

Methodological contributions of Social Neuroscience to the study of the multimodality of emotions in translation

Antonio Javier Chica Núñez

Universidad Pablo de Olavide, Spain
ajchinun@upo.es

This study analyses the interplay of various communication modes that enables emotions to be transmitted efficiently from source (ST) to target text (TT) in audio description (AD) as a multimodal text. It draws on existing experimental designs, including neutral or emotional conditions based on the congruency of stimuli from various modes (images, dialogue semantic content or prosody in a film, together with the semantic content of AD). This article reviews the methodological contribution that Social Neuroscience could make to the study of multimodal translation. To this end, some neurobiological models and studies are quoted regarding multimodal emotional information processing (Brück, Kreifelts, & Wildgruber, 2011), the impact of multimodal emotional processing on subjects' empathy (Regenbogen et al., 2012) and the dynamics of neural networks involved in human empathy and communication through the presentation of multimodal stimuli (Regenbogen, Habel, & Kellerman, 2013). Finally, an experimental design that focuses on the transfer of feelings and emotions in film AD, which would be suitable for a potential pilot study, is presented.

1. Introduction

1.1 Cognitive processes underlying multimodal translation

In recent times, Translation Studies has adopted the concept of multimodality as a new approach to tackling the complex cognitive reality present in any translation process, in particular in those processes involving changes in communication codes or the intervention of different modes of expression in the source (ST) as much as in the target text (TT). Audiovisual translation, in particular, is strongly defined by this notion of multiple communication modes' leading and conditioning the translation process.

The increasing use of audio description (AD) in the market of audiovisual translation has called on Translatology to propose theoretical and applied paradigms for the study of the multimodal nature of this translation practice (Braun, 2008, 2011; Jiménez Hurtado, 2010a, 2010b; Jiménez Hurtado & Seibel, 2012). However, to date, hardly any studies taking into account the cognitive and brain systems underlying translation processes more generally (Tymoczko, 2012; Schwieter & Ferreira, 2017), and multimodal translation processes in particular (Chica Núñez, 2015, 2016), have been proposed.

If we acknowledge that the brain processes reality through external multisensory inputs, the manner in which audio describers obtain access to knowledge and represent it is necessarily based on an active processing of different communication modes. This fact would redefine the translation process involved in AD as an eminently multimodal one and not simply as an intersemiotic one. In other words, the modes of transfer suggest managing modes of expression apart from images and spoken language: written language, prosody or music, among other communication inputs.

The notion of multimodality in film AD has previously been explored by, among others, Vercauteren (2012):

Audiovisual products such as films are inherently multimodal in nature, creating meaning from the combination of and interaction between different semiotic channels: images, dialogues, sounds and text-on-screen. ... So in order to render films accessible to the visually impaired, the semiotic channels to which they do not have access have to be translated into a mode that is accessible to them, i.e. the verbal mode (pp. 209–210).

Consequently, a systematic study of the multimodal translation and reception processes involved in AD should imply using suitable analytical methods to identify the influence of multisensory cognitive operations that occur in the translator's and the visually impaired recipient's brains, respectively.

I advocate applying empirical cognitive methods to measure not only the production of audio describers but also the "effects" of multimodal translation on visually impaired users, as put forward by Kruger and Kruger (2017) "in terms of the cognitive processing involved in readers' reception of translations" (p. 71). In turn, this would require dealing with the theoretical cognitive models that account for the brain's processing of multimodality, as presented in section 2 of this article.

1.2 Objectives

This article presents theoretical concepts that describe the cerebral processing involved in receiving multimodal messages. In other words, it concerns the comprehension of complex stimuli consisting of more than one semiotic code, which, as a consequence, are subjected to multisensory processing in different areas of the human brain. Therefore, such theoretical concepts would lead us to a methodological approach that can help with understanding the complex nature of AD as a type of multimodal translation. In particular, I aim to disentangle this intricate reality by focusing on the transmission of emotions in film AD and the way in which they can be measured using experimental reception studies inspired by cognitive and social neuroscience methods.

To this end, I resort to a number of perception theories. First, Zeki and Bartels' model (1999) on specialized vision processing is presented briefly. The information provided by their model is essential to understanding that external information stimuli—from a specific shade of blue to a sad look on the face of a film character—are processed by discrete brain modules or units that specialize in processing only specific stimuli. Secondly, Smith and Kosslyn's proposal (2008) illustrates the types of information-processing flow that occur in the brain: bottom-up signals—from senses to cognition—and top-down signals, in which cognitive structures such as memory or expectations flow, modulating the way in which perception occurs.

It is essential for any researcher applying experimental methods to the analysis of multimodal texts to be perfectly aware of these two main principles according to which brain information-processing operates: *functional specialization* and *bottom-up and top-down signal processing*. When they receive a multimodal discourse such as AD, certain brain nodes deal with specific communication modes (semantic content, pitch, intonation, etc.). Top-down processing—the influence of previous experiences on comprehending ironic speech, for instance—might lead the researcher to establish dependent variables or research questions that deal with the possible effects of such an element. This is explained in more detail in section 2.1.

