another language.

Visual nonverbal cues, TV anime and subtitling

Visual nonverbal cues

Most studies on the translation of nonverbal information in the field of audiovisual translation (AVT) have been based on dubbing, although some studies have also considered the relationship between subtitling and nonverbal items (De Linde and Kay, 1999; Pettit, 2004). This is perhaps due to the extra leeway often granted to the dubber in changing the spoken dialogue to fit the visual elements of the scene, though a similar leeway is often granted to subtitlers of ‘exotic’ languages, which have few speakers in the target culture. Unlike literary translation, where nonverbal items such as gestures and symbols are rendered from the source text verbal form to the target text verbal form, AVT constantly presents the subtitler with nonverbal items rendered through the aural or visual channel. Anime produced for a domestic Japanese TV audience can cause difficulties when being subtitled for a target Western audience, particularly when the target country is not familiar with the source culture’s visual conventions.

Despite the obvious relationship between the spoken/written text and the image, there have been few studies on the perception of nonverbal items in translated films. This could be due to the difficulty in analysing the relationship between translation and a lot of nonverbal items, or even in defining the terms (Wierzbicka, 1995). These include aspects of nonverbal communication which cross between language and
culture, such as gesture and idioms, and have come to be categorised as ‘borderline features’ (Bucaria and Chiaro, in press) and ‘lingua-cultural drops in translational voltage’ (Antonini, 2005: 214) by researchers into the perception of subtitling and dubbing.

Owing to the complexities involved in the concept of nonverbal communication (cf. Wierzbicka, 1995), a more restrictive term is necessary in the framework of this study. Here the term visual nonverbal cue (VNC) will be used, defined as a nonverbal item appearing in the image of an audiovisual text which has an intended secondary, connotative meaning.

**VNCs in TV anime**

Subtle emotional nuances can be difficult to animate, so animated characters’ emotions are often exaggerated and gestures more pronounced. When investigating VNCs in anime, it is vital that we consider not only their own visual grammar, i.e. the role of the visuals in telling the story, but also that of manga. In both anime and manga, specific iconography is used to denote certain emotions. The complex and reciprocal relationship between these two popular exports of Japanese culture can be seen in their combined history. The first major TV anime hit in Japan was 鉄腕アトム (tetsuwan atomu) in 1963, based on the manga of the same title. This gave life to a trend in the creation of TV anime based on manga, as well as to the simultaneous release of both manga and anime (Natsume, 2000). This interweaving of anime and manga has seen an equal sharing and influence on the visual aspects of both forms, with visual idioms, symbols and icons being shared by both. While some of these are universally comprehensible, others may present more difficulties for the non-Japanese viewer.

Nonverbal communication with the viewer occurs all the time, particularly in animation, where the number of VNCs tends to be very high. This necessitated the scope of VNC to be narrowed for the present study. Thus, VNCs with a secondary meaning in the source culture were chosen, due to the difficulties they present for viewers and subtitlers alike. The use of ‘culture-specific’ as a term to describe them will be avoided in this study, as, although the intended meaning of a VNC may be specific to the source culture and to those familiar with it, the cue itself is not necessarily culturally specific. After all, a cut of a character sneezing when she is being spoken about may seem like an unnatural cut to a Western viewer, but it is certainly not culturally specific. The connotative meaning, which can be somehow identified with the notion of ‘my ears are burning’, however, is specific to certain cultures. For this reason, the term ‘culturally marked’ as used by Howell (2005), may be more suitable because, although the VNCs do have a particular imprint of the source culture (in their connotative meaning), they may not necessarily be specific to the culture.

