



# Enough is enough: how much intonation is needed in the vocal delivery of audio description?

Anna Jankowska, Joanna Pilarczyk, Kinga Wołoszyn & Michał Kuniecki

**To cite this article:** Anna Jankowska, Joanna Pilarczyk, Kinga Wołoszyn & Michał Kuniecki (2023) Enough is enough: how much intonation is needed in the vocal delivery of audio description?, Perspectives, 31:4, 705-723, DOI: [10.1080/0907676X.2022.2026423](https://doi.org/10.1080/0907676X.2022.2026423)

**To link to this article:** <https://doi.org/10.1080/0907676X.2022.2026423>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 08 Mar 2022.



Submit your article to this journal [↗](#)



Article views: 1786



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



# Enough is enough: how much intonation is needed in the vocal delivery of audio description?

Anna Jankowska <sup>a\*</sup>, Joanna Pilarczyk <sup>b</sup>, Kinga Wołoszyn <sup>b</sup> and Michał Kuniecki <sup>b</sup>

<sup>a</sup>Chair for Translation Studies, Jagiellonian University, Kraków, Poland; <sup>b</sup>Institute of Psychology, Psychophysiology Laboratory, Jagiellonian University, Kraków, Poland

## ABSTRACT

In this article we report on an experiment examining audience experience when two different audio description (AD) intonation types are used. Following the classification put forward by Cabeza-Cáceres (2013), we tested twenty participants with vision loss who were asked to listen to twenty film clips with AD voiced following the adapted and emphatic intonation. All clips were rated as evoking emotions, either positive or negative. Participants' experience was measured through heart rate variability, a self-assessment emotional response (Self-Assessment Manikin), a self-assessment presence questionnaire (ITC-SOPI) and an interview. Results show that participants strongly preferred adapted intonation, particularly when presented clips were evoking negative emotions. Higher rating of intonation type was also linked to greater intensity of the emotional experience and immersion, both self-reported and measured as psychophysiological reaction. The qualitative analysis of the participants' reports on their experience indicated that adapted intonation is considered a golden mean between emotionless and melodramatic intonation. Film genre, voice type and emotional valence can be influencing factors. Emphatic intonation could be beneficial for certain genres, but a poorly-matched voice can distract the audiences.

## ARTICLE HISTORY

Received 10 May 2020  
Accepted 2 January 2022

## KEYWORDS

Audio description; emotions; heart rate variability; presence; voicing; intonation

## 1. Introduction

Audio description (AD) for films is an additional audio track woven into the soundtrack (Szarkowska, 2011). In order to prepare this track, a script is first written and then voiced by human or synthetic voice.

Audio description voicing (ADV) refers to how AD is delivered. This broad and still undefined term is explored by researchers and practitioners to refer to aspects that can be assigned to the following areas: speed (Cabeza-Cáceres, 2013; Jankowska et al., 2017) voice type (human vs. synthetic; male vs. female) (Szarkowska, 2011), intonation (Cabeza-Cáceres, 2013; Machuca et al., 2020) and volume (Machuca et al., 2020; Udo et al., 2010).

**CONTACT** Anna Jankowska  [anna.m.jankowska@uj.edu.pl](mailto:anna.m.jankowska@uj.edu.pl)

\*Present address: University of Antwerp, Antwerp, Belgium.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Common consensus is that ADV is an important aspect of AD reception (e.g., Fryer, 2016; Machuca et al., 2020) that has so far received insufficient attention from the academia and practitioners. While advice on ADV is provided in most AD guidelines (e.g., Bittner, 2010; Independent Television Commission, 2000; Remael et al., 2015; Snyder, 2010; Szymańska & Strzyński, 2010), it is limited to general instructions that the voicing should be neutral and clear, and at the same time not monotonous, adequate to the tone of the original, and delivered at a rate that can be understood. Additionally, some guidelines contain specific reading speed recommendations (e.g., Netflix; Snyder, 2010). So far there have only been a few experimental studies that explore its influence on audience reception (Cabeza-Cáceres, 2013; Fernández i Torné & Matamala, 2015; Walczak & Fryer, 2018; Walczak & Szarkowska, 2012). ADV research has so far concentrated mainly on testing the reception of synthetic voices vs. human voices. Studies were conducted in Poland (Drożdż-Kubik, 2011; Mączyńska, 2011; Szarkowska & Jankowska, 2012; Walczak & Fryer, 2017; Walczak & Szarkowska, 2011), Japan and the USA (Kobayashi et al., 2010), as well as Spain (Fernández i Torné & Matamala, 2015).

Audience reception, which nowadays is an important line of AD research, has facilitated a thematic and methodological shift in the field (Jankowska, 2019). Initially, when investigating audience reception, researchers concentrated on preferences and comprehension, measured with subjective methods such as self-assessment questionnaires (Jankowska, 2019). Now we are witnessing a growing interest in the area of presence and emotional response (e.g., Fryer & Freeman, 2014; Ramos Caro, 2015; Walczak & Fryer, 2017). Also, researchers start to employ psychophysiological measures of emotional and cognitive processes associated with AD reception (e.g., Fryer, 2013; Matamala et al., 2020). In this article we follow these new methodological developments in AD audience reception studies to look into the effect that ADV, and in particular intonation, might have on the AD audience.

## 2. Audio description intonation

Intonation, which lies at the core of this article, arouses particular interest in the context of voice neutrality. Fryer (2016, p. 88) notices that ‘a neutral delivery has come to be recognized as “the norm”’ since describers have been traditionally ‘encouraged to use a particularly neutral way of speaking’. This has been reflected in the available guidelines; however, as Machuca et al. (2020, p. 59) notice, guidelines lack clarity in this aspect as they ‘seem to promote a neutral AD while advising to take into account the nature of the material’ which makes them open to interpretation. This interpretation was made visible by Cabeza-Cáceres (2013) who analyzed AD guidelines and recordings from Spain, Germany, the UK and the USA, and put forward three intonation types: *uniform*, *adapted* or *emphatic*. He describes the first one as neutral, the second as adapted to the content of the original, and the third as emphasizing the emotional context of the original content (Cabeza-Cáceres, 2013).

Four studies have been of particular importance in this context. Kobayashi et al. (2010) measured audience preference for human, synthetic neutral, and synthetic emotional (happy and sad) voices applied to two videos: a comic cartoon and tragic drama. The authors point out a significant interaction between video type and voice type. In the case of the cartoon, human and synthetic neutral voices were rated

significantly better than emotional synthetic voices, both happy and sad. In the case of the drama, the human voice was rated significantly better than synthetic voices and the synthetic happy voice was rated significantly worse than other voices. The authors conclude that inappropriate use of emotional voices may have a negative effect on video experience.

A similar conclusion was reached by Cabeza-Cáceres (2013), who tested the preference for three conditions of human intonation: uniform, adapted, and emphatic. Results show that while intonation does not affect comprehension, uniform intonation and emphatic intonation are associated with lower enjoyment as both intonation types provoke rejection to a similar extent. This is especially interesting given the fact that Spanish audiences are used to uniform intonation (Cabeza-Cáceres, 2013).

Fryer and Freeman (2014) found out that, contrary to the synthetic voice, human voice can actively enhance presence and emotional response for some emotions. They concluded that prosody is an essential part of audio description as emotions can be ‘effectively conveyed via the paralinguistic content of the describer’s voice rather than the semantic content of the AD script’ (pp. 105) This finding was corroborated by Walczak and Fryer in terms of presence (2018), which was rated higher for drama with human narrated AD.

In summary, these studies show that intonation is an influencing factor in audience experience, whether it is measured as preference, presence, or emotional reaction.

### 3. Measuring audience experience: presence, emotional response, and psychophysiological reaction

Audience studies, which are now an important research avenue in AD, are experiencing a visible shift from audience reception to audience experience studies that measure aspects such as presence (Fryer & Freeman, 2013; Walczak & Fryer, 2017, 2018), emotional (e.g., Fryer & Freeman, 2014), and psychophysiological response (e.g., Ramos Caro, 2016; Rojo et al., 2014).

