Artículos

# La agradabilidad de las voces de los audiodescriptores: estudio acústico y perceptivo

# The pleasantness of audio describers' voices: acoustic and perceptual study

# Agréabilité des voix d'audio-descripteurs: une étude acoustique et perceptive

Machuca Ayuso, María Jesús; Ríos, Antonio

**María Jesús Machuca Ayuso**
mariajesus.machuca@uab.cat
Universitat Autònoma de Barcelona, Perú

**Antonio Ríos**
Antonio.Rios@uab.cat
Universitat Autònoma de Barcelona, Perú

**Boletín de la Academia Peruana de la Lengua**
Academia Peruana de la Lengua, Perú
ISSN: 0567-6002
ISSN-e: 2708-2644
Periodicidad: Semestral
vol. 71, núm. 71, 2022
boletin@apl.org.pe

Recepción: 20 Abril 2021
Aprobación: 11 Febrero 2022
Publicación: 15 Junio 2022

**URL:** http://portal.amelica.org/ameli/journal/497/4973443009/

**DOI:** https://doi.org/10.46744/bapl.202201.009

Esta obra está bajo una Licencia Creative Commons Atribución 4.0 Internacional.

**Resumen:** Las recomendaciones de las guías indican que la voz de las audiodescripciones no debe transmitir emociones para evitar, así, influir sobre la interpretación de una obra. El objetivo de este trabajo es identificar las características de la voz de diferentes audiodescriptores para determinar qué voces perciben los usuarios como más agradables. La caracterización de una voz agradable puede contribuir a la mejora de este servicio evitando la fatiga de aquellas personas, generalmente con problemas de visión, que necesitan de dicho servicio para compensar la falta de información visual. Partiendo de un corpus de audiodescripción (uab.cat/viw), se analizan el tono, la velocidad de elocución y el volumen de voz de 10 audiodescriptores, hombres y mujeres, en tres lenguas: español, catalán e inglés. Posteriormente, las voces son evaluadas por un total de 208 sujetos, con y sin dificultad visual, nativos de estas lenguas. Aunque partíamos de la hipótesis de que las características que definen una voz agradable en una lengua sirven también para otra lengua, aunque con diferencias en función de los problemas de visión de los usuarios, los resultados muestran que la percepción de una voz agradable no depende tanto de la discapacidad visual, sino más bien de la lengua nativa en la que se produce la audiodescripción.

**Palabras clave:** audiodescripción, voz, prosodia, agradabilidad, discapacidad visual.

**Abstract:** The recommendations of the guidelines indicate that the voice of the audio descriptions should not convey emotions to avoid influencing the interpretation of a work. The aim of this work is to identify the characteristics of the voice of different audio describers to determine which voices users perceive as more pleasant. The characterization of a pleasant voice can contribute to the improvement of this service, avoiding the fatigue of those people, generally visually impaired, who need this service to compensate for the lack of visual information. Starting from a corpus of audio description (uab. cat/viw), tone, speed of elocution and voice volume of 10 male and female audio describers are analyzed in three languages: Spanish, Catalan, and English. Subsequently, the voices are evaluated by a total of 208 subjects, with and without visual impairment, native speakers of these languages. Although we started from the

hypothesis that the characteristics that define a pleasant voice in one language also apply to another language, although with differences depending on the vision problems of the users, the results show that the perception of a pleasant voice does not depend so much on the visual impairment, but rather on the native language in which the audio description is produced.

**Keywords:** audio description, voice, prosody, pleasantness, visual impairment.

**Résumé:** Les guides recommandent que la voix des audiodescriptions ne transmette pas d'émotions, pour éviter ainsi d'influencer l'interprétation d'une œuvre. L'objectif du présent travail est d'identifier les caractéristiques de la voix de divers audio-descripteurs, afin d'établir quelles voix sont plus agréables d'après la perception des usagers. La caractérisation d'une voix agréable peut contribuer à l'amélioration de ce service, en évitant la fatigue des usagers, en général des malvoyants qui ont besoin de ce service pour compenser le manque d'information visuelle. En partant d'un corpus d'audio-descriptions (uab.cat/viw), nous avons analysé le ton, le débit et l'intensité de la voix de 10 audio-descripteurs, hommes et femmes, en trois langues : espagnol, catalan et anglais. Postérieurement les voix ont été évaluées par 208 sujets, avec et sans difficultés visuels, natifs de ces langues. Alors que nous partions de l'hypothèse que les caractéristiques que définissent une voix agréable dans une langue, servent aussi pour d'autres langues, même avec des différences obéissant aux problèmes de vision des usagers, nos résultats montrent que la perception d'une voix agréable ne dépend pas tellement de la malvoyance, mais plutôt de la langue de l'audio-description.

**Mots clés:** audio-description, voix, prosodie, agréabilité, malvoyance.

## 1. Introducción

La audiodescripción es una modalidad de la traducción intersemiótica que convierte un contenido visual en palabras. Su objetivo es permitir que aquellas personas con pérdida de visión puedan acceder a un contenido audiovisual gracias a esa información sonora adicional. Las recomendaciones que se han propuesto para las audiodescripciones consideran que la locución ha de ser neutra (Iglesias *et al*., 2015; Fryer, 2016, entre otros). Esta se recoge también en los estándares del español, en la norma *UNE 153020* (Comité Técnico de Normalización 153, 2005).

A un audiodescriptor se le pide una pronunciación correcta de su lengua, claridad en la emisión oral y determinadas destrezas articulatorias, como son el control del volumen de voz, un uso adecuado de las pausas, de las inflexiones y del ritmo, y una velocidad de habla que permita la comprensión de los enunciados (véase, por ejemplo, American Council of the Blind, 2009). Todas estas recomendaciones hacen referencia al modo de realizar la audiodescripción, pero no sabemos cómo han de ser las características intrínsecas de la voz de los audiodescriptores (hombres y mujeres) para que resulten agradables a los usuarios.

La agradabilidad de la voz ha sido un tema de interés en diversos campos, como la interpretación (Iglesias, 2007), los medios audiovisuales (Rodero, 2002) y la síntesis del habla (Coelho *et al*. 2011), pero no es un concepto fácil de definir desde un punto de vista objetivo, pues no deja de estar sujeto al gusto particular de cada persona. Un acercamiento a la definición de agradabilidad de la voz de los audiodescriptores,

relacionando los parámetros acústicos con los perceptivos, puede contribuir a la mejora de los productos audiovisuales. A pesar de la inevitable subjetividad en el juicio de los usuarios de la audiodescripción, es posible —y es, por tanto, nuestro objetivo— cuantificar los valores de los parámetros prosódicos que definen la voz de los audiodescriptores considerados más agradables y observar determinadas tendencias generales en la elección.

El objetivo de este trabajo es analizar la agradabilidad de las voces de los audiodescriptores en diferentes lenguas. Se estudia la caracterización de sus voces en relación con los parámetros acústicos de tono, velocidad de habla y volumen de voz para determinar cuáles son los indicios acústicos que corresponden a la percepción de una voz agradable y precisar si los resultados dependen del sexo del audiodescriptor y si diferentes sujetos, personas con y sin discapacidad visual, perciben la agradabilidad de igual forma. Las hipótesis de las que parte este trabajo son las siguientes:

1. En la audiodescripción, se consideran más agradables las voces masculinas y femeninas con un tono grave.
2. La preferencia por una voz agradable será diferente en función de la lengua materna de los usuarios de la audiodescripción.
3. La preferencia por una voz agradable será diferente en receptores con y sin capacidad visual.

