

Recibido / Received: 21/05/2019  
Aceptado / Accepted: 30/09/2019

Para enlazar con este artículo / To link to this article:  
<https://doi.org/10.6035/MonTI.2020.12.02>

Para citar este artículo / To cite this article:

Machuca, María J.; Anna Matamala & Antonio Ríos. (2020) "Prosodic features in Spanish audio descriptions of the VIW corpus." In: Richart-Marset, Mabel & Francesca Calamita (eds.) 2020. *Traducción y Accesibilidad en los medios de comunicación: de la teoría a la práctica / Translation and Media Accessibility: from Theory to Practice*. *MonTI* 12, pp. 53-77.

## PROSODIC FEATURES IN SPANISH AUDIO DESCRIPTIONS OF THE VIW CORPUS<sup>1</sup>

MARÍA J. MACHUCA

MariaJesus.Machuca@uab.cat  
Universitat Autònoma de Barcelona

ANNA MATAMALA

Anna.Matamala@uab.cat  
Universitat Autònoma de Barcelona

ANTONIO RÍOS

Antonio.Rios@uab.cat  
Universitat Autònoma de Barcelona

### Abstract

The aim of this study is to analyse the prosodic features of a corpus of audio descriptions in Spanish to determine the user preferences, both sighted and persons with sight loss. The analysis is contextualised by a thorough review of the guidelines and recommendations on voicing audio description. The corpus analysis is based on 10 audio descriptions produced by Spanish professionals. The audio descriptions are

- 
1. Acknowledgements: This research is part of the NEA project, funded by MINECO/FEDER UE, reference code FFI2015-64038-P and RAD (Researching Audio Description: Translation, Delivery and New Scenarios), reference code PGC2018-096566-B-I00 (MCIU/AEI/FEDER, UE). The authors are members of TransMedia Catalonia, a research group funded by the Catalan Government under the SGR funding scheme (2017SGR113). We would like to thank Anna Jankowska for her translation of the Polish recommendations.



Esta obra está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional.

analysed in terms of duration, pitch and amplitude. Then a perception test involving 60 participants (31 persons with sight loss and 29 sighted persons) is carried out, using both male and female voices. This research has provided new knowledge in this field and has shown how preferences correlate with certain prosodic values. The results provide food for thought for service providers when selecting the most suitable voices for audio description.

**Keywords:** Audio description; Prosody; Media Accessibility; Speech Analysis; Perception.

## Resumen

El objetivo de este estudio es analizar los rasgos prosódicos en un corpus de audiodescripciones en español para determinar las preferencias de usuarios con y sin pérdida de visión. El análisis supone una revisión de las recomendaciones sobre la locución que aparece en las guías. El corpus de análisis está constituido por 10 audiodescripciones de locutores profesionales masculinos y femeninos. Se ha analizado la duración, el tono y la amplitud que se emplean en los enunciados emitidos. Posteriormente, se ha llevado a cabo un test de percepción en el que han participado 60 sujetos (31 personas con pérdida de visión y 29 sin pérdida de visión). Esta investigación ha servido para establecer las preferencias de los hablantes relacionadas con los valores prosódicos. Los resultados pueden contribuir a una mejor selección de las voces de los audiodescriptores.

**Palabras clave:** Audiodescripción; Prosodia; Accesibilidad a los medios; Análisis de habla; Percepción.

## 1. Introduction

Audio description (AD) is an intersemiotic translation in which visuals are translated into words. The aim of an AD is that a person who cannot access the visuals can actually understand, enjoy and engage with an audiovisual content, thanks to this additional audio information. Research into AD has been developed within the field of audiovisual translation and media accessibility with descriptive and experimental approaches (for example, Villoslada, 2018, in Spanish). Topics that have been dealt with in the literature include the AD of diverse filmic, linguistic and cultural elements (Maszerowska, Matamala and Orero 2014), often relying on case studies but also with some corpus-based approaches (Salway 2007; Jiménez Hurtado & Seibel 2012;

Matamala to be published shortly). More recently experimental studies have approached the reception of diverging AD strategies (Mazur & Kruger 2012; Igareda & Matamala 2012) using different methodological tools, such as questionnaires, surveys or eye-tracking studies.

Despite this increasing research, investigations into voicing are still scarce. The main focus in this regard has been the comparison of human-voiced with text-to-speech AD (Szarkowska 2011; Fernández-Torné & Matamala 2015), the application of sound techniques (López, Kearney & Hofstädter 2016), and the reception by end-users of diverging strategies (Matamala et al. to be published shortly). Iglesias-Fernández, Martínez Martínez & Chica Núñez (2015) have also carried out a small-scale study in which they have shown how the congruence of the describer's voice with the scene together with the quality of the voice favour more positive assessments by users. However, to the best of our knowledge, there has been no thorough prosodic analysis of AD features despite its importance which has already been highlighted by Sánchez Mompeán (2018). Fryer (2016) also stresses the importance of prosody so that the describer and the description are perceived as trustworthy and authentic.

This article aims to describe the prosodic features of a corpus of Spanish AD. The corpus is made up of 10 AD in Spanish of the same short film, “What happens while—”, by Núria Nia. Although all 10 producers were given the same instructions –to produce an AD based on professional standards- the result is a corpus of diverging AD not only in terms of content selection (Matamala 2018) but also in terms of voicing. This article focuses on this last aspect by describing, first of all, the main prosodic features of each AD i.e. intonation groups, duration, pitch (average F0), and amplitude. Then, six of them (three male and three female) with diverging prosodic features are used to design a perception test with the aim of determining preferences by users with both normal sight and persons with sight loss (PSL).

The article begins with an overview of how standards, guidelines and handbooks approach voicing in AD. Section 2 describes the methodology and corpus used in the descriptive part of the study, as well as the results (acoustic analysis). Section 3 presents and reports on the results obtained in our second phase (perceptive analysis). The article, which aims to start to fill a gap in AD research, concludes with suggestions for future research.

