1. Introduction

1.1 Interdisciplinarity

As an “intersemiotic, intermodal or cross-modal translation or mediation” (Braun, 2008: 2), audio description (AD) is a resource valued by blind and partially sighted audiences, many of whom depend on it to assist their understanding and enjoyment of entertainment, information and education services (Greening & Rolph, 2007). Yet AD offers more than a supplementary audio landscape of transmediated visual cues: it is a lens through which we may view human meaning-making in action, the nature of shared viewing experiences and (conversely) the heterogeneous nature of human cognitive approaches. Audio description research operates at the intersection of many disciplines from linguistics (Salway, 2007), semiotics (McGonigle, 2013) and cognition (Boeriis & Holsanova, 2012; Holsanova, 2016; Holsanova, 2020), to narratology (Kruger, 2010), psychology (Holsanova, 2012; Fryer & Freeman, 2012; Wilken & Kruger, 2016) and latterly, artificial intelligence (Braun et al., 2021). By taking an interdisciplinary approach, audio description researchers have capitalised on the opportunity to explore the functioning of the human brain beyond the development of visual acuity, exploring concepts such as audience immersion and presence (Fryer & Freeman, 2012; Walczak & Fryer, 2017), emotion description and interpretation (Salway & Graham, 2003; Igareda & Maiche, 2009; Starr, 2018), focalisation (Kruger, 2012) and attention tracking (Orero & Vilarò, 2012; Di Giovanni, 2014; Mazur & Chmiel, 2016). Video content description also facilitates the exploration of sound and images as both independent and integrated sensory channels, helping us to understand how human beings engage in meaning-making using simultaneous streams of information, our ability to salvage relevance (Braun, 2016) and saliency from a multimodal assault on the senses and the manner in which we satisfy our need to order such events into a meaningful narrative, using life experience and common knowledge (Starr et al., 2020: 165).

1.2 What does AD mean in a non-blind audience context?

The adaptable nature of AD lends itself to addressing audiences with diverse accessibility needs and expectations. One of the more interesting developments in the field of audio description
over recent years has been its emergence as a testbed for extending media accessibility to audiences who would benefit from content descriptions for reasons other than sight loss (Franco et al., 2015; Starr, 2018; Bernabé & Orero, 2021; Braun et al., 2021). Consequently, while the commercial production and distribution of AD at the feature film, video streaming and broadcast television levels tends to be undertaken with sight-impaired audiences to the fore, other forms of AD are being trialled which have the potential to become more commercially sustainable as digital services and post-production techniques evolve. Therefore, as a creative, mediatorial strand of audiovisual translation, AD affords describers the freedom to adapt and develop scripts to address the needs of a range of audiences, limited only by their own inventiveness.

In deliberating the use of audio description for non-blind audiences it is perhaps useful to begin by considering the characteristics of AD which differentiate it from other audio intercessions. This may prove helpful in the present context since audio description is generally synonymous with sight impairment, and other applications might at first appear indistinguishable from information obtained through other, more commonly available, audio sources. Modern life is replete with vocal cueing and prompts designed to guide our behaviours, whether for the non-blind or sight-impaired. These commands enlist our cooperation, “passengers are reminded smoking is not allowed on platforms”, ensure our compliance with health and safety requirements, “please mind the gap” and guide our interactions with technology, “place your hands on the screen, matching your fingers to the outline”. While these cues are assistive, audible and descriptive, they are not audio description. AD can be distinguished from standard audio commands by dint of the fact that it generally takes the form of a partial text (Braun, 2008), typically enhancing access to the visual aspects of multimedia assets and artefacts, while the consumer retrieves the remainder of the narrative material through sound or other sensorial channels. In the case of film and television productions, this means paying attention to the dialogue, ambient noise and sound effects. The notion of complementarity suggests that an act of selection is necessary, as everything that occurs in the visual domain is not automatically subject to transmediation via AD from sight to sound. Furthermore, the selective nature of the process poses an important question in respect of new forms of AD for non-blind audiences: if the definition of audio description is broadened to include texts created for purposes other than helping those with sight loss, for example, delivering cultural references (Orero & Wharton, 2007), educational material (Walczak, 2016) or access to affective content (Starr, 2018), what is the impact on the method of creation, selection of material and style of delivery? Regardless of the answer to these questions, it is clear that even in a non-blind AD (N-BAD) context, complementarity is the factor which sets it apart from other streams of audible information. Furthermore, in the case of audiences for whom supplementary visual cues are a “nice to have” rather than a “need to have”, for instance where sighted viewers listen to a TV show while committing their visual resources elsewhere, their AD preferences are likely to differ somewhat from those of the traditional blind and partially sighted (BPS) audience, for whom AD is essential to their enjoyment of the audiovisual artform.

