

Los audiodescriptores: voces neutras y voces agradables¹

María J. Machuca¹, Anna Matamala¹, Antonio Ríos¹

¹Universitat Autònoma de Barcelona
MariaJesus.Machuca@uab.cat ORCID: <https://orcid.org/0000-0003-4377-288X>
Anna.Matamala@uab.cat ORCID: <http://orcid.org/0000-0002-1607-9011>
Antonio.Rios@uab.cat ORCID: <https://orcid.org/0000-0001-5262-7153>

Recibido: 7/10/2021; **Aceptado:** 25/10/2021; **Publicado en línea:** 18/01/2022

Citación / Cómo citar este artículo: María J. Machuca, Anna Matamala, Antonio Ríos (2020). Los audiodescriptores: voces neutras y voces agradables. *Loquens*, 7(2), e076. <https://doi.org/10.3989/loquens.2020.076>

RESUMEN: El objetivo de este estudio es analizar los rasgos prosódicos de un corpus de audiodescripciones en español para determinar la neutralidad y la agradabilidad en la prosodia que emplean los locutores. El análisis supone una revisión de las recomendaciones sobre la locución que aparecen en las guías. Según la Asociación Española de Normalización (UNE 153020), la locución del guion de la audiodescripción (AD) debe ser neutra, debiendo evitarse la entonación afectiva.

El corpus de análisis está constituido por las audiodescripciones de locutores profesionales masculinos y femeninos. Se han analizado la duración, el tono y la amplitud de los enunciados emitidos. Posteriormente, se ha llevado a cabo un test de percepción para evaluar la neutralidad y la agradabilidad de las voces de aquellos locutores seleccionados a partir de los resultados obtenidos en el análisis de los parámetros acústicos considerados. Las respuestas del test de percepción fueron tratadas tomando por separado dos grupos de sujetos: un grupo formado por las personas que presentaban una discapacidad visual y otro grupo constituido por personas que no tenían esa limitación. Los resultados de este estudio pueden contribuir a una mejor selección de las voces de los audiodescriptores dependiendo de la tarea que deban realizar, y a establecer las diferencias en función de las limitaciones visuales que puedan tener los usuarios.

Palabras clave: audiodescripción, prosodia, usuarios con pérdida de visión, percepción.

ABSTRACT: *Voices in audiodescription: Neutrality and pleasantness.*— The aim of this study is to analyse the prosodic features of a corpus of audio descriptions in Spanish in order to determine the neutrality and the pleasantness in the prosody used by the speakers. The analysis involves a revision of the recommendations found in guidelines and standards on the voices of the audio describers. According to the Spanish standard by Asociación Española de Normalización (UNE 153020), speech in audio description (AD) must be neutral, and emotional intonation should be avoided.

The corpus is obtained from professional male and female speakers. The duration, pitch and amplitude of AD units have been analysed. Then, a perception test has been carried out to assess the neutrality of the voices of the describers, which were chosen by means of an acoustic analysis, and the relationship between neutrality and pleasantness. Perception tests results were analysed taking into account two separate groups: on the one hand, persons with sight loss and, on the other hand, persons without sight loss. The results of this study can help to better select audio description voices depending on the specific task to be performed and can also help to identify differences depending on the user profile.

Keywords: audio description, prosody, users with sight loss, perception.

¹ Parte de los datos de este trabajo se presentaron en el VIII Congreso Internacional de Fonética Experimental, celebrado en Girona del 28 al 30 de junio de 2021.

1. INTRODUCCIÓN

La audiodescripción (AD) es una modalidad de traducción intersemiótica que convierte un contenido visual en palabras. Se trasladan los elementos visuales, así como ciertos elementos sonoros difíciles de comprender sin las imágenes, a elementos lingüísticos sonoros. La AD se aplica a distintos tipos de contenidos, como pueden ser, por ejemplo, películas, obras de teatro y óperas. La AD permite que personas sin acceso a las imágenes como las personas con discapacidad visual puedan acceder a contenidos audiovisuales (Matamala, 2019). En el estándar UNE 153020, se recomienda que la voz de los audiodescriptores en español sea neutra (AENOR, 2005). En cambio, autores como Cabeza-Cáceres (2013), por ejemplo, afirman que hay tres maneras de llevar a cabo una audiodescripción: una más uniforme y plana, en línea con la entonación habitual de la tradición española o alemana; una adaptada, tradicionalmente más propia del entorno británico, y la enfática, más expresiva y habitual en los contenidos de Estados Unidos; la elección de cualquiera de ellas no afecta a la comprensión de lo que se está describiendo. La primera manera se corresponde con una voz más neutra, ya que apenas presenta cambios prosódicos; la segunda se relaciona con el hecho de que una película sea infantil, o que se esté audiodescribiendo un partido de fútbol, es decir, depende de la tipología de las piezas que se están audiodescribiendo; la tercera está relacionada con una voz que puede resultar más o menos agradable debido a la expresividad con la que se emite.

De la misma forma que no se encuentra una definición propia de la neutralidad, tampoco existe una definición precisa para el concepto de agradabilidad, pues, como señala Iglesias Fernández (2006, 2013), se ha definido de forma subjetiva, combinándolo con diversos parámetros, como la entonación, la fluidez, el tempo y la dicción.

El análisis de las voces de los audiodescriptores puede ayudarnos a caracterizar las voces neutras y determinar, así, si la agradabilidad de estas voces depende de la neutralidad que el usuario haya percibido en ellas.

