JoSTrans
The Journal of Specialised Translation

The dubbing effect: an eye-tracking study on how viewers make dubbing work

Pablo Romero-Fresco, University of Vigo and University of Roehampton

ABSTRACT

Although dubbing is regularly criticised for its artifice and its manipulation of film sound, it has proved to be the preferred modality of audiovisual translation for millions of viewers. Research in this area has explored at length the way in which the professionals involved in dubbing make it work. What has been overlooked so far is the cognitive process undergone by the viewers to make it work. In order to explore this issue, this paper starts with a discussion of several aspects that may be relevant to the perception and overall reception of dubbing, including cultural arguments on habituation, psychological and cognitive notions of suspension of disbelief and perceptual phenomena such as the McGurk effect. It then goes on to compare, with the help of eye-tracking technology, the eye movements of a group of native Spanish participants watching a clip dubbed into Spanish featuring close-ups with (a) the eye movements of a group of native English participants watching the same clip in English and (b) the eye movements of the Spanish participants watching an original (and comparable) clip in Spanish. This analysis is complemented by data on the participants' comprehension, sense of presence and self-perception of their eye movements when watching these clips. The findings obtained point to the potential existence of a dubbing effect, an unconscious eye movement strategy performed by dubbing viewers to avoid looking at mouths in dubbing, which prevails over the natural way in which they watch original films and real-life scenes, and which allows them to suspend disbelief and be transported into the fictional world.

KEYWORDS

1. Introduction

Despite being, perhaps along with voiceover, the most criticised (and even vilified) audiovisual translation (AVT) mode, there is little doubt that, generally speaking and from different viewpoints, dubbing works. It is still the preferred form of access to foreign-language audiovisual content for millions of viewers in countries such as Spain, Italy, France and Germany and the preferred choice to translate cartoons and children's films in subtitling countries (Chaume 2013). Its success is not only commercial, as recent research shows that dubbing is also a very effective translation mode from a cognitive point of view (Wissmath *et al.* 2009; Perego *et al.* 2016). Despite the artifice involved in replacing the original actors' voices for other voices in another language, it seems that (habituated) dubbing viewers still manage to suspend disbelief and become immersed in the fiction of film (Palencia 2002).

Research in AVT has devoted a great deal of attention to describing and analysing the work of professionals in the dubbing industry, such as translators, dialogue writers, actors and directors (Ávila 1997; Chaume

2004; Pavesi 2009; Sánchez Mompeán 2017; Spiteri Miggiani 2019). However, very little has been written from the point of view of dubbing viewers (Ameri *et al*. 2017; Di Giovanni 2018). How do we watch a dubbed film? How do we manage to suspend disbelief without being distracted by its artificial nature and by the mismatch between audio and visual elements? In short, what cognitive mechanisms do we activate to make dubbing work?

The aim of this paper is to answer these questions by analysing, with the help of eye-tracking technology, the viewing patterns of spectators watching dubbed and original films. This analysis is complemented by a discussion of other aspects that may be relevant to the perception and overall reception of dubbing, including cultural arguments concerning habituation, psychological and cognitive notions of suspension of disbelief and perceptual phenomena such as the McGurk effect (McGurk and MacDonald 1976).

2. Habituation and threshold of acceptability

Zabalbeascoa's (1993: 248) view that "dubbing and subtitles […] are a question of national habit and taste" still applies a few decades later. Research shows that audiences seem to prefer the translation method they are most familiar with (Luyken *et al.* 1991; Kilborn 1993; Koolstra *et al.* 2002; Eppler and Kraemer 2018). This may also explain the harsh criticism directed at dubbing by scholars and professionals who have not been brought up with this practice, which has been described as "a kind of cinematic netherworld filled with phantom actors who speak through the mouths of others" (Rowe 1960: 116), "a stepchild of translation at best and no true son of literature at all" (*ibid.*: 117), "the wedding of the phonetic beast to the literary beauty" (*ibid.*: 117) and "a monster which combines the splendid features of Greta Garbo with the voice of Aldonza Lorenzo" (Borges 1945: 88). However, important as it may be, the notion of habit has often been used as a blanket argument that prevents us from having a more in-depth knowledge of the factors that account for the dubbing viewers' acceptance of this type of translation.

A useful concept in this regard is Gunning's (2003) idea of habituation. Referring to the first exposure to new technologies, and thus applicable to film, Gunning (*ibid.*: 44) explains that audiences go from wonder to knowledge to habituation and automatism, with the outcome of this habituation being "to render us unconscious of our experience." Wonder, "the first of all passions" (*ibid.*: 15), draws our attention to the new technology as something that astounds us by performing in a way that seemed unlikely or magical before. This gives way to curiosity to understand how it works (knowledge), habituation after frequent usage and finally unconscious automatism. The case of film may be slightly different.

If consumed from a very early age, the sense of wonder is not necessarily followed by knowledge. When they are first exposed to film, children normally have no knowledge of the artifice involved in cinematic fiction, which means that they go straight from wonder into habituation and automatism. By the time they learn about the prefabricated nature of cinema, film viewing has already settled as an unconscious experience whose enjoyment requires not questioning the reality of what they are seeing, that is, suspending disbelief. Crucially, dubbing audiences are exposed to both original and dubbed films from an early age. They are astounded by the magic of cinema (wonder), regardless of whether or not it is dubbed. The artifice of dubbing (the mismatch between audio and visuals, the almost inevitable lack of total synchrony, even in high-quality dubbing, etc.) is overlooked along with the artifice of cinematic fiction, as they go from wonder to habituation and unconscious automatism. By the time dubbing audiences learn about dubbing (just as when they learn about film), they have already internalised how to watch it without questioning it. In other words, getting used to dubbing, when it happens at an early age, is simply part of the (unconscious) process of getting used to film.

Closely linked to the notion of habit or habituation is that of tolerance. Even if a particular audience is used to dubbing, there is a tolerance threshold that must be respected with regard to at least two of the key dubbing constraints: synchrony and the naturalness of the dialogue. According to Rowe (1960: 117), this tolerance threshold may vary across countries:

> American and English audiences are the least tolerant, followed closely by the Germans. […] The French, staunch defenders of their belle langue and accustomed to the dubbing process since those early days when rudimentary techniques made synchronization a somewhat haphazard achievement, are far more annoyed by slipshod dialogue than imperfect labial illusions. To the Italians, the play's the thing and techniques take the hindmost, as artistically they should.

A more updated take is provided by Chaume (2013: 15), who refers to a "threshold of acceptability" that should not be overstepped, for which it is necessary to adhere to a series of quality standards (Chaume 2007): compliance with synchronisation norms, the writing of credible and realistic dialogue, coherence between what is heard and what is seen, fidelity to the source text, technical adequacy of sound recording and appropriate performance and dramatisation of the dialogue. If these quality standards are met, the illusion of authenticity, or the illusion of an illusion involved in dubbing (Caillé 1960: 108), can still be maintained, thus allowing the dubbing audience to suspend disbelief and become immersed in the fiction.

3. The suspension of disbelief

The notion of suspension of disbelief was originally coined in 1817 by the poet and philosopher Samuel Taylor Coleridge (in Parrish 1985: 106), who suggested that if a writer could provide a fantastic tale with a "human interest and a semblance of truth", the reader would suspend judgement

concerning the plausibility of the narrative. This term has since been used for film (Allison *et al*. 2013) and AVT (Bucaria 2008). Pedersen (2011: 22) applies it to subtitling, calling it a "contract of illusion" or tacit agreement between the subtitler and the viewers where the latter agree to believe "that the subtitles *are* the dialogue, that what you read is actually what people say." In his definition of suspension of disbelief, Chaume (2013: 187) makes explicit reference to dubbing:

> a term referring to the reader/viewer's ability or desire (or both) to ignore, distort or underplay realism in order to feel more involved with a videogame, a film, or a book. It might be used to refer to the willingness of the audience to overlook the limitations of the medium (for example, that a film is dubbed), so that these do not interfere with the acceptance of those premises.