Thirdly, Brück, Kreifelts and Wildgruber (2011) propose a model of multimodal affective information-processing which, among other things, integrates emotional information that originates in several communication modes (verbal vocal, verbal non-vocal and facial visual cues) in brain areas associated with the processing of these signals. Finally, the research of Regenbogen et al. (2012, 2013) has come to identify certain nodes in the cortex that seem to be associated with integrating multimodal information (facial

expressions, prosody and verbal discourse) and may also be linked to the “empathy appraisal model” (De Vignemont & Singer, 2006). Since our methodological proposal relates to the emotions expressed in AD via different modes, both theoretical models would be fundamental to this approach, as will be seen in section 2.2.

Connecting these theoretical models to the study of multimodality in translation also constitutes the main objective of this article. Following the studies of Brück et al. (2011) and Regenbogen et al. (2012), I hereby create a methodological proposal for analysing the effects of multimodality on translation production from the perspective of the audio describer and translation reception—taking into account the influence of the multiple communication modes involved in AD on the comprehension of the target discourse by blind users.

The models presented above have in common the use of research methods applied in fields that deal with the study of the human brain, such as Cognitive Neuroscience. Here I refer to complex non-invasive techniques such as functional magnetic-resonance imaging (fMRI), magnetoencephalography (MEG) or electroencephalography (EEG). fMRI is a technique that measures brain activity by detecting changes associated with blood flow: when an area of the brain is in use, blood flow to that region also increases. MEG is a neuro-imaging technique for mapping brain activity by recording magnetic fields produced by electrical currents that occur naturally in the brain. In turn, EEG is an electrophysiological monitoring method used to record electrical activity in the brain by means of electrodes placed on the scalp. Other techniques applied to detect cognitive patterns in vision (eye-tracking) or emotional arousal (galvanic skin response) are increasingly being used in fields such as Psychology and even Translation Studies (see the use of eye-tracking techniques by Orero and Vilaró (2012), for instance).

This article examines the suitability of these techniques to the study of multimodality in analysing emotional information encoded in various semiotic systems. Furthermore, it describes the types of experimental design implemented in studies that apply the techniques mentioned above. These kinds of experiment and technique could help researchers in Translation Studies to analyse multimodal translation from a more rigorous perspective and shed some light on the mental processes that take place in the translator’s and the recipient’s brain.

The existence of a variety of multimodal elements in AD contributes significantly to compartmentalizing the object of study. For example, as is proposed in the experimental approach presented in section 3.3, one could consider whether the co-presence of several modes in AD—film visual content, AD semantic content, AD prosodic content, among others—is essential to the adequate transmission of emotions to blind users. In other words, by means of experimental techniques such as stimulus-response timing the researcher may be able to assess the relevance of the different modes of expression to the adequate receipt of a translation in cognitive terms. It would be useful not only to determine the level of understanding of the emotions expressed by characters in the film but also to assess whether blind and visually impaired users experience emotional contagion, that is, empathy with characters’ emotions.

In a way, these kinds of experimental study of AD reception will contribute to checking the effectiveness of the translation process itself, hence its quality, by assessing the effects of a translation on the recipient. In this sense, I agree with Kruger and Kruger’s (2017) request for translation research to apply both theoretical and empirical work to methodological approach promoted in this work. It is intended to be a method for studying the multimodal translation of emotions in AD. In this sense, this proposal will be posited merely as an approach to the study of multimodality in translation and will require testing via a pilot study prior to any final implementation. It is based on Neuroscience studies that address the processing of affective multimodal information through neuro-imaging techniques.

Since this study aims to analyse the transmission of emotions through different communication modes involved in cinema AD, the subject sample should consist of a

group of visually impaired people. In essence, the main objective of this work is to develop an experimental design based on the theoretical models presented in section 2. This design should be suitable for determining the impact of different communication modes on the transmission of emotions in film AD.

Methodologically, such an empirical and behavioural study based on cognitive theories would make it feasible to assess and quantitatively measure the effectiveness of AD and hence of multimodal translation to convey the emotions expressed in a given film to the targeted visually impaired recipients.

2. Theoretical and methodological neuroscientific models for studying emotions in multimodal communication

2.1 General brain processing of multimodal stimuli

As one might expect, multimodal stimuli processing adjusts to the standard functioning of the human brain. This is presented in a condensed manner as follows: signals from the physical world are perceived through our senses and turned into sensory inputs that activate a network of functionally specialized brain areas for processing. To give an example, processing an image activates visual regions of the brain exclusively devoted to recognizing certain image features such as form (area V3 of the visual brain), colour (area V4) or motion (area V5). This functional specialization is executed at such a microscopic level that, in the case of colour, certain columns of neurons in area V4 exclusively fire with specific light wavelengths (e.g., 450 nm (dark blue)), whereas other groups of neurons, at the same node, activate with a wavelength of only 600 nm (yellow).