**Subtitling VNCs**

The strategies used by the subtitler in the excerpts analysed will be categorised according to the three main translation strategies for dealing with nonverbal items in dubbing outlined by Chaume Varela (1997) as *equivalence*, *explicitness* and *total
substitution. Equivalence occurs when the subtitlers leave the VNC as it is, explicitness when they explain it in some way, and total substitution is when the original meaning is flaunted to fit the target text. Although subtitling usually tends to avoid total substitution, due to the constant presence of the source language dialogue, there is a certain freedom in the case of a language such as Japanese. As mentioned above, where there is little exposure to the source language in the target culture, the subtitler often has more freedom, as the viewer is more likely to be fully dependent on the subtitles to understand the dialogue. Moreover, the foreign verbal audio track is there to increase the 'exoticism’ of the viewing experience, which is like ‘watching dangerous animals from behind an armoured glass screen in the zoo’ (Gottlieb, 2005a), although viewers are also made aware of aspects such as the pitch, accent and timbre of various characters. An example of this occurs in The Passion of the Christ (2004), where a language extinct in its spoken form is subtitled. This may seem to be a boon to the subtitler, who is granted extra freedom in their translation of the text, but this freedom can be counterbalanced by the constraints presented by an increased number of cultural signs often found in the image of the ‘exotic’ source text. Here, the subtitler must make a choice whether to tackle or ignore these items, which can often prove more difficult for the subtitler than finding lexical equivalents, particularly when they are referred to in the dialogue or are important to the plot.

We see this issue arising in a scene from the series Doki Doki School Hours. A tall teenager hugs her short teacher. The teenager has previously explained her attraction to her teacher, partially due to the teacher’s petite stature, and as she hugs her, another student comments:

<table>
<thead>
<tr>
<th>Japanese dialogue</th>
<th>あ　海老で鯛は釣れた</th>
</tr>
</thead>
<tbody>
<tr>
<td>[English translation]</td>
<td>Talk about a big fish</td>
</tr>
<tr>
<td>[English subtitle]</td>
<td>and a little shrimp.</td>
</tr>
</tbody>
</table>

As the student says this, the image of a shrimp and a bream appear above the hugging characters. The Japanese dialogue is the equivalent of the English proverb ‘to use a sprat to catch a mackerel’. However, in this case, the image of the shrimp and bream would not allow for the direct translation of the dialogue in a subtitle. Instead, the subtitler explains the presence of the image by drawing attention to the size comparison and avoiding the proverbial meaning of the shorter character attracting the taller one.

**Scene and frame semantic model**

The use of scene and frame semantics (Fillmore, 1977) as a suitable theoretical model for analysing translation strategies in AVT has been established by several scholars (Kussmaul, 2005; O’Hagan, 2006) because:

It allows consideration of the proximity between the image (scenes) evoked by the translator upon reading the source text and that intended by the originator of
the source on the one hand, and how the difference affects the resultant translation (frames) on the other.

(O’Hagan, 2006)

Scene and frame semantics suggests that language users associate real-world scenes with prototypical scenes in their mind and then choose a certain linguistic frame to interpret them. The presence of images and sounds in audiovisual (AV) texts should theoretically help the subtitler create a mental scene in her own mind close to that intended by the source text, thus enabling a translated frame close to that of the original.

Here we will take an excerpt as an example to try to explain the translation and perception process using the scene and frame model. To this purpose, let us imagine that the subtitler watches a scene where a character has a nosebleed, which has come to symbolise sexual arousal in anime and manga, particularly for male characters. Unlike reading literature, where the reader goes from a frame to a scene, for the viewer of an AV product (particularly the viewer of a foreign film) the scene dominates and, despite advances in digital editing processes, it usually must be considered as untouchable by the subtitler. Here the subtitler must decide how to alter the frame in which the scene is presented to communicate the aspects of the original meaning considered relevant to the non-Japanese viewer, and which may not be compensated for in other semiotic channels.

Depending on the translation strategy chosen by the subtitler, the frame presented to the viewer may differ. However, the scene remains constant. This may sometimes result in a clash between the scene and the frame presented, between the visual verbal (in the form of subtitles) and the visual nonverbal (in the form of the image) channels.

One possibility to elicit a viewer’s interpretation of the subtitled text is to ask them to respond to a question about the VNC present in it. So, in the nosebleed example below, if the subtitler does not explain it in any way, the viewer will have to use other semiotic channels to create an interpretative frame of their own. However, a viewer familiar with the scene of a nosebleed symbolising sexual arousal will not need a translation to be provided and their interpretative frame will more likely show awareness of the reasons for the nosebleed. Asking the viewer to answer questions about VNCs is one way of eliciting an interpretative frame which can be compared with the intended meaning of the original.