Closely related and also applied to dubbing are the notions of suspension of linguistic disbelief, i.e. "the process that allows the dubbing audience to turn a deaf ear to the possible unnaturalness of the dubbed script while enjoying the cinematic experience" (Romero-Fresco 2009: 68-69), and that of suspension of paralinguistic disbelief (Sánchez Mompeán 2017), the same process applied to the unnatural intonation sometimes found in dubbed films. According to these views, then, the suspension of disbelief works at different levels to allow viewers to be immersed in the dubbed fiction without being distracted by its prefabricated nature or by the potential lack of naturalness of its language or intonation.

From a psychological standpoint, the notion of suspension of disbelief is normally tackled as one more factor in a complex set of elements used to describe our involvement in different kinds of narratives. Drawing on Busselle and Bilandzic (2009), Fresno (2017) explains that in order to engage with a story, viewers must understand it and become immersed in it (Figure 1). In turn, this requires two prerequisites: interest in the filmic experience, and suspension of disbelief or willingness to participate in it. Comprehension does not involve understanding every element in a film plot but just enough for the viewers to create and update mental representations of the fictional world, which, as per Johnson-Laird's (1983) Mental Model Theory, leads to comprehension. The immersion or psychological involvement with the narrative is facilitated by feelings of flow (Csikszentmihalyi 1990), when the viewer is absorbed in the act of watching fiction; transportation, when "all mental systems and capacities become focused on events occurring in the narrative" (Green and Brook 2000: 701); presence (Biocca 1997), the feeling of being in a mediated space different to where your body is located; and finally the viewers' disposition towards the characters (Zillmann 1994; Raney 2004). While it is harder to argue, as per identification theories (Cohen 2001), that viewers experience the events on screen as if they were characters (which would mean that they have the urge to phone the police if a character is in danger), disposition theories explain how viewers develop affective propensity towards the characters,

depending on whether they like them more or less or feel closer to or further away from them.
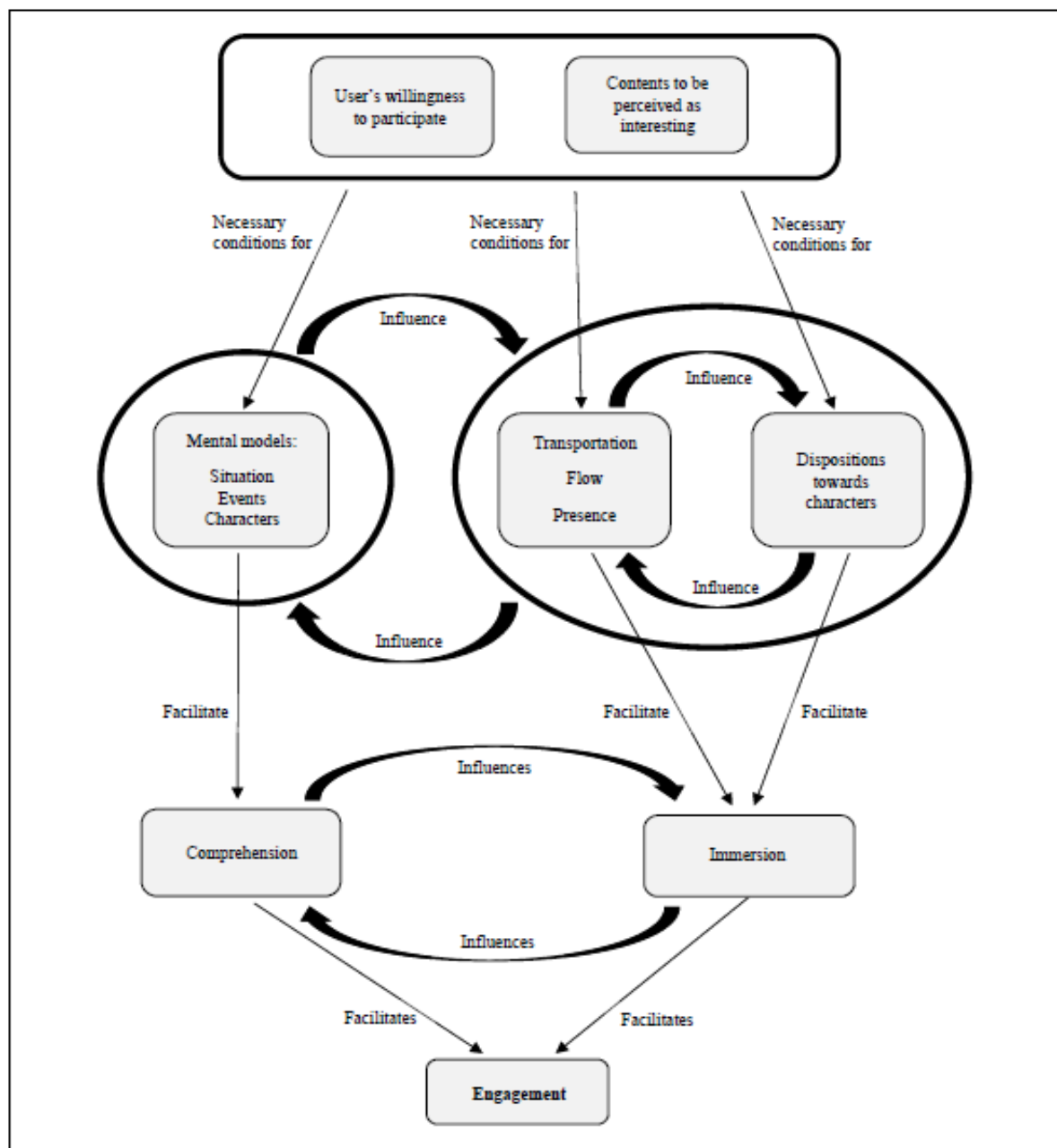

Figure 1. Facilitators of engagement (Fresno 2017: 22)

It would thus seem that, when first exposed at an early age to dubbed films, viewers may feel a sense of wonder that leads to habituation and to an automatic and unconscious engagement with the dubbed fiction, facilitated by their ability to suspend disbelief, their interest in the story, some degree of comprehension of the plot and a sense of immersion that involves feelings of flow, transportation and presence. This process of engagement is not affected by the discovery, years later, of the artifice involved in dubbing, since by then this path to engagement has already been unconsciously internalised as part and parcel of the process of watching film. However,

the question remains as to whether this process can explain why and how dubbing viewers do not seem to be put off by the general discrepancy between audio and visuals or by the lack of synchrony in the dubbing actors' lip movements, which is likely to occur to a greater or lesser degree even in high-quality dubbed films. The McGurk effect, tackled in the next section, would suggest otherwise.

4. The McGurk effect

The McGurk effect (1976) is generally regarded as one of the most powerful perceptual phenomena demonstrating the interaction between hearing and vision in speech perception. It is described by Smith *et al*. (2013) as "an auditory illusion that occurs when the perception of a phoneme's auditory identity is changed by a concurrently played video of a mouth articulating a different phoneme." A typical example would involve the audio of a given phoneme (such as */ba/*) dubbed over a speaker whose mouth is visually articulating another phoneme (such as */ga/*). Most subjects will report hearing */da/* even though the only sound that is heard is */ba/*. Discovered by Harry McGurk and John MacDonald in 1976, this phenomenon shows that speech perception is multimodal and that vision can often be more important than audio in the perception of sounds. From a neurological standpoint, the McGurk effect shows that information from the visual cortex instructs the auditory cortex which phoneme to 'hear' before an auditory stimulus is received (Smith *et al*. 2013). This is generally regarded as a robust effect, i.e. knowledge about it does not seem to eliminate its illusion. The effect has been shown to apply under very different conditions, including different viewers' profiles (Rouger *et al.* 2008), audiovisual cross-dressing (combination of female faces and male voices) (Green *et al*. 1991), cross-cultural comparisons (Rosenblum 2010) and even speakers standing on their heads (Green 1994). Partly due to this phenomenon and to the prevalence of vision in the perception of sound, Navarra (2003) shows that even full sentences are difficult to process when there is a mismatch between visuals and audio, given that the viewer's attention is unavoidably directed to the asynchronous lip movements.