1.3 Cognitive assistance for non-blind audiences

Perhaps one of the more interesting developments in audio description research in recent years is the exploration of AD for cognitive accessibility and enrichment (Starr, 2018; Bernabé & Orero, 2021; Starr & Braun, 2021). While a single, optimised solution for interpreting and mediating narrative texts for all may not be wholly realisable, audio description is sufficiently flexible that it can be used to mediate the challenges faced by a number of cognitively
heterogeneous audiences. For example, atypical cognitive frameworks such as autism, learning difficulties, PTSD and attention deficit hyperactive disorders (ADHDs) each present non-visual oriented barriers to accessibility (Fellows, 2011; Franco et al., 2015). It is also possible that AD could be used to mitigate inattentional blindness (Kuhn & Tatler, 2011; Mack & Rock, 1998). In this way, AD is slowly emerging as a platform for delivering information, education and entertainment to individuals who would benefit from support to make sense of the world around them, whether on a temporary basis or in the longer term.

Thus, as a complementary text, audio description enhances and mediates the many layers of meaning within both static (sculpture or other fine art works) and dynamic (moving image) artefacts and texts. Its application raises questions about the way the human mind makes sense of complex narrative, how we build mental models (Johnson-Laird, 2010) from existing schemata and paradigms and in the context of non-blind AD, the nature of meaning-making from the perspective of different cognitive frameworks and lived experience. In parallel, the highly subjective nature of AD creation, where information is prioritised for inclusion or omission for a wide range of reasons (time constraints, relevance, usefulness, duplication or accessibility through other channels etc.), lends itself to great adaptability but also creates a greater burden of strategic complexity.

Little is really known about the describer’s role as a high-level cognitive processing task. Plotlines consist of a collection of cues assembled by the screenwriter and interpreted by the director in order to construct meaning, leaving the audience with the task of decoding (Bordwell, 2010) and re-assembling those cues in some narratively meaningful fashion. Where new, (non-blind) audiences require assistance in this regard, the describer’s choice of translation strategy will be challenging, not least because they face largely uncharted waters. However, the more we understand about cognitive engagement as a mediation process, the more we will learn about human meaning-making and how we may facilitate greater access to those who require support. A greater understanding of cognitive narratology will not only benefit human AD production, but will also provide a valuable contribution to the development of future video description computer models, making the wider proliferation of AD material more attainable (Braun & Starr, 2019; Starr et al., 2020; Braun et al., 2021).

In the following sections, consideration will be given to current perspectives on AD for non-blind audiences with an overview of the types of applications and adaptations presently being investigated; this is followed by a look at issues critical to the future development of N-BAD, including commercial viability for minority audiences and the forging of two distinct approaches, the bespoke and the generalist. Under the heading “current research” these optional pathways are discussed in greater detail, through the lens of two case studies. The chapter concludes with discussion of future directions, focusing on the opportunity to strike a balance between bespoke and generalist approaches to N-BAD using the latest advances in machine learning and platform technologies.

Film and television source text formats have been selected as the primary focus, since this is the forum where most N-BAD research and practice has taken place to date.

2. Current perspectives

In recent years, tentative steps have been taken towards wider inclusivity with the adoption of universal design (UD) principles and the concept of accessibility as a human right (Greco, 2016) recognised in law (Council Directive 2010/13/EC, 2010; United Nations, 2006, 2014a, 2014b; European Commission, 2012; European Commission, 2019). Artificial boundaries
previously drawn between levels of ability in society are being challenged and a more holistic approach taken to creating physical spaces, as well as digital and artistic assets (Taylor & Perego, 2021). According to the founders of the UD movement, their purpose is “to reduce the physical and attitudinal barriers between people with and without disabilities” (Story et al., 1998: 11; Ellis, 2016). Conceptually speaking, universally designed assets from which technical workflows filter and extract information salient to each audience, moving from the “general” (i.e., broadly drawn video descriptions) to the “particular” (i.e., descriptions which resonate with each audience in line with their specific needs) would appear to be the ideal solution. Advances in “red button” personalisation options, audio object creation (Shirley & Oldfield, 2015) and the rapid development of artificial intelligence in processing audiovisual material (Braun et al., 2021) makes this a more realistic goal than it might at first appear. Training computer models to simplify standard audio description texts for language learners, or to extract information relating only to the description of emotions, for example, is already within reach.

Experimental studies trialling AD with non-blind audiences (Perego, 2016) and those with alternative (non-visual) accessibility requirements (Franco et al., 2015; Starr, 2018; Jankowska, 2019) have been a recent development in the timeline of audiovisual translation, and remain relatively rare. For commercial expediency, the production of AD for film and television productions remains restricted to the BPS domain. High AD production costs (Szarkowska, 2011: 143), given the labour intensive nature of scripting and voicing a supplementary soundtrack in post-production, currently present an insurmountable barrier to repurposing AD for multiple audiences as a bespoke service, on a project by project basis. While text-to-speech may offer a partial solution, this process still requires human scripting and a machine or human to identify the gaps in dialogue (Szarkowska, 2011). Nevertheless, this position is fluid and likely to change with the advent of machine learning, artificial intelligence and the potential to introduce (semi) automation of video descriptions into the audiovisual translation landscape (Braun et al., 2021).