Furthermore, these specialized processing systems are wired to work in parallel and in a distributed manner within the whole sensory brain cortex. Therefore, there are bottom-up and top-down connections. If we continue to use the example of colour, both columns of neurons will fire in parallel when an object reflects wavelengths corresponding to dark blue and yellow at the same time. This kind of parallel and functionally segregated processing seems to correlate with the creation of a conscious experience of the attribute or stimulus processed too. Zeki coined the term *microconsciousness* to refer to this fact in relation to visual perception:

... each node of a processing-perceptual system creates its own microconsciousness. We propose that, if any binding occurs to give us our integrated image of the visual world, it must be a binding between microconsciousnesses generated at different nodes (Zeki & Bartels, 1999, p. 225).

Another property of information processing systems is the direction of information flow. When perceptual operations are led by the incoming data provided by sensory stimulation, it is said that data processing takes place in an upward fashion (ascending or bottom-up processes) in a way that perceptual representations become more and more complex as processing occurs. Conversely, processing is done in a downward fashion (descending or top-down processes) when the information direction is inverted and mental representations are complex. For instance, our previous knowledge, objectives or expectations lead and modulate the operations of the system, actively searching and extracting sensory information from lower processing areas. This type of top-down processing might help to differentiate a human face in an abstract painting that does not, in fact, contain the representation of a face. As pointed out by Smith and Kosslyn (2008), “each perception act involves both types of processing” (p. 57).

According to the interpretation of each specific stimulus, either one type of processing or the other will be more relevant; but most of the time both ascending and

descending processes work together to recognize visualized objects more efficiently. The fundamental fact of this kind of processing, which may also be valid for multimodal perception, is that both information flows take place in a dynamic interaction in which bottom-up and top-down transmission constantly interrelate.

Research on multimodal translation requires being conversant with the basic principles according to which the brain processes external inputs, that is, functional specialization and ascending–descending information processing, as revealed by Zeki and Bartels (1999), and Smith and Kosslyn (2008), respectively. In methodological terms, these models are convenient for identifying specific modes of expression affecting the comprehension of AD by the blind user—for example, the voice pitch of the describer and the possible influences based on previous experiences. In case the user despises high-pitched voices, our design would need to take this into account by analyzing the scores on AD comprehension for such users. Besides this, it is useful for researchers to be well versed in these theoretical models in general if they intend to design imaging experiments involving the identification of active brain regions while multimodal information is processed.

2.2 Multimodal affective information processing

Bearing in mind that most people can almost immediately assess and decode the emotional states of other human beings through the different cues they produce (vocal and facial expressions, body gestures, etc.), all of us could be considered *experts in emotions*, so to speak.

One of the communication modes that provides much affective information is that of vocal expression, articulated through a series of sub-modes such as verbal vocal speech (spoken or sung), and especially through prosody, but also by means of non-verbal vocal expressions (laughter, screaming, crying, sobbing, sighing, etc.). This section presents Brück et al.'s model (2011), which focuses on the cognitive processing of the different aural and visual modes to express emotions and on the multimodal integration of such modes within the brain.

In particular, regarding the processing of affective information received through different sensory channels (aural and visual) and realized as several expression modes (vocal verbal and non-verbal, and facial), Brück and colleagues' model describes the functions of the brain areas involved in perceiving and integrating audiovisual stimuli linked to the recognition of emotions in human beings. If researchers are to develop methodological tools to study the generation of emotions by AD, it is essential that they at least be conversant with the way in which such operations take place in the brain.

This model is based on the similarities presented by emotional processing in both sensory modalities, aural and visual, which are determined by these facts (Brück et al., 2011, p. 395):

- a) The most basic stages of perceptual analysis take place in the fusiform face area (FFA), which is responsible for facial recognition and is located at the fusiform gyrus of the brain cortex; the voice areas are responsible for recognizing vocal messages and they are located at the auditory cortex (temporal lobe).
- b) The dorsolateral and orbitofrontal brain structures are involved in explicitly assessing visually and vocally expressed emotions.
- c) Emotional information is processed through two routes: the explicit or “cortical” and the implicit or “limbic”, in agreement with the classic model of LeDoux (1998).

Taking this into account, the authors describe three general processing steps which take place in the brain areas that manage the perception of audiovisual, non-verbal emotional signals—for example, a facial expression accompanied by a vocal sound implying anger:

- Visual and vocal signals extracted in the specialized primary cortices are associated with each sensory channel. This means that the image of an angry face will initially be decoded by general visual brain areas (V1 and V2) and later on by the specialized FFA, which will perform a facial recognition task. The same will apply to the vocal sound, but in the specialized auditory cortex.
- Auditory and visual information is integrated into a single percept in the posterior superior temporal cortex (p-STC). The signals will converge to form an integrated piece of relevant affective information from which meaning can be extracted either consciously or through unconscious processing.
- Affective information related to the activation of dorsolateral and orbitofrontal brain structures is cognitively managed and subjected to controlled evaluation.

These steps indicate that the brain may control such processes in a cognitive fashion, that is, they are modulated from higher areas; alternatively, the processes may occur in an implicit or uncontrolled manner, being implemented via limbic pathways (such as the amygdala). In the latter case, aspects such as expectations or memory do not have an impact on the process of receiving and decoding such sensorial information.