Este artículo muestra, en primer lugar, una visión general de las opiniones sobre lo que se considera una voz agradable, un tema poco estudiado en el campo de la fonética experimental e inédito, a nuestro saber, en el campo de la audiodescripción. A continuación, se describe la metodología de nuestro experimento, que consta de dos partes: un análisis acústico, que permite caracterizar la prosodia de las voces de los audiodescriptores que conforman el corpus de trabajo, y un test perceptivo en el que los participantes en el experimento valoraron las voces que se seleccionaron a partir de los datos acústicos. Finalmente, se discuten los resultados obtenidos en el test de percepción y se muestran las conclusiones del trabajo.

## 2. Marco conceptual

### 2.1. El estudio de la voz desde distintas disciplinas

El tono de voz producido por una persona cuando habla lo caracteriza e individualiza frente a los otros hablantes (Keller, 2005). En los estudios fonéticos se han propuesto diversas categorizaciones de los diferentes tipos de calidad de voz. Se pueden citar, entre otras, las de Ladefoged (1971) y Laver (1980), pero en ninguna de esas clasificaciones se han relacionado las descripciones de la voz con la percepción que los oyentes pueden tener sobre la agradabilidad de los tipos de voz. Este hecho puede deberse a que el concepto de agradabilidad no está exento de subjetividad en su definición.

La agradabilidad de la voz es uno de los parámetros que se tienen en cuenta en la evaluación de la calidad de la interpretación. Iglesias (2006) señala que no existen criterios objetivos medibles para definir dicha agradabilidad a pesar de que una voz agradable se considera un requisito relevante en la valoración de los intérpretes, pues, unida a una prosodia acertada, puede contribuir a una mejor transmisión del mensaje. La definición de voz agradable no es clara, y se ha basado en la entonación, la fluidez, el tempo y la dicción (Iglesias, 2013). No obstante, se ha demostrado que un tono y un volumen de voz muy altos son juzgados negativamente por parte de los usuarios e influyen en el reconocimiento de la calidad de las prestaciones de los intérpretes (Iglesias, 2006).

En el ámbito de los profesionales de la comunicación en los medios orales, pueden encontrarse artículos relacionados con este concepto, pues, como afirma Rodríguez (1989), es el medio por excelencia para analizar la expresividad oral. Este mismo autor ya señalaba que era necesario precisar una teoría sobre los matices sonoros de la voz para dominar el funcionamiento expresivo de esta en aquellos procesos comunicativos en los que se verbaliza un texto (p. 13). Los resultados de sus experimentos muestran que una voz radiogénica

puede definirse como una voz agradable, armónica, relajada, cálida y transparente, pero las conclusiones a las que llega están más relacionadas con la imagen que se crea el oyente del locutor que con las características de una voz agradable, aunque señala que los tonos más graves tienden a ser mejor aceptados (p. 258). Rodero (2002) define también la voz radiogénica como la más agradable para los oyentes y la que utiliza todos sus recursos expresivos de acuerdo con el contenido del discurso informativo. Según esta autora, esta voz debe tender a registros graves, aunque en un estudio experimental anterior (Rodero, 2001) señala que, si bien la voz agradable en hombre se corresponde con una voz grave, no ocurre lo mismo con las voces femeninas, en las que un tono grave-medio es lo más adecuado. Y que, tanto para hombres como para mujeres, las voces agudas se rechazan como voces agradables. Un tono grave no es el único requisito para obtener una buena locución radiofónica, pues también se ha de controlar la intensidad de la voz, que no ha de ser muy elevada, y la velocidad de habla, ni muy lenta ni muy rápida, para que los oyentes asimilen el mensaje que se transmite (Rodero, 2003).

La agradabilidad es un aspecto que preocupa también en la creación de voces sintéticas, pues el usuario debe percibir una voz agradable cuando interactúa con aplicaciones en las que se emplea un sistema de síntesis de voz. En este sentido, Coelho *et al.* (2011) han demostrado que las evaluaciones de las voces sintéticas en las que se han considerado estos aspectos de calidad de la voz indican una mejora del 7 % para las voces femeninas y del 3 % para las masculinas respecto a los sistemas que no los consideran. La mejora de las voces sintéticas constituye un reto realmente importante para usuarios que tienen problemas de visión; no obstante, no existen demasiadas investigaciones sobre este grupo de usuarios, a pesar de que hay estudios que indican que las personas con dificultades de visión prefieren la voz natural a la voz sintética a la hora de elegir, por ejemplo, las voces en una audiodescripción (Walczak y Fryer, 2018; Tor-Carroggio, 2020). Una mejora en la calidad de las voces sintéticas podría modificar la aceptabilidad de estas voces en usuarios con dificultades para ver. En este sentido, Podsiadło y Chahar (2016) observaron que los usuarios con problemas de visión presentaban gustos diferentes a la hora de seleccionar una voz sintética en comparación con los que no tenían dificultades visuales. Este grupo prefería un tono de voz grave, una voz madura y que no expresara ningún tipo de emociones, este tipo de voz minimizaría la fatiga auditiva y aumentaría la inteligibilidad de las producciones orales sintéticas.

La dificultad de definir propiamente una voz agradable a causa de una interacción entre diferentes parámetros relacionados con la voz ya ha sido tratada en diferentes investigaciones, así Collados (1998) y Collados *et al.* (2007) ponen de manifiesto la interrelación entre entonación y agradabilidad, una entonación plana es percibida como poco agradable. Lo mismo ocurre con el tono, con la fluidez (Iglesias, 2013) o con la pronunciación correcta de los sonidos (Blasco y García, 2007). Este hecho no resulta extraño teniendo en cuenta que los parámetros que caracterizan la voz de un hablante están relacionados con su tono de voz, la velocidad con la que habla y el volumen de voz que tiene, rasgos que se alteran también para expresar determinado tipo de emociones (Frick, 1985).

## 3. Metodología

A partir de un corpus de audiodescripciones, detallado en el apartado siguiente, llevamos a cabo, por un lado, un análisis observacional — análisis acústico de las audiodescripciones— y, por otro lado, un estudio perceptivo, en el que los sujetos participan como potenciales usuarios de las audiodescripciones para evaluar la agradabilidad de las voces de los audiodescriptores seleccionados.

## 3.1. Corpus

El corpus en el que se basa este trabajo está formado por las audiodescripciones (AD) en tres idiomas (subcorpus del inglés, subcorpus del español y subcorpus del catalán) del cortometraje *What happens while*, filmado por la directora Núria Niu para el proyecto Visuals Into Words. El objetivo de contar con una filmación propiedad de nuestro grupo de investigación (TransMedia Catalonia, https://grupsderecerca.uab.cat/transmedia/) era tener la autorización de los derechos de autor para sucesivas investigaciones. El cortometraje original, rodado en inglés, fue doblado al español y al catalán por los mismos actores. Las audiodescripciones fueron encargadas por la coordinadora del proyecto a empresas especializadas y en ellas intervinieron locutores profesionales, 10 en cada lengua (5 hombres y 5 mujeres en español, 6 mujeres en catalán y 4 hombres en inglés), 30 en total. El VIW (uab.cat/viw), Visuals Into Words, es un corpus de acceso abierto y, por sus características, permite comparar la audiodescripción del mismo contenido visual por parte de diferentes profesionales (Matamala, 2019).