2. ESTADO DE LA CUESTIÓN

Cualquier tipo de emoción o afectividad en la locución pasa por el empleo de unos rasgos prosódicos determinados. Según Alter (2002), la prosodia de las emociones viene caracterizada por los rasgos suprasegmentales de los valores de la F0 (*pitch* o frecuencia de vibración de las cuerdas vocales del hablante), del tempo y de la intensidad. Otros autores (véase, por ejemplo, Kraxenberger et al., 2018) han establecido que se puede predecir el contenido emocional de la lectura de poemas a través de los valores de la F0 y de la velocidad de elocución. Hay también muchos estudios que caracterizan los rasgos prosódicos utilizados para expresar la alegría y la tristeza. En el caso de la alegría, se manifiesta a través de valores más altos de la F0 media y de la intensidad media, así como de un tempo más rápido; la expresión de la tristeza, por el contrario, se ha asociado con valores más bajos de

estos parámetros (Scherer, 1986, 2013; Banse y Scherer, 1996; Ververidis y Kotropoulos, 2006; Pell et al., 2009 y Stolarski, 2015, entre otros).

Opuesto a los términos emoción, afectividad o expresividad existe el término “neutro”. No obstante, la definición de este término no es muy clara y, a veces, va ligada a un parámetro prosódico determinado, como es el caso de la entonación “neutra”, que hace referencia a una entonación no marcada propia de una oración declarativa. En el Diccionario de Lingüística y Fonética de Crystal (2008) se alude al término “neutro” como algo no marcado, es decir, a una ausencia de rasgos. No obstante, si en la entonación se puede hablar de la ausencia de rasgos, una entonación declarativa frente a una interrogativa, por ejemplo, esto no se puede hacer extensivo ni al tono, ni a la velocidad de elocución, ni a la intensidad. Cada hablante tiene un tono y una velocidad de elocución intrínseca a la hora de hablar y un volumen de voz determinado, con lo que sería muy difícil decidir qué significa neutralidad en el empleo de estos rasgos prosódicos.

De la misma forma, Hidalgo Navarro (2017) señala que no existe en lo coloquial una entonación neutra en sentido estricto, pues las palabras siempre están en un contexto situacional y siempre queremos transmitir algo de una forma determinada. ¿Qué determina lo que es neutro en la velocidad de elocución o cuándo podemos decidir que un hablante tiene un volumen neutro a la hora de hablar? En este trabajo partimos de una definición de “voz neutra” como aquella que no muestra valores extremos, ni demasiado altos, ni excesivamente bajos: cuanto menos rango haya entre los valores máximos y mínimos, menos diferencia habrá de esos valores y menos posibilidades de modulación de la voz.

No obstante, la tendencia a registros más graves, tanto entre hombres como entre mujeres, parece ser que es lo que caracteriza a una voz radiogénica en español (Rodero, 2001, 2002), definida esta como una voz agradable, armónica, relajada, cálida y transparente (Rodríguez Bravo, 1989). Los usuarios con problemas de visión también prefieren en las audiodescripciones de voces sintéticas una voz grave (Podsiadło y Shweta, 2016), pero sin que exprese ningún tipo de emociones, lo que permitiría reducir la fatiga auditiva y aumentaría la inteligibilidad cuando se tratara de producciones orales sintéticas.

En este trabajo, se han analizado en un corpus de audiodescripciones los parámetros acústicos relacionados con el tono, la intensidad y la velocidad de elocución de los audiodescriptores, masculinos y femeninos, que han sido considerados por separado, ya que, por razones fisiológicas, los valores de la F0 de la mujer siempre son más altos que los de los hombres. Finalmente, se ha preparado una prueba de percepción en la que las participantes, con o sin pérdida de visión, debían decidir, entre tres voces masculinas y tres femeninas, qué voz consideraban más neutra y cuál de ellas les parecía más agradable. Existen investigaciones que indican que las habilidades en la percepción auditiva de las personas con problemas de visión no son las mismas que las de las personas que tienen este tipo de impedimento. Las personas con pérdida de visión

perciben mejor y más rápido los cambios relacionados con la voz, compensan así su falta de visión (por ejemplo, Gougoux et al., 2004; Gougoux et al., 2009). En este trabajo se han distinguido estos dos grupos en el test de percepción para observar si existen diferencias debidas a las capacidades visuales a la hora de seleccionar la voz agradable y la voz neutra. El objetivo es determinar la relación entre neutralidad y agradabilidad, y saber si los hablantes con pérdida de visión perciben estas voces de la misma manera que los que no presentan tal pérdida.

Nuestra hipótesis de partida considerando los artículos consultados es que las voces, tanto masculinas como femeninas, que se consideren agradables se relacionarán con registros de voz graves y, en este sentido, se percibirán como neutras, pues en las frecuencias graves hay menor posibilidad de modulación. Por lo tanto, habrá una relación entre neutralidad y agradabilidad. Además, se supone que las mejoras en las habilidades auditivas de los hablantes con dificultades de visión no alterarán la selección de lo que consideran neutro o agradable en comparación con los hablantes que no tienen ese problema visual.