Becoming familiar with the way in which the brain integrates and assesses multimodal emotional information can prove very helpful for the study of multimodal translation. If the focus were placed on cinema AD as a particular translation modality, it would be advisable to analyse the role of multimodality in perceiving and comprehending the emotional aspects present in the ST. For this reason, the research team needs to follow the usual path that leads to the use of an effective study method, and consequently they have to pose several research questions, establish a hypothesis, select an experimental method, define variables for a study, prepare an experimental design, have a controlled experimental environment, sample an adequate group of subjects, etc. These issues are addressed in section 3, where the experimental stages of a pilot study are presented. Such a study method would presumably be effective in assessing the effects of several communication modes on blind and visually impaired users' reception of cinema AD.

As Brück et al. (2011, p. 384) indicate, several studies have shown that the superior temporal cortex (STC) of both brain hemispheres significantly and specifically responds to human vocal stimuli (Belin, Zatorre, Lafaille, Ahad & Pike, 2000; Fecteau, Armony, Joannette, & Belin, 2004; Grossman, Oberecker, Koch, & Friederici, 2010). Accordingly, this area of the auditory cortex (in locations distributed at the anterior, middle and posterior regions of the STC) could be described as a functionally specialized system for processing voice qualities.

Besides anatomical characterizations, studies have focused on describing the cognitive functions and competences developed by the temporal areas in voice processing. Take, for example, the extraction, assessment and categorization of non-linguistic vocal information, such as the ability to decode affective information, that is, recognizing emotions from vocal expressions. Studies using neuro-imaging techniques have shown the conscious response of the temporal areas during the perception of voices expressing an emotion in comparison with the response to neutral voices (Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009; Wiethoff et al., 2008).

If we consider that recognizing emotions from vocal expressions is key for blind and visually impaired people in their daily interactions with others, as well as through mediated information inputs (radio, TV, cinema, etc.), an appropriate subject for developing any experimental research on the quality and cognitive adequacy of current audio description practices may be that of *emotional prosody*. By this term is meant that vocal attributes provide the listener with information about the mental and emotional state of a speaker, which might not necessarily corresponds to their semantic speech content.. Consider, for example, the friend who told you they were doing quite well, but that from their voice tone you noticed a shroud of sadness that words could not disguise.

In this regard, extensive research on the acoustics of emotions has identified four prime acoustic parameters that signal emotions in voices: voice intensity, speech rate, voice pitch and energy distribution in the frequency spectrum (Johnstone & Scherer, 2000; Scherer, Johnstone, & Klasmeyer, 2003). The manner in which these parameters combine to express emotions (anger, sadness, fear and joy) has been systematically studied, as Brück and colleagues indicate:

While *anger* generally has been described to be indexed by increasing voice pitch paired with increases in loudness; *sadness*, for instance, has been characterized to show decreases in voice pitch, speech rate and loudness of a speaker's voice. *Fear*, on the other hand, has been suggested to be communicated by increases in voice pitch combined with increases in speech rate, while *joy* has been related to increases in voice pitch, loudness and speech rate (Brück et al., 2011, p. 395).

In the same vein, the line of research developed by Regenbogen et al. (2012, 2013) is similar to that illustrated in the present section. Brück et al. (2011) aim at developing an all-encompassing

cerebral network model ... that proposes the contribution of various cortical and subcortical brain structures to the processing of voice signals both in isolation and in context of accompanying (facial and verbal) emotional cues (p. 383).

To this end, they reviewed the literature of the field extensively. For their part, Regenbogen and colleagues approach multimodality through the study of empathy, recognition of emotion and individual responses to visual and vocal cues conveying emotional information. Following the outcomes of two fMRI studies, the authors offer a preliminary profile of the brain network dynamics in human emotional communication and empathy (Regenbogen et al. 2013, p. 1). The basic structure of their experimental design, in terms of emotional and neutral conditions and different stimuli sources (facial expression, semantic vocal sound and prosody), will be resorted to in order to create a proposed methodology for a potential reception study, as described in section 3. This experimental design should contribute to measuring the degree of empathy and recognition of emotion in blind users' reception of AD.

2.3 Multimodal communication in human beings: multimodal processing of emotional messages and empathy generation

In our communicative exchanges, most of the time unconsciously, we evaluate the signals emitted by other people constantly and reciprocally. The successful assessment of all signals is considered essential to generating empathy and social coherence. De Vignemont and Singer (2006) have approached this issue in their *empathy appraisal model*, which shows that empathy may be generated and inferred through the exchange of dynamic information via different communication channels and modes of expression.