In AVT, the scene presented for translation is going to be the same as the scene presented to the viewer, with of course the altered frame provided by subtitles. However, depending on factors including the viewer’s knowledge of the source language and culture, the interpretive frame can differ. The interpretive frame is made by comparing the original text plus the subtitles with prototypical scenes in their mind. The strategy chosen by the subtitler can have a strong effect on the viewer’s cognitive scene and resulting interpretative frame. Where a subtitler uses total substitution and the original meaning of the VNC is altered, the target language viewer will more than likely experience a significantly different interpretative frame than the source language viewer of the original text, although missing elements may be compensated for later in a text.
Here we should note that viewers from the same culture can also have different interpretations of various items in an AV text. However, visual idioms and symbolism, which tend to be widely understood within the same culture, may present serious comprehension difficulties to viewers unaware of their connotative cultural or genre-related meaning. By studying the frame of the viewer’s interpretation of nonverbal cues, the scene and frame model allows us to hypothesize how the subtitler’s interpretative frame affects the viewer’s mental scene created while watching the anime.

**Perception studies in AVT**

There has been little research into audience perception of dubbing and subtitling, although mention must be made of several studies into the perception of translated humour (Antonini, 2005; Chiaro, 2004; Fuentes Luque, 2003) and language features in dubbing (Bucaria and Chiaro, in press), as well as more general studies into cinemagoers’ reactions to subtitling (Widler, 2004).

While a large amount of research into the perception of some Japanese VNCs has been carried out in the field of intercultural studies, there have yet to be any studies carried out into the perception of the subtitling of nonverbal cues in anime. Here the viewer is presented with VNCs which have been, in one way or another, mediated by the subtitler, therefore the approach taken by the latter in translating culturally marked VNCs is expected to have an impact on viewer perception of them.

Some may question the need for perception studies of AVT in the area of anime, especially considering their commercial success verified by the high income levels. However, as Bucaria and Chiaro (in press) point out, viewer perception could be a good means of evaluating subtitle quality, especially in relation to such a successful genre. When carrying out a viewer perception study, the intended viewer should also be kept in mind. Having to cater for the needs and expectations of a very vocal fan community, subtitlers of anime products may be forced to comply with different guidelines compared with those of other media genres, where ‘authenticity’ may not be of such importance to the fan. Perhaps their primary goal is not for the subtitled anime to be easily appreciated by all those who view it, but instead for the fan community, whose heightened familiarity with the subject matter may not necessitate as much explanation of culturally and genre marked items in the AV text. After all, their cognitive scene will differ greatly from that of a viewer unfamiliar with the visual grammar of anime and the culture of Japan. Thus, the frame desired by the fan may be the one which the distributor aims to create, keeping ‘Japan familiar’ viewers in mind.

On a pragmatic level, studies on the perception of subtitled anime also help to determine whether the cultural and aesthetic sensibilities being portrayed by the subtitles are those which were intended by the original product, because as viewers ‘we have a right to know how translators do what they do, as specialist professionals in conveying and in part recreating, other people’s messages’ (Candlin, 1990: ix–x). This issue can gain importance in the case of cultures less familiar to the viewer, where the subtitler’s role as mediator between the viewer and the source text, indeed the source culture, is more pronounced.
Case study: an experiment

The experiment here presented started out by adapting the design of several studies (mentioned above) into the perception of translated humour, which were carried out under Professor Delia Chiaro at Bologna University with the use of a questionnaire. The purpose of the experiment was to investigate the perceived and actual understanding of culturally marked VNCs in subtitled scenes from a TV anime series. The study’s focus on VNCs was a factor in editing the questionnaire, and some elements were adopted from investigations into the perception of nonverbal items in the field of dubbing and intercultural studies (Bucaria, 2005; Jungheim, 2004).

Methods and tools

Participants

The sample frame of the questionnaire was obtained from a selection of students from Dublin City University. A group of Japanese learners from beginners’ courses were surveyed as well as non-Japanese speakers, so as to make a comparison between their perceptions of VNCs. The Japanese learners were expected to have a higher level of cultural awareness with regards to Japan, while their language level was low enough to ensure reliance on the subtitles. There were a total of 58 respondents, with 30 Japanese learners and 28 non-Japanese speakers.