Presence is a theoretical concept defined as ‘perceptual illusion of non-mediation’ (Lombard & Ditton, 1997) or in other words ‘a subjective experience of being in one environment, while physically situated in other’ (Walczak & Fryer, 2017, p. 8). It is one of the quality measures used in advanced broadcast and virtual environments (Lessiter et al., 2001). Both subjective (e.g., presence questionnaires, continuous assessment, qualitative measures, psychophysical measures, and subjective corroborative measures) and objective measures (e.g., psychophysiological measures, neural correlates, behavioral measures, and task performance measures) can be used to evaluate presence (van Baren & Ijsselstein, 2004). It is, however, argued that since presence is a subjective sensation, it should be primarily assessed using subjective measures (Sheridan, 1992). Currently, the most frequently used measures of presence are self-assessment questionnaires, applied post-test (van Baren & Ijsselstein, 2004).

Another important aspect of audience experience is the emotional reaction to a presented film. Here, we adopt the dimensional concept of emotions, following many other studies involving the presentation of emotion-inducing material, be it emotional images, film clips, or audio recordings (Bradley & Lang, 2007). According to the dimensional concept of emotions (Russell & Barrett, 1999), each emotional episode can be described

on two basic dimensions: valence and arousal. Valence determines if the stimulus is pleasant or unpleasant, while arousal determines the intensity of the emotional response it evokes. For example, highly positive stimuli might be calming, therefore evoking low arousal, e.g., forest scenery, or energizing, therefore evoking high arousal, such as extreme sports. Both arousal and valence dimensions of a stimulus, be it an image, sound, or word, can be conveniently assessed using the Self-Assessment Manikin (SAM) Scale, a graphical 9-point rating scale devised by Bradley and Lang (1994) which became a standard tool for the evaluation of emotional stimuli.

In this work, we also examine the psychophysiological measure of the cognitive and emotional state of a person, namely the heart rate variability (HRV), i.e., the irregularities in the time that passes between consecutive heart beats (for review see Shaffer & Ginsberg, 2017). In healthy people, beat-to-beat intervals are not even, but are constantly changing within an optimal range, and can be described as complex, chaotic fluctuations (Goldberger, 1991). The HRV provides information not only on physical health, but also on the psychophysiological state of the individual. Stress and other negative emotional states may result in reduced HRV, while high HRV values indicate higher self-regulatory capacity, including emotional regulation (for reviews see Appelhans & Luecken, 2006; McCraty & Shaffer, 2015). The reduction of HRV can be observed when participants are exposed to films inducing negative emotions. Specifically, viewing fear-evoking film clips is linked to decreased HRV (for review see Kreibig et al., 2007). However, studies on sadness-inducing clips provided mixed results, with one showing reduced HRV, one showing increased HRV, and five others showing no effect (for review see Kreibig et al., 2007).

Another psychological factor linked to HRV involves cognitive functions. The relation between attention and HRV has been well-documented across the entire lifespan. A study on infants showed that directing attention to the presented stimuli is related to decrease in HRV (Richards & Casey, 1991). In children and adolescents higher HRV is related to poorer performance in the sustained attention task (Griffiths et al., 2017). Also in adult populations a suppressed HRV was observed during engaging tasks with a high working memory load (Aasman et al., 1987; Hansen et al., 2003; Tattersall & Hockey, 1995). Together, it shows that low HRV might be indicative of enhanced attention engagement with the task.

In summary, among others, HRV may provide insight into emotional and cognitive states and into the engagement with stimuli like emotion-evoking pictures or films, which consists of both directing attention and emotional response. Moreover, it can be measured over longer periods of time, for example lasting 5 min (Shaffer & Ginsberg, 2017), which makes it an appropriate psychophysiological index in the studies using film clips, like the one presented here.

#### 4. Overview of the current study

In the present study, we analyze how intonation might affect participants' evaluation of audio description. The primary goal of the current study was to explore audience reaction to two different intonation types put forward by Cabeza-Cáceres (2013), that is adapted and emphatic intonation.

From a validated database of emotional films (Schaefer et al., 2010) we selected 20 short clips – 10 clips evoking positive emotions and 10 clips evoking negative emotions

(see section 4.2 for more information on materials). Clips were presented with either adapted or emphatic intonation. We expected that the emphatic intonation would be rated higher, since it would correspond better with the emotional content of the clips. We also aimed to determine which intonation type would be perceived as more appropriate depending on whether the clip was negative or positive. We expected that more intense emotional experience and higher presence would result in a higher evaluation of AD. To this end, after each clip, we included a presence questionnaire. We also measured heart rate variability, as an index of emotional and attentional engagement, during each clip. To assess the impact of all those factors on AD ratings, we used mixed models as a statistical tool to check the direction and significance of the relation between each factor (intonation type, emotional valence, presence, heart rate variability) and AD evaluation. Additionally, after the entire experimental session, we conducted short interviews with the participants, as they could provide valuable information, allowing us to better understand the obtained results.

#### 4.1. Participants

Overall 20 participants (6 female and 15 male), aged between 21 and 60 ( $M = 33.5$ ), took part in the study. All participants had self-declared vision loss and were native speakers of Polish. We did not collect data on the type of vision loss as we follow the current market practice – AD is produced in one version for all its audiences. The study followed the ethical rules of empirical research with human participants and was approved by the Universitat Autònoma de Barcelona Ethics Committee. Participants were recruited by two NGOs based in Poland through advertising on its social media and newsletters. Each participant gave their informed consent prior to the experiment. All data collected during the study have been anonymized.

#### 4.2. Materials

For the sake of the ecological validity of the experiment, we used excerpts from feature films that contain both non-verbal and verbal soundtrack. Clips for the experiment were selected from a validated database of emotional films (Schaefer et al., 2010). All clips in the database were divided into seven emotional categories: neutral state, amusement, anger, disgust, fear, sadness, tenderness, and rated for arousal on a 9-point SAM scale (Schaefer et al., 2010). For our study, we chose twenty clips, including ten evoking positive emotions (e.g., a moment of tenderness between lovers) and ten evoking negative emotions (e.g., a murder scene) (see Table 1). Duration of the selected clips varied between 1 m 09s and 5 m 45s. Total duration of negative and positive clips was very similar ( $M = 26$  m 30s and  $M = 26$  m 13s, respectively). While choosing the clips, we ensured that each of them had enough space for AD, that is to say, that there was enough space to insert AD between dialogues and/or important sounds. Regarding arousal, the clips selected for the experiment had ratings between 3.55 and 5.66 ( $M = 4.7$ ,  $SD = 0.64$ ) on a 9-point SAM (Bradley & Lang, 1994) scale ranging from calm (1) to excited (9). The arousal ratings did not differ significantly between valence categories;  $t(18) = 0.71$ ,  $p = 0.48$ .