Regenbogen et al. (2012) use this model to conduct an experimental fMRI study on the recognition of emotions and affective responses (empathy) in subjects when three communication modes are presented in different combinations: facial expressions, speech and prosody. Their initial assumption is that empathy levels will be higher when the type of emotional information expressed via these three modes is congruent and, conversely, the ability to recognize and share the emotional state presented to the subject will be affected when one of the three modes conveys neutral information. This notion is based on the following hypothesis by Regenbogen et al. (2013):

Within the empathy appraisal theory, if one of the multiple modes from which information is extracted to generate an emotion is incongruent with the rest, that may lead to a failure of sufficient embedding of an emotion, which would prevent successful multimodal integration and unequivocal empathic response. (p. 2)

The results of the experiment conducted by Regenbogen et al. (2012, 2013) confirmed that multimodality represents a facilitative effect for recognising emotion. This was proved at a behavioral level—that is, through subject ratings of self and other emotions—and also at a neural level, with the activation of brain areas involved in the processing of information pertaining to the three channels:

... emotions displayed via facial expressions, prosody, and speech content were associated with bilateral activations in fusiform gyri (FFG), auditory cortices (AC), and left angular gyrus (AG), respectively ... The identification of these four different network-nodes associated with the perception of human interactions constituted a prerequisite for understanding the dynamics that underlie multimodal integration. (Regenbogen et al., 2013, pp. 2–3)

The models and studies presented by Regenbogen et al. (2012) depend on neuro-imaging techniques whose implementation is both complex and costly. Unfortunately, the field of Translation Studies usually lacks the resources to gain access to this type of scientific equipment. Other instruments and techniques such as EEG and eye-tracking have been applied in cognitive studies related to multimodal translation (Órero & Vilaró, 2012, p. 295). In the Órero & Vilaró study (2012), the authors intended to examine whether subtle details which may go unnoticed by the viewer should be audio described by conducting a perception experiment using eye-tracking technology and questionnaires. Nonetheless, there is still much room to exploit the potential of neuro-imaging techniques (fMRI or MEG) in order to retrieve information on the activation of brain areas when integrating multimodal stimuli in translation. Such experimental conditions have not been applied in Translation Studies or Interpreting Studies, for several reasons: the high cost of functioning and maintenance; ignorance about their handling and application in traditionally humanistic fields; the intrinsic complexity involved in creating and implementing experimental designs satisfactorily and in analysing and interpreting results (Chica Núñez, 2016).

3. Methodological application: experimental design proposal to study emotion transmission in multimodal translation

3.1 Multimodality effects on emotion reception in AD

Certainly, one of the aspects that might become useful for the study of the emotional aspects of AD would be the prosodic features of the narrator's voice. At this point, it would be advisable to address the matter of whether actual AD practices allow for emotional experiences to be transmitted (cognitive understanding) and generated (affective contagion) effectively to blind users.

In Spain, as well as in other countries with a similar tradition in this practice, professional AD follows a series of norms and guidelines that usually coincide in recommending neutral and objective descriptions, which are limited to reflecting just the most relevant elements of the visual narrative. This can be seen perfectly in the recent Spanish AD of the TV series *Games of Thrones* produced by ONCE (the Spanish National Organization for the Blind), where we could find almost no traces of emotional semantic content to assess in order to consider its suitability as a viable empirical study.

As expressed by several codes of good practice in AD, the wording of descriptions ought to be as impartial as possible and avoid using technical terms or subjective points of view (Rai, Greening, & Leen, 2010). In this sense, standard AD scripts follow Snyder's "what you see is what you say" rule (WSIWYS). In the words of Snyder (2005), "qualitative judgments get in the way of a good AD, since they constitute a subjective interpretation on the part of the describer and are therefore unnecessary and unwanted" (p. 195).

In cinema AD, the opinions voiced by users on this matter (Fernandez Iglesias, Martínez Martínez, & Chica Núñez, 2011) led me to set out the hypothesis that if we want AD actually to work and make a deep impression on the recipient, it is necessary to go beyond formal traditional descriptions and use aesthetic criteria that generate vivid sensations and feelings in blind people.

A number of research studies have also found that alternative approaches to AD style, using subjective and emotional vocabulary, metaphors, neologisms and description of emotions, and even specific film terminology, resulted in a more engaging, enjoyable and entertaining experience for blind and visually impaired users, as indicated by Walczak and Fryer (2017). For instance, in cinema AD, Ramos Caro (2013) compared the reception of neutral descriptions with descriptions containing an emotional value by adding literary and stylistic devices to the language of AD, which implies a semantic approach. As measuring tools, she used Likert scales and heart-rate recordings to gauge discrete emotions such as disgust, fear and sadness. Her results point to significant statistical differences in the case of fear and sadness, but not in the case of disgust.

Previously in this article, it was stated that neurobiological studies of vision indicate that the emotional processing of images (made by the limbic system) is essential to extracting meaning from the object visualized (Damasio, 1999; LeDoux, 1998). This is because the visual signal connects with the amygdala (one of the main emotion-processing centres in the brain) both before and after it reaches specialized areas of the visual cortex, where the percept becomes conscious. In this sense, it can be asserted that the describer's voice can enhance and intensify the plot of a film as much as the characters' intentions by fitting in with the emotional landscape of the film (Fernández Iglesias, Martínez Martínez & Chica Núñez, 2011, p. 78). Provided that multimodality proves itself essential to conveying emotions in AD reception, a valid methodological approach to studying its effects on blind and visually impaired users has to be put into practice in future research on this topic. Consequently, the following sections present an experimental design proposal to carry out efficient studies on the reception of emotions through the multimodal artifacts delivered by AD, such as image content, the semantic content of dialogues, AD semantic content and AD prosody.