## 2. Voicing in AD: an overview

Fryer (2016: 88) acknowledges that describers have been traditionally “encouraged to use a particularly neutral way of speaking” and “a neutral delivery has come to be recognized as ‘the norm’”. However, it is often advised to take into account the specific features of each production. Snyder (2014: 47) considers vocal skills to be one of the four fundamental elements of AD: “We make meaning with our voices”, he states, and adds that the “voicer’s delivery should be consonant with the nature of the material being described”. Generally speaking, there seem to be slightly different approaches to voicing which may depend on the tradition: Cabeza-Cáceres (2013) describes Spanish and German voicing as “uniform”, British AD voicing as “adapted” and USA AD voicing as “emphatic”. The first one is flat, the second one adapts the prosodic features to the original content, and the third one is more expressive. Cabeza-Cáceres compared these three styles with users and found that the choice of style does not affect comprehension. He also observed that there is no user agreement as far as enjoyment is concerned: the same number of users liked and rejected the uniform and emphatic intonations (Cabeza-Cáceres 2013: 331).

Our corpus of study is in Spanish, and the standard which governs its production is UNE 153020 (AENOR 2005). In terms of voicing, the standard indicates that the particular voices must be selected according to the types of voice needed (male, female, adult, young) and using the appropriate tone for each work. The standard recommends that voices must be clear and voicing must be neutral with appropriate intonation, rhythm and vocalization, without further clarification of what this exactly means.

It is also interesting to see different approaches in other standards, guidelines and recommendations regarding voicing. The international ISO (2005) standard indicates that the narrator should have good native language skills and the ability to articulate. It advises that the same style of narration should be followed consistently within the same content and recommends that the voice of the narrator complements the content being described. There is an acknowledgement that “often trained actors are employed as narrators and use their talent to infuse the description with appropriate emotive characteristics” (ISO 2015: 12). As regards AD styles of narration, it differentiates

between newsreader style (which relays information in a serious manner), commentator style (which provides entertainment), first person (taking a first person role) and third person.

At the European level, the ADLAB guidelines (Remael, Reviers & Vercauteren 2015: 57) explain that, in the AD process, the client will choose a voice talent with voice qualities that match the film's genre and style. They also acknowledge the lack of research about which voices match each film genre, but state that the choice is often based on the contrast with the voices in the dialogues or on the genre or style of the film. This approach is not always transferred into national guidelines or recommendations.

The Broadcasting Authority of Ireland guidelines (BAI 2012: 2) recommend using a neutral voice but also consider it important “to add emotion at different points in different films to suit the mood and the plot development. [...] The description should not, however, become a performance in its own right”. Neutral speech is also the recommendation in the Greek guidelines (Georgakopoulou 2008), although this neutrality seems to refer more to the language being used rather than the actual voicing. German guidelines by Benecke & Dosch (2004) devote a specific section to AD voicing. They explain that they prefer describers from within the author's production team, rather than outsiders, because the latter are prone to adding interpretations. Nevertheless, it is interesting to note how German guidelines also refer to a change of attitude: although they initially considered that the describer should remain neutral, it seems that humorous productions have forced them to revise this approach. This balance between neutrality and adaptation to the film is also suggested in the French guidelines by Morisset & Gonant (2008), who specifically indicate that the voice should be adjusted to the emotion of the scene and to the rhythm of the action but should maintain a certain neutrality. ITC guidance on standards for AD (2000: 8) recommends a “clear, pleasant and expressive voice” for the describers and indicate that PSL “tend to hold strong opinions about people's voices. If they do not like the voice, they will not listen”. They expand on the features of the AD delivery in the following terms: “Good audio description should be unobtrusive and neutral but not lifeless or monotonous and the delivery should be in keeping with the nature of the programme” (AD 2000: 10). In Poland the Polish National Broadcasting council guidelines do not

include any suggestions on voicing but the recommendations produced by Szymańska, & Strzymiński (2010) advise that the speaker's voice should correspond to the nature of the programme and should not distract users. They consider that good AD should be neutral and discreet but not monotonous and uniform. Excessive modulation is also to be avoided according to the recommendations by the Polish Foundation for Culture Without Barriers (Künstler et al. 2012).

The American Council of the Blind (2009), in its 3.1. updated version of the AD-ACB-ADP guidelines, stresses that meaning is communicated by voices through pronunciation, enunciation, breath control, volume, pause, inflection, pace, tempo, phrasing, and tone. These guidelines advocate speaking clearly and at a speed that can be understood, adapting vocal delivery to the nature of the material described. LARRS guidelines (n.d) also consider that the narrator's voice should match the product. Netflix (n.d.) is rather more specific and, in a section dedicated to voicing, gives advice on the speech rate (160 words per minute), audio mix (AD should be mixed to sound as though it was part of the original content), describer consistency (the same voice should be used for all episodes or movie sequels) and vocal approach. In this regard, the recommendation is that the "delivery of the description should match the volume, pace, tone and rhythm of the content", the voice should be distinguishable from other voices but should not be "distracting or animated in such a way as to disrupt the objectivity of the narration by becoming the voice of a performer". In the Canadian Described Video Best Practices (AMI & CAB 2013) there is a specific section on style and tone (delivery and narration), with no specific items or emphasis on neutrality, intonation or prosodic features. In fact, the approach focuses more on language use. The only reference to voicing states that, when a content includes a describer and a narrator, they should be easily differentiated. The other recommendations concern technical specifications of the recordings.

Finally, in Australia, Mikul (2010: 11) states the AD should not draw attention to themselves, and the describer "should blend seamlessly with the rest of the audio".

To sum up, there are slightly different approaches to voicing in AD but they all have in common indications that are often vague and open to

interpretation. Many of them seem to promote a neutral AD while advising to take into account the nature of the material. However, it is not totally clear what it is meant by this. There is no doubt that more research on prosodic features is needed using established measurements in phonetics, so that future guidelines or revisions of current guidelines can incorporate research results.