El corpus está compuesto por unidades audiodescriptivas (AD), es decir, segmentos textuales relacionados con la representación visual. Cada unidad se ha dividido en grupos de entonación, considerando los límites prosódicos de las pausas y las inflexiones tonales, mediante el programa de análisis de habla PRAAT (Boersma y Weenink, 2020). En la tabla 1 se pueden observar, en los subcorpus de cada una de las lenguas, el número de unidades de audiodescripción y de grupos de entonación de los audiodescriptores masculinos y femeninos de cada una de las empresas participantes en el proyecto.

TABLA 1
Características de los subcorpus del español del catalán y del inglés

| Español | | | | Catalán | | | | Inglés | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Empresa | Sexo | AD | GE | Empresa | Sexo | AD | GE | Empresa | Sexo | AD | GE |
| Aptent | F | 69 | 150 | Access friendly | F | 31 | 122 | Ericsson-UK | F | 31 | 116 |
| Aristia | F | 60 | 84 | Aptent | F | 67 | 170 | Bridge-US | F | 71 | 211 |
| CEIAF | F | 52 | 103 | SBT | F | 42 | 95 | BTI-UK | F | 44 | 142 |
| Kaleidoscope | F | 43 | 112 | SDI Media | F | 33 | 66 | Ericsson- AU | F | 35 | 117 |
| SDI Media | F | 63 | 112 | Subtil | F | 44 | 98 | Mennell-CA | F | 52 | 157 |
| Edsol Producciones | M | 41 | 96 | Blasi | M | 62 | 169 | MindsEye-UK | F | 45 | 194 |
| Ericsson | M | 46 | 92 | Descriptik | M | 42 | 103 | Able-NZ | M | 37 | 131 |
| Navarra de Cine | M | 40 | 108 | Multisignes | M | 46 | 156 | Deluxe-UK | M | 35 | 155 |
| Soni2 | M | 43 | 128 | Plural | M | 40 | 131 | SDI-Media | M | 43 | 113 |
| Trágora | M | 38 | 131 | Soni2 | M | 59 | 152 | ADassociates | M | 46 | 133 |

Nota. F: femenino; M: masculino; AD: audiodescripción; GE: grupos de entonación.

## 3.2. Análisis acústico y selección de los informantes

Se han medido los valores medios de la frecuencia fundamental, F0 (Hz), de la amplitud (dB) y de la duración (ms) de cada grupo de entonación, también mediante el programa PRAAT. Estos parámetros nos han permitido identificar las características prosódicas de los audiodescriptores: el tono, que es un indicador de la cualidad de voz (cuanto mayor es el valor de la F0, más aguda es la voz), el volumen de voz (cuanto mayor es la amplitud, mayor es el volumen de voz y más intensa se la percibe) y la velocidad de habla (cuanto mayores son los valores de duración en los grupos de entonación, menor será la velocidad con la que se emite el fragmento). La figura 1 muestra un ejemplo del análisis de un grupo de entonación.

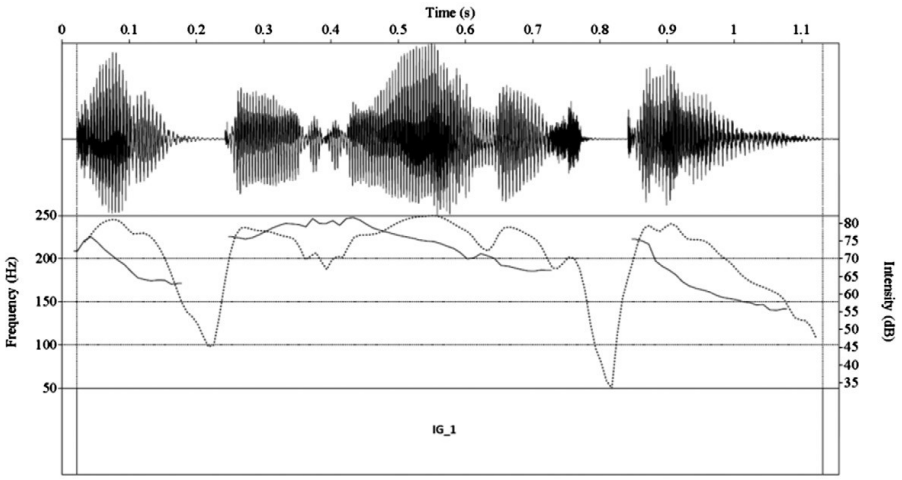


FIGURA 1
Ejemplo de un grupo de entonación analizado mediante el programa PRAAT
Nota. La línea punteada se corresponde con la intensidad y la línea continua con el tono del hablante.

Una vez obtenidos los datos acústicos (véase la tabla 1 del anexo), se procedió a la selección de los audiodescriptores masculinos y femeninos, 3 por sexo, en cada una de las lenguas (para el análisis del español, véase Machuca *et al*., 2020). Se consideraron aquellos que, en comparación con los de su grupo, presentaban valores extremos, máximos y mínimos, de los parámetros analizados, y también los que presentaban valores medios entre esos extremos. Cuando no había una diferencia relevante, se tomaron los que tenían un valor de F0 más alto, ya que este es el parámetro asociado a la cualidad voz y con el que se manifiesta claramente la individualidad de los hablantes. En la tabla 2 se muestran las voces seleccionadas para el test de percepción.

TABLA 2
Voces seleccionadas para el test de percepción

| Voces femeninas del español | Velocidad de habla | F0 | Amplitud |
|---|---|---|---|
| Audiodescriptora 1 (Aptent) | + | + | + |
| Audiodescriptora 2 (Aristia) | - | - | med |
| Audiodescriptora 3 (Kalleidoscope) | + | med | - |
| Voces masculinas del español | | | |
| Audiodescriptor 1 (Edsol) | med | - | med |
| Audiodescriptor 2 (Ericsson) | - | + | - |
| Audiodescriptor 3 (Soni2) | + | med | + |
| Voces femeninas del catalán | | | |
| Audiodescriptora 1 (SBT) | - | + | med |
| Audiodescriptora 2 (Subtil) | + | med | + |
| Audiodescriptora 3 (SDI Media) | med | - | - |
| Voces masculinas del catalán | | | |
| Audiodescriptor 1 (Blasi) | - | + | - |
| Audiodescriptor 2 (Soni2) | + | med | + |
| Audiodescriptor 3 (Descriptik) | med | - | med |
| Voces femeninas del inglés | | | |
| Audiodescriptora 1 (BTI-UK) | + | - | + |
| Audiodescriptora 2 (Ericsson-UK) | - | + | + |
| Audiodescriptora 3 (MindsEye-UK) | med | med | - |
| Voces masculinas del inglés | | | |
| Audiodescriptor 1 (Able-NZ) | + | - | med |
| Audiodescriptor 2 (Deluxe-UK) | + | + | + |
| Audiodescriptor 3 (SDI Media-UK) | - | - | - |

Nota. (+) indica los valores máximos; (–), los valores mínimos; y (med), los valores medios.