From the above participants, two female and two male undergraduate multimedia students from the non-Japanese speakers group were recruited to explore attention distribution between subtitles and VNCs with eye-tracking.

The study was carried out using three interdependent instruments, which will now be explained under the following headings:

- TV anime excerpts
- Eye tracker
- Questionnaire

TV anime excerpts

*Doki Doki School Hours* (DDSH) was first broadcast in 2004 as sensei no ojikan on TV Tokyo, a channel that broadcasts over half of the weekly animation shows being aired on terrestrial television networks in Japan, as well as some of the world’s most successful TV anime series like *Pokémon*, *Yu-Gi-Oh!* and *Naruto* (TV Tokyo Corporation, 2006). Volume one of the North American DVD release from 2005, distributed by Geneon with an English adaptation script by Gabrielle Gulko and an English subtitle script by Kevin McKeown, was used for the study. It contains the first four episodes of the series, which are based on the first collected volume of the manga of the same name published by Takeshobo in 1999.

DDSH is a Japanese high school comedy. It revolves around a class whose teacher, Suzuki Mika, is in her twenties but looks like a young child. The series is made up of
short sketches surrounding the teacher and her students, who conform to a variety of high school stereotypes. The rationale for choosing this series was the likelihood of participants being unfamiliar with it and the large volume of culturally and genre marked visual language it contains. This can be linked back to the series relation with the manga predecessor, including the transference of the manga tradition of onomatopoeic and mimetic expressions appearing as text on-screen.

Choosing the excerpts for investigation was a crucial aspect of the study. As the series is quite closely linked to the manga it is based on, scenes tend to be short and equivalent to one or two *yon-koma* strips, making for convenient extraction. When choosing excerpts, three principle categories of nonverbal item were considered, two of them noted alongside paralanguage by Chaume Varela (1997) as possibly causing translation difficulties:

1. **Cultural signs.** Nonverbal cultural signs can be found in vast amounts in audiovisual texts of this kind, so that the task of the subtitler is to preserve as much as possible the coherence between world knowledge expressed by these signs, which can include colours, places, foods, etc., and words. Chaume Varela (1997: 324) points out the difficulties which can be encountered when dealing with cultural signs noting that ‘exotic cultures may find real problems to understand certain signs, scripts or schemata’.

2. **Kinesics.** The fact that many kinesic signs can be culturally marked means that the subtitler may have difficulties in reconciling on-screen actions with the subtitle translation. These signs include gestures and facial expressions, both of which may have significantly different meanings depending on the culture. As the current investigation is concerned with culturally marked cues, only those kinesic signs with a specific connotative meaning in a Japanese context were considered.

The paralanguage factor was disregarded while gathering the excerpts, as it was considered irrelevant for this study. However, the highly specific visual characteristics of anime led to the creation of a third category for investigation:

3. **Genre signs.** When observing the role of VNCs in this series, it was vital to consider the visual grammar of anime, but also that of its print media partner, manga. This category was created to include those VNC that have their origins in anime or manga and which have a specific connotation with these genres.

The excerpts selected for inclusion in the study were chosen according to their content, following the procedure as detailed. Firstly, 15 excerpts were extracted from the series and shown to a panel of five non-Japanese speaking postgraduate students. On the basis of preliminary feedback, a decision was made to use 10 excerpts for the pilot study. These excerpts were then digitised for use with the eye-tracking monitor.

Questions were then elaborated for these 10 excerpts and an online pilot questionnaire was created and tested on two participants for usability on the campus-wide online platform, Moodle, in order to check for ambiguities and problems in the structure or content of the questions. Feedback from the questionnaire brought to
light the low download speeds of excerpts as an issue for the participants, therefore a
decision was made to decrease the number of excerpts to increase the probability of
questionnaire completion. Bearing this in mind, the initial 10 excerpts were further
reduced to six, to equally represent the three categories (a–c) outlined above.

As well as encompassing the three categories of VNC under study, the VNCs in
each scene had undergone one of the three translation strategies mentioned above: (i)
equivalence, (ii) explicitness or (iii) total substitution. It was expected that, the more
explicative the strategy, the higher the number of correct answers would be, and that
the use of total substitution would lead to a lower number of correct answers. This is
in line with the core concerns of scene and frame semantics, where more explicative
strategies should make it possible for the viewer to experience a cognitive scene closer
to that experienced by a source culture viewer. However, where the original meaning
is changed significantly, there is room for possible misunderstanding with the viewer’s
cognitive scene and the resulting interpretative frame differing significantly from what
was intended in the source text.