AD for all clips was drafted by a professional Polish describer. Sixteen of the selected clips contained foreign language dialogue. The remaining four clips did not contain any

**Table 1.** The description of the clips used in the experiment, as provided in Schaefer et al. (2010).

Title	Emotion	Duration	Arousal
The visitors	amusement	2 m 09s	3.55
There is something about Mary		1 m 57s	4.02
Benny and Joon		2 m 07s	4.09
When Harry met Sally	tenderness	2 m 51s	4.67
When a man loves a woman		1 m 41s	3.71
Ghost		3 m 34s	5.00
The eight day		2 m 10s	5.25
Life is beautiful (3)		4 m 08s	5.44
The professional (1)	anger	2 m 50s	5.53
The dead poets society		2 m 46s	5.66
Leaving Las Vegas		2 m 35s	5.00
Schindler's list (3)		1 m 34s	5.06
Schindler's list (2)		1 m 59s	5.13
Hellraiser	disgust	1 m 30s	3.93
Trainspotting (2)		1 m 48s	4.84
Saving private Ryan		5 m 45s	5.73
Child's play II	fear	1 m 09s	4.63
Copycat		2 m 00s	4.76
Scream II		3 m 45s	4.81
The shining		4 m 25s	5.11

dialogue. For those clips which contained dialogue voice-over (VO)<sup>1</sup> was prepared by a professional Polish audiovisual translator. Both AD and VO were voiced in a professional recording studio. Following national AD guidelines and research findings (Szarkowska & Jankowska, 2015; Szymańska & Strzyński, 2010; Żórawska et al., 2011), ADs in all clips were voiced by a female voice-talent and VO by a male voice-talent. Regarding the AD intonation types, the studio was instructed to record two ADs for each clip: one following the standard Polish AD intonation and one acting out emotions, following intonation used in dubbing. The standard AD voicing in Poland draws heavily from voice-over, which assumes a certain degree of adaptation of intonation to the original content<sup>2</sup> (see Jankowska et al., 2017 for a detailed description of AD voicing in Poland). In our experiment, following the classification of Cabeza-Cáceres (2013), this type of intonation will be referred to as *adapted intonation*. The second intonation type used in our experiment will be referred to as *emphatic intonation* as it emphasized the emotional context of the original content (Cabeza-Cáceres, 2013). In total we prepared 40 clips: 20 following the adapted intonation and 20 following the emphatic intonation.

After the recording, the clips were mastered so that the original soundtrack and the additional soundtracks (VO and AD) had the same mean volume level. Since our sample included both blind and partially sighted participants, we presented only the audio track of the clips to minimize the possible variables and equalize experimental conditions for all participants, in order to make the results more clear in interpretation in terms of the reception of ADV itself, rather than involving a potential interaction between access to video track and ADV.

### 4.3. Procedure

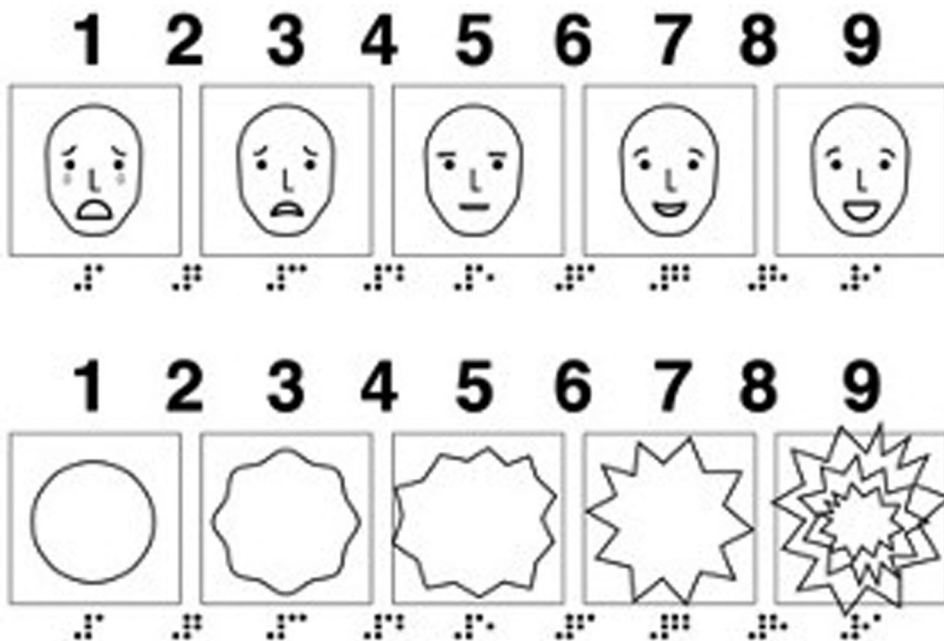
All participants were emailed information regarding the procedure and an informed consent form. Immediately prior to the experiment, a researcher read out loud the detailed information about the study, presented the device for measuring



psychophysiological reactions and – upon request – provided additional explanations. Then, the informed consent form was read, and participant's verbal responses were recorded.

The experimental procedure lasted approximately one hour and involved the presentation of 20 clips and a set of questions following each clip. Each participant listened to a given clip once – in only one, pseudo-randomly chosen, AD version (adapted or emphatic intonation). Out of 10 positive and 10 negative clips, half were presented with the emphatic intonation and half with the adapted intonation. The pseudo-random choice also ensured that the summed durations of the emphatic and the adapted clips in each experimental session did not differ by more than 5 min (on average 2 m 7s), and hence all participants spent approximately half of the experimental session listening to the emphatic intonation, and the other half to the adapted intonation. Moreover, each intonation version of every clip was presented to approximately half of the participants (no less than 8 and no more than 12). The pseudo-random choice of voicing version per each clip and per each participant was performed using an in-house MATLAB (MathWorks, Inc., Natick, MA) script, while randomization of the clips' order during the experimental sessions was controlled by the PsychoPy (Peirce, 2007).

After each clip, participants were asked 10 questions which concerned the rating of valence and arousal elicited by the given clip, the feeling of presence, and preferences regarding intonation type. Emotional experience was rated using the tactile version of SAM (Bradley & Lang, 1994) created by Iturregui-Gallardo & Méndez-Ulrich (2020) (Figure 1), which had already been used with participants with vision loss in previous experiments on emotional reactions to movie material (Iturregui-Gallardo & Méndez-Ulrich, 2020; Matamala et al., 2020). The tactile SAM is a simplified version of the



**Figure 1.** Tactile Self-Assessment Manikin (Iturregui-Gallardo & Méndez-Ulrich, 2020).



SAM scale prepared in the relief printing technique. Participants rated emotional valence and arousal.

As the last step of the procedure, we collected demographic data (age, gender, experience in watching films with AD) and conducted a short interview to see whether the participants spotted any difference between the clips. Finally, we asked participants to state their preference for emphatic or adapted intonation and encouraged them to justify their choice. Sessions ended with debriefing.

#### **4.4. Measurements**

To measure the sense of presence, we used a modification of the short version of the ITC-Sense of Presence Inventory (ITC-SOPI). The short version of the ITC-SOPI had previously been used to measure presence in AD related research (Fryer & Freeman, 2013, 2014; Walczak & Fryer, 2017, 2018). The questionnaire was shortened to limit the fatigue of the participants. We chose two out of three positive subscales: Sense of Physical Space ('I felt I was visiting the places in the scenes'; 'During the clip I had a sense of being in the scenes'; 'I felt surrounded by the scenes') and Engagement ('I felt myself being drawn in'; 'I lost track of time'; 'I paid more attention to the scenes than to my own thoughts'). The ITC-SOPI uses a 5-point Likert-type scale (1 = strongly disagree and 5 = strongly agree). Intonation type ('I liked the way AD was read') was rated on a 5-point Likert-type scale (1 = strongly disagree and 5 = strongly agree).

Heart rate was measured with the Biosignalsplux Explorer device (Biosignalsplux, Lisbon, Portugal) by ECG sensor supplied with the device. Disposable AgCl electrodes were placed on the left side of the thorax. Signal was digitized with frequency of 2 kHz using 16 bit wide analogue to digital conversion with the gain ratio of 1000 and 0.5–100 Hz bandwidth. R peaks of the QRS complex (Einthoven, 1912) were detected online using the Pan-Tomkins algorithm implemented into custom-written LabVIEW recording software (Lascu & Lascu, 2007). Time intervals between two consecutive R peaks (RR intervals) were converted to heart rate and stored for offline analysis. Then, the heart rate (in beats per minute, bpm) was cleaned for artifacts. Firstly, we compared consecutive values and excluded data points for which the difference in value between one data point and the previous one exceeded the set threshold (more than 20 units). Secondly, we excluded too large (heart rate over 115 bpm) and too small values (heart rate under 40 bpm). The heart rate variability (HRV) was assessed using the standard deviation of the interbeat intervals of normal sinus beats, called SDNN measure (Shaffer & Ginsberg, 2017), calculated in milliseconds. The heart rate data was preprocessed using MATLAB.