This begs the question of how dubbing can possibly work for the viewers if (a) the McGurk effect is so robust that it works across cultures, languages and in the most varied contexts and (b) the mismatch between visuals and audio in onscreen faces (which to a greater or lesser degree is inevitable in dubbing) normally draws the viewers' attention to the asynchronous lips and makes it difficult to understand and process single words and even sentences. As suggested by Evan (2011: online), "experimental psychologists should investigate how viewers manage to switch off the lip-reading without even being aware of what they are missing." Could it be that dubbing viewers are amongst the few individuals who have managed to switch off the McGurk effect so as not to be distracted by the asynchronous combination of sound and image? Have they found a way to avoid being put off by the mismatch between lips and audio or do they

simply not look at the lips? Should the latter be true, is this an unconscious mechanism and can the above-mentioned early-acquired habit of viewing dubbed films and the ability to suspend disbelief account for this?

The following sections and the experiment presented in this article aim to provide answers to these questions.

5. Eye tracking and face viewing

Before presenting the experiment conducted for this article, it is important to review what has been learnt so far regarding how we normally look at faces, and especially the viewers' distribution of attention between eyes and mouth.

Early studies (Buswell 1935; Yarbus 1965/1967) and also more recent research on face processing and the perception of gaze (Langton *et al*. 2000; Birmingham and Kingstone 2009) have shown that we tend to focus on faces and, more specifically, on eyes, when looking at other human beings. This may be partly explained by the visual saliency and social importance of eyes (Senju and Hasegawa 2005; Senju *et al*. 2005). However, most of this research has focused on static images, rather than dynamic viewing. Recent research performed on dynamic face viewing suggests that this attention bias may be task-dependent and not exclusive to the eyes (Gosselin and Schyns 2001). Buchan *et al*. (2007) found that their participants' gaze was directed to the eyes when asked to perform emotion judgements and to the mouth when asked to recognise speech. In a recent study aiming to identify what controls gaze allocation during face perception, Võ *et al.* (2012: 12) concluded that there is no such thing as a general bias to look at someone's eyes and that, at least during dynamic face viewing, "gaze follows function." In other words, we seem to adjust our gaze allocation dynamically "for the purpose of seeking information on an event-to-event basis" (*ibid.*: 11). In their study, conducted with eighty-eight participants watching videos with close-ups of different people speaking, the mouth attracted as much as 34% of the gaze allocation. This is in line with the findings obtained by Foulsham and Sanderson (2013), who found a distribution of 71% on the eyes and 29% on the mouth in dynamic face viewing with speaking faces. The percentage of time fixating the mouth has been shown to increase when there is background noise (Buchan *et al*. 2012), low linguistic competence (Robinson *et al*. 2015) or poorly synched lips (Smith *et al* 2013), which is not too dissimilar to what happens in dubbing.

In contrast with the intense scholarly activity devoted to the analysis of static and dynamic face viewing, the application of eye tracking to dubbing is still in its infancy. Vilaró and Smith (2011) compared the gaze behaviour of viewers watching an animated film in the original English audio condition, a Spanish language version with English subtitles, an English language version with Spanish subtitles and a final version dubbed into Spanish

without subtitles. The participants were English speakers who did not know Spanish. The results of the study show evidence of subtitle reading in all conditions (even when they were in Spanish and therefore unhelpful for the participants) and a great deal of similarity in the exploration of peripheral objects. Visual and verbal recall proved similar across the different conditions except for the version dubbed into Spanish, whose poorer results are to be expected given that the participants could not understand the Spanish dialogue. In a recent study, Perego *et al*. (2016) used eye tracking and behavioural measurements to analyse the differences in the visual, cognitive and evaluative reception of two subtitled and dubbed films with two different degrees of complexity. Their results confirm the cognitive efficiency and positive reception of both AVT modalities but also that complex audiovisual material may require extra effort from the viewers so as to accelerate their reading process. To our knowledge, no research has yet analysed and compared how viewers watch faces in original and dubbed films. This is the aim of the experiment presented in this article, whose findings, along with the above-included discussions on habituation, suspension of disbelief and engagement, intend to provide a picture of how viewers make dubbing work.

On the one hand, it may be reasonable to expect dubbing viewers to allocate an unusually high amount of attention (perhaps more than the above-mentioned 30%) to the characters' mouths, as suggested by (a) what has been learnt so far about the perception of speaking faces, (b) the mouth bias triggered by speaking faces with imperfectly synched audio and (c) the focus on lips caused by the McGurk effect in a situation of mismatch between image and audio.

However, excessive focus on the characters' mouths may also put off dubbing viewers, making it difficult for them to suspend disbelief and engage with the film. As a result, the hypothesis for this experiment is that given our tendency to (a) lip read and be confused by asynchrony as per the McGurk effect and (b) look at both eyes and mouth in moving faces, we have adopted an unconscious strategy not to look at mouths in dubbing (because there is no useful information to obtained from there) in an attempt, aided by an early acquired and subconsciously internalised dubbing viewing habit, to suspend disbelief and be engaged with the dubbed fiction.

6. The experiment

In order to test this hypothesis, an analysis is presented here of the effect of dubbing on the viewers' eye movements and, more specifically, on their distribution of attention between eyes and mouth when watching faces in close-ups. For this purpose, the eye movements of a group of native Spanish participants watching a clip dubbed into Spanish featuring close-ups was compared with (a) the eye movements of a group of native English participants watching the same clip in English and (b) the eye movements

of the Spanish participants watching an original, and comparable, clip in Spanish. This analysis on the distribution of attention was complemented by data on the participants' comprehension, sense of presence and awareness/perception of eye movement when watching these clips.

6.1. Materials

The first stimulus video was the 6-minute final scene (from 1:36:00 to 1:42:29) of *Casablanca* (Michael Curtiz, 1942) dubbed into Spanish, of which 2 minutes (from 1:36:12 to 1:38:12) were closely analysed to detect eye movements in close-ups. A second stimulus video consisted of the original English version of the same excerpt, which was used with the control group of native English participants. Finally, the third stimulus video, used to analyse native Spanish viewers' eye movements when watching an original film in Spanish, was a 6-minute scene (from 0:29:15 to 0:35:23) from *Todo sobre mi madre* (Pedro Almodóvar, 1999), of which 2 minutes (from 0:30:01 to 0:32:01) were closely analysed to detect eye movements in close-ups. Drawing on Perego *et al.* (2016), the videos were compared regarding their audiovisual complexity. Despite the significant difference in production year (1942 and 1999) and format (black and white vs. colour), the videos proved to be remarkably comparable regarding duration, speech rate (measured in words per minute), type-token ratio (degree of lexical variation), lexical density, syntactical complexity and number of close-ups, as shown in Table 1:

Comparison of stimulus videos		
	Casablanca	*Todo sobre mi madre*
Duration	2:02	1:58
Speech rate	189 wpm	185 wpm
Type-token ratio	0.58	0.60
Lexical density	41.36%	47.2%
Syntactical complexity (average sentence length)	10.53	10.26
Number of close-ups	10	7
Percentage of close-ups	77%	70%
Average shot length (of close-ups)	5.7s	3.6s

Table 1. Complexity indices for the two stimulus videos used in the experiment (adapted from Perego *et al*. 2016)