### 3. Describing the prosodic features of a corpus of AD

In the descriptive stage of our research a corpus of 10 AD in Spanish was analysed. The methodological aspects and results will now be discussed.

#### 3.1. Methodological aspects

The 10 professional AD were obtained from the Visuals Into Words (VIW) corpus (Matamala 2018), the only existing open-access corpus that allows to compare different audio descriptions on a single content. VIW aimed to develop a multimodal and multilingual corpus of AD using a single stimulus. The corpus is built upon a short film, *What happens while*. The film was especially created for the project by the Catalan film director Núria Nia in order to have copyright clearance (Matamala & Villegas 2016). The short film is 14 minutes long and portrays how different characters, namely a student, a businessman and a retiree, approach time. The film was originally shot in English, and it was then dubbed into Spanish and Catalan by the same dubbing actors in a professional studio in Barcelona. All the versions are available open access on the project website: <http://pagines.uab.cat/viw/>. Audio descriptions were commissioned from professionals in all three languages. They produced an .mp4 file containing the final mix, a time-coded script, as well as the sound files. Apart from the professional AD, some students also volunteered and provided written AD without a recorded version. The corpus has a total of 47 AD: 30 professional (10 per language) and 27 non-professional (10 in Spanish and 7 in Catalan).

The subcorpus used for our analysis is made up of the Spanish AD produced by professionals, as described in Table 1, which also indicates whether the voice talent was male or female. The corpus is made of 6,191 words spread across 480 AD units. Each AD unit has been divided into intonation

groups. AD units can be defined as the textual segments related to the visual representation. In other words, AD units are intersemiotic translations of the visuals into spoken words, and they are inserted in the gaps where there is generally no dialogue and no relevant music or sounds. An intonation group can be defined as the minimal fragment delimited by pauses or tonal changes, usually corresponding to punctuation marks. They are also known as intonational phrases (IP) in the field of Phonology (Cruttenden, 1997). In this sense, there are more intonation groups than AD units. The ratio shows the relationship between AD units and intonation groups and, therefore, whether the describers are following the punctuation marks available on the open access script to establish prosodic boundaries. The results show that describers deliver more intonation groups than AD units and do not really take punctuation into account to introduce prosodic boundaries (silent pauses and tonal inflections).

Service provider	Female/Male voice	AD units	Intonation groups	Ratio
Aptent	Female	67	150	0.44
Aristia	Female	57	84	0.67
CEIAF	Female	51	103	0.49
Edsol Producciones	Male	40	96	0.42
Ericsson	Male	45	92	0.49
Kaleidoscope	Female	42	112	0.37
Navarra de Cine	Male	39	108	0.36
SDI Media	Female	60	112	0.53
Soni2	Male	42	128	0.33
Trágora	Male	37	131	0.28

Table 1. Corpus features

The companies are representative of the AD service providers in Spain and have different profiles. There are Spanish companies which specialise in access services: Aptent Soluciones is a Madrid-based company created in



2011 which provides accessibility solutions with a strong technological component. Arístia is a Madrid-based company which has specialised in providing AD and audioguides since 1993, especially for the Spanish national blind organisation ONCE. CEIAF, created in 2002 and based in Seville, provides AD and subtitling for the deaf and hard-of-hearing, especially for the Spanish RTVE television service. Other companies, such as Trágora Traducciones, offer different types of translation service including AD. There are audiovisual production studios that also provide AD: this is the case with Edsol Producciones, based in Madrid, and Soni2, in Córdoba. There is a different profile with Kaleidoscope, a Granada NGO created in 2013, that promotes universal accessibility and specialises in museum and cultural heritage accessibility as well as training. Navarra de Cine is a company which promotes accessibility with an emphasis on film festivals. Two international players are SDI Media and Ericsson, with offices in Spain, that provide AD as well as many other local services.

In order to carry out the prosodic analysis, the 10 audio files were segmented into intonation groups. Figure 1 shows an example of speech analysis with PRAAT (Boersma & Weenink 2018). The AD units were divided into intonation groups and F0 values (Hz), average amplitude (dB) and duration (ms) were measured for each group. These parameters allowed us to identify differences among speakers.

F0 values measure pitch, which is considered to be an indicator of voice quality. Average amplitude is related to volume: the greater the *amplitude*, the greater the *amount* of energy carried by the wave and the more intense the sound will be. Intensity is perceived as the *loudness* of the sound. Duration is related to the length of the units and to the speech rate of the describers. The higher the average duration values in an intonation group, the lower the speech rate will be. Figure 1 shows the intensity with a dotted line and the F0 frequency (melodic contour) with a continuous line.

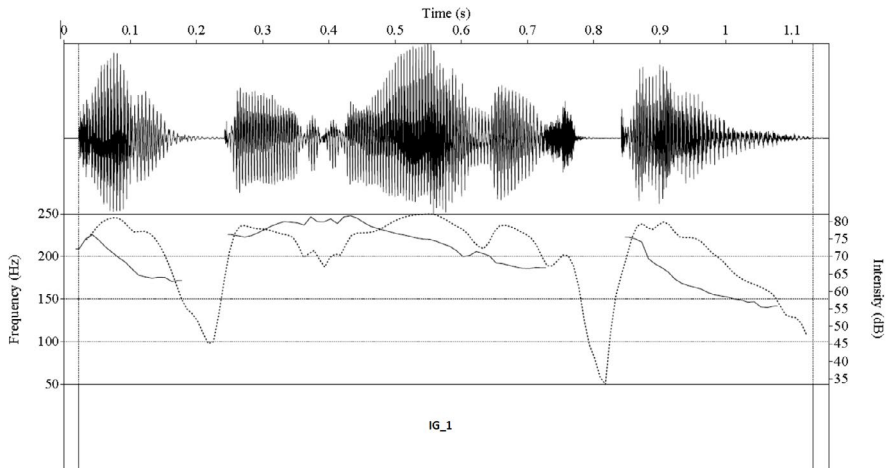


Figure 1. Example of speech analysis with the PRAAT program

### 3.2. Results

Although no instructions were given to service providers, 50% chose a female voice and 50% chose a male voice. The choice of a male or female voice is sometimes connected to the characters that appear on the scene. If there are more female characters, it seems that male voices are preferred, and viceversa, although in some countries the choice may also be related to the genre. In this particular 14-minute film there are two male characters and two female characters, as well as a female off-screen voice. This probably explains no clear preference for a female or male voice by the providers.

