Psycholinguistics and perception in audiovisual translation

Louise Fryer

Introduction

This chapter explores audiovisual translation (AVT) in relation to models of psycholinguistics and cognitive perception. It introduces the concept of media accessibility as embracing various types of translation—namely sign language interpreting, subtitling for the deaf and hard of hearing (SDH) and audio description (AD) for people who are blind or partially sighted—that enable users with a sensory disability to access audiovisual media content. It focuses on the subjective nature of perception, and addresses Gambier’s (2009) concerns about loss in AVT, demonstrating how language and other forms of knowledge can compensate for the loss of information in sensory modes that cannot be perceived directly by the user. It argues that access modes of AVT such as AD and SDH are no less prone to loss or subjectivity than lexical modes of translation. It also discusses how the concept of immersion or presence can be of use in measuring the efficacy of AD and SDH and how Gibson’s (1979) concept of affordances and an understanding of visual perception can help describers make choices in content selection. This chapter therefore addresses Vermeer’s (1989) concern as to how translators choose between different translation options that may appear to be equally possible and appropriate.

Media accessibility

Most AVT modes are interlingual, translating the foreign source language into a different target language so that viewers who do not speak the former will be able to understand the dialogue by means of subtitles, dubbing or voice-over. However, the access modes of AVT are not only intralingual, but also intermodal. They translate information from one sensory mode to another, leaving the target language unchanged. For this reason, some scholars have wondered whether access modes are really translations at all. Gambier (2009) for example suggests these types of translation may be better thought of as adaptation, manipulation, transfer or remake. However, since the 1970s, people who are D/deaf or who have impaired hearing have been able to access the audio component of an audiovisual product through visually presented words via subtitling and sign language interpreting. Since the mid-1990s, the visual elements
have been made audible for people who are blind or whose vision is impaired through AD. Accessibility is now acknowledged as an umbrella term that encompasses all modes of translation for minorities with sensory disabilities (Díaz Cintas and Anderman 2008).

It is also recognized that the access modes go beyond their primary purpose. They are useful not only for people with specific needs but also more broadly for anyone in a disabling situation who is unable to access the full AV content. For example, hearing people trying to listen in a noisy environment might welcome subtitles and sighted people might wish to listen to an AD version of a film while driving. Young children acquiring language, migrants or other non-native speakers might benefit also from the added context and rich vocabulary that AD provides. Equally, non-native speakers might find the dialogue easier to follow if they watch the same language subtitles (captions) as they watch a television programme or a film. Yet, if much of the research in the field of media accessibility has been concerned with discovering ways in which to maximize the media experience for people previously excluded, it has also shed light on the way media are experienced by people who are able to access both audio and visual channels.

**Current debates**

In his book *Enforcing Normalcy* (1995), Lennard Davis examines how cultural assumptions govern our conception of people with disabilities and drive the proliferation of ‘ableist’ discourses in society. ‘[N]ormalcy’, Davis argues, ‘is constructed to create the “problem” of the disabled person’ (1995: 1) and boundaries between the abled and the disabled are constructed and maintained through a process of cultural construction. Ultimately, understanding disability requires a deep understanding of the ‘able’ or ‘normal’. Against this background, a recent survey by the Erasmus-plus-funded research project ADLAB PRO (2017) sampled tutors and trainers of 192 AD courses and found that ‘knowledge of the needs of the visually impaired’ was ranked as the third most important competence that should be taught to trainee describers. Following Davis’s logic, in order to understand blindness, it is necessary to expand our understanding of sight. Snyder (2005) urges describers to develop their powers of observation, while Fryer (2016: 60) advocates ‘that describers should reflect the casual observation of the average member of the sighted audience but develop their writing skills and grasp of language so as to be able to convey that visual information effectively.’ In order to do so, it is important that describers have a model of how ‘normal’ sight works.