6.2. Apparatus

Participant's eye movements were recorded using the standalone Tobii T120 eye tracker (Tobii Technology AB, Stockholm, Sweden) integrated in a 17-inch monitor with a 1024×768 resolution that allowed the maximisation of the stimulus display to cover the entire screen. Both the eye-tracking server and the client display application ran on Windows PCs connected via 1 GB Ethernet. This eye tracker, which operates at a sampling rate of 60 Hz with

an accuracy of 0.5°, is unobtrusive, as it allows for a large degree of head movement and ensures natural behaviour, which is important in order to obtain ecologically valid results. During the recording time, the Tobii T120 eye tracker collects raw gaze movement data every 16.6 ms, using a filter to parse the coordinates of the movements into fixations and saccades. For the analysis, two areas of interest were drawn on those shots of the videos that featured close-ups, one covering the characters' eyes and the other covering their mouths. When using the eye-tracking data to test the above-mentioned hypotheses, the focus was placed on three types of measurements that are relevant to gain knowledge of visual attention distribution: number of fixations, mean fixation duration and percentage amount of time spent on the defined areas of interest. A distinction was made between close-ups with dialogue and silent close-ups in order to ascertain whether the presence of dialogue has any impact on the viewers' eye movements. Thus, the means of those values were obtained for the total duration in which the characters were shown on screen, either speaking or listening.

6.3. Participants

This study involved 42 participants (31 female and 11 male), mostly postgraduate students and young professionals. None of them received course credits or payment for participation. Of those 42 participants, 18 were native English and 24 were native Spanish. All of them had normal or corrected-to-normal vision. A total of 31 participants reported that they did not wear corrective lenses of any sort, seven reported that they wore contacts, and four reported that they wore glasses. Due to poor calibration and other data collection issues, the data from seven participants were discarded from the final analysis, bringing the total down to 35 (15 native English and 20 native Spanish) and dropping the number of males and females to 8 and 27, respectively. The ages of participants ranged from 25 to 60 ($M = 28.00$; $SD = 8.55$).

6.4. Procedure and experimental design

Participants sat in front of the eye tracker at a distance of 60-70 cm, the eye tracker camera's focal length. Calibration was performed once for each participant before viewing the first video and required following nine dot targets displayed sequentially on the screen, each shrinking in diameter from 30 to 2 pixels.

Before starting the eye-tracking test, all participants were asked to fill in a demographic questionnaire with information about their age, gender, occupation, foreign language level and viewing habits. They were then told that they would be watching two video clips (in the case of the native Spanish participants) or one (in the case of the native English participants) and that there would be a further post-test questionnaire about, amongst other aspects, their comprehension of the clip.

The post-test questionnaires, filled in with pen and paper after every clip, included questions about visual and verbal comprehension (assessed with a 5-point scale, 1 for no comprehension and 5 for full comprehension), sense of presence (for which the ITC-Sense of Presence Inventory, developed at Goldsmiths University, was used) and perception of eye movement when watching these clips, to determine the extent to which participants are aware of how they distribute their attention between eyes and mouths when watching faces in close-ups, with and without dialogue. Regarding the latter, rankings were represented with a five-point scale (1 for no time spent on eyes or mouth; 5 for all the time on eyes or mouth), with open-ended questions asking for reasons for a given rating. Finally, a brief post-test qualitative interview was held with each participant about the purpose of the study and the comparison between their perceived distribution of attention and the results shown by the eye tracker.

6.5. Results

In order to analyse the data obtained in the study, between-group analyses of variance were performed. The results are shown in Table 2. Although subgroup data for the English and Spanish cohorts were not consistent with normal distribution, a parametric analysis was performed and presented; firstly because the compared subgroups were statistically equinumerous (χ^2 (2, N = 55) = .91; p = .635), since parametric analyses are relatively resistant when groups are equal; secondly, because the nonparametric analyses performed gave exactly the same results in both the main and post-hoc tests. For this reason, the analysis was more powerful and accessible for the reader. The Saphiro-Wilk test was used to estimate the normality of distributions, a general one-way ANOVA was selected for the estimation of effects and post-hoc tests were performed using the Dunnet's T3 method. In cases of violation of the assumption of equality of variance, non-homogeneous Welch tests were used. The calculations were carried out in the SPSS 22.0 statistical package.

There was a significant difference between groups in percentage of time spent on characters' eyes in close-ups with dialogue [F(2, 23.44) = 26.26; p < .001; η^2 = .338]. Spanish participants watching *Casablanca* (M = 95.00; SD = 3.54) spent a significantly (p = .008) greater percentage of time looking at eyes than English participants watching *Casablanca* (M = 76.19; SD = 20.07) and that the same group of Spanish participants watching *Todo sobre mi madre* (M = 76.56; SD = 12.07; p < .001). The groups were also different in percentage of time spent on characters' mouths in close-ups with dialogue [F(2, 23.44) = 26.24; p < .001; η^2 = .338]. The Spanish participants watching *Casablanca* (M = 5.00; SD = 3.54) focused significantly (p = .008) less on the characters' mouths than the English participants watching *Casablanca* (M = 23.81; SD = 20.07) and the Spanish participants watching *Todo sobre mi madre* (M = 23.44; SD = 12.07) (p < .001).

	I: *Casablanca* English Participants (N = 15)		II: *Casablanca* Spanish Participants (N = 20)		III: *Todo sobre mi madre* Spanish Participants (N = 20)				
	M	**SD**	**M**	**SD**	**M**	**SD**	**F(2)**	**p**	**Post-hoc**
movie comprehension	5.00	---	4.90	.31	4.60	.50	7.20	.003**	I > III II > III
% of time on									
characters' eyes with dialogue	76.19	20.07	95.00	3.54	76.56	12.07	26.26	< .001**	I < II II > III
characters' mouths with dialogue	23.81	20.07	5.00	3.54	23.44	12.07	26.24	< .001**	I > II II < III
characters' eyes with no dialogue	82.65	14.63	85.84	6.72	---	---	.75	.394	ns
characters' mouth with no dialogue	17.35	14.63	14.15	6.72	---	---	.75	.394	ns
declarative time (1 = no time; 5 = all the time)									
characters' eyes (5-point scale)	4.73	.46	4.65	.49	4.65	.49	.16	.849	ns
characters' eyes (percentage)	94.6%		93%		93%				
characters' mouths (5-point scale)	3.13	.52	3.35	.49	3.05	.76	1.28	.286	Ns
characters' mouths (percentage)	62.5%		67%		61%				