Questionnaire

Procedure

Participants either took an online electronic form of the questionnaire (n = 6) or did a
paper version (n = 52). They were shown the excerpts and, after each one, they were
asked to rate their understanding on a Likert scale from zero to six, with zero
representing complete lack of understanding and six representing total understanding.
They were then asked a short open question to test their understanding, which was
subsequently coded yes or no.

Results

Overall difference between perceived and actual understanding. Looking at Figure 1, we
can see that generally declared understanding was higher than actual understanding.
Declared understanding was measured as follows: participants who self-rated their
understanding of the VNC in question between 3 and 6 were classed as having
declared understanding, while those who selected between 0 and 2 were classed as
not having declared understanding.

The category which shows the highest discrepancy between declared and actual
understanding is cultural items. Total substitution was used in both of the excerpts in
the cultural items category, which may go some way in explaining the discrepancy
between perception in understanding and actual understanding, as the semantic frame
presented by the subtitles alters the original meaning.

The first excerpt containing a cultural item revealed the highest discrepancy of all,
with over half of the participants stating that they understood the game that was being
played, but in fact only 7% answering correctly. In the scene, a student wants to
arrange seating by playing the king game, a game where all the participants choose a
numbered stick, bar one, who chooses the ‘king stick’, and subsequently gives orders
to various numbers, not knowing who has which stick. The subtitler chose a
substituting strategy, replacing the king game with drawing straws, which looks somehow similar. In doing so, he also disregarded the Japanese dialogue: 

<table>
<thead>
<tr>
<th></th>
<th>Japanese dialogue</th>
<th>[Literal translation]</th>
<th>[English subtitle]</th>
</tr>
</thead>
<tbody>
<tr>
<td>王様ゲームで決めよう 王様ゲーム</td>
<td>Let’s choose with the king game, the king game</td>
<td>Let’s choose by drawing straws! The winner decides!</td>
<td></td>
</tr>
</tbody>
</table>

However, the Japanese text which appears on-screen simultaneously with the dialogue is captioned with the equivalent of ‘King Game’. However, the low amount of answers which actually said king game is an indicator that participants relied more on the subtitle than the caption, or were perhaps more confident in the familiar cognitive scene presented by the subtitle-induced frame of drawing straws.

**Difference between Japanese learner and non-Japanese speaker groups for perceived and actual understanding.** A comparison was then made between both groups, students learning Japanese (group A, n = 30) and non-Japanese speaking students (group B, n = 28), in terms of the percentages of declared and actual understanding. The data in Figure 2 shows that the Japanese learner group had a higher level of perceived understanding for each VNC bar one, and had a higher level of actual understanding for each VNC. We find that only one VNC was understood correctly by more than half of the nonspeaker group: the second kinesic item. All of the others were correctly understood by less than a third of the non-Japanese speaking participants.

For both groups, the percentage of declared understanding was higher than actual understanding, except for the second kinesic excerpt. This excerpt showed an example of explicitation, where the subtitle text explained the kinesic VNC in question. The scene explains how one student got the nickname ‘Old Man’. He is

![Figure 1](image-url)
shown reading the horse betting page of a newspaper, sticking his finger in his ear and then rubbing his face with a hot towel, all actions stereotypical of a middle-aged salary man. A student points at him, saying:

[Japanese dialogue]
ああ、おやじだろう

[English subtitle] See? Just like an old man!

Here the Japanese dialogue, literally meaning ‘Well, an old man, isn’t he’, is changed to ‘Just like an old man’, with the function of further explaining the similarity of the actions of the student to those of an older man. This excerpt scored the highest percentage of correct answers overall (72%), and was the only one which over two-thirds of participants answered correctly. This could be to do with the global nature the actions portrayed as those of an ‘old man’, with the exception perhaps of the use of a hot towel, or maybe the use of explicitation in the subtitle.