As far as prosodic features are concerned, Table 2 summarises the values obtained for each female describer, with average values and standard deviations for duration, pitch and amplitude.

Service provider	Duration		Pitch		Amplitude	
	IG (ms)	sd. (ms)	mean (Hz)	sd. (Hz)	Mean (dB)	sd. (dB)
Aristia	1736.95	764.31	135.85	7.2	67.4	0.9
Aptent	1590.84	894.20	200.12	10	75.4	1.4
CEIAF	1814.93	1057.34	191.41	14.12	59.8	2.6
Kalidoscope	1052.72	486.8	164.21	19.7	58.8	3.0
SDI	1616.58	868.7	171.37	15.6	66.65	6.8

Table 2. Prosodic parameters of female describers: mean values and deviations

When the average F0 values, amplitude and duration are compared, one can observe that Aristia's describer has the lowest pitch (F0) and a high volume. Her speech rate compared to the rest of the group is slightly higher (11.9% above the mean value, 1552 ms of all the female describers). Aptent's describer has the highest pitch, the highest volume and her mean duration value is quite close to the female average (1552 ms). Kaleidoscope's describer has a low pitch and also has the lowest voice volume and lowest duration (hence, the highest speech rate). CEIAF's describer shows a high pitch with a low voice volume and the highest duration. Finally, the SDI describer has duration, pitch and amplitude values very close to the average of each parameter (1552 ms, 172 Hz and 65 dB, respectively). Based on these features, voices which were further from the average in terms of duration, pitch and amplitude were chosen in designing the perception test.

As far as the male describers are concerned, Table 3 shows the results obtained in the prosodic analysis.

Service provider	Duration		Pitch		Amplitude	
	IG (ms)	sd.(ms)	mean (Hz)	sd. (Hz)	Mean (dB)	sd. (dB)
Ericsson	1801.97	1040	101.71	14.30	64.27	2.23
Soni2	1230.17	688	121.33	8.44	69.82	2.42
EdSol	1363.66	627	89.29	7.22	65.24	1.78
Navarra	1324.26	569.7	101.03	15.11	65.36	2.76
Trágora	1974.62	1182	100.04	8.93	65.31	1.32

Tabla 3. Prosodic parameters of male describers: mean values and desviations

The average duration of intonation groups shows clear differences between two groups of describers: on the one hand, those with a higher duration are the Ericsson and Trágora describers, which implies a slower speech rate; on the other hand, Soni2, Edsol and Navarra's describers, with an average lower duration, read the audio description scripts faster. Average F0 values show that the Edsol describer has the lowest-pitch voice while the highest-pitch voice belongs to the Soni2 describer. The other describers have similar values. As regards amplitude values, there are almost no differences among the male describers, although the Soni2 shows the highest volume and Ericsson has the lowest volume. There does not seem to be a correlation between the type of company and the type of voicing. According to these characteristics, voices which are further from the average in terms of duration, pitch and amplitude were chosen for the perception test.

#### 4. Perception research

The second step was a perception test with users aiming to elicit user preferences in terms of voices. Section 4.1. presents the methodological design and 4.2. reports on user feedback.

##### 4.1. Methodological aspects

The perception test followed a procedure approved by UAB's ethical committee and lasted approximately 30 minutes. The perception test was completed

by 62 participants. Two male participants were excluded to avoid a possible gender bias. Finally, data from all 60 female participants were analyzed. In order to determine the effect of sight, they were divided into two groups: 29 sighted participants, 31 persons with sight loss. Sighted participants were contacted through email and PSL were contacted through a user association. They were asked to give their informed consent and were asked whether they were sighted or PSL.

18 stimuli were created for the test. Each of them was composed by two voices. 3 males and 3 females had been selected. The voices were chosen using the descriptive analysis outlined above (Section 2). The three voices that offered the greatest contrasts from a prosodic point of view were selected: Aptent, Aristia and Kaleidoscope for female voices, and Edsol Producciones, Ericsson and Soni2 for male voices. Female and male voices were not mixed in the experiment. Table 4 summarises the features of the female voices selected, and Table 5 summarises the features of the male voices selected. 1,080 answers were analysed, 60 for each comparison (60\*3 female comparison + 60\*3 male comparison).

Data collection was performed online. Participants responded to the test using an online form. They had to choose in each pair the voice they preferred. Answers were collected in an Excel file for further statistical analysis. Data were submitted to a SPSS software program (25 v.). Chi-square tests were performed. In all these tests, the independent variable was the answers and the dependent variable was the group (sighted vs. PSL).

Each sample contained a comparable segment of AD. Although it could be argued that AD must be assessed in the context of an audiovisual product, a basic experimental approach was prioritised and samples were shown in isolation. Background knowledge is needed before designing more complex experiments in ecologically valid settings, which will be a necessary follow-up of this investigation.

	Speech rate	Pitch	Amplitude
Speaker_1 (Aptent)	+	+	+
Speaker_2 (Aristia)	-	-	+
Speaker_3 (Kalidoskope)	+	-	-

Table 4. Female voices: prosodic features

Speaker 1 has a higher speech rate and a higher volume and pitch. Speaker 2 is characterised by her volume, and Speaker 3 features a higher speech rate.