## 3.3. Prueba de percepción

Tras elegir las voces a partir de los parámetros acústicos, se procedió a preparar los test de percepción, uno por lengua. Se seleccionaron tres fragmentos breves con sentido de cada voz para formar los estímulos de los test. Los estímulos están formados por pares de fragmentos con un contenido similar, para contrastar dos voces, o masculinas o femeninas, de modo que todas se comparan unas con otras. A modo de ejemplo, los fragmentos elegidos de las voces masculinas del español, de las empresas Edsol, Ericsson y Soni2, se combinan del siguiente modo: Edsol-Ericsson, Edsol-Soni2 y Ericsson-Soni2. Para cada lengua se crearon 9 estímulos con las voces masculinas y otros 9 con las femeninas. Las dos voces de cada estímulo están separadas por un silencio de 800 ms, y ante la primera voz se incluyó un silencio de 1000 ms.

La prueba de percepción siguió un procedimiento aprobado por el comité de ética de la Universidad Autónoma de Barcelona, y cada una tenía una duración aproximada de 20 minutos. Los participantes, nativos de la lengua en cuestión, usando un formulario de Google en línea, debían dar primero su consentimiento para participar en la prueba y después seleccionar, de las dos voces de cada estímulo, la que consideraban más agradable. Para determinar el efecto de la discapacidad visual, los participantes se dividieron en dos grupos: sin pérdida de visión y con pérdida de visión.

Las respuestas se recopilaron en un archivo de Excel para su posterior análisis estadístico. Se realizaron pruebas de chi-cuadrado con el *software* SPSS. En todos los casos, la variable independiente fueron las respuestas y la variable dependiente cada uno de los grupos de sujetos (con pérdida de visión vs. sin pérdida de visión).

# 4. Resultados

## 4.1. Resultados de la agradabilidad del subcorpus del español

En el cuestionario, para identificar las preferencias vocales en español, han participado 31 personas con pérdida de visión y 29 sin pérdida, y se han obtenido un total de 540 respuestas para las voces de cada sexo (60*9 estímulos), 279 de los participantes con pérdida de visión y 261 de los participantes sin pérdida.

*Voces femeninas*

Según la prueba de chi-cuadrado ($\chi(2) = 540$, $p = 0{,}03$, $< 0{,}05$), existen diferencias significativas entre las preferencias por las voces femeninas del español por parte de los participantes con pérdida de visión y de los participantes sin pérdida.

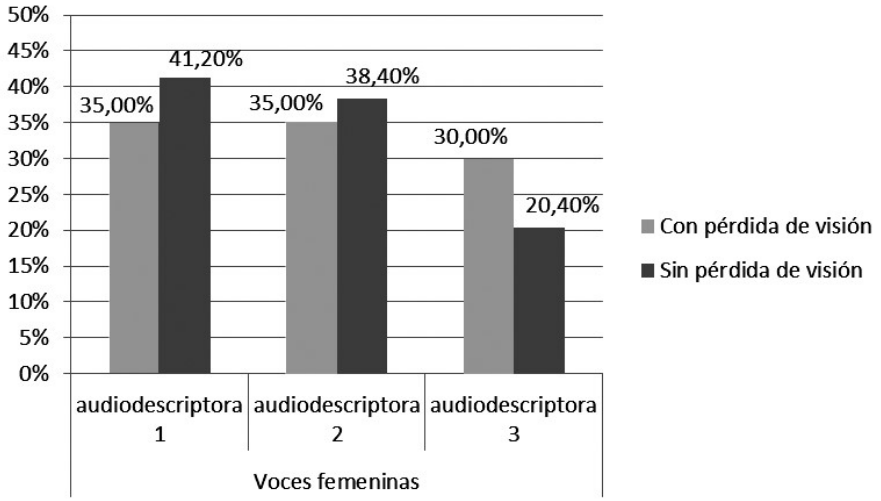


FIGURA 2
Preferencias de agradabilidad de las voces de las audiodescriptoras del español en porcentajes

La Figura 2 muestra los resultados de percepción de las voces de las audiodescriptoras del español. Se puede observar que los participantes con pérdida de visión prefieren por igual las voces de las audiodescriptoras 1 y 2 (35 %). La audiodescriptora 3 tiene un porcentaje algo menor (30 %). Las dos voces preferidas presentan los valores extremos de velocidad de habla y de F0 (la audiodescriptora 1 tiene los valores máximos y la 2, los mínimos), por lo que estos dos parámetros no pueden ser la causa de las preferencias. Además, la audiodescriptora 3 comparte un valor máximo de velocidad de habla con la audiodescriptora 1 y su valor medio de F0 se sitúa entre los de las dos favoritas. En los valores de amplitud, la descriptora 1 tiene el valor máximo; la 2, un valor medio; y la 3, el valor mínimo. Las preferencias, por tanto, se dirigen a un volumen de voz alto medio de las voces femeninas.

Los participantes sin pérdida de visión prefieren, en primer lugar, la voz de la audiodescriptora 1 (41,20 %), seguida de la audiodescriptora 2, con un porcentaje algo menor (38,40 %). La audiodescriptora 3 se sitúa a una gran distancia de ambas (20,4 %). Las diferencias entre los dos grupos de hablantes se deben más a la diferencia de los porcentajes que al orden de valoración de las audiodescriptoras. En este caso, es más evidente la preferencia por el volumen de voz alto medio de las audiodescriptoras 1 y 2. Ambos tipos de participantes rechazan, por tanto, las voces con un volumen de voz bajo como voces agradables.

*Voces masculinas*

En las preferencias por las voces masculinas del español, no existen diferencias significativas entre los participantes con pérdida de visión y sin pérdida ($\chi(2) = 540$, $p = 0{,}598$, $> 0{,}05$). Como se puede observar en la figura 3, los porcentajes de valoración son semejantes en ambos grupos. La voz elegida es la del audiodescriptor 2, por el 45,60 % de los participantes con dificultades de visión y por el 47,30 % de los participantes sin pérdida. La diferencia es notable sobre los otros dos audiodescriptores, que presentan porcentajes similares.