**Audio description as constrained translation**

AD is a form of constrained translation (Mayoral *et al.* 1988). Users cannot rely exclusively on the description, and need to make use of other types of information that they receive directly from the source text. Crucially, the AD must fit within the time constraints of the source soundtrack: descriptions can only be incorporated at those points where they do not cause overlap with the other channels on the soundtrack, especially the dialogue. In screen modes such as TV and film, this means that the describer should generate ‘in times’ and ‘out times’ in their script to tie the descriptions to particular moments in the soundtrack. In live AD modes—e.g. theatre, opera and sport—this timing is not absolute, but relative and putative. The describer will select two cues usually from the dialogue, but sometimes from non-verbal sound effects or visual cues, such as lighting; it should be noted, however, that the describer should be prepared to improvise, should the timing change. The fact that describers need to maintain synchrony with the source text in various ways means that they will have less freedom than lexical translators.
Unsurprisingly, in light of these constraints, much of the literature on AD has focused on the process through which describers decide which of the many elements in the visual array should be selected or omitted from their description (Rai et al. 2010, Greening and Rolph 2007, Remael and Vercauteren 2007). Writing about interlingual lexical translation, Vermeer notes that ‘[w]e still know too little about the functioning of the brain, and hence of culture and language, to be able to rely on much more than intuition when choosing between different variants which may appear to the individual translator to be equally possible and appropriate’ (1989: 231). Recent advances in neuroscience, however, mean that much more is now known about the brain, although there is still a long way to go before we are able to fully understand how language is processed (Feldman 2008). Access modes of AVT, in particular, have raised questions about user reception. Although the translator and the target audience are ‘part of the same speech community’ (Piety 2004), AD and SDH users may have no direct sensory experience of what is being subtitled or described. Consequently, neuroscientists, psychologists and translators are beginning to recognize that the audio and visual components of AV media do not function independently; instead, the meaning conveyed through each of these communicative channels influences and interacts with the other.

**Multisensory integration**

Humans are skilled at combining different sensory codes of meaning into a coherent whole through a process known as multisensory integration. The effect on people with no sensory impairment is thought to be cumulative, such that each sensory mode adds an extra level of understanding. For example, in the struggle to hear comments of an individual against the background chatter of a cocktail party, watching the speaker’s lips moving corresponds to an increase in the volume of the sound of interest of 15–20 decibels (Sumby and Pollack 1954). However, there is a hierarchy amongst the senses. Sight is prioritized over sound, such that auditory information is misattributed to a visual stimulus within a certain temporal window. For example, the ‘ventriloquist effect’ (e.g. Choe et al. 1975, Bertelson 1998) occurs when we mistakenly attribute the speech we hear as coming from the mouth we see moving, that is the mouth of the dummy rather than from the motionless mouth of the ventriloquist. In other words there is a visual bias of auditory location. A more complex explanation of the process thought to underlie multisensory integration is given below.

Temporal simultaneity is important for audiovisual binding to take effect, such that the audio and visual signals are recognized as belonging to the same event. It has been recently discovered that semantic congruency—loosely understood as the harmony or agreement between the meanings of several stimuli—also plays a role (Laurienti et al. 2004) in facilitating this recognition. We are more likely to recognize that a sound and an image are part of the same event if it makes sense that they should be so. For example if we hear a gun and see a body crumple, we will infer that the person has been shot, especially if we are watching a Western, the victim is a bank robber and the gunman a sheriff.

Eramudugolla et al. (2011: 60) have shown that speech is a particularly important example of multisensory integration because of its behavioural relevance to humans and the existence of specific brain regions that appear to be specifically tuned to auditory speech and lip gestures. Most modes of AVT are concerned with translating speech, and that is, to some extent, also the case with some of the access modalities. As Kruger explains, ‘[a]udio description . . . is substantially different from other AVT modes like dubbing and subtitling, primarily because the focus in other modes is on dialogue’ (2012: 232).
Yet, because AD is delivered orally and is a text to be listened to, rather than read, it shares many qualities with speech—as will be further elaborated below.