	Speech rate	Pitch	Amplitude
Speaker_1 (Edsol)	+/-	-	+/-
Speaker_2 (Ericsson)	-	+	-
Speaker_3 (Soni2)	+	+	+

Table 5. Male voices: prosodic features

To select male describers, the pitch was the key feature, as there are no big differences in amplitude values (see Section 2.2). Speaker 1, Edsol's describer, has the lowest-pitch voice and an average speech rate. Ericsson's voice talent has a high pitch, a low volume and a low speech rate. Finally, Speaker 3 (Soni2) has the highest pitched voice as well as a higher amplitude (and, therefore, volume) and a higher speech rate.

#### 4.2. Perception Results

As regards female voices, the results show differences only when Speaker 2 is compared to Speaker 3 (see Table 6). In spite of the percentage differences, both types of informants (sighted users and PSL) prefer Speaker 2. The same results are obtained when Speaker 1 is compared to Speaker 3. Both types of users do not like Speaker 3 (Figure 2). Speaker 3 was characterised by a low pitch, the lowest voice volume and the highest speech rate.



Figure 2. Speaker 3 (right column always) in comparison to the other two speakers

Comparing sighted users with PSL, there are no statistical differences between speakers 1 and 2 (see Table 6). It seems that sighted users find Speaker 1 more

pleasant, and PSL prefer Speaker 2 (Figure 3). The data seem to indicate a preference for a voice featuring no high pitch or speech rate among visually impaired users and a preference for voices with higher prosodic values for sighted users.

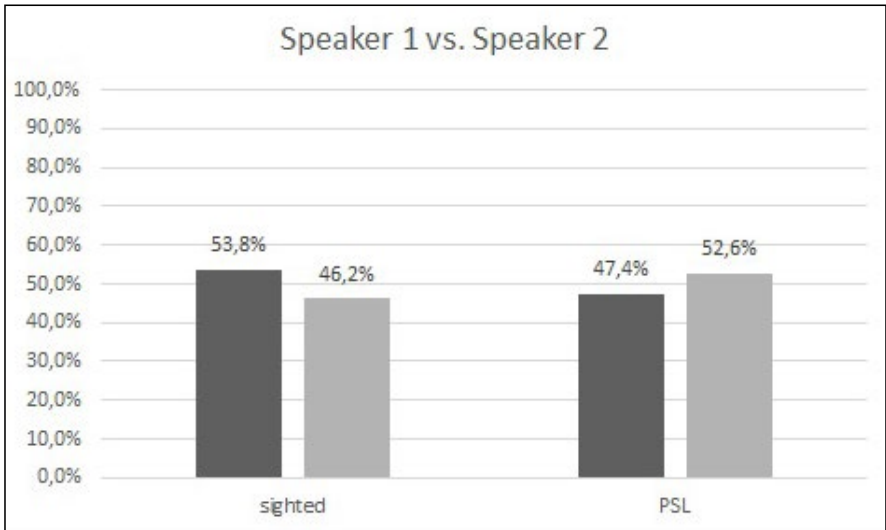


Figure 3. Speaker 1 (left column) compared to Speaker 2 (right column)

Comparison	Chi2
Speaker 1 vs Speaker 3	p=0.07
Speaker 2 vs Speaker 3	p=0.03***
Speaker 1 vs Speaker 2	p=0.3

Table 6. Significance Level (users\*speakers) in the comparison of female voices

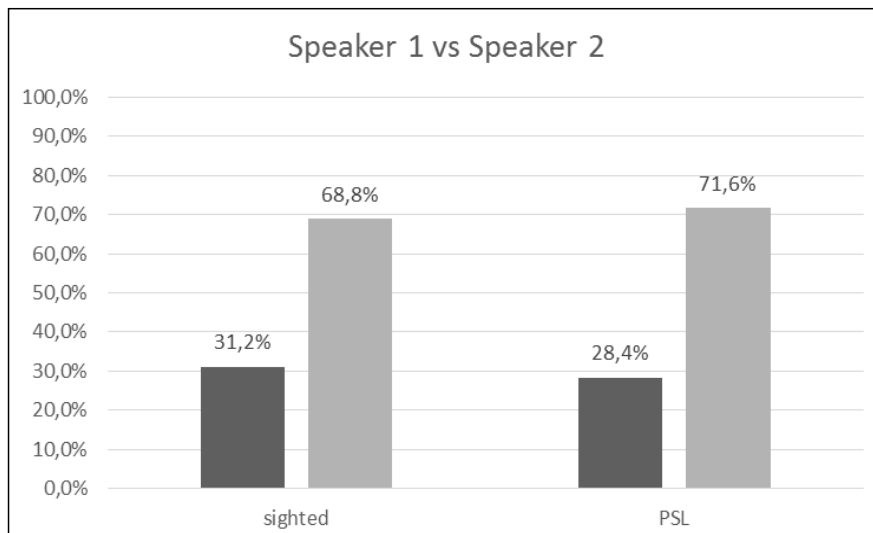
As regards male voices, there are no differences between sighted participants and PSL when selecting the voice that they find more pleasant (Table 7). They both prefer Speaker 2 and do not like Speakers 1 and 3. This response can be



observed in Figure 6. A higher pitch with the lowest volume and a low speech rate (actually the lowest in the male speakers selected for the perception test) characterised the voice preferred by users. Maximum prosodic values are rejected in a male describer.