3. PROCEDIMIENTO EXPERIMENTAL

Las grabaciones de los audiodescriptores profesionales empleadas para preparar los estímulos de la prueba de percepción se han obtenido del corpus Visuals Into Words (VIW) (Matamala, 2018). Este corpus es de acceso abierto y permite comparar la audiodescripción (AD) de diferentes profesionales sobre el mismo contenido visual. El corpus se basa en un cortometraje, *¿Qué pasa mientras?* La película fue creada especialmente para el proyecto

Visuals into Words (VIW, pagines.uab.cat/viw) con el fin de disponer de un corpus de diferentes voces de audiodescriptores profesionales (Matamala y Villegas, 2016). La versión original en inglés fue doblada al español y al catalán. Las audiodescripciones fueron encargadas a profesionales en los tres idiomas. El subcorpus utilizado para nuestro análisis está compuesto por 10 audiodescriptores profesionales del español, cinco voces masculinas y cinco femeninas. El corpus está compuesto por 480 unidades audiodescriptivas. Estas unidades pueden definirse como segmentos textuales relacionados con la representación visual. Cada unidad se ha dividido en grupos de entonación, y se han medido, mediante el programa PRAAT (Boersma y Weenink, 2019), los valores medios de F0 (Hz), de amplitud (dB) y de duración (ms) de cada grupo. Estos parámetros nos han permitido identificar diferencias entre las voces de los hablantes. La Figura 1 muestra un ejemplo del análisis del habla realizado con PRAAT. La línea punteada se corresponde con la intensidad y la línea continua con el tono del hablante.

Después de extraer los datos acústicos de los 10 locutores en un total de 1116 grupos de entonación, se seleccionaron las tres hablantes femeninas y los tres hablantes masculinos que presentaran más diferencia entre ellos, en los valores de los rasgos prosódicos considerados, con respecto al valor medio de todos los hablantes de cada sexo. En la Tabla 1 se muestran las voces seleccionadas para el test de percepción. Se indica, tomando como referencia la media de cada grupo analizado (hombre/mujer), con el signo “+” los valores máximos, con “-“ los valores mínimos y con “med”, los valores medios. En el caso de las hablantes femeninas, siempre hay algún parámetro

Figura 1: Oscilograma, curva de F0 y curva de intensidad del fragmento *el perro olisquea* producido por la informante de la empresa Aptent.

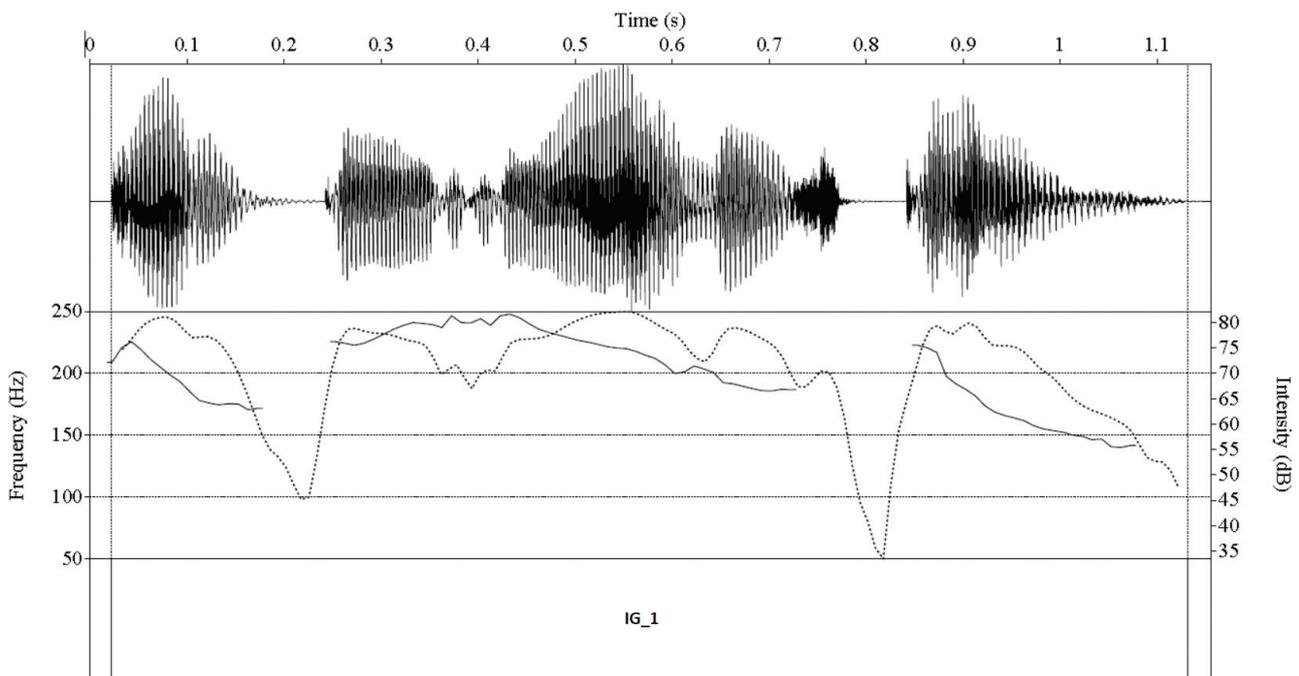


Tabla 1: Diferenciación de los audiodescriptores a partir de los valores de los parámetros acústicos.