**Context**

Translation of any kind and, in particular, AVT, does not exist in a vacuum, but occurs within a given context. The context might be physical, as is the case with AD taking place in a cinema, museum, theatre or sports stadium, and will vary along technical and social lines according to the medium being described:

- AD in a cinema is pre-recorded, mixed with the soundtrack, and closed—i.e. the user chooses to hear it via headphones or an application on his or her mobile phone. By contrast AD on TV is pre-recorded and mixed with the soundtrack and can be open—such that, when selected, it is heard by everyone in the room. AD in UK theatres is usually live, closed and received via a headset that does not fully cover the ears, so that the sounds from the stage are heard naturally, whereas in French theatres, it may be recorded but triggered live by a technician to keep it in synchrony with the dialogue. Finally, AD in a museum might be delivered live and heard naturally by anyone within earshot.

- The context might also vary along social lines. AD users might be attending a play with friends, watching TV at home with their family or via video on demand (VoD) on their computer, alone. Users might know the describer personally—as happens often in the theatre—or not—in the case of film or TV AD that, in many countries, is delivered by a voice talent.

At any rate, context is important because it is a ‘cognitive phenomenon’ (Aijmer and Simon-Vandenbergen 2004: 1783). The user’s understanding of the AD message will be affected not only by what the describer says, but also by ‘the mental states (what is known or believed)’ (ibid.) that the user brings to the experience.

**Mental models**

Psychologists (Schubert *et al.* 2001, Jones 2007, Pinchbeck and Stevens 2005) have postulated that data is organized in the brain through schemata. A schema is a shortcut by which the brain groups together everything a person has discovered, either directly through their senses (bottom up) or indirectly through their intellect (top down), about a person, place or object. The notion of *Umwelt* (von Uexküll 1940), understood as the perceptual world in which an organism exists and acts as a subject, usefully captures the influence of our environment on our experiential understanding and, ultimately, the expectations that individuals derive from the latter. Attending a Shakespeare tragedy, for example, AD users would not expect to laugh; should laughter feature in a performance of *Hamlet*, for example, it would be harder to process because it would be unexpected.

Schemas are said to create a reference frame into which we can slot the incoming data. They also present us with scripts (Schank and Abelson 1975), which activate for us a set of words we are likely to hear in that context. This can, for example, ease the potential confusion that polysemous words that have multiple meanings would have otherwise presented. In a written translation these will be homographs: words that look the same, but sound different, such as ‘tear’ (as in a drop of fluid excreted from the lacrimal gland when we weep) and ‘tear’ (as in the verb to rip or pull apart). In a description, which is orally delivered, confusion may more likely
stem from homophones that sound the same even if written differently. In a forest setting we might climb a ‘beech’. We are unlikely to climb a ‘beach’ even at the seaside. In the absence of a description of the location we may struggle to understand the word ‘beech’, upsetting the effectiveness of the translation by delaying or obstructing our ability to process it.

**Congruency**

One of the reasons why mental models are important is because the AD must be congruent with the context. Bearing this in mind can help describers select which elements to prioritize in their AD. For example, the AD of a horror film should prioritize ingredients essential to the horror genre in appropriate language. To describe a ‘witch’s familiar’ as a pussycat would be incongruent and conflict with the user’s expectation. But schemata can also save the describer valuable time. For example, the soundtrack of a horror film will convey scary audio information through the music, the creaking of floorboards, the hooting of an owl, the amplified heartbeat of the protagonist, that may need little explanation from the describer, while contributing tension (Fryer 2010). This may be because the user fills in the missing associated visual information from their *Umwelt*.

**AD and emotion**

Research by Ramos (2015) demonstrated that sound was able to induce symptoms of fear (measured by increase in heart rate) in blind participants, supporting research by Rickard (2004), which showed that music and sound cues in horror films increase galvanic skin response (sweating) in sighted participants. Rickard cites this as evidence supporting Schachter and Singer’s (1962) cognitive theory of emotion, according to which the type of emotion is determined cognitively, while the strength to which that emotion is felt is determined by bodily arousal. In the same way Fryer and Freeman (2014) found that blind people were able to state which emotion a film clip was trying to induce, even with no AD, but they only actively felt that emotion when listening to the film clip with AD. However, this was not the case for films designed to induce disgust. Ramos (2015), who has conducted relevant research on this issue, argues that ‘disgust-evoking scenes are not associated with any film genre in particular, and therefore lack concrete sound cues’ (2015: 83). It could therefore be said that viewers exposed to this type of scene lack specific scripts and schemata.