Anecdotally, it has long been acknowledged that sighted viewers turn to standard BPS-AD to retrieve visual information where access to television screens, tablets or other audiovisual devices is limited (Jankowska, 2015; Calvello, 2021). Indeed, individuals using audio description to access media streaming services such as Netflix, Amazon Prime Movie and Disney+ while driving or performing domestic tasks has recently given rise to the term “to audiobook” (Loftus, 2020; Rana, 2020). The comparison between audiobooking and audio description is an interesting one, since it offers the advantages of BPS-AD from a different user perspective: in this case, sighted audiences opting out of the visual experience. Nonetheless, audiobooking using standard audiovisual material is somewhat of a misnomer, since it is fundamentally a different activity from listening to an audio book which has been designed as a piece of audio performance from inception. Audio books provide a recitation, generally performed in a single narrative voice, of a written text; they lack the additional constraint of fitting visually descriptive vocalisations into the audio hiatuses created by breaks in film dialogue (Fryer, 2016: 33). In fact, audio books provide a complete text, albeit often abridged, whereas audiobooking with AD makes use of an alternative, partial translation (Braun, 2008). In the case of the latter, the consumer chooses to disengage from the visual aspects of storytelling, focusing instead on audio sources (dialogue, sound effects, scoring, narrator, AD) for narrative immersion. In terms of cognitive processing, monomodal audiobooking is likely to place a greater burden on the media consumer than an audio book, since a minimum of two audio narrative channels must be assimilated simultaneously in order to derive meaning, rather than one. This has yet to be scientifically tested, but could produce valuable insights into the cognitive challenges
of integrating multiple audio channels (audiobooking with standard BPS-AD), as opposed to entertaining just one (audio book).

Further mainstreaming of audio description with non-blind audiences has been observed in the case of second language learners (Walczak, 2016). Subtitles have long been regarded as a useful tool to enhance language learning in terms of lexicon acquisition (Vermeulen & Moreno, 2013) and to reinforce standard syntactical and grammatical structures (Talavan & Lerota, 2016: 68). When applied in the native language (L1), rather than as interlingual subtitles (L2), they have proved particularly successful for early learners and those with limited exposure to L2 (Pujadas & Muñoz, 2020). In language learning more generally, AD may be regarded as a companion service, providing language students with an opportunity to engage further with the spoken word (in L2), pairing on-screen visual representations with corresponding verbal (audio) descriptions to improve lexical competence, pronunciation and oral comprehension skills, as well as enhancing naturalistic phraseology (Vermeulen & Moreno, 2013).

Non-linguistic pedagogy has also emerged as a non-blind application for AD in the contexts of both audio learning styles and the growth in popularity of the “flipped classroom” teaching method (Bergmann & Sams, 2012). According to Calvello (2020): “[f]or those who gravitate towards auditory learning, audio descriptions can be a helpful aid when retaining new information in an educational setting”. As an experiment in teaching physical science, Cozendey and Da Piedade (2016) applied audio description to freeze-frame slide presentations during a laboratory class. Three groups of students participated in the experiment, one of which contained only sighted participants. Cozendey and Da Piedade’s findings showed that audio description played out over paused visual presentations (“extended AD”) assisted the focus of sighted participants by reinforcing visual data and improved focus and attentiveness. This is a significant finding for the 30% of the population who consider themselves to be auditory learners (Calvello, 2020). Audio cueing for narrative comprehension has also been explored in the case of children with autism spectrum disorders (Starr, 2018; Starr & Braun, 2021) with the implication that it could also be used for reinforcement learning. In this case, N-BAD strategies proved more successful than those generated specifically with sight-impaired audiences in mind (see Current Research for further details).

Finally, brief mention should be made of the newly emerging trend amongst sighted viewers of electing to watch films with audio description for the sheer aesthetic pleasure to be derived from the audio descriptions themselves (Lewis, 2021). Citing movies she considers worth watching simply to appreciate the artistic merits of the AD, Lewis (2021) states, “hearing Olaf described as ‘a carrot nosed snowman shuffling up to a purple flower’ is enough of a reason for me to watch Frozen with audio description”. This latest development, when taken in parallel with audiobooking, suggests that BPS-AD may slowly be “mainstreaming” as it evolves as an independent literary artform enjoyed by audiences universally.