#### **4.5. Statistical analysis**

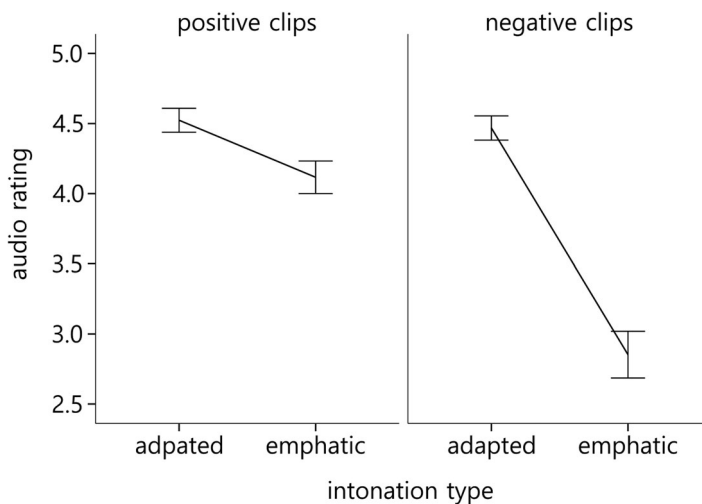
We have chosen a mixed model to analyze our data as it makes possible to account for individual differences in responses to experimental conditions by explicit specification of random effects in addition to fixed ones. In other words, the mixed model is well suited to the situation where idiosyncrasy in individual responsiveness to experimental conditions plays a significant role. Indeed, we expected that ratings of AD would not only be different for emphatic and adapted intonation (fixed effect), but also that the participants

would exhibit unique patterns in their preferences for ADV (random effect). Moreover, the mixed model, unlike repeated measures ANOVA, has advantage of allowing for the seamless integration of nominal and continuous predictors, both present in our analysis, into one statistical model. Consequently, using SPSS version 24 (Armonk, NY: IBM Corp.), we performed a linear mixed model on the audio rating as a dependent, to be explained, variable with following four predicting variables (fixed factors): intonation type (adapted, emphatic) and valence (negative, positive), treated as categorical repeated measures factors as well as continuous factors of presence scale score and SDNN. We included a random intercept for each participant and a random slope for each participant for the factor of intonation (accommodating individual differences in response magnitude to voicing type). We assumed a variance components covariance matrix structure for the random factors and diagonal covariance matrix structure for the repeated measures factors. Effect sizes for each of the fixed effects were estimated with partial  $R^2$  (Edwards et al., 2008). Selected post-hoc comparisons were conducted using Bonferroni corrected t-tests.

## 5. Results

### 5.1. Mixed model

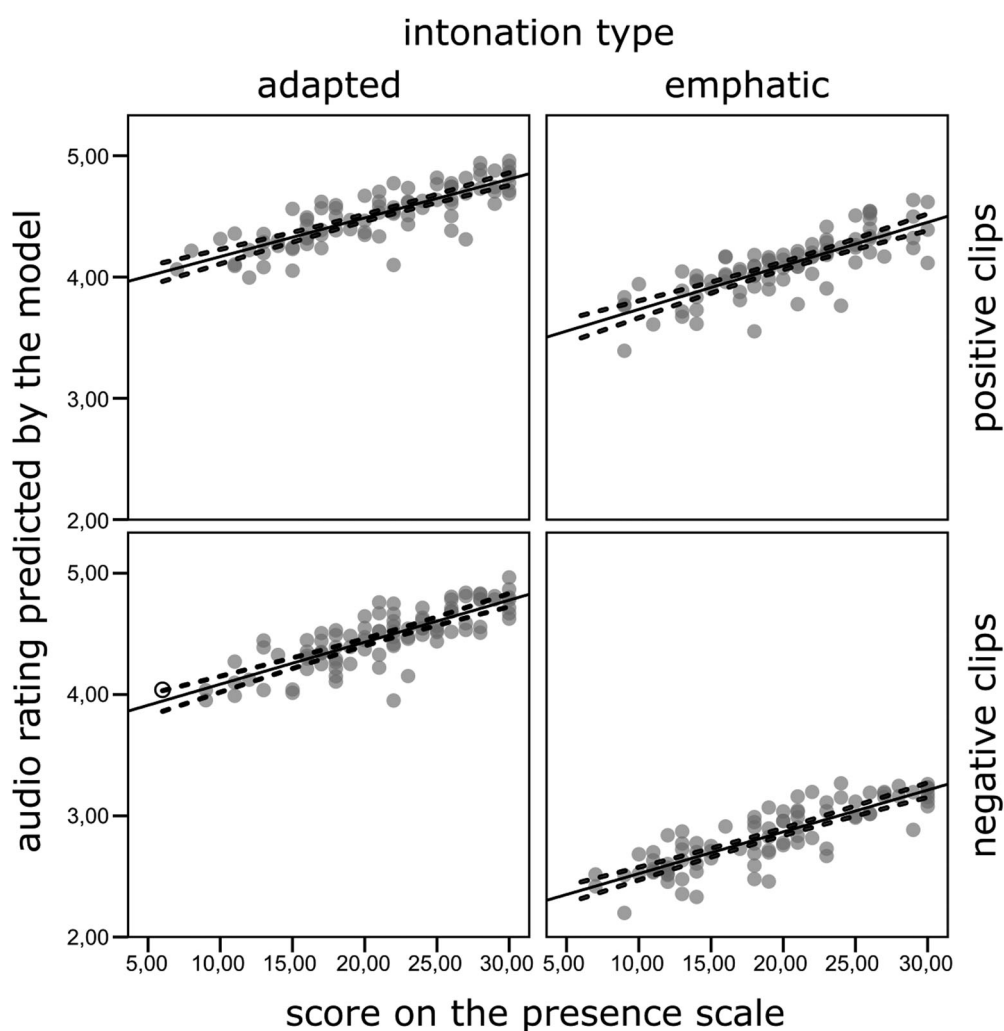
Ratings of the adapted intonation were systematically higher ( $M = 4.47$ ,  $SE = 0.17$ ) than the emphatic voicing ( $M = 3.49$ ,  $SE = 0.18$ )  $F(1, 21.1) = 40.7$ ,  $p < .001$ ,  $R^2 = .66$ . Also, the emotional valence of a stimulus affected the ratings of AD with the positive clips being rated higher ( $M = 4.31$ ,  $SE = 0.15$ ) than the negative clips ( $M = 3.65$ ,  $SE = 0.16$ )  $F(1, 184.9) = 60.49$ ,  $p < .001$ ,  $R^2 = .25$ . Crucially, the results yielded a significant interaction between ADV type and emotional valence of the clips ( $F(1, 184.4) = 46.34$ ,  $p < .001$ ,  $R^2 = .20$ ). The usage of the emphatic, as compared to the adapted intonation, resulted in low AD ratings of the negative clips, as compared to the positive clips (Figure 2). Post hoc comparisons conducted separately for each valence condition revealed that AD



**Figure 2.** Interaction of intonation type with valence. Bars represent standard error.

ratings depended on the ADV type, within both emotional valence categories, i.e., positive ( $p = .019$ ) and negative ( $p < .001$ ). The post hoc comparisons within the intonation category, however, indicated that AD-rating differed between positive and negative clips only in the case of the emphatic intonation ( $p < .001$ ). There was no significant difference for the adapted intonation ( $p = .27$ ). These results indicate that the significant main effect of valence in terms of AD-rating can be entirely attributed to differences observed in emphatic intonation. When adapted intonation was used, participants rated it as equally adequate in both positive and negative clips.