	Sexo	Tempo	F0	Amplitud
Locutor 1 (Aptent)	Mujer	+	+	+
Locutor 2 (Aristia)	Mujer	-	-	med
Locutor 3 (Kalleidoscope)	Mujer	+	med	-
Locutor 1 (Edsol)	Hombre	med	-	med
Locutor 2 (Ericsson)	Hombre	-	+	-
Locutor 3 (Soni2)	Hombre	+	med	+

que las identifica por tener un valor máximo o por tener un valor mínimo; en el caso de los hablantes masculinos, no siempre los valores de los parámetros acústicos nos servían para identificarlos porque no había diferencias significativas entre ellos. Esto es lo que ocurre con el Locutor 1, de la empresa Edsol, pues presenta un tempo y un volumen de voz que no lo caracteriza frente a los otros dos hablantes, pero el valor de la F0 nos permite diferenciar a los tres. Para una información más detallada acerca de cómo se clasificaron los locutores según los rasgos prosódicos, véase Machuca, Matamala y Ríos (2020).

Una vez seleccionados los audiodescriptores que se iban a emplear en la parte perceptiva, se elaboró la prueba de percepción con el objetivo de obtener las preferencias de los oyentes en términos de neutralidad y de agradabilidad. La prueba de percepción siguió un procedimiento aprobado por el comité de ética de la UAB, y tuvo una duración aproximada de 30 minutos. Aunque, en un principio, la prueba fue completada por 62 participantes, como únicamente dos de ellos eran hombres, se excluyeron para evitar un posible efecto de género, por lo que, al final, solo se analizaron los datos de las 60 mujeres participantes. Para determinar el efecto de la discapacidad visual, las participantes se dividieron en dos grupos: 29 sin pérdida de visión y 31 con pérdida de visión. Se crearon 9 estímulos para la prueba. Cada uno de ellos estaba compuesto por dos voces del mismo sexo. Se obtuvieron en total 540 respuestas, 60 para cada comparación (60 sujetos * 9 estímulos). 279 de las participantes con pérdida de visión y 261 de las participantes sin pérdida.

Las participantes, usando un formulario Google en línea, debían seleccionar de las dos voces de cada estímulo, primero, la que les parecía más neutra, y, segundo, la que les parecía más agradable. Para finalizar la prueba, los sujetos debían definir brevemente y con sus propias palabras qué consideraban una voz neutra; de esta forma partíamos de su definición sobre voz neutra para observar si en ella había alguna referencia a la agradabilidad.

Las respuestas se recopilaban en un archivo de Excel para su posterior análisis estadístico. Se realizaron pruebas de chi-cuadrado con el programa de software SPSS (v. 25). 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

Los resultados se presentarán en función de las preguntas que debían contestar las participantes en el formulario: sobre la neutralidad y sobre la agradabilidad.

4.1 Neutralidad

Como ya hemos mencionado, los resultados se presentan por separado para las voces masculinas y para las femeninas.

En cuanto a las voces femeninas (Figura 2), no se encontraron diferencias entre las participantes con o sin pérdida de visión a la hora de elegir la voz que sonaba más neutra. Los porcentajes entre la locutora 2 y la 3 no muestran diferencias significativas ($\chi(2) = 540$, $p = 0,42$, $> 0,05$), aunque el porcentaje de participantes que seleccionan la voz de la locutora 2 como más neutra es ligeramente mayor para las participantes sin pérdida de visión. Por otro lado, al comparar la locutora 1 con las otras audiodescriptoras, las respuestas indican que las participantes, independientemente de sus capacidades visuales, no aprecian esa voz como neutra. Recordemos que la locutora 1 presentaba los valores máximos de tono, la velocidad de habla más alta y la intensidad más alta en el volumen de la voz (ver Tabla 1).

Las participantes, tal como se observa en la Figura 3, donde se ilustran los resultados para los audiodescriptores masculinos, no clasifican la voz del locutor 3 como neutra, con porcentajes menores en comparación con los otros dos. El locutor 3 presenta valores máximos para dos de los parámetros analizados: el tiempo empleado en articular los grupos de entonación y el volumen con que los ha realizado. Los resultados también muestran que las participantes prefieren la voz del locutor 2 como representante de una voz neutra. No se encuentran diferencias significativas en los audiodescriptores masculinos si consideramos las condiciones visuales de las participantes ($\chi(2) = 540$, $p = 0,26$, $> 0,05$). Además, ha sido elegido como representante de la voz neutra el mismo locutor tanto para sujetos videntes como para no videntes.

4.2 Agradabilidad

Los resultados sobre agradabilidad muestran que no siempre existen diferencias entre los sujetos que presentan pérdida de visión y los que no presentan esta carencia a la hora de elegir una u otra voz, si consideramos los contrastes de las tres voces femeninas y de las tres masculinas (Tabla 1). Eso es lo que sucede con las voces femeninas, según la prueba de chi-cuadrado. Determinar la agradabilidad en las voces femeninas depende de si las participantes tienen o no problemas de visión ($\chi(2) = 540$, $p = 0,03$, $< 0,05$), no obstante, como veremos más adelante, la locutora que se rechaza como agradable es la misma en los dos casos. La selección de las voces masculinas no presentan diferencias debido a este factor ($\chi(2) = 540$, $p = 0,598$, $> 0,05$).

Figura 2: Preferencias de neutralidad de las voces de las audiodescriptoras del español (en porcentajes).