3. Critical issues (for non-blind AD)

Despite these encouraging signs, many societal and practical barriers impact the potential for AD to be harnessed as an accessibility service for the non-blind community, starting with awareness. The notion that standard BPS-AD might be beneficial to media consumer groups other than those with sight-impairment is not a new idea, but remains a minority viewpoint. Furthermore, at the most fundamental level, the non-blind AD use-case raises questions about the nature of accessibility and universal access more generally. With definitions of accessible media focused on those with supplementary physical and more recently, cognitive accessibility
needs, there is perhaps a case for re-visiting the nature of media access from a position of universal design (Story et al., 1998). Accessibility needs vary greatly in line with an individual’s unique life circumstances and extend far beyond the physiological or cognitive; for example, a second-language learner may be less able to access audio information in their non-native language than a partially-deaf native language speaker. In this context, we are all more or less able, dependent on our innate talents, life circumstances and learning experiences. Given this scenario, we can assume with some confidence that the individual who can communicate and comprehend information without need of additional assistance in any situation, environment or geographical location, simply does not exist. In this context, the challenge becomes to provide universal access to media streams, where the accessibility remit is framed in terms of all users facing some additional needs (physical, cognitive, language literacy, access to resources etc.). Since these issues will range widely between individuals, the solutions need to be sufficiently agile to meet the needs of the greatest number with the least additional resource. Within this ontological framework, the re-purposing of audio description for new applications and audiences raises two fundamental issues: the feasibility of creating AD for smaller-scale special interest groups and audiences and the dichotomy of opinions between adopting either purpose-driven (bespoke) or generalist (universal) approaches to AD creation.

3.1 The viability of AD targeted at smaller interest groups

According to the most recent government statistics, almost two million people experience sight loss in the UK, of whom 360,000 are registered as blind or sight-impaired (NHS, 2018). Yet, it has been estimated that while 60% of blind individuals are aware of audio description services (Ofcom, 2009: 33), uptake is considerably lower with a fraction over one-fifth (21%) of those with visual impairment saying they have used AD at least once (Ofcom, 2009). Audio description services are invaluable to those BPS audiences who depend upon them, but commercially, they present a high cost to consumer ratio. With the likelihood that an ageing world population increases the percentage of individuals experiencing sight impairments, the BPS-AD political lobby should gain momentum and lobbying “muscle” over time. Nevertheless, encouraging the continued adoption of AD by alternative interest groups (language learners, educators, cognitively challenged audiences, audiobookers) would reinforce the use-case argument. In order for this to happen, AD must be regarded as a resource of wider relevance and societal value and marketed as such. This would surely represent “good news” for the BPS community for whom greater financial resourcing would be a welcome outcome.

However, the problem of defining a series of diverse and fragmented audiences for AD is that many of these groupings will have specific requirements which are particular to their needs and which fall outside the standard BPS audio description brief. In a world where universal design is regarded as the gold standard, this leaves N-BAD practitioners facing a fork in the road: whether to champion a form of AD which is generalist (universally designed) in nature and therefore not optimised for any one user group; or alternatively, to favour a purpose-driven (bespoke) approach, which meets the particular needs of each minority user group but which is currently costly and resource-intensive to produce.

3.2 Generalist and bespoke approaches to AD for non-blind audiences

In terms of research, two approaches to expanding the reach of AD have been running in parallel: (i) exploring the use of traditional BPS-AD with non-blind audiences to determine whether
a coincidental cross-over applicability effect is evident (Franco et al., 2015; Walczak, 2016; Jankowska, 2019) and, (ii) creating bespoke (non-BPS) forms of AD designed for different purposes and audiences, with innovative translation strategies and methodologies adopted to meet the changing brief (Starr, 2018; Oncins & Orero, 2020; Bernabé & Orero, 2021). Both strands of research have their merits and present different challenges. Brief consideration is given subsequently to each category. One example of each approach will subsequently be examined in greater depth, under “Current Research”.

3.3 BPS-AD approaches for non-blind audiences

The benefits of expanding the reach of BPS-AD are clear: AD strategies which have been successfully trialled with sight-impaired viewers and subsequently applied in commercial production, are already available to all. Incidental gains achieved by assisting viewers with alternative, non BPS-AD needs, require no additional funding or human resources. However, certain limitations are inherent to BPS-AD (e.g., the visual remit) and a lack of optimisation for the non-BPS end-user will inevitably result in a lack of descriptive focus. This was evident in research conducted with audiences experiencing difficulties accessing the emotions portrayed in film. While BPS-AD may reference the affective state of certain protagonists, this is not delivered with any degree of consistency; conversely, sighted audiences found descriptions of protagonists’ appearance, actions and location both superfluous and distracting (Starr, 2018).

Nevertheless, the availability of public service (e.g., BBC iPlayer) and pay-per-view (e.g., Netflix, Amazon Prime Movies, Apple TV) audio description, particularly for film and television productions, makes BPS-AD an attractive first stop resource for audiences new to AD. The non-specialist approach to supplementing lost narrative cues for reasons other than sight impairment with visually assistive AD will always require compromise; the fact that it may also result in a failure to communicate effectively, or introduce additional confounds, strengthens the case for a more bespoke approach.