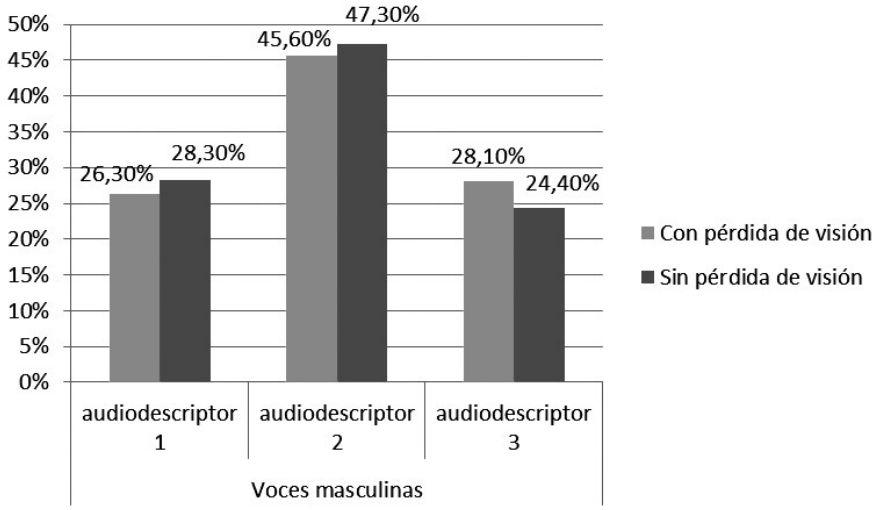


FIGURA 3
Preferencias de agradabilidad de las voces de los audiodescriptores del español en porcentajes

El audiodescriptor 2 tiene un valor máximo de F0, en oposición al audiodescriptor 3, con un valor medio, y al 1, con un valor mínimo. Por tanto, los participantes con pérdida y sin pérdida de visión prefieren, de las tres voces masculinas presentadas en el test, la que tiene el tono más agudo. No obstante, también influyen en la valoración del audiodescriptor 2 su velocidad lenta y su amplitud baja, que lo diferencian de los otros dos audiodescriptores, el 1 con valores medios y el 3 con valores máximos en esos dos parámetros.

## 4.2. Resultados de la agradabilidad del subcorpus del catalán

En el test de percepción del catalán han tomado parte 77 personas, 31 con pérdida de visión y 46 sin pérdida. El número total de respuestas obtenidas ha sido 693, 279 (31*9 estímulos) de los participantes con pérdida de visión para cada uno de los dos tipos de voces (femeninas y masculinas) y 414 (46*9 estímulos) de los participantes sin pérdida, también para cada tipo de voz.

*Voces femeninas*

La prueba de chi-cuadrado muestra que existen diferencias significativas entre ambos grupos de participantes en sus preferencias por las voces femeninas del catalán ($\chi(2) = 693$, $p = 0{,}000$, $< 0{,}05$). La figura 4 muestra los resultados de percepción de la valoración de agradabilidad para las voces femeninas. Los participantes con pérdida de visión consideran más agradable la voz de la audiodescriptora 1 (40,1 %); en segundo lugar, la de la audiodescriptora 2 (33 %); y, finalmente, la de la audiodescriptora 3 (26,9 %). Las diferencias entre esos porcentajes son similares.

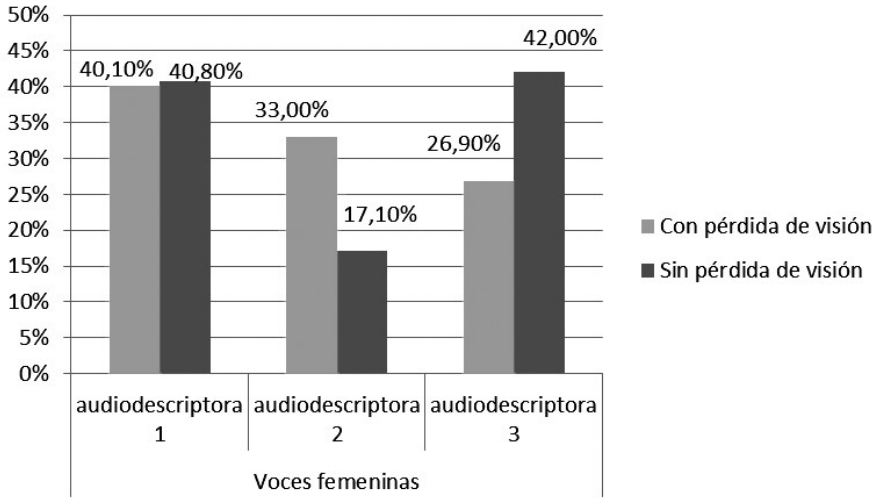


FIGURA 4
Preferencias de agradabilidad de las voces de las audiodescriptoras del catalán en porcentajes

La voz de la audiodescriptora 1 se caracteriza por tener el tono más alto y la velocidad de elocución más rápida. En el tono, se observa que existe una gradualidad de mayor a menor en los valores de las audiodescriptoras según la escala de preferencia: audiodescriptora 1 (+), audiodescriptora 2 (med.) y audiodescriptora 3 (-). De la velocidad de elocución, solo se puede deducir que prefieren un habla lenta, ya que la voz femenina que consideran menos agradable tiene una velocidad media y la que está situada en segundo lugar es la más rápida. La amplitud de la voz considerada más agradable se sitúa en un rango medio, entre los valores extremos de la audiodescriptora 2 (máximo) y de la audiodescriptora 3 (mínimo).

Los participantes sin pérdida de visión muestran unas preferencias distintas en el test. La voz mejor valorada es la de la audiodescriptora 3 (42 %), seguida —con un porcentaje algo menor— por la audiodescriptora 1 (40,8 %), ambas muy distanciadas de la audiodescriptora 2 (17,1 %). La audiodescriptora mejor valorada tiene una velocidad de habla media, pero los participantes toleran el habla lenta de la audiodescriptora 1 antes que el habla rápida de la audiodescriptora 3. Los valores de amplitud presentan una gradualidad de menor a mayor en el orden de preferencia por las audiodescriptoras. Los participantes prefieren un volumen de voz bajo (audiodescriptora 3) y toleran un volumen de voz medio (audiodescriptora 1) por encima del volumen de voz alto de la audiodescriptora 2. En cuanto al tono, prefieren una voz femenina grave, ya que la audiodescriptora 3 tiene el valor de F0 más bajo, pero el de la audiodescriptora 1, situada en segundo lugar en las preferencias de los participantes, tiene el más alto, por lo que este parámetro no parece influir en la valoración de las voces de las audiodescriptoras del catalán por parte de los participantes sin pérdida de visión.

*Voces masculinas*

En las preferencias por las voces masculinas del catalán, no existen diferencias significativas entre los dos grupos de los participantes en el test de percepción, como indica el resultado de la prueba de chi-cuadrado ($\chi(2) = 693$, $p = 0{,}262$, $> 0{,}05$). En ambos casos, la voz considerada más agradable es la del audiodescriptor 1, seguida de la voz del audiodescriptor 2; la menos agradable es la del audiodescriptor 3. Los porcentajes obtenidos son similares (ver figura 5). En el grupo de los participantes con pérdida de visión, 41,2 % prefirió la voz del audiodescriptor 1; el 36,9 %, la del audiodescriptor 2; y 21,9 %, la de audiodescriptor 3. En el grupo de los participantes sin pérdida de visión, 46,9 % optó por la voz del audiodescriptor 1; 35,3 %, por la del audiodescriptor 2; y 17,9 %, por la del audiodescriptor 3.