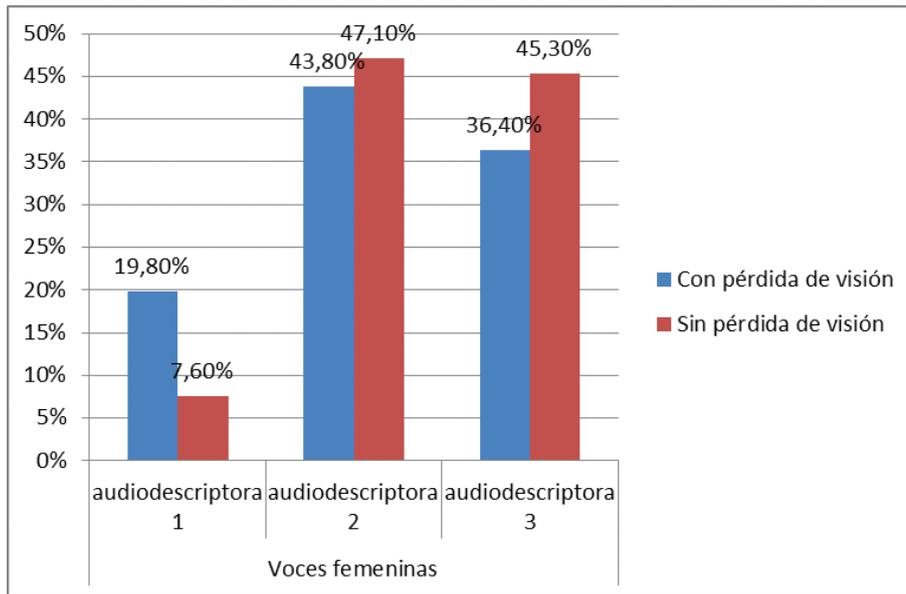
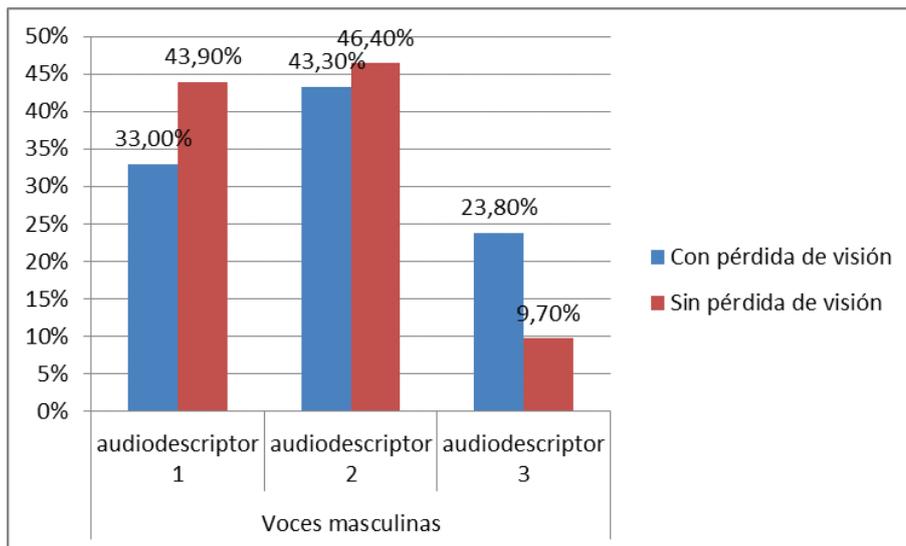


Figura 3: Preferencias de neutralidad de las voces de los audiodescriptores del español (en porcentajes).



Como se puede observar en la Figura 4, que nos muestra los porcentajes de percepción de qué voz femenina les resulta más agradable a las participantes, las que presentan déficit visual no aprecian una distinción clara entre las voces de las locutoras 1 y 2, ambas presentan un porcentaje del 35 %. La locutora 3 posee un porcentaje un poco más bajo, del 30 %. Las dos voces preferidas, aunque con porcentajes que no difieren mucho, presentan los valores extremos de velocidad de habla y de F0 (la locutora 1 tiene los valores máximos y la 2, los mínimos), por lo que podríamos señalar que estos valores, por su oposición, no serían la causa de las preferencias. Además, las locutoras

1 y 3 presentan un valor máximo de velocidad de habla. En cambio, los valores de amplitud pueden servir para diferenciar las tres locutoras, la 1 con un valor máximo, la 2 con un valor medio y la 3 con un valor mínimo. Podríamos señalar que la agradabilidad viene determinada por un volumen de voz más bien alto de las voces femeninas.

En relación con las participantes que no tienen problemas de visión, los resultados son muy similares, aunque la locutora 1 muestra claramente un porcentaje más alto que la 2. En este caso, es evidente que hay un rechazo de la voz de la locutora 3, la que se corresponde con una voz

Figura 4: Preferencias de agradabilidad de las voces de las audiodescriptoras del español (en porcentajes).