* $p < .05$; ** $p < .01$

Table 2: Intergroup differences

As an additional analysis, a mixed ANOVA model was made to estimate whether there is an interactive effect associated with differences both within groups and between them. This analysis allowed us to capture differences of a more complex type. Participants watching *Casablanca* were compared (English and Spanish group, treated as a between-group factor) regarding the percentage of time spent on eyes during silent close-ups and close-ups with dialogue (which was treated for each group as a within-group factor). A strong interaction effect of this factor with the group was obtained: $F(1, 33) = 13.86$; $p = .001$; $\eta^2 = .296$, which shows that when close-ups were silent both groups spent a similar amount of time looking at the eyes. Nonetheless, as indicated above, in the case of dialogues the Spanish participants spent more time looking at the eyes than the English participants as illustrated in Figure 2:
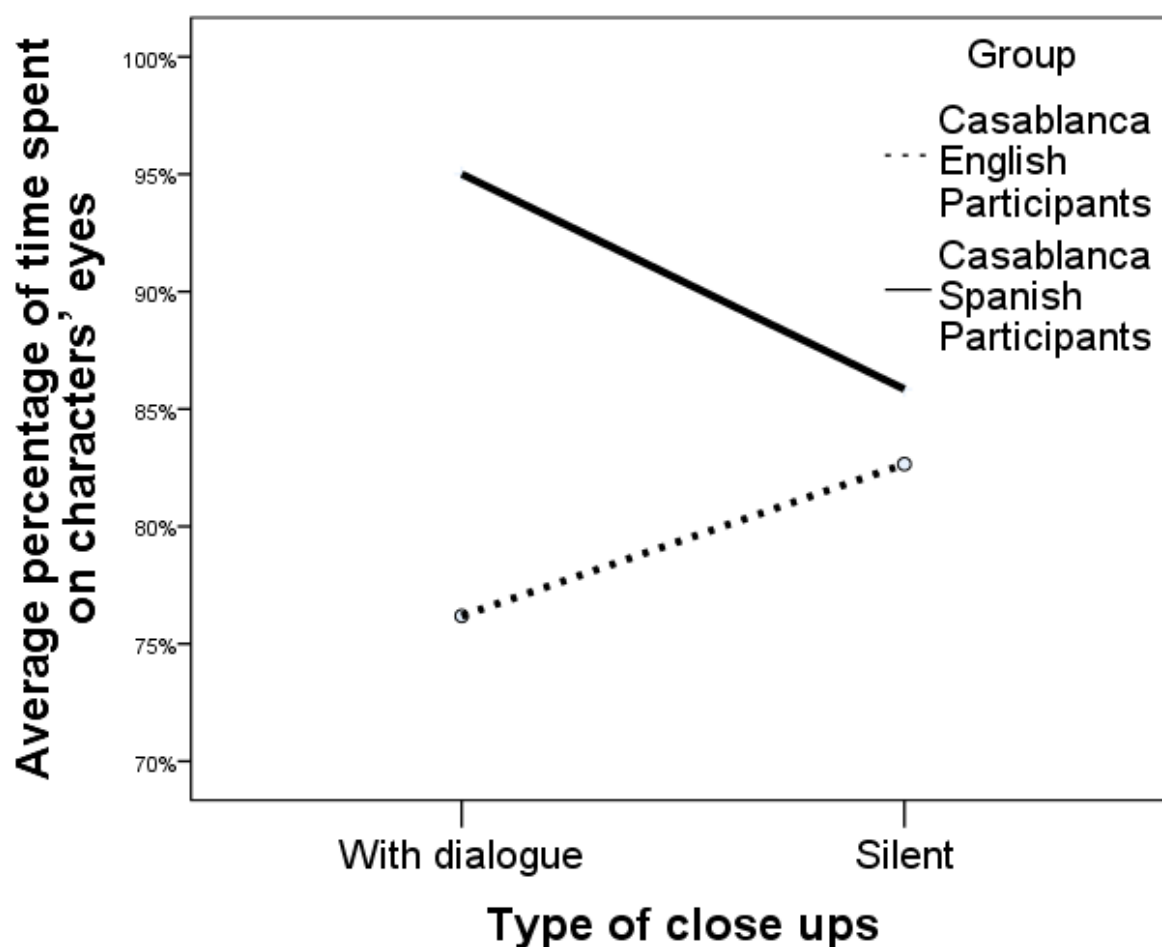
Figure 2. Average percentage of time spent on character eyes by group

The differences were then determined between the real (eye-tracking) and the perceived (subjective questionnaires) percentage of time spent on characters' eyes in close-ups. Since for the perceived percentage of time the participants used a five-point scale, this variable was transformed into a percentage distribution (Table 2), and then real and perceived data were treated as a within-object factor in mixed model (2 x 3), with the research group as between-subject factor. A significant within-factor effect, $F(1, 52)$ = 14.90, $p < .001$, $\eta^2 = .18$, shows that the real time ($M = 83.16$, $SD = 15.55$) was shorter than the time perceived by participants ($M = 91.82$, $SD = 11.84$). A significant interaction effect, $F(2, 52) = 7.70$, $p = .001$, $\eta^2 = .19$, indicates clearly that in all groups the perceived time was equally high, while real time was shorter for the English participants watching *Casablanca* and the Spanish participants watching *Todo sobre mi madre*. Exactly the same model was made for percentage of time spent on characters' mouths and a similar main effect was obtained $F(1, 52) = 206.77$, $p < .001$, $\eta^2 = .73$, where real time ($M = 16.83$, $SD = 15.55$) was shorter than perceived time ($M = 54.55$, $SD = 15.28$). A significant interaction effect, $F(2, 52) = 11.33$, $p < .001$, $\eta^2 = .08$, indicates that all groups perceived their time spent on characters' mouths as similarly high, while real time was definitely lower, especially for the Spanish participants watching the dubbed scene from *Casablanca*, as shown in Figures 3 and 4:
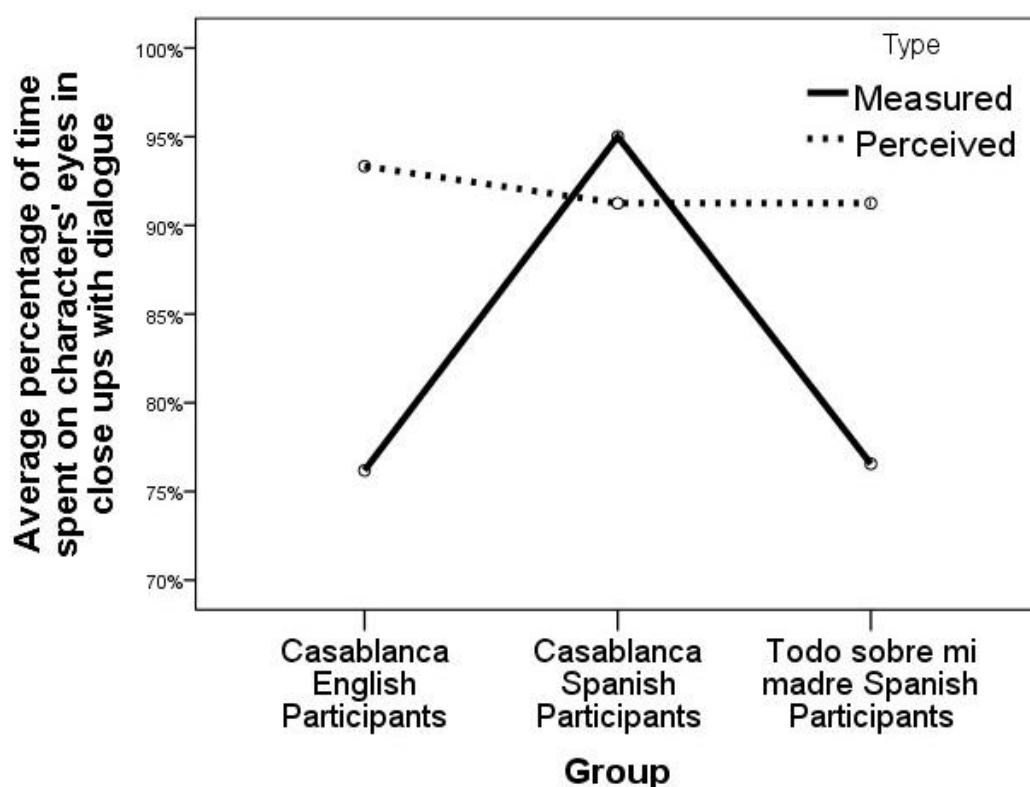
Figure 3. Average percentage of time spent on characters' eyes in close-ups with dialogue
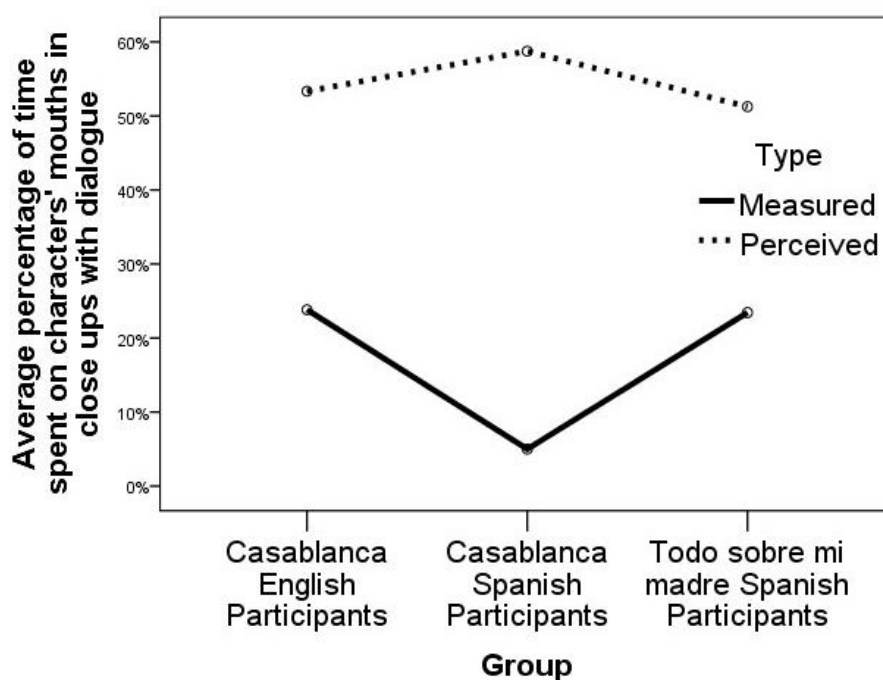