Also, two continuous covariates were linked to the AD rating, that is, presence and HRV (measured as SDNN index). Presence scale was positively related to the AD rating  $F(1, 234.0) = 25.43$ ,  $p < .001$ ,  $R^2 = .09$ . The model predicted that with one point on the presence scale, AD ratings increase by 0.035 points (Figure 3). That means

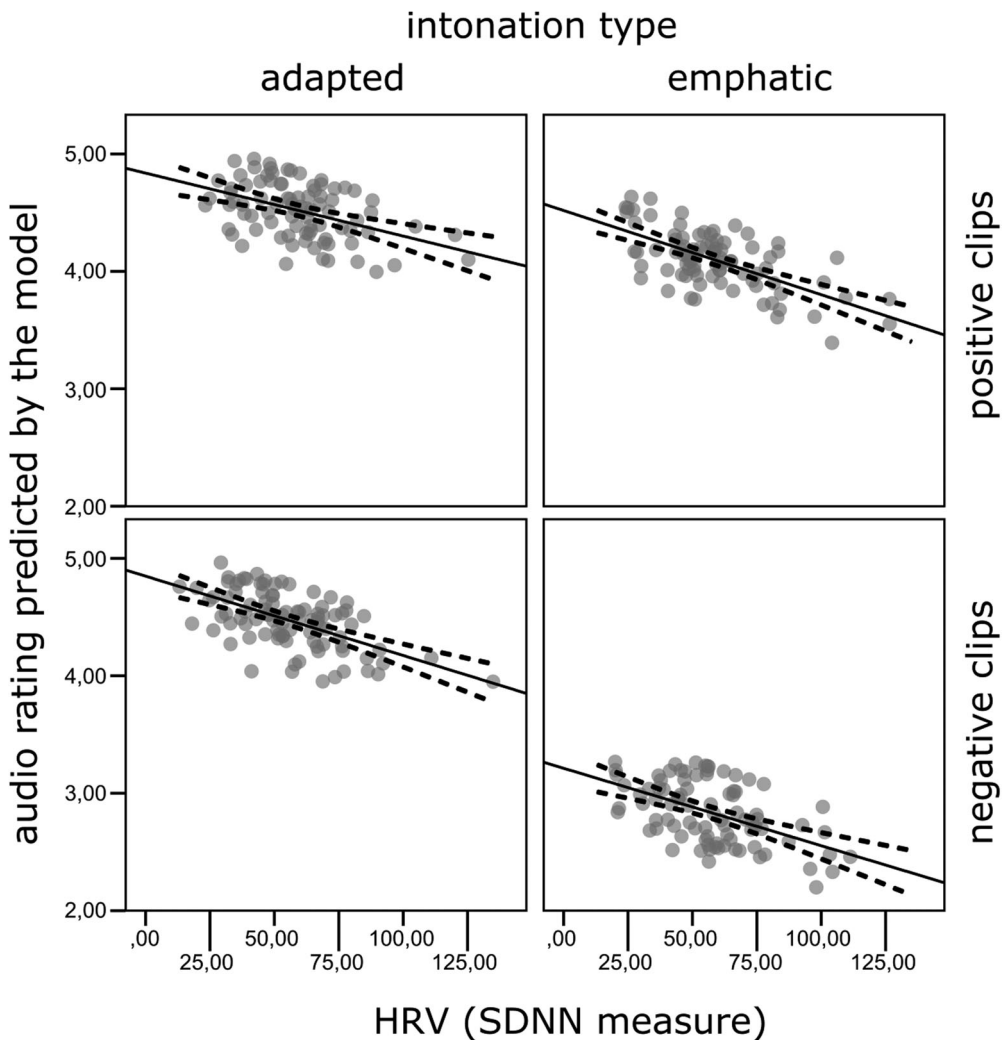


**Figure 3.** Relation between presence scale and audio ratings predicted by the model plotted separately for every condition with overlaid linear regression (solid lines) and 95% CIs (dashed lines).

that, in line with our expectations, higher presence resulted in higher evaluation of AD. Additionally, the SDNN index was significantly negatively related to the AD ratings  $F(1, 253.4) = 8.9$ ,  $p = .003$ ,  $R^2 = .03$ . According to the obtained statistical model, the AD ratings decreased by 0.007 points for each unit increase in the SDNN (Figure 4). This means that the lower the heart rate variability, the better the AD ratings.

## 5.2. Interviews

All participants were able to detect a difference between the clips and attribute it to the intonation type. During the discussion, participants noticed that as a general rule, there seem to be three AD intonation types in Poland. They described the first one as ‘too



**Figure 4.** Relation between heart rate variability (SDNN measure) and audio ratings predicted by the model plotted separately for every condition with overlaid linear regression (solid lines) and 95% CIs (dashed lines).

emotional' and 'exaggerated' or 'acted'. The second as 'flat', 'synthetic' and 'robot-like'. And the third one was compared voice-over, which, according to the participants, shows some interpretation, but does not reveal emotions. As one of the participants said:

I really like when AD is read with a little bit of interpretation. I mean I don't like it when it's acted, but I also don't like it when it's completely flat. Lektor-like<sup>3</sup> is the best way.

We believe that the intuitive categories put forward by the participants of our study can be fitted into the threefold classification put forward by Cabeza-Cáceres (2013), i.e., emphatic, uniform and adapted.

Participants were very clear and passionate about their preference for the AD intonation type. Most of them pointed to the adapted intonation as their preferred option since they found it unobtrusive and engaging at the same time, e.g.:

I like it when AD is slightly interpreted. Not exaggerated, not actor-like but also not completely flat [...]. Exaggerated acting draws too much attention, but with some acting and some intonation the AD is nicer to listen to.

Some participants literally referred to the adapted intonation as the golden mean. As two of them said:

If I had to choose between flat, slightly interpreted or acted, I would definitely choose the slightly interpreted one. I noticed this style in some of the clips I saw today. It's a good golden mean.

I would say that we should aim for the golden mean. It will fit with every movie. Both acted and flat reading can really destroy the film. It's irritating and discouraging.

While some adapted intonation in ADV is desirable, most participants noticed that the voice should be well matched with the film, and the intonation type should match the character of a scene. As one of the participants put it:

The voice should be well suited to the film and also to the genre. To give you an example of such adaptation – I think that a more cheerful voice in romantic comedies would be good. Or in action or military film – the voice could be more soldiery.

At the same time, all participants were very critical about the emphatic intonation. As some of them explained, it is disturbing and draws their attention away from the film:

I noticed that the female voice-talent was getting too emotional sometimes. The script was excellent, but the reading was flustering. It was taking up my entire attention.

An important issue raised by many participants was the fact that emphatic intonation does not allow them to experience the film in their own way:

I want to concentrate on what is happening and live it my way. I don't want the voice talent to tell me what I am supposed to feel. If AD is well written, we do not need voice-talent's emotions. Some interpretation in AD voicing is good, but it shouldn't be overdone.

## 6. Discussion

Results obtained in the experiment show that participants prefer the adapted intonation – it is considered to be the golden mean. What is more, our results show that film genre,

voice type, and emotional valence of a scene can affect the preference for a particular ADV. The interviews with the participants suggest that emphatic intonation could be beneficial for certain genres (e.g., romantic comedies, war films), but a poorly-matched voice (e.g., cheerful or comedy voice to a horror) can distract the audiences. When it comes to emotional valence, it seems that the issue of intonation type is especially sensitive when dealing with negative emotions. Participants judged emphatic intonation in negative clips as the least adequate. This could be due to the fact that emphatic intonation unnecessarily enhances emotions that are already very clearly expressed through the soundtrack or dialogues. As stressed by participants, in this particular case, emphatic intonation not only distracts them but also makes them feel as if they were told what they should feel.

All in all, we might conclude that similarly to scripting, the choice of intonation type should not be considered to be a 'binary opposition of objective vs. subjective' (Mazur & Chmiel, 2012); these are two extremes of a continuum.