Fryer and Freeman (2014) further showed that AD users were more likely to feel frightened if the AD for a horror film was delivered by a human voice than by a synthesized voice. A human voice was able to add another layer of meaning beyond the semantic, through prosody.

**Prosody**

Wildgruber *et al.* (2006) point out that information about emotional states is generally given through non-verbal signals. These may be apprehended visually, as is the case with facial expression, eye-gaze, gesture, body posture and relative spatial position between two or more characters (for example, facing each other or turned away). Non-verbal signals may also be apprehended aurally—e.g. through intonation, inflection, pitch and pace, namely prosody. AD users can use prosody to understand a character’s emotional frame of mind, not simply from the dialogue but also from the cadences that the describer or voice talent uses to deliver the AD. Although some research has already highlighted the importance of prosody in oral forms of translation (Malmkjær 2005, Kübler and Volanschi 2012), recent
research in neuroscience (Leonard and Chang 2014) has highlighted its importance in our cognitive architecture by identifying specific groups of neurons that are primed to be activated by changes in a speaker’s tone. These particular neurons—which respond to relative shifts in pitch, either upwards or downwards—are located in the superior temporal gyrus of the brain together with other neurons that are believed to process context-dependent aspects of speech. It is clear then, that audio describers should pay as much attention to the oral as to the lexical-semantic aspect of the AD message, as a series of words can have its meaning completely reversed by a change of inflection. Take for example a small boy who wants to eat some brightly coloured sweets. His parents tell him that he cannot have the blue ones because they make him cross. The little boy thinks for a moment, then he says: ‘It’s not HAVING the blue sweets that makes me cross; it’s NOT having the blue sweets that makes me cross’. This illustration further feeds into this analysis of mental models, psycholinguistics and AVT from its reference to colour.

**Colour and cognition**

AD guidelines in different countries diverge on several points (Rai et al. 2010) but all agree that describing colour is important. This may seem curious for a translation mode primarily aimed at people with little or no visual perception. Bertrand Russell (1917) distinguished between ‘knowledge by description’ and ‘knowledge by acquaintance’ arguing there is no way of going from one to the other. It is a question fundamental to psychology and philosophy (and to AVT) whether propositional and experiential knowledge are of equal value. As Hayhoe (2017) points out in his discussion of blind people’s experiences in museums, the importance of art—which I extend here to include the visual element of AV material—goes beyond the directly perceptual to include personal, cultural and symbolic meanings from which blindness need not automatically exclude, providing that the same information is presented in a different (i.e. verbal) mode. Although a person who has been blind since birth cannot know what red looks like, they may still understand that culturally red is associated with heat, passion and danger. It would be important to know if, for example, a character was wearing a red dress. Yet the meaning of red would depend on the cultural context of the film as red is associated with good luck in Asian cultures but with the communist revolution in Russia.

These cultural variations of colour have been of particular interest to psycholinguists. The Sapir-Whorf hypothesis (see Kay and Kempton 1984 for a review) which is also known as ‘the doctrine of linguistic relativity’ *(ibid.:* 65) proposes that language is culturally determined and even our very thought processes are constrained by the sounds or visual appearance of the words with which we think. For example, the Western categorization of colour into the ‘seven colours of the rainbow’ is not universal but culturally determined: Isaac Newton, the seventeenth-century English scientist who viewed the spectrum of light waves through a prism, chose to divide it into seven because that was regarded as a ‘magic’ number at the time. The number of colours ‘seen’ in a rainbow varies today across cultures. People in the Baltic States ‘see’ a rainbow as having two colours, red and blue, while in China a rainbow is considered to have five colours (Oldfield and Mitchinson 2013). Other studies argue for universal colour categories (e.g. Berlin and Kay 1969, Regier et al. 2015). Interestingly, Gilbert *et al.* (2006) found that the speed of identification of a green or blue square positioned with others of the alternate colour, was dependent on whether the square was presented in the left or right half of the visual field. The researchers concluded that people ‘view the right (but not the left) half of their visual world through the lens of their native language’ (2006: 489). Roberson *et al.* sought to ‘shed light on whether language and
cognition are coupled or separable in the domain of color categorization and perception’ (2005: 4). By presenting different coloured stimuli to participants and asking them to name them, they showed that the Himba in Northern Namibia have five basic colour terms which more or less equate to the English terms red, white (including beige and yellow) black, blue (although these overlap) and green (including colours in the blue/green/purple range). They noted that the Himba’s ability to discriminate, or at least the number of terms they have available to describe different hues is severely reduced. The researchers also found a relation between colour vocabulary and memory performance. Comparing speakers of different languages, they found that participants were more likely to remember stimuli that were most consistently named in their own language. Roberson et al. conclude that ‘[w]hat appears to be universal, in this case, is the tight link between naming and memory’ (2005: 21).