Figure 4. Average percentage of time spent on characters' mouths in close-ups with dialogue

No effects were observed regarding gender or age. Comprehension was on average very high (5/5 for *Casablanca* in English, 4.9/5 for *Casablanca* in Spanish and 4.6/5 for *Todo sobre mi madre*) and the sense of presence may be regarded medium-high (3.6/5 for *Casablanca* in English, 3.7/5 for *Casablanca* in Spanish and 3.8/5 for *Todo sobre mi madre*). No significant between-groups differences were observed regarding comprehension and sense of presence.

6.6. Discussion

This is, to our knowledge, the first eye-tracking study to compare how viewers watch faces in close-ups with and without dialogue in original and dubbed films. On the one hand, it would make sense to expect an intense focus on characters' mouths by dubbing viewers given the mouth bias commonly associated with close-ups with dialogue (Buchan *et al.* 2007), close-ups with imperfectly-synched video and audio (Smith *et al*. 2014) and the McGurk effect (Navarra 2008). On the other hand, because of these very same reasons, and given that no useful information can be obtained from mouths in dubbing, dubbing viewers may instead be expected to adopt an unconscious strategy to avoid looking at them, focusing instead on the characters' eyes. This is the hypothesis tested in this study, which is supported by the statistical evidence obtained in the analysis.

As can be seen in Figures 5 and 6, while the English participants watching the original version of *Casablanca* showed a very similar distribution of attention (76% on eyes vs 34% on mouth) to the one obtained by Võ *et al*. (2012) (76% vs 34%) and Foulsham and Sanderson (2013) (71% vs 29%), the viewing patterns of the Spanish participants watching *Casablanca* dubbed into Spanish are significantly different: 95% on eyes and 5% on mouths. This extreme focus on the eyes/negative mouth bias is unlike anything found so far in the literature and very different to the way in which the Spanish participants view faces in the original Spanish film used in the experiment (Figure 7), where, after watching the dubbed clip, they show the same distribution (76% vs 24%) found in the literature and in the English group watching the original version of *Casablanca.*[1]

Figure 5. Distribution of attention between eyes and mouth by the English group watching an original clip from *Casablanca*
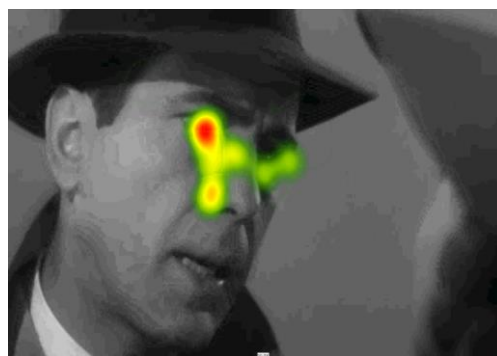

Figure 6. Distribution of attention between eyes and mouth by the Spanish group watching a dubbed clip from *Casablanca*

Figure 7. Distribution of attention between eyes and mouth by the Spanish group watching an original clip from *Todo sobre mi madre*

This pattern changes in close-ups with no dialogue, where the eye movements of English and Spanish participants watching *Casablanca* converge. English participants move away from the mouth and focus more on the eyes (82.6% on eyes vs 17.4% on mouth), which are likely to convey most of the meaning now that the mouth is not moving, whereas the Spanish participants finally look down to the mouth (85.8% vs 14.2%). It is as though the dubbing viewers, aware of the mismatch between images and sound in dubbed close-ups with dialogue, made a point of not looking at mouths, a phenomena that is not observed in original films or in dubbed films when there is no dialogue. Interestingly, this intricate strategy does not seem to be conscious, as there is no relation between perceived and real distribution of attention between eyes and mouth. The Spanish viewers believe they spent 67% of their time looking at mouths in the dubbed version of *Casablanca* and 61% in the case of *Todo sobre mi madre*, while in fact they spent 5% and 23%, respectively. In the qualitative interviews after the study, many of them insisted that they had spent almost half of their time looking at the mouths in the dubbed clip of *Casablanca*, until they were confronted with the replay of their fixations. This means that (a) there is a significant discrepancy between where the participants think they are

looking and where they are actually looking and (b) the strategy to avoid looking at mouths in dubbed close-ups with dialogue but not in original films, called here the dubbing effect, is an unconscious one.

Two more aspects may be worth discussing, even though the lack of statistical evidence means that they must remain as no more than anecdotal data or avenues for future research. Firstly, while most of the Spanish participants live in Spain and have regularly watched dubbed films throughout their lives, one of them, excluded from the analysis as an outlier, has lived in London for the past 20 years, a time during which she has not been exposed to dubbing. Her distribution of attention (55% on eyes vs 45% on mouth) shows an unusually high focus on the characters' mouths, not only more than the other Spanish participants watching *Casablanca* but also than the English participants and than the evidence found so far in the literature. Her lack of regular exposure to dubbed films may be preventing her from ignoring the imperfectly-synchronised lips and from becoming immersed in the film (her sense of presence is only 2.8/5, as compared to an average of 3.7/5 from all participants). Should this data be replicated and verified with other participants in the same situation, it could suggest that for dubbing to work (and for the dubbing effect to apply), it is necessary to have continuous exposure to dubbed products. Or, more accurately, prolonged lack of exposure to dubbing may cause viewers to lose the habit of focusing mainly on the characters' eyes, thus drawing their attention to the asynchronous mouths and having a negative impact on immersion and suspension of disbelief.

Secondly, it was observed that when Humphrey Bogart's character Rick says the line 'Here's looking at you, kid', the English participants largely turn their attention to the character's mouth (65% on the mouth, as opposed to 24% in the rest of the clip). Again, more data would be needed to draw solid conclusions, but it may be that the recognition of such an famous line, regularly cited as one of the most iconic quotes in the history of cinema (Time 2010) and so often written and read in other contexts, has drawn the participants' attention to the signifier rather than the signified, to the physical form of the words and the place where they are uttered, the characters' mouths, rather than to their meaning and the emotions they express, which are more often identified with the characters' eyes. This pattern has not been found in the Spanish dubbed version, which may be explained by the fact that the Spanish translator opted for four different translations for the four key moments in which this line is said in the film (*por nosotros* [here's to us]*, toda la suerte* [best of luck], *por todos nosotros* [here's to all of us] and *ve con él* [go with him]), as a result of which there has never been an iconic equivalent in Spanish for 'Here's looking at you, kid'. In other words, there is no signifier to draw the viewers' attention. At any rate, this mouth bias caused by iconic lines is at this stage no more than just an assumption that needs to be verified with further research.