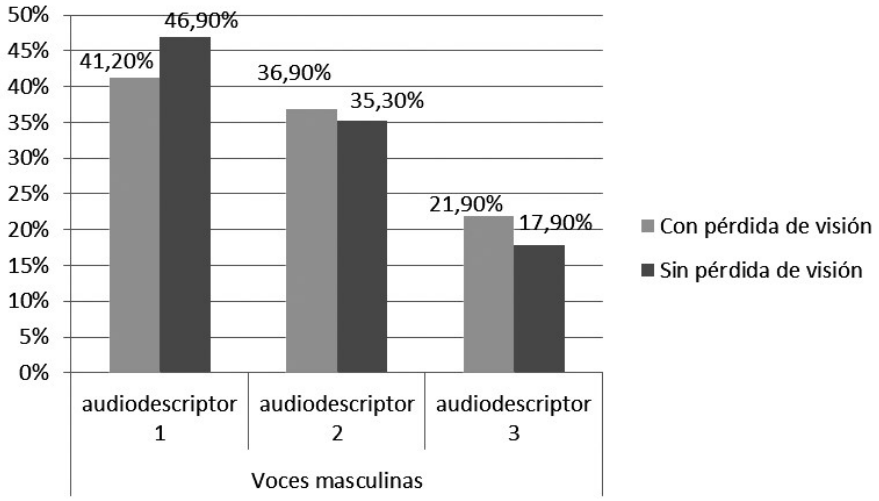


FIGURA 5
Preferencias de agradabilidad de las voces de los audiodescriptores del catalán en porcentajes

La voz masculina del catalán mejor considerada por su agradabilidad tiene el valor más alto de F0 de los tres audiodescriptores contrastados. En este parámetro se observa una gradualidad de mayor a menor en sus valores coincidente con las preferencias de los participantes, el audiodescriptor 2 tiene una F0 media y el 3, mínima. Esa gradualidad no está presente en la velocidad de habla y en la amplitud, que no parecen ser determinantes para valorar la agradabilidad. Los participantes prefieren un tono de voz agudo, con una velocidad de habla lenta y un volumen de voz bajo.

## 4.3. Resultados de la agradabilidad del subcorpus del inglés

Un total de 71 hablantes nativos de inglés americano han participado en el test de percepción de esta lengua, 40 con pérdidas de visión y 31 sin pérdida. Del primer grupo, se han obtenido 360 respuestas (40*9 estímulos) para cada uno de los tipos de voces del test. Del segundo grupo, se han recogido 279 respuestas (31*9 estímulos), también para cada tipo de voz. El número total de respuestas ha sido 639.

*Voces femeninas*

No existen diferencias significativas entre los dos grupos de participantes en la valoración de la agradabilidad de las voces femeninas del inglés, según la prueba de chi-cuadrado ($\chi(2) = 639$, $p = 0{,}654$, $> 0{,}05$). De hecho, ambos muestran las mismas preferencias y con porcentajes similares, tal como se observa en la figura 6: audiodescriptora 1 (con pérdida de visión: 40,6 %; sin pérdida: 38,4 %), audiodescriptora 3 (con pérdida de visión: 30,3 %; sin pérdida: 33,7 %) y audiodescriptora 2 (con pérdida de visión: 29,2 %; sin pérdida: 28,0 %).

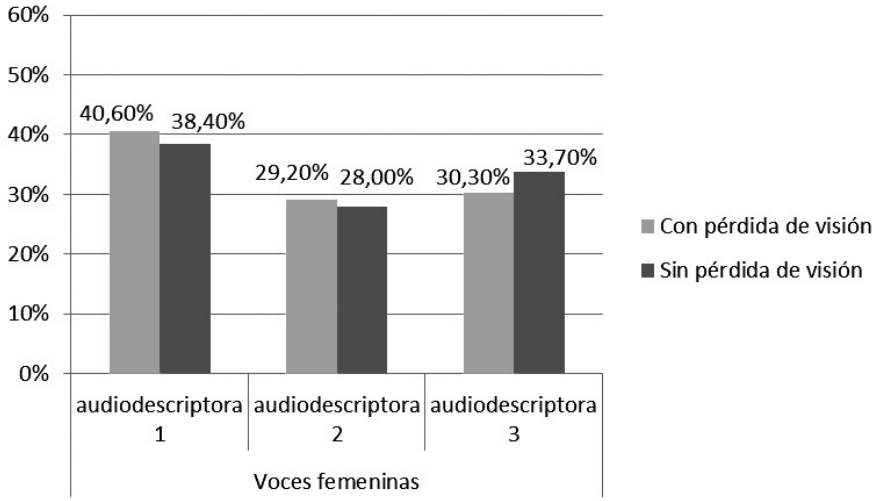


FIGURA 6
Preferencias de agradabilidad de las voces de las audiodescriptoras del inglés en porcentajes

Dos parámetros parecen intervenir en la identificación de la agradabilidad en el caso de las audiodescriptoras inglesas: la velocidad de habla y la F0. La agradabilidad disminuye conforme la velocidad de habla es menor y los valores de la F0 aumentan: audiodescriptora 1 (velocidad mayor, F0 menor), audiodescriptora 3 (velocidad media, F0 media) y audiodescriptora 2 (velocidad menor, F0 mayor). No se puede apreciar si la amplitud contribuye a determinar la agradabilidad de las voces femeninas del inglés, ya que la audiodescriptora más valorada y la menos valorada comparten una amplitud elevada.

*Voces masculinas*

El análisis de las respuestas sobre la agradabilidad de las voces masculinas del inglés muestra que existen diferencias significativas entre los dos grupos de participantes, ($\chi(2) = 639$, $p = 0{,}002$, $< 0{,}05$). No obstante, a partir de los datos de percepción observados en la figura 7, podemos señalar que ambos consideran que la voz más agradable es la del audiodescriptor 1 (49,7 % para los participantes con pérdida de visión y 56,3 % para participantes sin pérdida). Estos porcentajes son bastante superiores a los de las otras dos voces. Las diferencias se deben a la posición que ocupan en las preferencias por parte de los participantes de los audiodescriptores 2 y 3, y a sus porcentajes de aceptación.

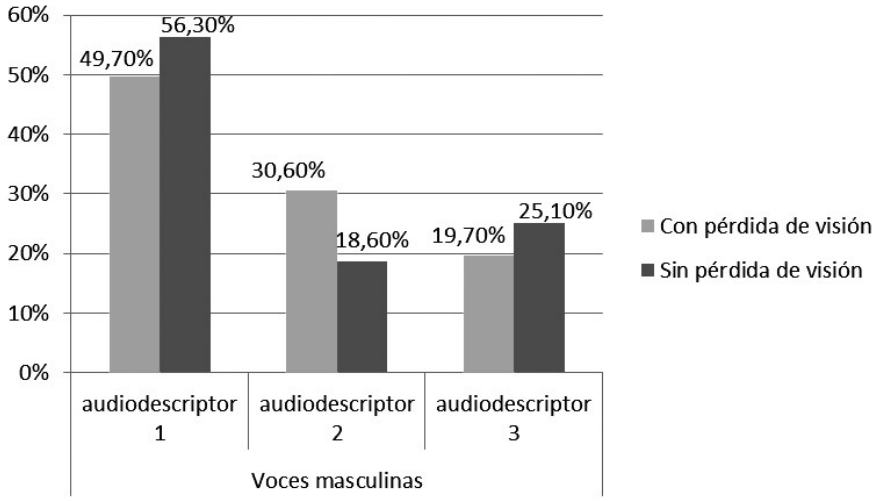


FIGURA 7
Preferencias de agradabilidad de las voces de las audiodescriptoras del inglés en porcentajes

El audiodescriptor 1 se caracteriza por un habla rápida y tener valores bajos de F0 y medios de amplitud. Los participantes con pérdida de visión sitúan en segundo lugar de sus preferencias al audiodescriptor 2 (30,6 %), que comparte con el 1 la velocidad de habla rápida, pero que difiere en la F0 y en la amplitud, con los valores más altos en ambos parámetros. El audiodescriptor 3, que presenta los valores mínimos en los tres parámetros analizados, es el menos valorado por este grupo de participantes (19,7 %, un porcentaje muy bajo) y se diferencia del 1 por la velocidad de habla y la amplitud, aunque comparte el mismo rango en la F0. Por tanto, parece que para los participantes con pérdida de visión prima la velocidad de habla como parámetro para seleccionar la agradabilidad.