In the examples that follow, the efficacy of BPS-AD has been tested with non-blind audiences.

Case study (I) Children with learning difficulties

In a small-group Brazilian study trialling audio description with children experiencing learning difficulties, including Down’s Syndrome, the visual orientation of BPS-AD proved rather problematic (Franco et al., 2015). The application of standard audio description to film material containing elaborate metaphorical referencing delivered only modest improvements in accessibility to and comprehension of the film material. In particular, it was observed that the cognitive nature of conditions experienced by the children, along with the specific manifestations of those conditions in the case of each child, meant that BPS-AD was unlikely to be either universally or indeed, individually, helpful. For example, the use of complex film metaphor, where parallel supernatural worlds and symbolic acts were used to signify key dramatic interludes, appeared confusing to some participants even with BPS-AD prompts. This pilot study, although small in scale, provides a useful insight into the limitations of BPS-AD for those with cognitive needs, suggesting that standard AD is far from a “one size fits all” solution for improving cognitive accessibility and reinforcing the notion that a universal approach to specific access issues may not present optimal results.
Case study (II) Second language learners

Outside of the classroom, AD is commonly used for language learning in a social setting. Language learners wishing to supplement classroom learning, or lacking the time or resources to attend formal tutoring, may elect to use film and television-based AD to enhance their lexical versatility and general comprehension (Vermeulen & Moreno, 2013). The benefits are manifest: complex conversations and sentence structures occurring in the film dialogue may be contextualised by the audio description, since, as one proponent maintains, “[b]y providing a great source of comprehensible input – language made understandable through context – audio description also makes the target language more accessible to language learners” (Beyond Language Learning, 2018). Aside from contextualisation, AD provides language learners background information that may not be visually sourced (e.g., “the following day, back at John’s house”) and which may have been lost in the quest to comprehend fast-moving dialogue when lexical gaps are present in the learner’s skillset. AD complements episodes of dialogue free soundtrack, giving the language student an opportunity to develop their listening skills further, and to gather information that may have been omitted previously by the describer, during moments of rapid dialogue.

The proliferation of research testing the efficacy of BPS-AD for language learning in the classroom reflects AD’s origin as a strand of intersemiotic translation (Jakobson, 1959 & 2000). Reporting on a primary-school level language learning experiment in a hybrid classroom in which both sighted and sight impaired students were present, Walczak observed positive outcomes for both groups, leading to the conclusion that there is, “immense potential for AD in educational settings” (Walczak, 2016: 201). Schaeffer-Lacroix reported on applying corpus exploration tasks in an intermediate level German language class, using the audio description of two television productions on corpus tool Sketch Engine. Students were invited to create new audio descriptions for the film material using the available corpora, as a way to, “foster language awareness of features and linguistic items needed for describing spatial events and visual perceptions” (Schaeffer-Lacroix, 2021: 173).

Taking a rather different approach, the ClipFlair Project created as part of the European Union’s Lifelong learning Programme (ClipFlair, n.d), delivered a system for enhanced foreign language learning through the application of revoicing and captioning activities relating to audiovisual extracts. In tests conducted with online foreign language learners, it was tentatively concluded that audio describing audiovisual material improved students’ oral fluency between pre- and post-study scores. Furthermore, the improvement in their oral production (speaking) scores was greater over the same period than in their scores for oral and reading comprehension and written production (Talavan & Lerota, 2016).

Similar systems for employing video description as a supplement to standard language learning methodologies can be found in the “Movie Talk” method (The Comprehensible Classroom, u.d.), where teacher-led storytelling activities revolve around communal video-viewing and a “pause and describe” technique which encourages immersion in naturalistic phrasing and extending existing lexicon (TPRS Storytelling, 2017).

Case study (III) Pedagogy

As touched upon earlier, the pedagogical benefits of AD have also been witnessed outside the language-learning remit. Effective teaching and learning requires the exploitation of the senses to improve cognitive processing of information on the part of the student. The value of applying audio channels to a multimodal learning experience in the classroom has been long
acknowledged: “[f]or sighted children, description offers a promise of a new way to promote literacy and learning . . .”, with the conclusion that, “. . . innovations and accommodations made for people with disabilities benefit many people without disabilities” (Described and Captioned Media Programme, 2021). Furthermore, academic research has shown that audio description can be used in scientific pedagogical environments where, when paired with think-aloud protocols, it offers cognitive insights revealing the way students apply different types of knowledge and competences to read and comprehend complex scientific journals (Holsanova, 2020). In science education, audio described media have been trialled in the training of nursing staff in Thailand (Srisuk & Witchit, 2020) and with physical science students in laboratory settings (Cozendey & Da Piedade, 2016), both with positive results.