Another finding of our study is the link established between the presence, HRV, and AD rating. The decrease in HRV with higher audio rating can be interpreted as engagement of attention in the presented clip, following previous studies that showed a lower HRV associated with sustained attention during a task (Griffiths et al., 2017; Richards & Casey, 1991). The attentional engagement in a clip might have also resulted in a deeper emotional engagement with the plot, leading to lower HRV. For example, studies showed that fear-evoking clips are associated with lower HRV (for review see Kreibig et al., 2007). The relationship between attentional and emotional engagement and its link to both physiological and self-reported indices of immersion should be addressed in further studies on audience reception of emotional clips. It seems that the more involved (higher presence scores) and focused (lower HRV) the audience was, the higher the AD rating was given. This observation is confirmed in the interviews with the study participants who repeatedly referred to the emphatic intonation as distracting, which might make presence more difficult to achieve.

Further research on ADV is needed since our results may be biased by some limitations. First of all, Polish audiences are used to voice-over and thus might prefer the *lektor*-like intonation, which, as study participants noticed, is *per se* neither flat nor emotionless. Different results might be obtained in countries with different reading traditions. However, results obtained by Cabeza-Cáceres (2013) may suggest that adapted intonation is the preferred option even for audiences used to uniform intonation. Another issue is that, in order to limit potential variables, all our clips were recorded by the same female voice-talent. As many of the participants noticed, the voice used was not well suited for all the clips. Participants perceived the voice as rather cheerful and claimed that it did not go well with the negative clips. An important limitation of our study is the choice of participants. As we already mentioned, we did not factor in the degree of sight loss or age of the participants, which varied quite considerably. This, however, is a common issue in studies involving participants with vision loss – it is a group that is relatively hard to reach. Last but not least is the duration of the clips used in the experiment, which might be seen as a limitation of the study, however, we believe that the issue of the length of the clips used is justified by the aims of our study. We chose short clips for a number of reasons. Firstly, in an experimental paradigm with multiple clips, it is necessary to expose the participants to a variety of stimuli while maintaining a reasonable duration

of the experiment. The use of longer clips could impact the number of stimuli employed and reduce the generalizability of the results. Secondly – to better control emotions to be elicited and to avoid co-elicitation of opposing emotions (e.g., fear vs. amusement). Thirdly, to reduce the impact of random variables, e.g., a peculiar reaction of a participant to a given clip. Short clips have been successfully used in psychophysiology and neuroscience (Bos et al., 2013; Schaefer et al., 2010) where it is claimed that clips that last between one and two minutes are long enough to provide the viewer with an understanding of the plot, engage their attention and change their affective state while clips longer than three minutes might prompt ‘carryover effects on the following excerpts’ (Maffei & Angrilli, 2019, p. 3). For what concerns the optimal duration of presence stimulus, there are no clues in the relevant literature – authors use stimuli of 100 s (Freeman et al., 2000), 10 min (Rigby et al., 2016) but also of 45 min (Troschianko et al., 2012). This issue is also not addressed by past authors in the field of audio description – they have used various number of clips lasting between 2 and 12 min (Fryer & Freeman, 2014; Walczak & Fryer, 2017, 2018). Our results show that stimuli used in the experiment resulted in a certain level of presence – of course, it could have been higher if we had used longer clips, but this does not invalidate the findings of our study.

## 7. Conclusions

The current article examined the effects of different AD intonation types on audience experience. The most important result is that the less expressive, adapted intonation was rated as more appropriate, regardless of the emotional content of the film clip. Even though participants were more willing to accept emphatic intonation for positive clips and argued that, in some cases, emphatic intonation could make a given scene more attractive, on the whole, they gave higher ratings to those positive clips that were read with adapted intonation. The emphatic intonation proved to be particularly distracting and evaluated as non-appropriate in the case of clips evoking negative emotions. This might be linked to the poor matching of the voice to the emotional content of the negative clips, corroborating the findings of Kobayashi et al. (2010) who concluded that inappropriate use of emotional voices may have a negative effect on video experience.

Moreover, we observed a link between the individual experience (i.e., the sense of physical space, engagement, emotional reaction, the focus of attention) and the evaluation of the intonation type. It seems that participants gave higher ratings to the intonation type which allowed them to focus more on the emotional content of the film. This pattern of results is confirmed by the interviews with the participants, who stated that they preferred non-intrusive intonation. Interestingly, our study clearly shows that intonation does not need to mirror the emotions of the film but rather should allow listeners to make their own interpretation. However, an important point to remember is that the study participants made it very clear that adapted intonation is not synonymous with flat and emotionless reading. This is in line with some of the AD guidelines (e.g., ITC, 2000).

It is also clear that further research is needed to fully understand the nature of different AD intonation types. Their current definitions put forward by Cabeza-Cáceres (2013) are rooted in the available AD guidelines, which, as already mentioned, are vague. We agree with Matamala et al. (2020, p. 71) that ‘we need to use linguistic tools to analyze prosodic values if we want to go beyond impressionistic suggestions and make research-based



recommendations'. Other issues that need to be addressed are the relationship between intonation type, type of voice, and film genre as well as replication of our experiment on speakers of different languages who are used to different ADV traditions. Another avenue to explore is the relationship between AD intonation type and the cognitive effort required to process it.

Last, but by far not least, is the burning issue of methodology in audiovisual translation and media accessibility. The common consensus is that while there has been a rise in experimental research in audiovisual translation and media accessibility, it is still not consolidated (Díaz Cintas & Szarkowska, 2020; Orero et al., 2018). One of the ways towards consolidation is inviting researchers from other domains (O'Brien, 2013). This study brought together experts from psychophysiology and audiovisual translation. What at times seemed like a methodological clash has been an invaluable experience which hopefully will contribute to shaping methodological paradigms of audiovisual translation.

## Notes

1. Voice-over is the predominant AVT modality on Polish television (Szarkowska, 2009).
2. Even though voice-over is usually described as monotonous and flat (Bogucki, 2010) some researchers and practitioners underline that intonation and modulation are important aspects of quality voicing in voice-over (Chłopicki in personal communication, October 15, 2020; Woźniak, 2012).
3. *Lektor* is the name used in Poland for voice-talents who specialize in voice-over reading.

## Acknowledgements

The authors would like to thank Weronika Janeczko, Zofia Kania and Radosław Sterna who assisted in data-gathering. Anna Jankowska would like to thank her co-authors, without whom this amazing journey could not and would not happen. It was an eye-opening experience that has changed my research paradigm. We would also like to thank the reviewers for their comments which allowed us to significantly improve this article.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work has been supported by the Spanish Ministry of Economy, Industry and Competitiveness (FFI2015-64038-P, MINECO/FEDER, UE) within the research group Transmedia Catalonia (2014SGR0027), by the Polish Ministry of Science and Higher Education (1311/MOB/IV/2015/0) under the Mobility Plus funding scheme, and by the National Science Centre in Poland (grant number 2014/15/N/HS6/04179).

## Notes on contributors

*Anna Jankowska*, PhD, is Research professor at the University of Antwerp and former Assistant Lecturer in the Chair for Translation Studies and Intercultural Communication at the Jagiellonian University in Krakow (Poland) and visiting scholar at the Universitat Autònoma de Barcelona

within the Mobility Plus program of the Polish Ministry of Science and Higher Education (2016–2020). Her recent research projects include studies on accessibility technologies, user experience, the viability of translating audio description scripts from foreign languages and multiculturalism in audio description. She is also the founder and president of the Seventh Sense Foundation which provides audio description and subtitles for the deaf and hard of hearing. Anna is member of ESIST and Editor-in-Chief of the Journal of Audiovisual Translation.

**Joanna Pilarczyk**, PhD, is a Post-doc at the Emotion and Perception Lab at the Institute of Psychology, the Jagiellonian University in Krakow, Poland. Her PhD thesis concerned the impact of image features on the viewers' emotional response. Her current projects focus on the role of beliefs and motivations in the processing of social and emotional stimuli. In her projects, she uses eye-tracking, pupillometry, electrocardiography, and neuroimaging.