Los participantes sin pérdida de visión sitúan en segundo lugar al audiodescriptor 3 (25,1 %), que comparte con el 1 los valores bajos de F0, pero que se diferencia por su velocidad lenta y su volumen de voz bajo. La voz peor valorada por este grupo es la del audiodescriptor 2 (18,6 %), que, como se ha indicado, tiene una velocidad rápida, como el audiodescriptor 1, pero que se diferencia de este por los valores de F0 y amplitud (tiene los valores más altos en los tres parámetros). Por tanto, el grupo de hablantes sin pérdida de visión prefiere la velocidad rápida del audiodescriptor 1, pero se inclinan por un tono grave y un volumen de voz medio bajo.

## 5. Discusión de los resultados

Los resultados de nuestro experimento indican la dificultad de relacionar los parámetros acústicos con una clara identificación de la agradabilidad de las voces, tanto en voces femeninas como masculinas. Iglesias (2007) ya indicaba la falta de criterios objetivos medibles a la hora de evaluar la agradabilidad de las voces. En este trabajo, hemos intentado establecer unas medidas objetivas en esa evaluación mediante la caracterización de las voces a partir de sus propiedades acústicas: velocidad de habla, tono y volumen de voz. No obstante, los resultados muestran que la percepción de agradabilidad no solo difiere en las lenguas analizadas sino también, en algunos casos, en el hecho de si los sujetos que oyen esas voces poseen una discapacidad visual o no. Se ha afirmado que, a pesar de la falta de objetividad a la hora de evaluar las voces más agradables, un tono agudo y un volumen de voz alto no era percibido como una voz agradable (Iglesias, 2013). En nuestros datos estas preferencias van ligadas a las lenguas, así, mientras que la tendencia en catalán y en español es a preferir voces con un tono alto (voces agudas), tanto para voces masculinas como para las femeninas, en inglés se valora como más agradable los tonos bajos, las voces más graves. El volumen de voz no es una característica definitoria

en el conjunto de nuestros resultados, pero, sí la velocidad de habla, que se prefiere con un tempo más bien rápido en inglés, pero no en las otras dos lenguas.

Así mismo, nuestros datos contradicen en cierto modo los resultados obtenidos para la caracterización de la voz radiogénica, en la que se rechazan los tonos agudos tanto para hombres como para mujeres (Rodríguez, 1989; Rodero, 2001; Rodero, 2002). Evidentemente, el objetivo de evaluar estas voces es diferente, pero, en todo caso, son voces que transmiten unos valores semánticos importantes, sobre todo, para las personas que tienen una discapacidad visual, cuya descripción es esencial para entender lo que se está interpretando.

Al contrario de lo que sucedía en el trabajo de Podsiadło y Chahar (2016) —que encontraban diferencias entre los gustos de los usuarios con problemas de visión y los que no tenían esta dificultad—, solo hemos encontrado diferencias en la evaluación de las voces femeninas en catalán: los sujetos con déficit visual prefirieron una audiodescriptora con una voz más aguda (audiodescriptora 1), mientras que los que no presentaban tal déficit prefirieron una voz grave (audiodescriptora 1). Todas las otras diferencias han venido dadas por el índice del porcentaje, pero no por la selección de la voz.

## 6. Conclusiones

Este trabajo contribuye a llenar un vacío bibliográfico existente sobre la caracterización de las voces en el ámbito de la audiodescripción, el instrumento esencial en esta profesión, pues los usuarios, generalmente con dificultad de visión, son capaces de interpretar correctamente una obra cinematográfica, lírica, teatral... a través de esta voz.

Podemos concluir, a partir de nuestros resultados, que la agradabilidad depende, sobre todo, de las lenguas, más que de la discapacidad visual que presente el individuo que está escuchando esas voces. La agradabilidad en las voces en catalán y en español se corresponde con voces de un tono agudo, excepto para los sujetos sin dificultad visual que prefieren en catalán una voz femenina de tono grave. Por el contrario, los sujetos nativos de inglés prefieren voces con un tono grave. Quizá en investigaciones futuras, debería profundizarse en las diferencias acústicas entre las lenguas analizadas, pues la caracterización del tono grave, o del agudo, desde el punto de vista acústico, puede no ser igual en español, en catalán y en inglés.

En cuanto a la velocidad de elocución, una tasa baja de habla se aprecia como una voz agradable en español y en catalán, pero no en inglés, en la que los sujetos prefieren un tempo rápido. La correspondencia entre la percepción de la intensidad de esa voz y la agradabilidad, analizada acústicamente mediante la amplitud, no arroja resultados claros, unas veces se prefieren voces con un volumen bajo, por ejemplo, para los hombres en español; otras, con un volumen alto, como el caso de las mujeres en español.

A modo de reflexión, cabe señalar que la audiodescripción es un servicio que sirve para suplir la falta de información visual de las personas con discapacidad visual, información que se considera importante para que el usuario perciba el producto audiovisual de forma similar a como lo percibirían las personas videntes; esta información se transmite a través de la voz, por lo que su descripción se hace necesaria si queremos mejorar el resultado.

## REFERENCIAS BIBLIOGRÁFICAS

American Council of the Blind. (2009). *Audio description standards*. http:// www.acb.org/adp/docs/ADP_Standards.doc

Blasco, M. J., y García, O. (2007). La incidencia del parámetro dicción. En A. Collados, E. M. Pradas, E. Stévaux, y O. García (Eds.), *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia* (pp. 175-194). Comares.

Boersma, P., y Weenik, D. (2020). *Praat* (version, v.6.1.37) [Software]. https://www.fon.hum.uva.nl/praat/

Coelho, L., Braga, D., Dias, M., y García-Mateo, C. (27-31 de agosto del 2011). *An Automatic Voice Pleasantness Classification System Based on Prosodic and Acoustic Patterns of Voice Preference* [Presentación en papel]. Interspeech 2011, Florence, Italy. https://doi.org/10.21437/Interspeech.2011-589

Collados, Á. (1998). *La evaluación de la calidad en la interpretación simultánea: la importancia de la comunicación no verbal*. Comares.

Collados, Á., Pradas, E., Stévaux, E., y García, O. (2007). *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia*. Comares.