**Visual and verbal meaning**

Some scholars (e.g. Braun 2008) have argued that translating images into words is more problematic than unimodal, lexical translation. However, as more is understood about psycholinguistics that position seems increasingly untenable. Mental models make clear the extent to which all translation users and translators are individual. As we each build up our own schemata based on our personal interactions and experiences, each of us develops our own mental lexicon. Jean Aitchison (2012) has argued cogently against what she calls the fixed meaning assumption: the idea that each word has a single uncontested meaning that is universally agreed. This contention is consistent with the views of American film critic, James Monaco: focusing on the relationship between pictures and meaning, Monaco advocates the iconicity of images, noting that whereas ‘a picture bears some direct relationship with what it signifies, a word seldom does’ (Monaco 2009: 176). However, that is not to say that an image has a single uncontested meaning. Monaco considers the image of a rose and explores the way its connotations vary depending on how it is shown by the filmmaker, whether it is in crisp or soft focus, with its thorns hidden or on display. A red rose has a particular meaning in Western culture where it is usually associated with love but in a historical film about the English Wars of the Roses, the colour of the rose would have a different connotation—a red rose being associated with the House of Lancaster which was at war with the House of York, whose own emblem was a white rose. The audio describer would need to focus on the colour rather than any other aspect of the rose, although they would not be expected to go into a long explanation regarding the socio-historical implication of the rose which may or may not be understood by every member of the audience, whether sighted or blind. The argument here is that the audiovisual translator, just as the translator of any other type of text, cannot control the reception of their translation. It is not possible to guarantee that the meaning one person attaches to specific words and phrases will be the same as another’s, not now and certainly not over time, as meanings drift and associations change. This is not, however, unique to receivers of audiovisual translations. Even audiences of the original will create their own personal variant of the film they are watching based on their knowledge, prejudices and assumptions (Remael 2012).

**Perception**

Perception is how we synthesize the multiple inputs we receive from our senses in order to, literally, make sense of the world. In the nineteenth century Hermann von Helmholtz (1977) defined perception as an active way of drawing unconscious inferences from the data we take in through
our eyes, ears, mouth, nose and skin. Bruner and Postman (1949) and Neisser (1976) developed the theory of perceptual hypothesis, according to which we make assumptions based on our experiences and test them against our sensory data. For example, if we see steam coming from a cup of coffee we hypothesize that it is too hot to drink. If we touch the cup, and a burning sensation on the pads of our fingers conveys confirmatory data, this reinforces the hypothesis. Next time, if we touch the cup, we will do so more carefully and perhaps blow on the liquid before taking a sip. Experience refines our hypothesis, such that we emphasize some aspects of the data (e.g. the density of the steam) and minimize others (e.g. the colour of the coffee). This function of filtering is designed to reduce cognitive load. Sensory data is inherently unstable and we constantly compare it against the permanent representations that are stored in the brain to allow us to infer and anticipate. The psychologist J. J. Gibson recognized that, despite our access to multiple sources of data, ‘[p]erception may or may not occur in the presence of information’ (1979: 56). In other words, although the information is there, we do not necessarily notice it.