3.2 Hypothesis and research questions

As has been mentioned in the previous section, traditional cinema AD practice usually follows a series of guidelines which state that the narrator's voice is not supposed to detract from the scenes or the characters' dialogues; for this reason, guidelines advise that the voice of the narrator should be neutral and free from emotion (AENOR, 2005; OFCOM, 2000). However, recent practice is moving away from this and proposing a more creative, subjective and emotional AD practice. Following this, the hypothesis is that using prosodic qualities of the voice congruent with the emotional mood of the film in AD does not detract from it but in fact brings value to it and improves the comprehension of AD discourse and the plot of a film generally. Moreover, it enhances the transmission of the emotions and aesthetic features of the ST to the users and, as a consequence, they should engage more actively with the original work.

Accordingly, the following research questions have been formulated:

- Is an effective transmission of the emotions generated by the film prevented by following traditional AD guidelines?
- What is the contribution of multimodality to the transmission of emotions in AD?
- How can we measure the effects of multimodality in the recognition of emotions and their empathic contagion to cinema AD users, i.e. experiencing the same kind of emotions as those of characters?

3.3 Experimental methodology, variables, research design and experimental procedure

With the intention of finding answers to these questions, a pilot experimental design is proposed in this article. It is based on psychophysics, that is, a recording methodology that involves stimulus presentation and subject reaction and which aims to collect data from a group of subjects composed of blind and partially sighted people, on the one hand, and from a group of sighted subjects, on the other. This choice of subjects is motivated by the necessity to compare the responses of both groups, since they have different access profiles to communication. In the case of people with problems obtaining access to visual codes, communication is entirely based on auditory modes and, to a lesser degree, on tactile ones. The size of the group of subjects should be representative enough of the population of both groups. However, bringing together sufficient blind and visually impaired participants could prove harder than expected in experimental research environments, as reflected by the review of Ricciardi, Bonino, Pellegrini and Pietrini (2013) in the field of neurobiology when they attempted to study the functional organization of the blind brain. The size of the groups of early and late blind individuals ranged from one to 21 subjects in the 39 studies they reviewed. For this reason, I would initially suggest recruiting a sample of 20 subjects altogether to ensure a valid level of representativeness.

The present experimental design focuses on transmitting the emotional aspects of the source work through AD. In line with Regenbogen et al. (2012), the dependent variables of the study are the recognition of emotions and the transmission or contagion of such emotion (empathy). These variables are transmitted through several modes of expressing emotion such as semantic speech content, prosody and, for blind and partially sighted subjects, facial expressions. In this way, the experimental design focuses on the influence of multimodality on these dependent variables.

In order to do so, the study could use a series of independent variables based on the three modes of expressing emotion selected plus their emotional congruence or incongruence with respect to the ST. For visually impaired subjects, the visual mode may not contribute to the effect of multimodal integration; however, it should not be disregarded, because participants with reduced vision could be included in the sample. As a consequence, the three modes or *channels* of presenting information to subjects should meet the following characteristics:

- Original emotional video clip (OEVC). Its content consists of a close-up of a facial expression of a given character clearly showing an emotional state. The emotional repertoire will be limited to four primary emotions: anger, sadness, fear and joy (Brück et al., 2011). Estimated duration: 10–12 seconds.
- Semantic content of the audio description (SCAD). Its emotional value will be either congruent or incongruent (neutral) with the emotion expressed in the OEVC.
- Prosodic content of the audio description (PCAD). Likewise, its emotional value will be either congruent or incongruent (neutral) with the emotion expressed in the OEVC.

These variables form an experimental design that presents four conditions, structured as follows:

- *All emotional*: OEVC + emotional SCAD + emotional PCAD. Both, SCAD and PCAD are congruent with the emotion expressed in OEVC.
- *Neutral SCAD/PCAD*: OEVC + neutral SCAD + neutral PCAD. Both, SCAD and PCAD are incongruent with the emotion expressed in OEVC.
- *Neutral SCAD/emotional PCAD*: OEVC + neutral SCAD + emotional PCAD. SCAD is incongruent with OEVC and PCAD is congruent with OEVC.
- *Neutral PCAD/emotional SCAD*: OEVC + emotional SCAD + neutral PCAD. PCAD is incongruent with OEVC and SCAD is congruent with OEVC.

Approximately 15–20 video clips for each of the four emotional values should suffice to conduct the pilot study, since the purpose of this type of study would be to verify that the experimental design and the conditions proposed are valid. The number of items used as stimuli could be modified in subsequent studies.

The experimental design will then be presented to subjects via a computer display system with technology that enables the subjects to input their responses. This will enable video, audio and texts or assessment scales (among other devices) to be presented in the display devices (screen and headphones or speakers) of the researcher's computer. Presentation will follow a continuous pattern in order to obtain quantitative data from the subjects' replies via keypad or mouse inputs. This type of technique is usually applied with software such as E-Prime, but for this study the software suite PsychoPy (Pierce, 2007) could be preferred, as it is easily programmable and open source.