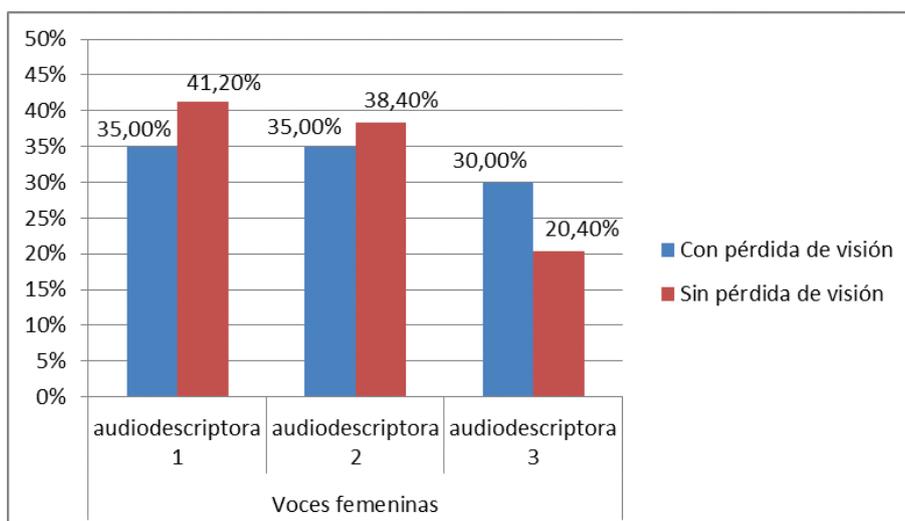
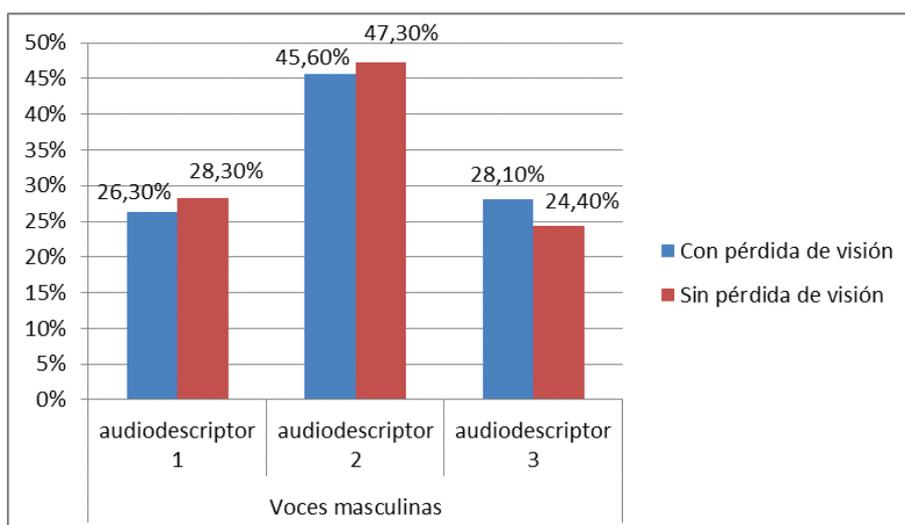


Figura 5: Preferencias de agradabilidad de las voces de los audiodescriptores del español (en porcentajes).



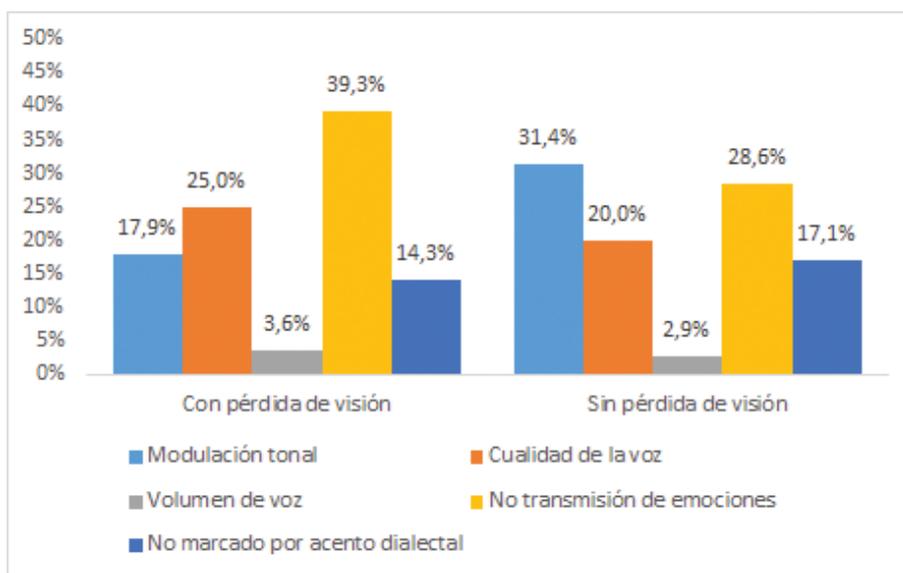
con valores mínimos de amplitud, por lo tanto, con una apreciación de baja intensidad cuando se oye esta voz.

En cuanto a las voces masculinas, la Figura 5, que ilustra los porcentajes de la selección de las participantes en ese tipo de voces, muestra que, tanto los sujetos que tienen problemas visuales como los que no, prefieren el locutor 2. Recordemos que este locutor posee la voz más aguda y el volumen más bajo; además, tarda más en realizar los grupos de entonación.

En el apartado de Metodología, habíamos mencionado que les pedíamos a los sujetos que indicaran al final de la prueba de percepción qué era para ellos una voz neutra. Se han clasificado las respuestas en relación a cinco tipos de categorías: la modulación tonal, la cualidad de la voz, el volumen de voz, la transmisión de emociones y el acento dialectal. Las respuestas no coinciden con el

número de sujetos que había realizado la encuesta, ya que algunos de los sujetos han mencionado en la definición más de un parámetro. Hay un total de 27 respuestas para los hablantes con problemas de visión y de 35 para los hablantes sin dificultades de visión. En la Figura 6 se recogen los porcentajes de frecuencia de aparición de las categorías mencionadas.