**Visual perception: the importance of affordances**

Gibson concentrated on visual perception and developed what he called an ‘ecological’ approach (1979) in recognition of the fact that we are part of our environment. We do not exist in a bubble, nor are we passive recipients of sensory data. Our pupils dilate and we look longer at things that we like. If we do not like what we see, we close our eyes or turn our heads or even walk away. Gibson distinguished between the qualities of objects (e.g. colour, texture, composition, size, shape, mass, elasticity, rigidity and mobility) and their ‘affordances’, i.e. what objects and environments enable us to do with them. He argues that our perception of things can be judged to be ‘true’ if we are able to engage with them successfully. For example, we may be able to describe to a blind person the qualities of a chair and tell them it is made of mahogany, with an ornate carved backrest and a seat upholstered in red brocade, when what they really want to know is whether or not anyone is sitting on it and whether it looks strong enough to take their weight, i.e. whether it affords them the opportunity to sit down.

Support for Gibson’s model of visual perception comes from brain-anatomy studies (Gawande 2008) showing that 80 per cent of fibres that connect to the primary visual cortex come from brain regions associated with higher function such as memory; only 20 per cent come direct from the retina. This relates to a newer model of perception referred to as the ‘brain’s best guess’ theory of perception (Gregory 2009), that likens the brain to a radio receiver as ‘the mind integrates weak, rudimentary signals from a variety of sensory channels, information from past experiences, and hard-wired processes, and produces a sensory experience full of brain-provided colour, sound, texture, and meaning’ (ibid.: 30). Gawande, like Gibson, concludes that ‘perception is inference’ (ibid.). Describers should therefore seek to give users the type and quantity of information such that they can infer and anticipate.

In line with the best-guess hypothesis, Kruger (2017) argues that when watching a film, viewers with no sensory impairment co-create the narrative using the same cognitive architecture by which they create their perception of reality. They ‘fill in the gaps, predicting and assuming continuities even when absent’ (Kruger 2017). This is clear from the developing language of film. Directors no longer show a tracking shot, following a visitor’s progress as they step up to a door, knock on it and wait for it to be opened by the host, before the visitor steps in over the threshold. They simply show the guest approaching the door then cut to the guest indoors with the host. The audience fills in the rest. Furthermore, viewers identify with characters and indulge in embodied simulation such that they map a character’s actions onto their own sensorimotor and neural representations (Gallese and Guerra 2012).
Guidelines produced by the American organization Audio Description International (ADI) propose that the ‘first rule of description’ is to ‘describe what you see’ (Rai et al. 2010, Annexe 4). Too often, the focus in AD is on the visual. From Snyder’s (2005) perspective, describers should make decisions on what to narrate on the basis of what can be seen around them, based on a WYSIWYS (‘What You See Is What You Say’) approach. According to blind academic Georgina Kleege (2015), this standpoint reveals what she calls a ‘severe visual dependency’, which is not unusual amongst the sighted given the extent to which our perceptual systems are dominated by our vision (see Posner et al. 1976 for a review). Yet Fryer (2010) and others (Remael 2012, Szarkowska and Orero 2014) have argued that ‘sound as much as vision influences choices in AD’ (Fryer 2010: 205). Although the goal of AD is to compensate verbally for the missing visual information, concentrating on what we see with our eyes may paradoxically not be the best way to do it. This raises the question of how to recognize the best way.

Using Gibson’s concept of affordances, it is worth asking what audiovisual products afford. They might offer us the chance to become immersed in another world (e.g. a sci-fi film), to find out about another country (a documentary), to examine our own world more closely (a nature documentary); or just sit back and enjoy ourselves (a light comedy). Beyond the global question addressed to the audiovisual product as a whole, the same question might be then applied to the visual elements of each scene. For example, the magnified image of a venomous spider on a fingertip might make us wonder at the impact of the creature in relation to its size. Following Gibson, the AD can be deemed to be ‘true’ if the user is able to engage with the AD product successfully. That is, to respond to it in the same way as an audience member with no visual impairment.