The final two excerpts both showed genre-marked VNCs in the form of a vein, signifying anger, in the first excerpt and a nosebleed in the second. As would be expected, the Japanese learners, who could be more familiar with the nonverbal genre conventions found in anime, had higher declared and actual understanding of these symbols. In both of these excerpts, declared and actual understanding were closer than in any other of the excerpts, so it seems that participants’ self-assessment of their own understanding was quite accurate in this area.
Perceived and actual understanding by subtitling strategy. Table 1 shows the strategy used by the subtitler to deal with the VNCs in each clip, as well as the percentages of declared and actual understanding for all participants ($n = 58$). Items where over half of participants answered correctly are shown in bold, and those where less than 10% answered correctly are shown in italics. Both items where total substitution was used saw the lowest scores and the largest discrepancy between perceived and actual understanding.

Eye-tracker

An eye-tracking monitor integrated into a TFT (thin film transistor) display was used to examine how viewer attention was distributed between the VNCs and on screen text. Owing to the vast amounts of data generated by the monitor, a significantly smaller sample of participants was used for this part of the evaluation. Viewers were shown the excerpts in the dedicated Advanced Translation Research laboratory at the School of Applied Language and Intercultural Studies in Dublin City University. The laboratory has controllable lighting and a comfortable setting to emulate a personal viewing setting.

Experimental design

The laboratory was set up for the experiment with a constant level of light and participants were asked to enter one at a time. Following the procedure of d’Ydewalle et al. (1991), the subjects were told they were involved in a study on emotional reactions to the excerpts, being measured through their pupillary contractions and dilations. They were then told a brief description of the series used. The experimenter assisted in calibrating the monitor to the participants’ eyes after they seated themselves. The latter were also told they would be shown some excerpts and asked to answer short questions after seeing them. After completion, it was ensured they had been naïve to the real purpose of the experiment before they were enlightened.

Results

Since the eye-tracking experiment was carried out on a small number of participants, the raw data for each participant is here reported instead of using summary statistics. The analyses of the excerpts began with the definition of two areas of interest: subtitle

<table>
<thead>
<tr>
<th>Excerpt</th>
<th>Strategy</th>
<th>Declared understanding</th>
<th>Actual understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinesic 1</td>
<td>Equivalence</td>
<td>84</td>
<td>52</td>
</tr>
<tr>
<td>Kinesic 2</td>
<td>Explicitation</td>
<td>55</td>
<td>72</td>
</tr>
<tr>
<td>Cultural item 1</td>
<td>Total substitution</td>
<td>52</td>
<td>7</td>
</tr>
<tr>
<td>Cultural item 2</td>
<td>Total substitution</td>
<td>34</td>
<td>3</td>
</tr>
<tr>
<td>Genre sign 1</td>
<td>Equivalence</td>
<td>38</td>
<td>31</td>
</tr>
<tr>
<td>Genre sign 2</td>
<td>Equivalence</td>
<td>41</td>
<td>36</td>
</tr>
</tbody>
</table>
area and VNC area. The only scenes from the excerpts which were analysed were those where a VNC was present. Although some eye data was missing from the recordings, this does not affect our results, as the gaze information is only required while in the VNC or subtitle area. Missing data occurs whenever the eye-tracking monitor cannot make a recording and can be caused by a variety of factors such as the participant looking away from the monitor or closing their eyes for extended amounts of time. The information we have used here is the fixation data. A fixation was defined as the gaze resting on an area of 30 pixels diameter for at least 80 ms.

Table 2 shows the fixation count for both areas of interest. Owing to the low number of participants, it is impossible to draw any decisive conclusions from these results. We should also note that length and amount of subtitles varied for each excerpt, as did the display time for each VNC, accounting for the variance in fixation figures between excerpts. However, in most cases, the participants fixated on the subtitle and VNC in the excerpts, and for all those questions which were answered correctly, both the subtitle and VNC were fixated upon. The low number of correct answers obtained by the four participants in the eye-tracking study also limited the amount of data which could be gathered on correct answers, with a total of 80% of clips misunderstood.