Comité Técnico de Normalización 153. (2005). *Norma UNE:153030. Audiodescripción para persones con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. Asociación Española de Normalización y Certificación.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97*(3), 412-429. https://doi.org/10.1037/0033-2909.97.3.412

Fryer, L. (2016). *An introduction to audio description: A practical guide*. Routledge.

Iglesias, E. (2006). La indefinición del parámetro "agradabilidad de la voz" y los estudios de calidad de la interpretación simultánea. En M.-J. Varela Salinas (Coord.), *La evaluación en los estudios de traducción e interpretación* (pp. 225-239). Editorial Bienza.

Iglesias, E. (2007). La incidencia del parámetro agradabilidad de la voz. En A. Collados, E. M. Pradas, E. Stévaux, y O. García (Eds.), *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia* (pp. 37-51). Comares.

Iglesias, E. (2013). Understanding variability in interpreting quality assessment: user's sex and judgements for pleasant voice. En C. Way, S. Vandepitte, R. Meylaerts, y M. Bartłomiejczyk (Eds.), *Tracks and Treks in Translation Studies: selected papers from the EST Congress, Leuven 2010* (pp. 103-125). John Benjamins.

Iglesias, E., Martínez, S., y Chica, A. (2015). Cross-Fertilisation between Reception Studies in Audio Description and Interpreting Quality Assessment. The Role of the Describer's Voice. En R. B. Piñero, y J. D. Cintas (Eds.), *Audiovisual Translation in a Global Context* (pp. 72-95). Palgrave Macmillan. https://doi.org/10.1057/9781137552891_5

Keller, E. (2005). The Analysis of Voice Quality in Speech Processing. En G. Chollet, A. Esposito, M. Faundez-Zanuy, y M. Marinaro (Eds.), *Nonlinear Speech Modeling and Applications* (pp. 54-73). Springer. https://doi.org/10.1007/11520153_4

Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*. University of Chicago Press.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge University Press.

Machuca, M. J., Matamala, A., y Ríos, A. (2020). Prosodic features in Spanish audio descriptions of the VIW corpus. *MonTI, 12*, 53-77. https://doi.org/10.6035/MonTI.2020.12.02

Matamala, A. (2019). The VIW corpus: multimodal corpus linguistics for audio description analysis. *RESLA, 32*(2), 515-542.

Podsiadło, M., y Chahar, S. (8-12 de setiembre de 2016). *Text-to-Speech for individuals with vision loss: a user study* [Presentación en papel]. Interspeech 2016, San Francisco, Estados Unidos. https://doi.org/10.21437/Interspeech.2016-1376

Rodero, E. (2001). *Locución Informativa Radiofónica* [Tesis doctoral, Universidad Pontificia de Salamanca]. Dialnet. https:// dialnet.unirioja.es/servlet/tesis?codigo=63245

Rodero, E. (2002). El tono de la voz masculina y femenina en los informativos radiofónicos: un análisis comparativo. En *Mujeres, Hombres y Medios de Comunicación* (Vol. 2, pp. 319-331). Lex Nova.

Rodero, E. (2003). Para hablar de manera correcta y agradable desde la cabina. La voz informativa radiogénica. *Revista Mexicana de Comunicación, 15*(79), 36-41.

Rodríguez, A. (1989). *La construcción de una voz radiogénica* [Tesis doctoral, Universitat Autònoma de Barcelona]. TDX. https://www. tdx.cat/handle/10803/4132;jsessionid=6966BDB724153606D97574EEE3177CDA#page=1

Tor-Carroggio, I. (2020). T(ime) T(o) S(tart) synthesising audio description in China? Results of a reception study. *The Journal of Specialised Translation*, (34), 171-191. https://jostrans.org/ issue34/art_tor.pdf

Walczak, A., y 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

## Apéndice

### *Anexo 1. Análisis prosódico del español, del catalán y del inglés*

Subcorpus del español

Voces femeninas

| Empresa | Duración/GE (ms) | Media F0 (Hz) | Media amplitud (dB) |
|---|---|---|---|
| Aristia | 1736.95 | 135.85 | 67.4 |
| Aptent | 1590.84 | 200.12 | 75.4 |
| CEIAF | 1814.93 | 191.41 | 59.8 |
| Kalidoscope | 1052.72 | 164.21 | 58.8 |
| SDI | 1616.58 | 171.37 | 66.65 |

Voces masculinas

| Empresa | Duración/GE (ms) | Media F0 (Hz) | Media amplitud (dB) |
|---|---|---|---|
| Ericsson | 1801.97 | 101.71 | 64.27 |
| Soni2 | 1230.17 | 121.33 | 69.82 |
| EdSol | 1363.66 | 89.29 | 65.24 |
| Navarra | 1324.26 | 101.03 | 65.36 |
| Trágora | 1974.62 | 100.04 | 65.31 |

Subcorpus del catalán

Voces femeninas

| Empresa | Duración/GE (ms) | Media F0 (Hz) | Media amplitud (dB) |
|---|---|---|---|
| Accessfriendly | 1404.65 | 195,03 | 68.67 |
| APTENT | 1503 | 207,79 | 70.51 |
| SBT | 1903.14 | 219,29 | 66.43 |
| SDI-Media | 1647.07 | 126,04 | 64.13 |
| Subtil-CA | 1596 | 180,29 | 69.97 |

Voces masculinas

| Empresa | Duración/GE (ms) | Media F0 (Hz) | Media amplitud (dB) |
|---|---|---|---|
| Blasi | 1594.33 | 137.03 | 63.00 |
| Descriptik | 1384.94 | 108.81 | 69.29 |
| Multisignes | 1542.29 | 135.34 | 69.40 |
| Plurals | 1341.76 | 123.01 | 72.85 |
| Soni2 | 1242.28 | 127.33 | 72.39 |

Subcorpus del inglés

Voces femeninas

| Empresa | Duración/GE (ms) | Media F0 (Hz) | Media amplitud (dB) |
|---|---|---|---|
| BTI-UK | 1160.27 | 140.40 | 70.59 |
| Bridge-US | 1304.52 | 196.04 | 61.99 |
| Ericsson-AU | 1300.79 | 165.35 | 67.03 |
| Ericsson-UK | 1725.39 | 186.88 | 68.27 |
| Mennell-CA | 1279.50 | 173.36 | 64.30 |
| MindsEye-UK | 1188.77 | 164.70 | 60.33 |

Voces masculinas

| Empresa | Duración/GE (ms) | Media F0 (Hz) | Media amplitud (dB) |
|---|---|---|---|
| ADAssociatesUS_synder | 1306.00 | 122.64 | 69.81 |
| Able_NZ | 1256.14 | 107.90 | 64.94 |
| Deluxe-UK | 1225.33 | 139.17 | 67.00 |
| SDI-Media-UK | 1584.81 | 99.05 | 63.07 |

## Notas

[1] Esta investigación forma parte del proyecto RAD (Researching Audio Description: Translation, Delivery and New Scenarios), código de referencia PGC2018-096566-B-I00. Los autores son miembros de TransMedia Catalonia, un grupo de investigación financiado por el Gobierno catalán bajo el esquema de financiación SGR (2017SGR113).