7. Conclusions

Despite the fact that dubbing is regularly criticised for its artifice and its manipulation of film sound, it has proved to be the preferred mode of AVT for millions of viewers. Although plenty of research has been conducted about the work carried out by the professionals involved in dubbing (translators, dialogue writers, dubbing actors, etc.), little is known about the process undergone by the viewers to make it work.

Drawing on the framework developed in this article, it is argued here that when first exposed to dubbed films at an early age, viewers may feel a sense of wonder that leads to habituation and to an automatic and unconscious engagement with the dubbed fiction, facilitated by their ability to suspend disbelief, their interest in the story, some degree of comprehension of the plot and a sense of immersion that involves feelings of flow, transportation and presence. This process of engagement is not affected by the discovery, years later, of the prefabricated nature of dubbing, since by then this path to engagement has already been unconsciously internalised. Getting used to dubbing, when it happens at an early age, is simply part of the (unconscious) process of getting used to film.

Yet, the question remains as to how dubbing viewers can manage to switch off the powerful McGurk effect and thus avoid being confused or distracted by the mismatch between lips and audio. A potential answer may lie in the results of the eye-tracking study presented in this article, which show that the Spanish participants watching a dubbed scene from *Casablanca* have an extreme negative mouth bias, with 95% of attention on the characters' eyes and only 5% on their mouths. This is in sharp contrast with their perception of how they watched this scene (58% on eyes vs 42% on mouths), with their own viewing patterns watching a comparable scene in Spanish (76% vs 24%), with the viewing patterns of the English participants watching the same scene from *Casablanca* (76% vs 24%) and with the data obtained so far in the literature for both film and real-life scenes.

Although in need of further research with larger and different samples, these results, which have subsequently been supported by those obtained in Di Giovanni and Romero (2018) with Italian participants,[2] point to the potential existence of a dubbing effect, an unconscious eye movement strategy performed by dubbing viewers to avoid looking at mouths in dubbing, which prevails over the natural way in which they watch original films and real-life scenes, and which arguably allows them to suspend disbelief and be transported into the fictional world. Although not conscious, this mechanism seems to be activated only with dubbed films and is then turned off when watching an original film, where the viewing pattern is aligned with eye movements in real life. From this point of view, there is a quasi-Darwinian quality to this effect, which enables viewers to adapt their viewing patterns in order to 'survive' in the dubbing environment, that is,

in order to overcome the danger of being put off by the asynchronous nature of dubbing, and thus achieve the ultimate goal of being engaged with the fictional story.

Acknowledgements

References

- **Allison, Robert S., Laurie Wilcox and Ali Kazimi** (2013). "Perceptual artefacts, suspension of disbelief and realism in stereoscopic 3D film." *Public 24*, 47(12), 149-160.

- **Ameri, Saeed, Masood Khoshsaligheh and Ali Khazaee Farid** (2017). "The reception of Persian dubbing: A survey on preferences and perception of quality standards in Iran." *Perspectives* 26(3), 435-471.

- **Ávila, Alejandro** (1997). *El doblaje.* Madrid: Cátedra.

- **Biocca, Frank** (1997). "The cyborg's dilemma: progressive embodiment in virtual environments." *Journal of Computer-Mediated Communication* 3(2). http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1997.tb00070.x/full (consulted 18.04.2018).

- **Birmingham, Elina and Alan Kingstone** (2009). "Human social attention." *Annals of the New York Academy of Sciences* 1156(1), 118-140.

- **Borges, José Luis** (1945). "On dubbing movies." *Sur* 128, 88:90.

- **Buchan, Julie N., Martin Paré and Kevin G. Munhall** (2007). "Spatial statistics of gaze fixations during dynamic face processing." *Social Neuroscience* 2(1), 1-13.

- **Busselle, Rick and Helena Bilandzic** (2009). "Measuring narrative engagement." *Media Psychology* 12, 321–347.

- **Buswell, Guy Thomas** (1935). *How People Look at Pictures: A Study of the Psychology and Perception in Art*. Chicago: University Chicago Press.

- **Caillé, Pierre-François** (1960). "Cinéma et traduction : Le traducteur devant l'écran. Le doublage. Le sous-tittrage." *Babel* 6(3), 103-109.

- **Chaume, Frederic** (2004). *Cine y traducción*. Madrid: Cátedra.

- **Chaume, Frederic** (2007). "Quality standards in dubbing: a proposal." *TradTerm* 13, 71-89.

- **Chaume, Frederic** (2013). *Audiovisual Translation: Dubbing*. Manchester: St. Jerome.

- **Cohen, Jonathan** (2001). "Defining identification: a theoretical look at the identification of audiences with media characters." *Mass Communication and Society* 4, 245-264.

- **Coleridge, Samuel Taylor** (1817). *Biographia Literaria.* http://www.gutenberg.org/files/6081/6081-h/6081-h.htm (consulted 3.12.2019).

- **Csikszentmihalyi, Mihaly** (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper & Row.

- **Di Giovanni, Elena** (2018). "Dubbing, perception and reception." Elena di Giovanni and Yves Gambier (eds). *Reception Studies and Audiovisual Translation*. Amsterdam: John Benjamins, 159-177.

- **Di Giovanni, Elena and Pablo Romero-Fresco** (2019). "Are we all together across languages? An eye tracking study of original and dubbed films." Irene Ranzato and Serenella Zanotti (eds). *Reassessing Dubbing: Historical Approaches and Current Trends*. Amsterdam: John Benjamins, 125-145.

- **Eppler, Eva and Mathias Kraemer** (2018). "The deliberate non-subtitling of L3s in a multilingual TV series: the example of *Breaking Bad.*" *Meta* 63(2): 365-391.

- **Evan, Dishy** (2011). "The McGurk effect… or why dubbing films should be banned." http://didishy.tumblr.com/post/11233675717/the-mcgurk-effect-or-why-dubbing-films-should (consulted 20.09.2018).

- **Foulsham, Tom and Lucy Anne Sanderson** (2013). "Look who's talking? Sound changes gaze behaviour in a dynamic social scene." *Visual Cognition* 21(7), 922–944.

- **Fresno, Nazaret** (2017). "Approaching engagement in audio description." *Rivista internazionale di tecnica della traduzione* 19, 13-32.

- **Gosselin, Frédéric and Philippe G. Schyns** (2001). "Bubbles: a technique to reveal the use of information in recognition tasks." *Vision Research* 41(17), 2261-2271.

- **Green, Kerry** (1994). "The influence of an inverted face on the McGurk effect." *The Journal of the Acoustical Society of America* 95, 3014.

- **Green, Kerry, Patricia K. Kuhl, Andrew N. Meltzoff and Erica B. Stevens** (1991). "Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect." *Perception and Psychophysics* 50(6), 524-536.

- **Green, Melanie C. and Timothy C. Brock** (2000). "The role of transportation in the persuasiveness of public narratives." *Journal of Personality and Social Psychology* 79, 701-721.

- **Gunning, Tom** (2003). "Renewing old technologies, astonishment, second nature, and the uncanny in technology from the previous turn-of-the-century." David Thorburn and Henry Jenkins (eds). *Rethinking Media Change: The Aesthetics of Transition*. Cambridge: MIT Press, 39-60.

- **Johnson-Laird, Philip N.** (1983). *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*. Cambridge: Cambridge University Press.

- **Kilborn, Richard** (1993). "Speak my language: Current attitudes to television subtitling and dubbing." *Media, Culture and Society* 15, 641–660.

- **Koolstra, Cees M., Allerd L. Peeters and Herman Spinhof** (2002). "The pros and cons of dubbing and subtitling." *European Journal of Communication* 17, 325–354.

- **Langton, Stephen R. H., Roger J. Watt and Vicki Bruce** (2000). "Do the eyes have it? Cues to the direction of social attention." *Trends in Cognitive Sciences* 4(2), 50-59.