Case study (IV) Literacy and creative writing

Creative writing skills are another area of teaching where students benefit from interaction with the audio description creation process (Kleege & Wallin, 2015). In a preliminary study to apply audio description skills to the development of inclusive and engaging prose in high school and university students, Vialard (2017) sought to turn AD “on its head”, regarding the intersemiotic shift between image and text as a “rich resource for educating students to be responsible and ethical readers, writers and designers of written communication” (Vialard, in Braun & Starr, 2021: 9). Testing this hypothesis with students at Texas Tech University, Vialard explored AD for the potential it offered practitioners to develop critical thinking, visual literacy, language analysis and close reading skills, believing that its rhetorical roots represented common ground with the discipline of creative writing and from which both fields of study could mutually benefit. Early results suggested that students trained in AD would “demonstrate a greater understanding of audience awareness” (Vialard, 2017: 13).

3.4 Alternative approaches to non-blind AD

As noted previously, bespoke or user-centric translation (Suojanen et al., 2015) as an approach applied to creating AD for new audiences affords the benefit of ultimate adaptability, both in terms of content selected for prioritisation in the AD text and the style of presentation. However, the principal disadvantage is that practitioners start from a “blank page”, both literally and metaphorically: considerable resource is required to create AD ab initio for each audience group and scant research exists to suggest what works for any given non-blind audience group. Furthermore, enhancing cognitive accessibility presents a significant challenge where multi-modal, multi-layered storytelling is concerned. Individuals experiencing atypical cognitive behaviours, whether the consequence of developmental difficulties, trauma or neurological pathologies, may require assistance both with accessing the information being relayed and with interpreting the causality and consequence of those cues and prompts (Starr, 2018). The describer is therefore presented with the difficult task of balancing these needs within the available dialogue hiatuses and in a manner which avoids overwhelming or overstimulating the viewer (Starr, 2018). Even so, the flexible nature of bespoke audio descriptions for non-blind audiences presents unlimited potential to offer up solutions for a wide range of cognitive accessibility challenges.

In the following section, the merits of two recent experiments in creating AD for non-blind audiences are discussed at length. The first case-study (Starr, 2018; Starr & Braun, 2021) is the only experiment known to the author at the current time, which re-positions AD as a bespoke service for a very particular non-blind audience (autistic individuals); the second re-defines
AD in more generalist terms, aiming to achieve a broader audience reach including both those with and without sight impairments (Bernabé & Orero, 2021).

4. Current research methods

4.1 “Bespoke” AD for non-blind audiences

One of the more celebrated aspects of AD is its infinite adaptability, both in the style of delivery (interpretive versus descriptive, florid versus prosaic stylings etc.) and in the information selection and prioritisation opportunities arising as content is parsed according to different translation strategies during the transmediation process. The first study involving the re-imagining of AD as a bespoke service for audiences with supplementary cognitive needs (Starr, 2018) was initiated after anecdotal material emerged suggesting some individuals experiencing difficulties reading the emotions of others (typically, through facial expression and body language) found that audio description provided assistance (Fellows, 2011). Emotion recognition deficiencies (ERD) occur in several personality disorder conditions, as well as situations such as post-traumatic stress and neural pathologies. However, in this specific case, the individuals taking part in the study were young people on the autistic spectrum for whom establishing the centre of narrative action and ascertaining where to concentrate their narrative focus can co-exist as accessibility problems alongside ERDs: “[i]ndividuals with autism often find that listening to audio descriptions paired with visuals helps them understand the emotional and social cues that are demonstrated through actions and facial expressions” (Calvello, 2020).

The bespoke AD for emotion recognition study (Starr, 2018; Starr & Braun, 2021) took the form of two experiments. The first was an exercise in applying a task-oriented translation framework, “skopostheorie” (Nord, 1997; Vermeer, 2012; Reiss & Vermeer, 2014), to the modelling of two new AD variants: one variant used to label the emotions occurring in short narrative film extracts (EMO-AD) and a second variant used to both label the emotions and provide an explanation as to their cause and/or consequence (CXT-AD). Each of these AD variants was compared with and controlled by standard audio description of the variety typically used to assist viewers with visual impairments (BPS-AD). The second experiment tested the efficacy of and engagement with all three variants (EMO-AD; CXT-AD; BPS-AD) with a sample group of young people on the autism spectrum. All participants registered average or above IQs and fell into the age range nine to 14 years.

Since the recognition of basic emotions (Ekman, 1999) in self and others tends to improve with age in autistic individuals providing there are no comorbid learning difficulties (Mencap, undated), the range of emotions and “states of mind” selected were relatively complex (e.g., furious, criticised, uncomfortable, unfriendly, annoyed, guilty) although some which might be considered more readily accessible (e.g., excited, scared, upset) were also included to allow for all ability levels. Participants viewed a series of short film extracts each containing one of these emotions and were informed where to place their attention by the use of a red frame drawn around the relevant protagonist. Each participant received all three variants of AD (EMO, CXT and BPS) in rotation across the extracts.