Los sujetos con pérdida de visión entienden por voz neutra la que no transmite emociones (39,3 %), mientras que no le dan tanta importancia a la modulación tonal del hablante (17,9%). En cambio, los sujetos que no presentan problemas visuales describen como una voz neutra la que no transmite emociones (28,6%) y no presenta cambios en la modulación tonal (31,4%). El volumen de voz es irrelevante para la descripción de voz neutra, ya que en ambos grupos de sujetos se obtienen los porcentajes más

Figura 6: Porcentaje de frecuencia de aparición de las categorías consideradas.

bajos (3,6 % para los que tienen pérdida de visión y 2,9 % para los que no la tienen).

5. DISCUSIÓN DE LOS RESULTADOS

Iglesias Fernández (2006) indicaba la falta de criterios objetivos medibles con los que evaluar la agradabilidad de las voces. En este experimento hemos pretendido subsanar esa carencia mediante la cuantificación de tres parámetros acústicos: velocidad de habla, tono y volumen de voz. No obstante, los resultados de nuestro experimento indican que no es fácil relacionar los parámetros acústicos con una clara identificación de la agradabilidad de las voces, tanto en voces femeninas como masculinas.

En español, según nuestros resultados, se tiende a preferir voces agudas, que se corresponden con unos valores elevados de frecuencia del tono, tanto para las voces masculinas como para las femeninas. En cambio, Iglesias Fernández (2013) ya había señalado que un tono agudo y un volumen alto, a pesar de la objetividad del concepto, no era percibido como una voz agradable. Lo mismo sucede con aquellos autores que han definido una voz radiogénica a partir de voces graves, tanto para hombres como para mujeres (Rodríguez Bravo, 1989; Rodero, 2001, 2002). Los resultados también muestran que no hay diferencias en la selección de los hablantes considerados agradables entre los dos grupos analizados, si bien, en el caso de las mujeres, la diferencia viene dada por el mayor porcentaje de preferencia y no porque se haya considerado una locutora diferente.

Los resultados también muestran que sigue sin haber una solución clara para la neutralidad, aunque hay un acuerdo entre las participantes de lo que no es una voz neutra: tanto en la voz femenina como masculina se seleccionan como “no neutras” las voces que presentan un valor máximo en los parámetros analizados. Si comparamos los dos

conceptos, no hay una relación entre neutralidad y agradabilidad en las voces femeninas ($\chi^2(2) = 540, p = 0,000, > 0,05$), tanto para los sujetos con dificultades de visión como para los que no las tienen. En este sentido, si nos fijamos en las voces femeninas, la locutora 3, que muestra los valores más bajos, se rechaza como una voz agradable y la locutora 1, que ofrece los valores más altos, se rechaza como una voz neutra, pero se acepta como una de las voces agradables. Podríamos señalar que en las voces femeninas los valores extremos se rechazan como voces neutras y los valores extremos altos se consideran voces agradables. En este caso, los dos grupos de hablantes, los que tienen dificultades visuales y los que no, se comportan de igual forma.

En cambio, para las voces masculinas, aunque los dos grupos de hablantes seleccionan el mismo locutor, tanto para la voz agradable como para la voz neutra, los resultados obtenidos en la prueba de χ^2 muestran que esos dos conceptos no se diferencian en el caso de los participantes con discapacidad visual ($\chi^2(2) = 540, p = 1,9, < 0,05$), pero son diferentes en los sujetos que no presentan problemas de visión ($\chi^2(2) = 540, p = 0,000, > 0,05$).

6. CONCLUSIONES

A la vista de los resultados obtenidos en la prueba de percepción, debemos señalar que para los sujetos es fácil decidir qué no es una voz neutra. En este sentido, siguiendo las definiciones que se comentaban al principio de este trabajo lo “no neutro” sería algo marcado, en oposición a lo que es neutro o no marcado. Los sujetos de los dos grupos analizados han señalado, tanto para las voces femeninas como para las masculinas, que las voces no neutras son las que se caracterizan por un elevado valor de F_0 , una mayor velocidad de elocución y un volumen de voz más alto (intensidad). Yendo un poco más allá en la interpretación de los resultados, se podría decir que

cuando compiten las voces de dos locutores que no se identifican muy bien en cuanto a la neutralidad o que hay muy poca diferencia entre ellos, el rasgo que destacan las participantes en una voz neutra es un mayor volumen de voz, si se trata de voz femenina, pero si se trata de una voz masculina, el rasgo es un tono más agudo.

En el caso de la agradabilidad, en las voces femeninas se identifica claramente que la voz que presenta un volumen bajo no es agradable, en cambio, en la voz masculina se escogen dos de los tres audiodescriptores seleccionados como agradables, la única característica común a los dos locutores es un tono de voz grave.