After the presentation of each video clip with AD, the subject will be asked two questions:

- How does the character feel? This question is intended to measure “emotion reception and recognition”.
- How do you feel in relation to what you have seen or heard? “Empathy” would be assessed with this question.

The aim of asking these questions is to retrieve information from the subjects about the emotional state of the character and their own emotional response using a seven-point scale that ranges from -3 (very negative) to $+3$ (very positive); 0 represents a neutral emotional state. The number and size of potential subject groups is to be determined by the researchers, but a balanced and representative sample is advisable if representative data are to be obtained.

The final purpose will be to check whether there is any recognition of emotion, on the one hand, and whether any empathy is evoked in the subjects in response to that emotion, on the other—that is, whether blind or visually impaired users feel the same emotion as that of the character in the film. Ultimately, the study aims to measure the effects of multimodal information integration at a behavioural level with respect to emotions in the case of AD. Accordingly, the reaction times of subjects will also be recorded in order to determine, at least behaviorally, the cognitive complexity of each task. Theoretically, the reaction times should be inversely proportional to the efficacy in decoding emotional states.

These experimental conditions could be expanded and enriched by adding further expression modes to the final message (lively, sad or thrilling music, for instance), and they too could, in turn, be either congruent or incongruent with the ST. However, for this pilot experiment, the design and conditions will be kept just as they were presented above.

4. Conclusions

This article presents a methodological proposal to conduct a pilot study on the multimodal integration of emotions in film AD. As the starting point for its development, various theoretical and methodological models from the field of Social Neuroscience were taken as reference. This discipline integrates theories and methods meant to help us understand both human cognitive and affective processes and the behaviour of individuals and groups in society.

Among such models we can find Brück et al.'s (2011) proposal, which addresses the following aspects from a comprehensive neuroscience standpoint: the acoustics of emotions; the brain areas that control the processing of speech and other vocal sounds, turning human beings into voice experts by their very nature. Their model also focuses on the cerebral processing of emotional prosody—"it is not what you say but how you say it"—or the integration of emotional information that originate in different communication modes (vocal verbal (speech), non-verbal vocal sounds (laughter, shouting, sobbing) and visual (facial and body gestures)) in the regions of the brain activated during the processing of these signals.

Besides this model, further studies based on fMRI techniques were described (Regenbogen et al., 2012, 2013), according to which some nodes in the brain cortex seem to respond to the integration of multimodal information (facial expressions, prosody and speech). Their functioning is related to the empathy appraisal model of De Vignemont and Singer (2006). Based on their outcomes, activation conditions in the superior brain cortex take place when the different expression modes involved in the message are congruent.

This same principle was applied to the experimental design presented in this article in relation to the notion of multimodal translation in AD. AD can no longer be defined merely as "the translation of images into words", because it also includes the translation of other, more subtle sub-modal elements, such as the emotions of film characters. These can be perceived via semiotic elements such as facial and body gestures and the original dialogues (i.e. semantic and prosodic content). All of them would qualify as under-studied aspects in the field of Translation Studies; however, their analysis might allow us to imbue the concept of *multimodal translation* with some clearer, more structured meaning. As these studies have shown, multimodality generally plays a facilitator role both in the comprehension of feelings and emotions and in the generation of empathy in the audience.

Finally, if the pilot study presented here is to be implemented at some point, I suggest that blind and visually impaired subjects constitute the main study population. If the results of the pilot study are deemed satisfactory, the researchers could also think about expanding it, first, by adding new, enriching experimental techniques such as galvanic skin response or EEG recordings and, subsequently, by supplementing them with more sophisticated techniques that capture neural correlation information, for example fMRI or MEG.

As a more general concluding remark, in line with Kruger and Kruger (2017), I would like to stress the importance of promoting this kind of empirical approach in Translation Studies. It should go a long way to helping practitioners better understand the cognitive features of translation production and reception, and also their effects on society and ideology as higher instances of the cultural concepts involved in the translation process.

Acknowledgements

This article was carried out thanks to the support of the OPERA project: Leisure and culture accessibility. Online portal for the dissemination and evaluation of accessible

audiovisual resources (FFI2015- 65934-R), funded by the Spanish Ministry of Economy, Industry and Competitiveness.