TABLE 2 Fixation count on subtitle and VNC areas of interest

<table>
<thead>
<tr>
<th>Excerpt</th>
<th>Kinesic 1</th>
<th>Kinesic 2</th>
<th>Cultural 1</th>
<th>Cultural 2</th>
<th>Genre 1</th>
<th>Genre 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male 1</td>
<td>Sub fixation</td>
<td>11</td>
<td>5</td>
<td>13</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>VNC fixation</td>
<td>14</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Male 2</td>
<td>Sub fixation</td>
<td>X</td>
<td>4</td>
<td>15</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>VNC fixation</td>
<td>X2</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Female 1</td>
<td>Sub fixation</td>
<td>43</td>
<td>5</td>
<td>13</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>VNC fixation</td>
<td>8</td>
<td>24</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Female 2</td>
<td>Sub fixation</td>
<td>38</td>
<td>8</td>
<td>12</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>VNC fixation</td>
<td>16</td>
<td>22</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

Conclusion

The results of this study empirically proved that, as expected, a higher percentage of the Japanese learning group perceived understanding and actually understood the VNCs than the non-Japanese speaker group. This is particularly noticeable in the scenes which showed kinesic and genre specific VNCs, where over half of the participants learning Japanese answered questions on both excerpts correctly. Also of note is that only one VNC was correctly recognized by more than one-third of the non-Japanese speaker participants, and equally that overall perceived understanding was higher than actual understanding. While the enhanced familiarity of Japanese learners with Japanese culture and the visual language of TV anime go some way towards explaining the higher scores in kinesic and genre-marked VNCs, cultural items were understood by no more than a tenth of all participants. If we link this back to the use of total explicitation in the subtitles, we seem to be able to support the
view that the altered semantic frame offered by subtitles can change the viewer’s perception of the scene in which they appear, and shows the power which the subtitler can have in shaping the viewer’s perception of the source text, and indeed the source culture.

The eye-tracking data showed that, in the majority of cases, when subtitles were present on screen with the VNC in question, the number of fixations on the subtitle area was higher than that in the VNC area. Therefore, the presence of subtitles seems to be a factor in distracting visual attention from the VNC area.

In conclusion, we should ask ourselves the importance of whether or not viewers comprehend these and other VNCs in a translated AV text. When asked their preferences for how culturally marked items should be dealt with in subtitling anime for DVD, just under half of the participants answered that they should be explained with a note at the top of the screen. This strategy is already being used in fansubs, versions of anime usually freely available online which are produced, translated and subtitled by fans (Díaz Cintas and Muñoz Sánchez, 2006). While these notes are certain to clear up ambiguities concerning the meaning of culturally and genre-marked VNCs, further research should be carried out into their effects on viewers’ reading speeds (Caffrey, 2007). An abundance of text on screen could place unwanted stress on the viewer and present difficulties in following the main narrative. And although this innovative method of subtitling has begun to make its way into commercially released TV anime on DVD, it has still to be transferred in any major way to feature-length anime or other audiovisual entertainment genres. Perhaps the key to a more culturally aware viewing experience could be provided by investigating and encouraging the use of other experimental subtitling methods.

Notes

1 This research is funded by the Irish Research Council for the Humanities and Social Sciences.
3 Tobii Eye Tracker 1750: sampling rate of 50 Hz, records the x- and y-axis of the subject’s point of regard on the computer screen as well as their pupil size.

References


**DVD**

*Doki doki school hours: 1st hour.* USA: Geneon Entertainment Inc., 2005.

---

**Colm Caffrey** is a doctoral research student at the School of Applied Language and Intercultural Studies (SALIS), Dublin City University, where he gained a first class honours, majoring in Japanese. Supported by an Irish Research Council for the Humanities and Social Sciences (IRCHSS) scholarship, his dissertation examines the viewer perception of nonverbal signs in subtitled TV anime (Japanese animation). Prior to taking up the doctoral study, Colm was involved in an Enterprise Ireland funded project at SALIS which looked into the feasibility of using example-based machine translation (EBMT) in translating subtitles for films on DVDs. He has recently translated Kei Oishi’s 死人を恋う (*In Love with the Dead*) for script development purposes for the first Irish–Japanese co-produced film. His research interests are anime and manga translation, horror translation, translation technology, perception of translated material and audiovisual translation. [email: colm.caffrey@dcu.ie]