- **Luyken, George-Michael, Thomas Herbst, Jo Langham-Brown, Helen Reid and Hermann Spinhof** (1991). *Overcoming Language Barriers in Television: Dubbing and Subtitling for the European Audience*. Manchester: European Institute for the Media.

- **McGurk Harry and John MacDonald** (1976). "Hearing lips and seeing voices." *Nature* 264, 746-748.

- **Navarra, Jordi** (2003). "Visual speech interference in an auditory shadowing task: The dubbed movie effect." *15th International Congress of Phonetic Sciences Proceedings*. Barcelona, 893-896.

- **Palencia Villa, Rosa María** (2002). *La influencia del doblaje audiovisual en la percepción de los personajes*. PhD Thesis. Barcelona: Universitat Autònoma de Barcelona.

- **Parrish, Stephen** (1985). "'Leaping and lingering': Coleridge's lyrical ballads." Richard Gravil, Lucy Newlyn and Nicholas Roe (eds). *Coleridge's Imagination*. Cambridge: Cambridge University Press, 102-117.

- **Pavesi, Maria** (2009). "Dubbing English into Italian: A closer look at the translation of spoken language." Jorge Díaz Cintas (ed.). *New Trends in Audiovisual Translation*. Bristol: Multilingual Matters, 197-209.

- **Pedersen, Jan** (2011). *Subtitling Norms for Television*. Amsterdam: John Benjamins.

- **Perego, Elisa, David Orrego-Carmona and Sara Bottiroli** (2016). "An empirical take on the dubbing vs. subtitling debate. An eye movement study." *Lingue e Linguaggi* 19, 255-274.

- **Raney, Arthur A.** (2004). "Expanding disposition theory: Reconsidering character liking, moral evaluation, and enjoyment." *Communication Theory* 14(4), 348-368.

- **Robinson, Jennifer, Jane Stadler and Andrea Rassell** (2015). "Sound and sight: An exploratory look at *Saving Private Ryan* through the eye-tracking lens." *Refractory: a Journal of Entertainment Media* 25. http://refractory.unimelb.edu.au/2015/02/06/robinson-stadler-rassell/ (consulted 20.09.2018).

- **Romero-Fresco, Pablo** (2009). "Naturalness in the Spanish dubbing language: A case of not-so-close *Friends.*" *Meta* 54(1), 49-72.

- **Rosenblum, Lawrence** (2010). *See What I'm Saying: The Extraordinary Powers of Our Five Senses*. New York: W. W. Norton & Company Inc.

- **Rouger, Julien, Bernard Fraysse, Olivier Deguine and Pascal Barone** (2008). "McGurk effects in cochlear-implanted deaf subjects." *Brain Research* 1188, 87-99.

- **Rowe, Thomas Louis** (1960) "The English dubbing text." *Babel* 6(3), 116-120.

- **Sánchez Mompeán, Sofía** (2017). *The Rendition of English Intonation in Spanish Dubbing*. PhD Thesis. Murcia: Universidad de Murcia.

- **Senju, Atsushi and Toshikazu Hasegawa** (2005). "Direct gaze captures visuospatial attention." *Visual Cognition* 12(1), 127-144.

- **Senju, Atsushi, Toshikazu Hasegawa and Yoshikuni Tojo** (2005). "Does perceived direct gaze boost detection in adults and children with and without autism? The stare-in-the-crowd effect revisited." *Visual Cognition* 12(8), 1474-1496.

- **Smith, Elliot, Scott Duede, Sara Hanrahan, Tyler Davis, Paul House and Bradley Greger** (2013). "Seeing is believing: Neural representations of visual stimuli in human auditory cortex correlate with illusory auditory perceptions." *PLoS ONE* 8(9), e73148.

- **Slaghuis, Walter L. and Alanda K Thompson** (2003). "The effect of peripheral visual motion on focal contrast sensitivity in positive- and negative-symptom schizophrenia." *Neurophsychologia* 4, 968-980.

- **Spiteri Miggiani, Giselle** (2019) *Dialogue Writing for Dubbing. An Insider's Perspective*. London: Palgrave MacMillan.

- **Vilaró, Anna and Tim J. Smith** (2011). "Subtitle reading effects on visual and verbal information processing in films." Published abstract in Perception. ECVP abstract supplement 40, 153. European Conference on Visual Perception. Toulousse, France.

- **Võ, Melissa L.-H., Tim J. Smith, Parag K. Mital and John M. Henderson** (2012). "Do the eyes really have it? Dynamic allocation of attention when viewing moving faces." *Journal of Vision* 12(3), 1-14.

- **Wästlund, Erik, Poja Shams and Tobias Otterbring** (2017). "Unsold is unseen … or is it? Examining the role of peripheral vision in the consumer choice process using eye-tracking methodology." *Appetite* 120, 49-56.

- **Wissmath, Bartholomaus, David Weibel and Rudolf Groner** (2009). "Dubbing or subtitling? Effects on spatial presence, transportation, flow, and enjoyment." *Journal of Media Psychology: Theories Methods and Applications* 21(3), 114-125.

- **Yarbus, Alfred L.** (1965/1967). *Eye Movements and Vision*. New York: Plenum Press.

- **Zabalbeascoa, Patrick** (1993). *Developing Translation Studies to Better Account for Audiovisual Texts and Other New Forms of Text Production*. PhD Thesis. Lleida: University of Lleida.

- **Zillmann, Dolf** (1994). "Mechanisms of emotional involvement with drama." *Poetics* 23, 33–51.

Biography

Pablo Romero Fresco is Ramón y Cajal researcher at the Universidade de Vigo, Spain, and Honorary Professor of Translation and Filmmaking at the University of Roehampton, London. He is the author of the books *Subtitling through Speech Recognition: Respeaking* (Routledge), *Accessible Filmmaking: Integrating Translation and Accessibility into the Filmmaking Process* (Routledge, forthcoming) and the editor of *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe* (Peter Lang). He is on the editorial board of the Journal of Audiovisual Translation (JAT) and is currently working with several governments, universities, companies and user associations around the world to introduce and improve access to live events for people with hearing loss. He has collaborated with Ofcom to carry out the first analysis of the quality of live subtitles on TV in the UK and is working with the Canadian Radio-television and Telecommunications Commission (CRTC) on a similar project in Canada. He is the leader of the international research centre GALMA (Galician Observatory for Media Access), for which he is currently coordinating several international projects on media accessibility and accessible filmmaking, including "Media Accessibility Platform" and "ILSA: Interlingual Live Subtitling for Access", funded by the EU Commission. Pablo is also a filmmaker. His documentary *Joining the Dots* (2012) was screened during the 69th Venice Film Festival and has been used by Netflix as well as film schools around Europe to raise awareness about audio description.

Email: p.romero-fresco@roehampton.ac.uk
promero@uvigo.es

Notes

[1] It is worth noting that eye tracking can only detect the central vision obtained by the fovea (Slaghuis and Thompson 2003). Foveal vision allows us to obtain detailed information typically within six degrees of our field vision, that is, spanning five words in a row when reading printed text at ordinary size at about 50 centimeters from the eyes. Parafoveal or peripheral vision, which can span up to 120 degrees, is thus not detected by eye trackers. However, even though peripheral vision can be used to differentiate movement from stillness and even certain types of rhythms and contrast, it cannot help to distinguish colours, shapes or details (Wästlund *et al*. 2017). For the purpose of this study, peripheral vision could potentially be used, given the right conditions, to differentiate whether a mouth is moving or not, but certainly not to discern the degree of (a)synchrony between moving lips and the dubbed audio. In other words, participants whose fixations are found

on the characters' eyes cannot be expected to be put off by the imperfect synchrony of lips and audio often found in dubbing.

[2] Although the study by Di Giovanni and Romero-Fresco (2019) was conducted after the experiment presented here, it has been published earlier. The Italian study did not focus on the dubbing effect, but it has found evidence to support it.