References

- AENOR (2005). Norma UNE: 153020. Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías. Madrid: AENOR.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309–312.
- Braun, S. (2008). Audiodescription research: State of the art and beyond. *Translation Studies in the New Millennium*, 6, 14–30.
- Braun, S. (2011). Creating coherence in audio description. *Meta: Journal des traducteurs/Meta: Translators' Journal*, 56(3), 645–662.
- Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, 8(4), 383–403.
- Chica Núñez, A. J. (2016). La traducción de la imagen dinámica en contextos multimodales. Granada: Tragacanto.
- Chica Núñez, A. J. (2015). Multimodality and multi-sensoriality as basis for access to knowledge in translation: The case of audio description of colour and movement. *Procedia - Social and Behavioral Sciences*, 212, 210–17.
- Damasio, A. R. (1999). *The feeling of what happens*. New York: Harcourt Brace.
- De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10(10), 435–441.
- Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology*, 19(12), 1028–1033.
- Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex?: An fMRI study. *Neuroimage*, 23(3), 840–848.
- Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, 65(6), 852–858.
- Fernández Iglesias, E. Martínez Martínez, S., & Chica Núñez, A. J. (2015). Cross-fertilization between reception studies in audio description and interpreting quality assessment: The role of the describer's voice. In R. Baños Piñero & J. Díaz Cintas (Eds.), *Audiovisual translation in a global context* (pp. 72–95). London: Palgrave Macmillan.
- Jiménez Hurtado, C., & Seibel, C. (2012) Multisemiotic and multimodal corpus analysis in audio description: TRACCE. In A. Remael, P. Orero, & M. Carroll (Eds.), *AVT and media accessibility at the crossroads. Media for All 3* (pp. 409–425). Amsterdam: Rodopi.
- Jiménez Hurtado, C. (2010a). Fundamentos teóricos del análisis de la AD. In C. Jiménez Hurtado, A. Rodríguez Domínguez, & C. Seibel (Eds.), *Un corpus de cine: Teoría y práctica de la audiodescripción* (pp. 13–56). Granada: Tragacanto.
- Jiménez Hurtado, C. (2010b). Fundamentos metodológicos del análisis de la AD. In C. Jiménez Hurtado, A. Rodríguez Domínguez, & C. Seibel (Eds.), *Un corpus de cine: Teoría y práctica de la audiodescripción* (pp. 57–110). Granada: Tragacanto.
- Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. *Handbook of Emotions*, 2, 220–235.
- Kruger, H., & Kruger, J. L. (2017). Cognition and reception. In J. W. Schwieter & A. Ferrerira (Eds.), *The handbook of translation and cognition* (pp. 71–89). Malden, MA: Wiley Blackwell.
- Lachaud, C. M. (2013). Conceptual metaphors and embodied cognition: EEG coherence reveals brain activity differences between primary and complex conceptual metaphors during comprehension. *Cognitive Systems Research*, 22, 12–26.
- LeDoux, J. (1998). *The emotional brain: The mysterious underpinnings of emotional life*. New York, NY: Simon and Schuster.

- OFCOM (2000). ITC Guidance on standards for audiodescription. Retrieved from http://www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/archive/audio_description_stnds/ (Accessed November 2017).
- Orero, P., & Vilaró, A. (2012). Eye tracking analysis of minor details in films for audio description. In R. Agost, P. Orero, & E. di Giovanni (Eds.), *Multidisciplinary in audiovisual translation*. MonTI 4, 295–319.
- Peirce, J. W. (2007). PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1), 8–13.
- Rai, S., Greening, J., & Leen, P. (2010). A comparative study of audio description guidelines prevalent in different countries. Retrieved from http://www.rnib.org.uk/professionals/Documents/International_AD_Standards_comparative%20study_2010.doc (Accessed July 2017).
- Ramos Caro, M. (2013). *El impacto emocional de la audiodescripción*. PhD Thesis, Universidad de Murcia.
- Regenbogen, C., Habel, U., & Kellermann, T. (2013). Connecting multimodality in human communication. *Frontiers in Human Neuroscience*, 7(774), 1–10.
- Regenbogen, C., Schneider, D. A., Gur, R. E., Schneider, F., Habel, U., & Kellermann, T. (2012). Multimodal human communication: Targeting facial expressions, speech content and prosody. *Neuroimage*, 60(4), 2346–2356.
- Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to understand the sighted one!: Is there a supramodal cortical functional architecture? *Neuroscience & Behavioral Reviews*, 41, 64–77.
- Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith, (Eds.), *Handbook of affective sciences* (pp. 433–456). Oxford: Oxford University Press.
- Schwietter, J. W., & Ferreira, A. (Eds.) (2017). *The handbook of translation and cognition*. New Jersey: John Wiley & Sons.
- Smith, E. E., & Kosslyn, S. M. (2008). *Procesos cognitivos: modelos y bases neurales*. Madrid: Pearson.
- Snyder, J. (2005). Audio description. *The Visual Made Verbal Across Arts Disciplines—Across the Globe*. *Translating Today*, 4, 15–17.
- Tymoczko, M. (2012). The neuroscience of translation. *Target*. *International Journal of Translation Studies*, 24(1), 83–102.
- Vercauteren, G. (2012). Narratological approach to content selection in audio description: Towards a strategy for the description of narratological time. *MonTI: Monografías de traducción e interpretación*, 4, 207–230.
- Walczak, A., & Fryer, L. (2017). Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment*, 35(1), 6–17.
- Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody: Influence of acoustic parameters and arousal. *Neuroimage*, 39(2), 885–893.
- Zeki, S., & Bartels A. (1999). Toward a Theory of Visual Consciousness. *Consciousness and Cognition*, 8, 225–259.