Comparison	Chi2
Speaker 1 vs Speaker 3	p=0.6
Speaker 2 vs Speaker 3	p=0.2
Speaker 1 vs Speaker 2	p=0.7

Table 7. Comparison of male voices



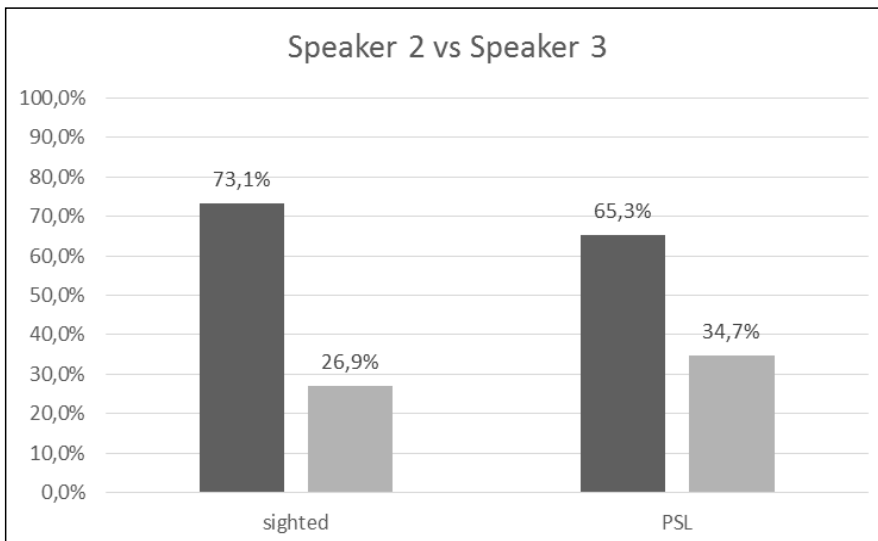
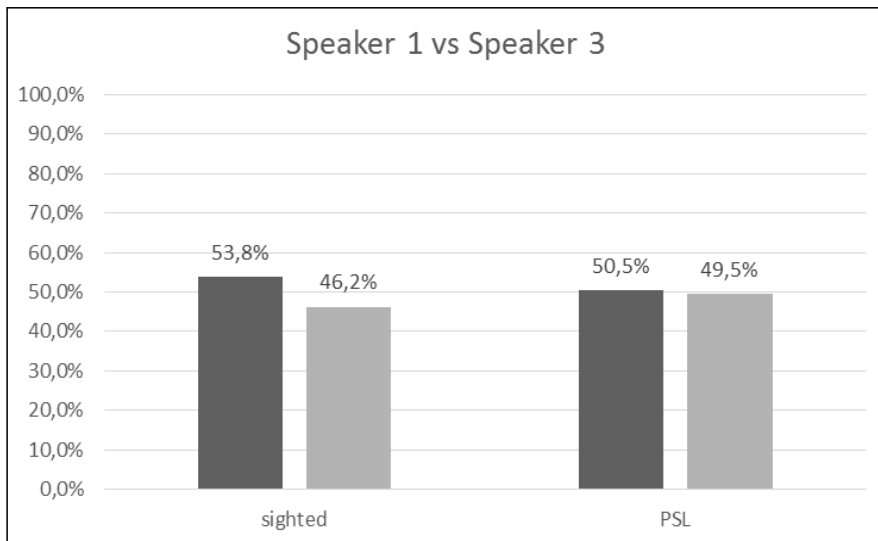


Figure 4. Percentages of pleasantness ratings related to the type of users (sighted and PLS listening to each speaker)

## 5. Conclusions

The way in which an AD is delivered is as important as the way in which it is written. Recommendations, guidelines and handbooks such as those mentioned in section 2 acknowledge this fact (Snyder 2014, ISO 2015, Fryer 2016). However, the advice given is often vague. We need to use linguistic tools to analyse prosodic values if we want to go beyond impressionistic suggestions and make research-based recommendations. This study is just a first step on a topic that merits more in-depth research: prosody in AD.

Using a corpus analysis our investigation has shown the prosodic values of both male and female professionals describing a short film in Spanish, after having received exactly the same instructions. Different approaches to voice selection have been found and analysed. In future analyses it would be helpful to gather additional qualitative data in order to have a better understanding of the choice of voices by service providers, an aspect which was not tackled in this paper.

In addition to the corpus analysis, the article reports on a perception test in which users indicated their preferences between both male and female voices. The comparison was carried out within each gender, as our interest lies in the prosodic features of the voices and not in gender aspects. Both male and female voices are used for audio descriptions, depending on various factors and the assessment of user preferences for both types is relevant from a research perspective. In this regard, users seem to reject female voices with low pitch, low volume and a high speech rate in Spanish. As for the preferences, differences between sighted participants and PSL were found: PSL preferred the voice with the lowest pitch, a low speech rate and a high volume, whereas sighted participants preferred the voice with the highest pitch, the highest volume and a slightly higher than average speech rate. As far as male voices are concerned, there were no differences based on the sight of the participants. They all seemed to prefer a voice with a high pitch, the lowest volume and a low speech rate and they rejected maximum prosodic values.

This research has provided new knowledge in this field and has shown how preferences correlate with certain prosodic values. However, since it is the first research of this nature to the best of our knowledge, it also has some limitations. Further research with full audio description excerpts, in which

the describer's voice is combined with the audio visual content voices, is needed in order to have a better understanding of the relationship between the different soundtracks in audio description productions. It would also be worthwhile to include familiarity as a key aspect in user preferences: in this regard, it is not unusual for users to complain when the voice they usually hear on audio description productions changes.

Another interesting issue that arises in the bibliography which needs further analysis is neutrality. The term "neutral" is often found to refer to the prosodic features of a voice, but there is still no clear definition of what it means. A definition of a "neutral voice" in terms of pitch, volume and amplitude is needed in order to have a better understanding of its meaning and produce useful guidelines and recommendations.

To sum up, prosody has been a forgotten aspect in AD research but one that merits further research. This will only be possible through the collaboration of media accessibility experts and phoneticians, which has been a feature of this investigation.