Assessing impact

For a translation to be successful it must deliver ‘an equivalent intended effect’ (Lee 2008: 168). The difficulty comes in deciding how that equivalence should be assessed. The studies by Ramos (2015) and Fryer and Freeman (2014) cited above reflect attempts to introduce a scientific method for assessing impact, with AVT catching up with the moves made towards ‘scientific’ investigation of translation in the 1960s (e.g. Nida 1964). Use of existing psychological measures such as the Emotion Elicitation Scale (Gross and Levenson 1995) and the ITC Sense of Presence Inventory (Lessiter et al. 2001) allows the impact on audiences with and without sensory disabilities to be directly compared (Fryer and Freeman 2013). These measures rely on questionnaires and have been complemented by physiological measures such as heart rate and galvanic skin response in order to circumvent the limitations of post-hoc questionnaires—that have been criticized for being subjective and dependent on the user’s memory. More complex measures that track psychophysiological responses in real time have also been explored. For example, Jan Louis Kruger and his colleagues (2016) have turned to neurophysiological signals using an electroencephalogram (EEG) that measures electrical activity in the brain. Specifically, they draw on a paradigm that discriminates between those who were more emotionally involved and those who were less so, as distinguished by the levels of coherence in electrical signals between the prefrontal cortex and posterior parietal cortex of the brain. The significance of those brain areas is that activity in the prefrontal cortex is associated with executive function (attention) and regulation of emotions while
the posterior parietal cortex is associated with the imagination, particularly when we imagine ourselves to be someone else. Increased levels of coherence suggest greater control or increased inhibition in response to emotive stimuli resulting in suppression of the emotional response, while decreased coherence is associated with increased immersion (less suppression of emotional response).

**Functional neural networks**

Models of the way the brain works have moved from the idea of specialist regions dedicated to the processing of unique types of sensory data in favour of functional networks. These temporary links are thought to arise from the synchronous firing of neurons, increasing the probability that different modes of sensory data bind. This is the mechanism believed to underlie cognitive processes such as multisensory integration described above. Recent thinking suggests that this electrical brain activity reflects an evaluation of whether incoming data from different sensory channels is complementary and thus deserving of attention, or is redundant and should be suppressed (Senkowski *et al.* 2008). Brain regions are linked via a network of connected pathways that communicate through synchronized electrical activity. Oscillations in neuronal electrical activity are categorized into five frequency bands: delta (0.5–3.5 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (13–30 Hz) and gamma (>30Hz), (Gray *et al.* 1989). Imaging techniques, such as EEG, can show coherent frequencies suggesting that two or more regions are connected (Singer and Gray 1995). Measuring such activity allows us to ‘see’ how the synchrony or coherence changes according to cognitive tasks. Increased coherence in the beta band of the electrical signal is linked to less emotional involvement, because it implies the activation of the amygdala, a brain structure associated with attention that can limit emotional response. Reiser *et al.* (2012) argue that decreased functional coherence between the prefrontal cortex and posterior structures indicates a decoupling of the amygdala, in other words, a disconnection of the top-down control that prevents us becoming emotionally involved. The more loose the connection, the less inhibited our response to emotion-stimulating data will be.

Beta-coherence therefore is believed to decrease with immersion and increase with less emotional involvement (Reiser *et al.* 2012). Kruger *et al.* (2016) found that beta-coherence was lower for students watching a subtitled than an unsubtitled film clip, suggesting that the subtitled version was more affective. This is in line with self-report measures used in other studies (e.g. Wissmath *et al.* 2009). The benefit of physiological measures is that they are continuous and can measure changing emotional states over the course of a film. They are also thought to be objective, as responses cannot be faked to please the researcher. The disadvantages are that physiological measures are complex to administer, analyse and interpret.

Tim Smith (2013, 2015) has used eye tracking to try to understand how our eyes engage with the moving image and eye tracking has been seized on by members of the AVT research community, such that translations can reflect more accurately the ‘natural’ process of visual perception (Orero and Vilaró 2012, Di Giovanni 2014). However, as Di Giovanni herself points out, eye tracking can help to identify which parts of the screen to focus on, but not how much of that information to convey. Her participants reacted most positively to the ADs that included the most detail, which were those constructed in line with the eye-tracking analysis. However it has also been argued that too much detail can overburden the user, contributing too much to their cognitive load. While eye tracking may have a role to play in guiding some describer decisions, it cannot compare the impact of original and translated versions on people with and without sight.
Presence