**Kinga Wołoszyn**, MA, obtained her Master's degree in psychology at the Jagiellonian University (JU) in Kraków. Her PhD thesis concerning the processing of emotional material, within the framework of embodied cognition, at the Psychophysiology Lab. With the team, she has been conducting research on the physiological and attentional aspects of the processing of emotional natural scenes and its neural bases. In her experiments, she uses various methods, including electroencephalography (EEG), functional magnetic resonance imaging (fMRI), eye-tracking, pupillary response, electromyography (EMG), and electrocardiography (ECG). She is also a collaborator in the New Approaches to Accessibility project.

**Michał Kuniecki**, PhD, is an Assistant Professor at the Institute of Psychology at the Jagiellonian University in Krakow (Poland). He studies emotion and visual perception. His interests include the role of formal features of visual stimuli and their meaning in engaging spatial attention and eliciting emotional responses. In his work, he utilizes a whole array of psychophysiological methods, such as EEG, fMRI, eye-tracking, pupillary response, and heart rate.

## ORCID

Anna Jankowska  <http://orcid.org/0000-0001-6863-5940>

Joanna Pilarczyk  <http://orcid.org/0000-0002-2654-7536>

Kinga Wołoszyn  <http://orcid.org/0000-0001-6910-8667>

Michał Kuniecki  <http://orcid.org/0000-0003-2059-9914>

## References

- Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors*, 29(2), 161–170. <https://doi.org/10.1177/001872088702900204>
- Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology*, 10(3), 229–240. <https://doi.org/10.1037/1089-2680.10.3.229>
- Bittner, H. (2010). *Audio description guidelines – A comparison*. [https://www.uni-hildesheim.de/media/\\_migrated/content\\_uploads/AD\\_Guidelines\\_Comparison\\_-\\_Read.pdf](https://www.uni-hildesheim.de/media/_migrated/content_uploads/AD_Guidelines_Comparison_-_Read.pdf)
- Bogucki, Ł. (2010). The demise of voice-over? Audiovisual translation in Poland in the 21st century. In B. Lewandowska-Tomaszczyk & M. Thelen (Eds.), *Meaning in translation* (pp. 415–424). Peter Lang.
- Bos, M. G. N., Jentgens, P., Beckers, T., & Kindt, M. (2013). Psychophysiological response patterns to affective film stimuli. *PLoS ONE*, 8(4), e62661. <https://doi.org/10.1371/journal.pone.0062661>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Bradley, M. M., & Lang, P. J. (2007). Motivation and emotion. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 581–607). Cambridge University Press.

- Cabeza-Cáceres, C. (2013). Audiodescrípció i recepció: efecte de la velocitat de narració, l'entonació i l'explicitació en la comprensió fílmica [Audio description and reception: effect of narration speed, intonation and explicitness on filmic comprehension] (PhD). Universitat Autònoma de Barcelona, <https://hdl.handle.net/10803/113556>
- Díaz Cintas, J., & Szarkowska, A. (2020). Introduction: Experimental research in audiovisual translation – Cognition, reception, production. *The Journal of Specialised Translation*, (33), 3–16.
- Drożdż-Kubik, J. (2011). Harry Potter i Kamień Filozoficzny słowem malowany - czyli badanie odbioru filmu z audiodeskrypcją z syntezą mowy [Harry Potter and the Philosopher's Stone painted with words: research into reception of the film with text-to-speech audio description]. (MA). Jagiellonian University in Kraków, <https://ruj.uj.edu.pl/xmlui/handle/item/175953>
- Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, 27(29), 6137–6157. <https://doi.org/10.1002/sim.3429>
- Einthoven, W. (1912). The different forms of the human electrocardiogram and their signification. *The Lancet*, 179(4622), 853–861. [https://doi.org/10.1016/S0140-6736\(00\)50560-1](https://doi.org/10.1016/S0140-6736(00)50560-1)
- Fernández i Torné, A., & Matamala, A. (2015). Text-to-speech vs. human voiced audio descriptions: A reception study in films dubbed into Catalan. *The Journal of Specialised Translation*, 24, 61–88. [http://www.jostrans.org/issue24/art\\_fernandez.pdf](http://www.jostrans.org/issue24/art_fernandez.pdf)
- Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E., & IJsselstein, W. (2000). Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators and Virtual Environments*, 9(2), 149–164. <https://doi.org/10.1162/105474600566691>
- Fryer, L. (2013). Putting it into words: The impact of visual impairment on perception, experience and presence. pp. 372.
- Fryer, L. (2016). *An introduction to audio description a practical guide*. Oxon: Routledge.
- Fryer, L., & Freeman, J. (2013). Cinematic language and the description of film: Keeping AD users in the frame. *Perspectives: Studies in Translatology*, 21(3), 412–426. <https://doi.org/10.1080/0907676X.2012.693108>
- Fryer, L., & Freeman, J. (2014). Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In A. Felhofer & O. D. Kothgassner (Eds.), *Proceedings of the international society for presence research* (pp. 99–107). Facultas Verlags- und Buchhandels AG.
- Goldberger, A. L. (1991). Is the normal heartbeat chaotic or homeostatic? *Physiology*, 6(2), 87–91. <https://doi.org/10.1152/physiologyonline.1991.6.2.87>
- Griffiths, K. R., Quintana, D. S., Hermens, D. F., Spooner, C., Tsang, T. W., Clarke, S., & Kohn, M. R. (2017). Sustained attention and heart rate variability in children and adolescents with ADHD. *Biological Psychology*, 124, 11–20. <https://doi.org/10.1016/j.BIOPSYCHO.2017.01.004>
- Hansen, A. L., Johnsen, B. H., & Thayer, J. F. (2003). Vagal influence on working memory and attention. *International Journal of Psychophysiology*, 48(3), 263–274. [https://doi.org/10.1016/S0167-8760\(03\)00073-4](https://doi.org/10.1016/S0167-8760(03)00073-4)
- Independent Television Commission. (2000). *ITC guidance on standards for audio description*. [http://audiodescription.co.uk/uploads/general/itcguide\\_sds\\_audio\\_desc\\_word3.pdf](http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf)
- Iturregui-Gallardo, G., & Méndez-Ulrich, J. L. (2020). Towards the creation of a tactile version of the Self-Assessment Manikin (T-SAM) for the emotional assessment of visually impaired people. *International Journal of Disability, Development and Education*, 67(6), 657–674. <https://doi.org/10.1080/1034912X.2019.1626007>
- Jankowska, A. (2019). *Audiovisual media accessibility*. pp. 231–260.
- Jankowska, A., Ziółko, B., Igras-Cybulska, M., & Psiuk, A. (2017). Reading rate in filmic audio description. *Rivista Internazionale di Tecnica della Traduzione = International Journal of Translation*, 19, 75–97. <https://doi.org/10.13137/2421-6763/17352>
- Kobayashi, M., O'Connell, T., Gould, B., Takagi, H., & Asakawa, C. (2010). Are synthesized video descriptions acceptable? *ASSETS'10 - Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 163–170. . <https://doi.org/10.1145/1878803.1878833>