The results of this study illustrated the versatility and adaptability of the AD translational form, showing that purpose-driven strategies for creating new variants of AD could be achieved by establishing “text type” (Nord, 1997) affiliations within the target text in each case. Thus, the BPS variant of AD was identified as having an “informative expressive” (1997: 37–38) remit, being principally concerned with referencing visually portrayed “objects and phenomena”; the EMO variant was classed as “operative” (1997: 38) such that “both content
and form are subordinate to the extra-linguistic effect that the text is designed to achieve” (1997: 38) and the CXT variant assumed an “operative expressive” brief, since it was required to be “operative” to the extent that it would elicit engagement with the source text emotions, while simultaneously invoking “expressive” elements to convey context whereby “the stylistic choices made by the author contribute to the meaning of the text” (1997: 38).

An example of the way this approach played out in practice is illustrated in Table 30.1. In this case, the descriptions correspond to the emotion “furious”. The scene, from the film *Paper Moon*, takes place in a café where a little girl, Addie and her conman father, Moses, are having a conversation about the way he stole her money.

In this example, the BPS-AD variant is centred on creating a visual representation of the scene including the precise situation (café, window, table), attitude (sitting, slumped) and activity cues (waitress, cake, coffee). The fact that in the BPS-AD the waitress is “smiling” implies benevolence, but not necessarily emotional undercurrent; whereas, in the EMO-AD variant the describer has chosen to interpret this smile as a sign of friendliness (a “state of mind”). As the young girl becomes “very angry” with Moses, the CXT-AD variant is used to express the reason for this emotion: “because he’s taken her money”.

While the experiment was conducted with a small sample of participants (*n*=12) and cannot therefore be generalised, there was a noted preference for both bespoke AD variants over the BPS variant, with the EMO version proving marginally more popular than the contextualised variant (CXT) across the group. This is likely to have been a function of the reduced cognitive load with the EMO variant. In addition, several children expressed their dislike for the BPS variant which they considered “annoying” and superfluous to their needs (“I don’t need to be told he’s walking down the street”); nevertheless, the integration of an audio described secondary soundtrack into the viewing experience, which it was anticipated might induce hyper-stimulation in children on the autism spectrum, was accepted without issue by all participants.

This early experiment in creating bespoke AD for non-blind audiences suggests there is demonstrable potential for adapting and re-purposing AD to assist audiences facing a range of barriers to accessibility. Further consideration is given to the production of non-blind AD later (“Future Directions”), including the practicalities of extracting relevant information streams from more generalist AD approaches and the use of AI and digital platforms to deliver non-standard AD solutions.

### 4.2 Generalist AD for blind and non-blind audiences

Although a comprehensively designed AD script could provide the material for machines (or indeed, human beings) to filter and re-version the information in a way that addresses
different audiences’ needs (for instance, extracting descriptions from a broadly-drafted AD where they contain emotion labelling, for use with ERD audiences), a more elegant solution would be to produce a universal script that delivers AD to non-blind and sight impaired audiences simultaneously.

The Easy AD project (Oncins & Orero, 2020; Bernabé & Orero, 2021) exemplifies this approach, adopting a text simplification strategy designed to create an audio resource which is both easier to listen to and more cognitively accessible than standard AD. In this case, the ambition is to improve comprehensibility and “contribute to mainstreaming AD” (Bernabé & Orero 2021: 56). Following the recommendations of Sade et al. (2012), Bernabé and Orero state that, “audio descriptions that are useful for audiences other than blind and low-sighted persons opens the opportunity to optimise them both in financial terms and for those who benefit from easier-to understand texts” (2021: 56). Certainly, the market for a more “generalist” AD is broad, ranging from the elderly experiencing some sight and/or hearing loss (since AD can provide access to low-level sound effects and reinforce context where some dialogue has been missed), to people with learning difficulties, those with low levels of literacy, second language learners, individuals impacted by neural pathologies like dementia and those affected by barriers to learning such as dyslexia and ADHD. As a call to action, the European Disability Strategy (European Commission, 2017), Inclusion Europe (2009) and instruments legislating for accessibility (European Union 2010) require that such groups are addressed as fully engaged citizens in modern society (Bernabé & Orero, 2021). Opening pathways to better communication is an essential and fundamental step on the road to equality and inclusive citizenship; text simplification is the natural route to finding a system which, while not suitable for all, will optimise accessibility for many.