## References

- AENOR. (2005) *Norma UNE:153030. Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. Madrid: AENOR.
- Accessible Media Inc (AMI) & The Canadian Association of Broadcasters (CAB). (2013) *Described Video Best Practices. Artistic and Technical Guidelines*. Electronic version: <https://ecfsapi.fcc.gov/file/7520940294.pdf>
- American Council of the Blind. (2009) *Audio description standards*. Electronic version: <[http://www.acb.org/adp/docs/ADP\\_Standards.doc](http://www.acb.org/adp/docs/ADP_Standards.doc)>
- BAI. (2012) *BAI guidelines – audiodescription*. Electronic version: <[https://www.bai.ie/en/media/sites/2/dlm\\_uploads/2018/05/20171220\\_AccessRules\\_MinisterReport\\_vFinal\\_DMCL.pdf](https://www.bai.ie/en/media/sites/2/dlm_uploads/2018/05/20171220_AccessRules_MinisterReport_vFinal_DMCL.pdf)>
- BENECKE, Bernd & Elmar Dosch. (2004) *Wenn aus Bildern Worte werden*. Munich: Bayerischer Rundfunk.
- BOERSMA, Paul & David Weenik. (2018) *Praat: doing phonetics by computer* [Computer program]. Electronic version: <<http://www.praat.org>>
- CABEZA-CÁCERES, Cristóbal. (2013) *Audiodescripció i recepció. Efecte de la velocitat de narració, l'entonació i l'explicitació en la comprensió filmica*. Barcelona:

- Universitat Autònoma de Barcelona. Tesis doctoral inedita. Electronic version: <<https://www.tdx.cat/handle/10803/113556>>
- CRUTTENDEN, Alan (1997) *Intonation* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166973.
- FERNÁNDEZ-TORNÉ, Anna & Anna Matamala. (2015) “Text-to-speech versus human-voiced audio descriptions: a reception study in films dubbed into Catalan.” *JosTrans. Journal of Specialised Translation* 24, pp. 61-88.
- FRYER, Louise. (2016) *An introduction to audio description*. London: Routledge.
- GEORGAKOPOULOU, Yota. (2008) “Audio Description Guidelines for Greek- A Working Document.” In: Rai, Sonali; Joan Greening & Leen Petré (eds.) 2008. *A Comparative Study of Audio Description Guidelines Prevalent in Different Countries*. London: RNIB.
- MORISSET Laure & Frédéric Gonant. (2008) *La charte de l’audiodescription*. Paris: Ministère des Affaires Sociales et de la Santé. Electronic version: <<https://www.sdcine.fr/wp-content/uploads/2015/05/Charte-de-laudio-description-1008.pdf>>
- IGAREDA Paula & Anna Matamala. (2012) “Variations on the Pear Tree Experiment: different variables, new results?” *Perspectives. Studies in Translatology* 20:1, pp. 103-123.
- ITC. (2000) *ITC Guidance on Standards for Audio Description*. Electronic version: <[http://audiodescription.co.uk/uploads/general/itcguide\\_sds\\_audio\\_desc\\_word3.pdf](http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf)>
- IGLESIAS FERNÁNDEZ, Emilia; Silvia Martínez Martínez & Antonio Javier Chica Núñez. (2015) “Cross-fertilization between reception studies in audio description and interpreting quality assessment: the role of the describer’s voice.” In: Baños Piñero, Rocío & Jorge Díaz-Cintas (eds.) 2015. *Audiovisual translation in a global context*. London: Palgrave Macmillan, pp. 72-95.
- ISO. (2015) *Information technology - User interface component accessibility — Part 21: Guidance on audio descriptions* (ISO/IEC TS 20071-21).
- JIMÉNEZ HURTADO, Catalina & Claudia Seibel. (2012) “Multisemiotic and multimodal corpus analysis in audio description: TRACCE.” In: Remael, Aline; Pilar Orero & Mary Carroll (eds.) 2012. *Audiovisual translation and media accessibility in the crossroads*. Amsterdam: Rodopi, pp. 409-425.
- KÜNSTLER, Izabela; Urszula Butkiewicz; Robert Więckowski & Anna Żórawska. (2012) *Audiodeskrypcja – zasady tworzenia* [Audio description

- rules for creation]. Electronic version: <<http://dzieciom.pl/wp-content/uploads/2012/09/Audiodeskrypcja-zasady-tworzenia.pdf>>
- LARRS (n.d.) *Guidelines and standards for LARRS audio describers*. Electronic version: <<http://www.larrs.org/guidelines.html>>
- LÓPEZ, Mariana; Gavin Kearney & Krisztián Hofstädter. (2016) “Enhancing audio description: sound design, spatialisation and accessibility in film and television.” *Proceedings of the Institute of Acoustics* 38:2. Electronic version: <[http://enhancingaudiodescription.com/wp-content/uploads/2016/11/RS2016-paper-Lopez-et-al\\_accessible.pdf](http://enhancingaudiodescription.com/wp-content/uploads/2016/11/RS2016-paper-Lopez-et-al_accessible.pdf)>
- MASZEROWSKA, Anna; Anna Matamala & Pilar Orero (eds.) (2014) *Audio description: New perspectives illustrated*. Amsterdam: John Benjamins.
- MATAMALA, Anna. (2018) “One short film, different audio descriptions. Analysing the Language of audio descriptions created by students and professionals.” *Onomazéin* 41, pp. 185-207.
- MATAMALA, Anna. (forthcoming) “The VIW corpus: Multimodal corpus linguistics for audio description analysis.” *Revista española de lingüística aplicada*.
- MATAMALA, Anna; Olga Soler-Vilageliu; Gonzalo Iturregui-Gallardo; Anna Jankowska; Jorge Luis Méndez-Ulrich & Anna Serrano. (forthcoming) “Electrodermal activity as a measure of emotions in media accessibility research: methodological considerations.” *Journal of Specialised Translation*.
- MATAMALA, Anna & Marta Villegas. (2016) “Building an audio description multilingual multimodal corpus: the VIW project.” In: Edlund, Jens; Dirk Heylen & Patrizia Paggio (eds.) 2016. *Proceedings: Multimodal Corpora: Computer vision and Language processing- MMC2016*. Paris: ELRA, pp. 29-32. Electronic version: <<http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-MCC-2016-proceedings.pdf>>
- MAZUR, Iwona & Jan-Louis Kruger. (2012) “Pear Stories and audio description: language, perception and cognition across cultures.” *Perspectives. Studies in Translation Theory and Practice* 20:1, pp. 1-3.
- MIKUL, Chris. (2010) *Audio Description Background Paper*. Electronic version: <<http://www.mediaaccess.org.au/sites/default/files/MAA%20-%20Audio%20Description%20Background%20Paper.doc>>
- Netflix (n.d.) *Netflix audio description style guide v1.0*. Electronic version: <<https://partnerhelp.netflixstudios.com/hc/en-us/articles/215510667-Audio-Description-Style-Guide-v1-0>>