- Kreibig, S. D., Wilhelm, F. H., Roth, W. T., & Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology*, 44(5), 787–806. <https://doi.org/10.1111/j.1469-8986.2007.00550.x>
- Lascu, M., & Lascu, D. (2007). *LabVIEW Event Detection using Pan-Tompkins Algorithm*. Paper presented at the Proceedings of the 7th WSEAS international conference on signal, speech and image processing.
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A Cross-Media Presence Questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, 10(3), 282–297. <https://doi.org/10.1162/105474601300343612>
- Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2). <https://doi.org/10.1111/j.1083-6101.1997.tb00072.x>
- Machuca, M. J., Matamala, A., & Ríos, A. (2020). Prosodic features in Spanish audio descriptions of the VIW corpus. *MonTI*, 12(12), 53–77. <https://doi.org/10.6035/monti.2020.12.02>
- Maffei, A., & Angrilli, A. (2019). E-MOVIE - Experimental MOVies for induction of emotions in neuroscience: An innovative film database with normative data and sex differences. *PLOS ONE*, 14(10), e0223124. <https://doi.org/10.1371/journal.pone.0223124>
- Matamala, A., Soler-Vilageliu, O., Iturregui-Gallardo, G., Jankowska, A., Méndez-Ulrich, J.-L., & Ratera, A. S. (2020). Electrodermal activity as a measure of emotions in media accessibility research: Methodological considerations. *The Journal of Specialised Translation*, 129–151. [https://jostrans.org/issue33/art\\_matamala.pdf](https://jostrans.org/issue33/art_matamala.pdf)
- Mazur, I., & Chmiel, A. (2012). Audio description made to measure: Reflections on interpretation in AD based on the pear tree project data | WA publications. In A. Remael, P. Orero, & M. Carroll (Eds.), *Audiovisual translation and media accessibility at the crossroads. Media for all 3* (pp. 173–188). Rodopi.
- Maćzyńska, M. (2011). Text-to-speech audio description with audio subtitling for a feature film. A case study of Borat: Cultural Learnings of America for Make Benefit Glorious Nation of Kazakhstan. Paper presented at the Points of view in language and culture, Kraków. <http://avt.ils.uw.edu.pl/files/2012/04/TTS-AD-with-audio-subtitling-M%C4%85czy%C5%84ska-%C5%81%C3%B3d%C5%BA-2011.pdf>
- McCraty, R., & Shaffer, F. (2015). Heart rate variability: New perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. *Global Advances in Health and Medicine*, 4(1), 46–61. <https://doi.org/10.7453/gahmj.2014.073>
- Netflix. Audio Description Style Guide v1.0 – Netflix | Partner Help Center. <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215510667-Audio-Description-Style-Guide-v1-0>
- O'Brien, S. (2013). The borrowers: Researching the cognitive aspects of translation. *Target*, 25(1), 5–17.
- Orero, P., Doherty, S., Kruger, J.-L., Matamala, A., Pedersen, J., Perego, E., Romero-Fresco, P., Rovira-Esteva, S., Soler-Vilageliu, O., & Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *The Journal of Specialised Translation*, (30), 105–126.
- Peirce, J. W. (2007). Psychopy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Ramos Caro, M. (2015). The emotional experience of films: Does audio description make a difference? *The Translator: Studies in Intercultural Communication*, 21(1), 68–94. <https://doi.org/10.1080/13556509.2014.994853>
- Ramos Caro, M. (2016). Testing audio narration: The emotional impact of language in audio description. *Perspectives*, 24(4), 606–634. <https://doi.org/10.1080/0907676X.2015.1120760>
- Remael, A., Reviers, N., & Vercauteren, G. (2015). *Pictures painted in words: ADLAB Audio Description guidelines*. pp. 99. <https://www.openstarts.units.it/handle/10077/11838>
- Richards, J. E., & Casey, B. J. (1991). Heart rate variability during attention phases in young infants. *Psychophysiology*, 28(1), 43–53. <https://doi.org/10.1111/j.1469-8986.1991.tb03385.x>
- Rigby, J. M., Brumby, D. P., Cox, A. L., & Gould, S. J. J. (2016). *Watching movies on netflix*. Paper presented at the Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct.

- Rojo, A., Ramos Caro, M., & Valenzuela, J. (2014). The emotional impact of translation: A heart rate study. *Journal of Pragmatics*, 71, 31–44. <https://doi.org/10.1016/j.pragma.2014.07.006>
- Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 76(5), 805–819. <https://doi.org/10.1037/0022-3514.76.5.805>
- Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition & Emotion*, 24(7), 1153–1172. <https://doi.org/10.1080/02699930903274322>
- Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5, 258. <https://doi.org/10.3389/fpubh.2017.00258>
- Sheridan, T. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1(1), 120–126. <https://doi.org/10.1162/pres.1992.1.1.120>
- Snyder, J. (2010). *Audio description guidelines and best practices*. <http://docenti.unimc.it/catia.giaconi/teaching/2017/17069/files/corso-sostegno/audiodescrizioni>
- Szarkowska, A. (2009). The audiovisual landscape in Poland at the dawn of the 21st century. In A. Goldstein & B. Golubović (Eds.), *Foreign language movies – dubbing vs. subtitling* (pp. 185–201). Hamburg: Verlag Dr. Kovač.
- Szarkowska, A. (2011). Text-to-speech audio description: Towards wider availability of AD. *The Journal of Specialised Translation*, 15, 142–162. [https://www.jostrans.org/issue15/art\\_szarkowska.pdf](https://www.jostrans.org/issue15/art_szarkowska.pdf)
- Szarkowska, A., & Jankowska, A. (2012). Text-to-speech audio description of voiced-over films. A case study of audio described *Volver* in Polish. In E. Perego (Ed.), *Emerging topics in translation: Audio description* (pp. 81–98). EUT Edizioni Università di Trieste.
- Szarkowska, A., & Jankowska, A. (2015). Audio describing foreign films. *The Journal of Specialised Translation*, 243–269.
- Szymańska, B., & Strzymiński, T. (2010). *Standardy tworzenia audiodeskrypcji do produkcji audio-wizualnych*. [http://avt.ils.uw.edu.pl/files/2010/12/AD-\\_standardy\\_tworzenia.pdf](http://avt.ils.uw.edu.pl/files/2010/12/AD-_standardy_tworzenia.pdf)
- Tattersall, A. J., & Hockey, G. R. J. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors*, 37(4), 682–698. <https://doi.org/10.1518/001872095778995517>
- Troscianko, T., Meese, T. S., & Hinde, S. (2012). Perception while watching movies: Effects of physical screen size and scene type. *I-perception*, 3(7), 414–425. <https://doi.org/10.1068/i0475aap>
- Udo, J. P., Acevedo, B., & Fels, D. I. (2010). Horatio audio-describes Shakespeare's Hamlet. *British Journal of Visual Impairment*, 28(2), 139–156. <https://doi.org/10.1177/0264619609359753>
- van Baren, J., & Ijsselstein, W. (2004). *Measuring presence: A guide to current measurement approaches*. [http://www8.informatik.umu.se/\\$\sim\\$jjworth/PresenceMeasurement.pdf](http://www8.informatik.umu.se/$\sim$jjworth/PresenceMeasurement.pdf)
- 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. <https://doi.org/10.1177/0264619616661603>
- Walczak, A., & Fryer, L. (2018). Vocal delivery of audio description by genre: Measuring users' presence. *Perspectives*, 26(1), 69–83. <https://doi.org/10.1080/0907676X.2017.1298634>
- Walczak, A., & Szarkowska, A. (2011). *Text-to-speech audio description to educational materials for visually-impaired children*. Paper presented at the Advanced Research Seminar on Audio Description, Barcelona. [http://avt.ils.uw.edu.pl/files/2012/04/Presentation\\_Barcelona.pdf](http://avt.ils.uw.edu.pl/files/2012/04/Presentation_Barcelona.pdf)
- Walczak, A., & Szarkowska, A. (2012). Text-to-speech audio description to educational materials for visually-impaired children. In S. Bruti & E. di Giovanni (Eds.), *Audiovisual translation across Europe. An ever-changing landscape* (pp. 209–234). Peter Lang.
- Woźniak, M. (2012). Voice-over or voice-in-between? Some considerations about the voice-over translation of feature films on Polish television. In A. Remael, P. Orero, & M. Carroll (Eds.), *Audiovisual translation and media accessibility at the crossroads. Media for all 3* (pp. 209–228). Rodopi.
- Żórawska, A., Więckowski, R., Kunstler, I., & Butkowska, U. (2011). *Audiodeskrypcja – zasady tworzenia*. <https://dzieciom.pl/wp-content/uploads/2012/09/Audiodeskrypcja-zasady-tworzenia.pdf>