For the Easy-to-Read study, the aim was to produce AD scripts which employed less structurally complex, sophisticated and elaborate language than might generally be present in the standard AD approach for sight-impaired audiences (Bernabé & Orero, 2021: 67). As an act of intersemiotic translation (Jakobson, 1959, 2000), text simplification strategies emulating the principles of plain language and easy-to-read included selecting, “simple, short and common words that do not contain difficult syllables . . . as well as avoiding abstract and foreign words” (2021: 67). Complex and cognitively demanding constructs like metaphor, simile, irony and idioms were excluded. Following the recommendations of Inclusion Europe (2009) and AENOR (2018), syntactic adjustments were applied to the language simplification strategy, with short sentences created by splitting and reordering more complex constructs. Pronoun usage was kept to a minimum to avoid confusion, cases and moods were simplified by avoiding passive and progressive structures and negatively ordered sentences prohibited (Bernabé & Orero, 2021: 68).

Full results from the Easy-to-Read reception studies are anticipated shortly. If the proposed simplification strategies prove effective in producing texts that are cognitively accessible to a wide range of individuals with competing accessibility needs and there is minimal impact from the inevitable loss of “colour” that this will represent for some audiences accustomed to more elaborate prose, then significant progress will have been made towards orienting AD as a universally designed (Story et al., 1998) accessibility service for all. In doing so, it will offer the benefits of audio description to large numbers of non-blind media consumers for whom AD is currently unhelpful.

These two case studies illustrate polarised approaches to designing AD for non-blind audiences: the bespoke (affective AD for ERD audiences) and the generalist (anodyne AD for many, if not all, audiences). While both offer different benefits to the consumer, it is fair to say that neither approach will be suitable in all situations or for all audiences. Essentially,
the former is likely to be sufficiently targeted towards its primary audience that it will prove irrelevant to everyone outside that group; the second, as a “catch all” AD, could prove too generalised and insipid for many. However, the harnessing of modern technologies, may offer a third route to extending AD “to all”, where a more comprehensive generic text is the point of departure for extracting the particular needs of individual groups. This concept is explored in the next section, as the focus shifts to the future of AD for the non-blind and the opportunities presented by advances in technology delivery platforms and computer intelligence.

5. Future directions

The proliferation of streaming platforms and digitised channels (Picard et al., 2016), as well as the production of video description for retrieval and accessibility by means of computer vision models (Braun et al., 2021), have introduced the possibility of offering audiences multiple viewing options concurrently, to match a range of accessibility needs. As discussed previously, in the case of non-blind viewers these options may be as simple as further adapting standard BPS audio description to encourage non-blind users to “audiobook” a mini-series while driving (e.g., adding audio introductions and audio described trailers). Mobile device applications also offer the flexibility to deliver new, dedicated audio description streams for non-blind audiences with specific requirements, the emotion description case-study being one such example.

However, these platforms can only be exploited if there is a means to script each AD variant in the first place. Broadcast and film BPS-AD is becoming more accessible as countries legislate for equality of access for all citizens, with some nations imposing minimum tariffs on AD production (UK, Spain) which force commercial producers to invest in script writing. Non-blind audiences such as language learners are free to benefit from these services where they are available. Where other forms of bespoke AD are more appropriate, for instance as a way to introduce emotion cues and contextualised affective descriptions for those experiencing emotion recognition deficiencies, the issue becomes one of cost and audience uptake. The same fate applies to universal AD derived from easy-to-read principles, until such time as resources are committed to making this kind of data stream widely available.

Nevertheless, it is possible that from the general, we may derive the particular. With the rapid growth of artificial intelligence within the audiovisual entertainment arena and the emergence of research focused on generating video descriptions of moving images using computer vision models and speech analysis techniques (Braun & Starr, 2019; Starr et al., 2020; Braun et al., 2021), the possibility that broadly drawn descriptions can be filtered and sub-divided into a number of discrete streams of information targeted at a range of diverse audiences, becomes a reality. If this were the case, comprehensively written computer- (or indeed, human-) generated video content descriptions could be parsed to extract phases describing emotions or “states of mind” in much the same way as machine translation is currently parsed using NLP; a similar process could be applied to post-editing the AD source text to simplify phraseology, lexicon and syntax, producing Easy-to-Read type descriptions. In these two examples alone, the needs of several audience groups would be met from one universally designed AD “master” script.

However, it is important to note that while we are witnessing a revolution in the development of media resources in an age of artificial intelligence, it is unlikely that it will be possible to produce fully automated video content descriptions which meet the divergent accessibility requirements of a range of non-blind audiences on a like-for-like basis with human-generated BPS-AD in the near future. Visual accuracy remains a significant issue affecting automated video description creation, and insufficient training data is available at present to remedy this problem in the short term. The fact that automating AD requires the detection of audio hiatuses
and a matching of these to the text-to-speech processing of AI video descriptions, suggests an additional layer of production complexity. For the time being, post-editing platforms and tools of the kind being developed as a by-product of research programmes like the MeMAD project (Limecraft, u.d.) offer the potential to bring together the speed and versatility of the machine, with the flair and sophistication of the human describer, enabling new streams of AD to be crafted to meet the needs of BPS and non-blind audiences alike.

6. Further reading


7. References