Instead both types of audience can and have been compared using another measure borrowed from psychology. One that is based on ‘presence’ or a feeling of immersion that is so strong that we forget that what we are witnessing is a mediated experience (Lombard and Ditton 1997). It is presence that is disrupted by any form of incongruence discussed above. Presence was developed as a measure to test the efficacy of virtual environments. Much like those who doubt the efficacy of the access modes of AVT, early presence researchers (e.g. Zelter 1992) assumed that the more sensory channels a virtual environment replicated, the more immersed the consumer of that environment would be. However, they discovered that there was no direct linear relationship. In what is known as ‘the book problem’, Gysbers et al. (2004) point out that even low-immersion media such as TV or books can induce a sense of presence (Bracken 2010, Bracken 2005, Jones 2007, Wirth 2007). This has been shown to be the case, for both SDH and AD, where levels of presence were higher for people with a sensory impairment watching the target text compared with people with no sensory impairment watching its original counterpart (Wissmath et al. 2009, Fryer and Freeman 2013).

Linguistic compensation

In 2006, Yves Gambier wrote about the widely held view among specialists that AVT involves a certain degree of loss, and goes on to state that, should that loss exist, ‘[i]t cannot be restricted to verbal elements. Is there not a certain loss in the meaning of pictures when one reads subtitles?’ (2006: 3). He particularly bemoans ‘language hypertrophy’ or the prioritizing of linguistic over non-verbal communication modes including ‘camera moves, viewing angles, editing, soundtrack, tone of voices, facial expressions, gestures, gazes, body movements, all of which are also meaningful’ (ibid.). In response to Gambier, Heyns (2009: 131) talks of the ‘serendipitous gain’ of translation, those happy accidents that arise from expressing a phrase differently, such that the new phrase brings its own associations (scripts and schemata). It could be further argued that language hypertrophy is deserved (particularly for spoken language with its additional channels of prosodic information) especially because research that shows that the primary visual cortex, i.e. the part of the brain devoted by the sighted to vision, is recruited by blind individuals for memory, verbal and auditory processing tasks and speech (Bedny et al. 2011, Collignon et al. 2011, Ricciardi and Pietrini 2011).

As this overview has shown, language is essential to the integration of sensory input and therefore to perception. Thus, speech replaces visual input for people with a visual impairment, and written text replaces audible speech for those who are hearing impaired. Fryer (2013: 43) has proposed a ‘linguistic compensation hypothesis’ that underpins the accessibility practices in AVT, suggesting that a missing sensory channel need not necessarily be replaced by another sensory mode, but can be effectively replaced with words. Unfortunately speech does not save us from the treacherous waters of translation with its subjective mental lexicons and culture-bound meanings, but it does at least put us all, regardless of sensory loss, in the same boat.

Conclusion

This chapter has drawn on models of visual perception in order to provide a framework for AD as an example of the access modes of AVT. It has shown that by understanding the interaction between the audio and the visual and by drawing on concepts relating to perception and presence, we are indeed able to rely on much more than intuition when choosing
between different variants that may appear to the individual translator to be equally possible and appropriate. The interconnection between arts and science disciplines has contributed useful research methods, resulting in reception studies using scientific methods that have moved translators away from measuring meaning (i.e. comprehension) to measuring experience, broadening our understanding of the purpose of translation.

Further reading


Maszerowska, A., A. Matamala, and P. Orero (eds) *Audio Description: New Perspectives Illustrated*, Amsterdam & Philadelphia: John Benjamins | *One of the best introductions to the complexities of AD, this book takes Tarantino’s film Inglourious Basterds as a starting point and discusses all the decisions an audio describer would need to take in order to make it accessible to audiences with a visual impairment.*


Related topics

8 Audio description: evolving recommendations for usable, effective and enjoyable practices
15 Narratological and audiovisual translation
17 Multimodality and audiovisual translation: cohesion in accessible films
18 Sociolinguistics and linguistic variation in audiovisual translation
22 Eye tracking in audiovisual translation research
23 Audiovisual translation and audience reception

References


chocinematics_inpress.pdf [last access 20 December 2017].

223


**Filmography**