- REMAEL, Aline; Nina Reviere & Gert Vercauteren. (2015) *Pictures painted in Words. ADLAB Audio Description guidelines*. Trieste: Edizione Università di Trieste.
- SALWAY, Andrew. (2007) "A corpus-based analysis of audio description." In: Díaz-Cintas, Jorge; Pilar Orero & Aline Remael (eds.) 2007. *Media for All. Subtitling for the Deaf, Audio Description, and Sign Language*. Amsterdam: Rodopi, pp. 151-174.
- SÁNCHEZ MOMPEÁN, Sofía. (2018) "Towards a quality model for paralinguistic in dubbed and audiodescribed dialogue." Presentation at UMAQ conference. Barcelona, 4-5 June 2018. Electronic version: <[https://www.academia.edu/37084299/UMAQ\\_International\\_Conference\\_Understanding\\_Quality\\_in\\_Media\\_Accessibility\\_](https://www.academia.edu/37084299/UMAQ_International_Conference_Understanding_Quality_in_Media_Accessibility_)>
- SNYDER, Joel. (2014) *The visual made verbal. A comprehensive training manual and guide to the history and applications of audio description*. Arlington, VA: American Council for the Blind.
- SZARKOWSKA, Agnieszka. (2011) "Text-to-speech audio description: towards wider availability of audio description." *JosTrans. Journal of Specialised Translation* 15, pp. 142-162.
- SZYMAŃSKA, Barbara & Tomasz Strzymiński. (2010) *Standardy tworzenia audiodeskrypcji do produkcji audiowizualnych* [Standards for creating audio description to audiovisual productions]. Electronic version: <<http://www.audiodeskrypcja.org.pl/standardy-tworzenia-audiodeskrypcji.html>>
- VILLOSLADA, Ana. (2018) "La lengua de la audiodescripción. Análisis del guion audiodescrito de Ocho apellidos vascos". *TRANS: revista de traductología*, 22, pp. 61-80.

## BIONOTE / BIONOTA

MARÍA J. MACHUCA received a PhD in Spanish Philology from the Universitat Autònoma de Barcelona (UAB), and she was appointed full-time lecturer in Spanish Language at the same university. Her research, publications and participation in national and international projects deal with the application of experimental phonetics to different research areas, such as Forensic Phonetics, speech technology, speaking styles, and foreign language acquisition. Her publications can be retrieved from: <http://files.cat.uab.cat/files/ maria-jesus-machuca-ayuso/>

MARÍA J. MACHUCA es doctora en Filología Hispánica por la Universidad Autónoma de Barcelona y profesora agregada en el Departamento de Filología Española. Sus investigaciones, publicaciones y proyectos en los que ha participado están relacionados con la aplicación de la fonética experimental en diferentes áreas, como la Fonética Judicial, la tecnología del habla, los estilos de habla y la adquisición de lenguas extranjeras. Sus publicaciones pueden consultarse en <http://files.cat.uab.cat/files/maria-jesus-machuca-ayuso/>

ANNA MATAMALA, BA in Translation (UAB) and PhD in Applied Linguistics (UPF), is an Associate Professor at UAB (Barcelona). Currently leading TransMedia Catalonia, she has participated and led projects on audiovisual translation and media accessibility. She has taken an active role in the organisation of scientific events (M4ALL, ARSAD), and has published in journals such as *Meta*, *Translator*, *Perspectives*, *Babel* or *Translation Studies*. She is currently involved in standardisation work. [gent.uab.cat/amamatamala](http://gent.uab.cat/amamatamala)

ANNA MATAMALA es licenciada en Traducción e Interpretación por la Universitat Autònoma de Barcelona (UAB) y doctora en Lingüística Aplicada por la Universitat Pompeu i Fabra. Es profesora titular en la UAB. Es directora del grupo TransMedia Catalonia. Ha participado y coordinado proyectos en traducción audiovisual y en accesibilidad a los medios de comunicación audiovisuales. También ha colaborado en la organización de congresos científicos (M4ALL, ARSAD) y publicado en revistas científicas, tales como *Meta*, *Translator*, *Perspectives*, *Babel* o *Translation Studies*. Actualmente, colabora en proyectos de estandarización: [gent.uab.cat/amamatamala](http://gent.uab.cat/amamatamala)

ANTONIO RÍOS MESTRE received his PhD degree in Spanish Phonology from UAB. He holds a position of Associate Professor at the Department of Spanish Studies. He teaches undergraduate courses in Translation, Oral Expression, Phonetics, and Phonology. His research focuses mainly on the studies related to oral language, particularly in the prosodic and segmental aspects.

ANTONIO RÍOS MESTRE es doctor en Filología Hispánica por la Universidad Autónoma de Barcelona y profesor agregado en el Departamento de Filología Española de dicha universidad, donde imparte clases de lengua española



aplicada a la traducción, de fonética y fonología y de expresión oral. Su labor investigadora se centra, fundamentalmente, en el estudio de la lengua oral, en particular en los aspectos fonéticos prosódicos y segmentales y en fonética judicial.