Por otro lado, en cuanto a la búsqueda de una definición de “voz neutra”, se puede observar que los parámetros que destacan más en esa definición son los mismos para los dos tipos de sujetos. Sin embargo, a los sujetos con pérdida de visión no les preocupa demasiado la modulación tonal; lo que realmente le interesa a este grupo de sujetos es que la voz del audiodescriptor no transmita emociones. En cambio, al sujeto que no presenta dificultades en la visión le preocupa más la modulación tonal, debido, seguramente, a que las emociones que transmiten los personajes le llegan tanto por el canal visual como por el auditivo. Sería necesario, no obstante, llevar a cabo estudios con distintos tipos de audiodescripciones en contexto para confirmar estas observaciones. En cuanto a la hipótesis inicial conforme considerábamos que las voces agradables podrían ser también las seleccionadas como neutras, hemos visto, a partir de los resultados, que no es así y que depende de que el audiodescriptor sea femenino o masculino: el volumen parece ser más importante en la identificación de neutralidad y agradabilidad en la voz femenina, mientras que el tono es más característico en la identificación de una voz masculina. El hecho de que los sujetos con dificultades visuales, a pesar de las mejoras en sus habilidades auditivas, hayan seleccionado, tanto para la neutralidad como para la agradabilidad, los mismos locutores que los sujetos que no presentan este problema nos confirma que hay rasgos en la voz que se identifican de la misma forma, pese a la discapacidad visual.

7. AGRADECIMIENTOS

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).

BIBLIOGRAFÍA

- AENOR. (2005). *Norma UNE:153030. Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. Madrid: AENOR.
- Alter, K. (2002). Suprasegmental Merkmale und Prosodie, En H. M. Müller (Ed.), *Arbeitsbuch Linguistik* (pp.148–169). Paderborn: UTB-Ferdinand Schöningh.
- Banse, R. y Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614-636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Boersma, P. y Weenik, D. (2019). *Praat: Doing Phonetics by Computer* [Computer program]. Electronic version: <<http://www.praat.org>>
- Cabeza-Cáceres, C. (2013). *Audiodescripció i recepció. Efecte de la velocitat de narració, l'entonació i l'explicitació en la comprensió filmica*. Barcelona: TDX. Versión electrónica: <https://www.tdx.cat/handle/10803/113556>
- Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell (6.ª edición).
- Gougoux, F., Belin, P., Voss P., Lepore, F., Lassonde, M. y Zatorre, R.J. (2009). Voice perception in blind persons: A functional magnetic resonance imaging study. *Neuropsychologia*, 47, 2967–2974. <https://doi.org/10.1016/j.neuropsychologia.2009.06.027>
- Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J. y Belin, P. (2004). Neuropsychology: Pitch discrimination in the early blind. *Nature*, 430, 309. <https://doi.org/10.1038/430309a>
- Hidalgo Navarro, A. (2017). Nuevas aportaciones al estudio funcional de la entonación coloquial: propuesta ecléctica de integración de modelos de análisis. *Estudios Filológicos*, 60, 127-150.
- Iglesias Fernández, 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). Sevilla: Bienza.
- Iglesias Fernández, 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). Amsterdam: John Benjamins.
- Kraxenberger, M. et al. (2018). Prosody-Based Sound-Emotion Associations in Poetry. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2018.01284>
- Machuca, M. J., Matamala, A. y Ríos, A. (2020). Prosodic features in Spanish audio descriptions of the VIW corpus. En M. Richart-Marset y F. Calamita (Eds.), *Traducción y accesibilidad en los medios de comunicación: de la teoría a la práctica / Translation and Media Accessibility: From Theory to Practice. MonTI 12*, (en prensa).
- Matamala, A. (2018). One short film, different audio descriptions. Analysing the language of audio descriptions created by students and professionals. *Onomazéin*, 41, 185-207.
- Matamala, A. (2019). *Accessibilitat i traducció audiovisual*. Vic: Eumo.
- Matamala, A. y Villegas, M. (2016). Building an audio description multilingual multimodal corpus: the VIW project. En J. Edlund, D. Heylen y P. Paggio (Eds.) *Proceedings: Multimodal Corpora: Computer vision and Language processing- MMC2016* (pp. 29-32). Paris: ELRA. Electronic version: <<http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-MCC-2016-proceedings.pdf>>
- Pell, M. D., Paulmann, S., Dara, C., Alasserri, A. y Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: a comparison of four languages. *Journal of Phonetics*, 37, 417–435. <https://doi.org/10.1016/j.wocn.2009.07.005>
- Podsiadło, M. y Shweta, Ch. (2016). Text-to-Speech for Individuals with Vision Loss: A User Study. 347-351. <https://doi.org/10.21437/Interspeech.2016-1376>.
- Rodero, E. (2001). *Locución informativa radiofónica*. Tesis doctoral. Universidad Pontificia de Salamanca.
- 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* (pp. 319-331). Madrid: Lex Nova
- Rodríguez Bravo, A. (1989). *La construcción de una voz radiogénica*. Tesis doctoral. Universitat Autònoma de Barcelona
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143. <https://doi.org/10.1037/0033-2909.99.2.143>

Scherer, K. R. (2013). Emotion in action, interaction, music, and speech. En A. A. Micheal (Ed.), *Language, Music, and the Brain: A Mysterious Relationship* (pp. 107–139). Cambridge: MIT Press.

Stolarski, Ł. (2015). Pitch patterns in vocal expression of “Happiness” and “Sadness” in the reading aloud of prose on

the basis of selected audiobooks. *Research in Language*, 13, 141–162. <https://doi.org/10.1515/rela-2015-0016>

Ververidis, D. y Kotropoulos, C (2006). Emotional speech recognition: resources, features, and methods. *Speech Communication*, 48(9), 1162-1181. <https://doi.org/10.1016/j.specom.2006.04.003>