EDICIONES
COMPLUTENSE

CC BY

Círculo de Lingüística Aplicada a la Comunicación
ISSN 1576-4737

VARIA

Voces neutras y alteración tonal

Inés Fernández Alonso
Universitat Autònoma de Barcelona (España)
María J. Machuca
Universitat Autònoma de Barcelona (España)
Anna Matamala
Universitat Autònoma de Barcelona (España)

https://dx.doi.org/10.5209/clac.84236

Enviado: 8 de septiembre de 2022 • Aceptado: 17 de octubre de 2022

ES **Resumen.** Algunas guías recomiendan que la locución en la audiodescripción sea neutra (por ejemplo, AENOR, 2005). Sin embargo, hay pocas investigaciones que aborden qué es una voz neutra desde un punto de vista prosódico. Este artículo presenta los resultados de una prueba de percepción en español en la que se manipula el tono de voz de un hablante con el objetivo de observar si deja de percibirse esa voz como neutra, tal como se había considerado en estudios previos. En el experimento, en el que participan 26 personas con conocimientos de audiodescripción y 29 sin conocimientos, también se pregunta a los participantes cómo definirían una "voz neutra". Ninguno de los participantes presentaba discapacidad visual. Los resultados muestran que las definiciones de voz neutra pueden clasificarse en 10 categorías. También indican que las voces graves se consideran más neutras que las agudas, si bien los participantes siguen identificando como voz neutra la voz original en la que no se ha modificado el tono.
Palabras clave: Prosodia, neutralidad, audiodescripción, accesibilidad, percepción

ENG Neutral voices and tonal modification

Abstract. Current recommendations say that the voicing of audio description should be neutral (for example, AENOR, 2005). However, little research has been carried out into what a neutral voice is from a prosodic point of view. This paper presents the results of a perception test in Spanish in which the pitch of a speaker's voice that was considered in previous studies as neutral is manipulated to determine if the voice continues to be perceived as neutral. The experiment involved 26 people with audio description skills and 29 without, and participants were asked how they would define a "neutral voice." None of the participants had sight loss. The results show that the definitions of a neutral voice can be classified into 10 categories. In addition, the low-pitched voices were considered more neutral than the high-pitched ones, although the participants continued to identify the original voice with no pitch modifications as the most neutral.
Keywords: Prosody, neutrality, audio description, accessibility, perception.

Sumario. Introducción. 1. Marco teórico. 2. Marco metodológico. 2.1. Test de percepción. 3. Resultados. 3.1. Definiciones de "voz neutra". 3.2. Prueba de percepción. 4. Discusión. Conclusiones. Agradecimientos. Contribución de autoría CREdIT. Referencias bibliográficas. Anexo: Formulario empleado para realizar el test de percepción.

Cómo citar: Fernández Alonso, I.; Machuca, M. J. y Matamala, A. (2024). Voces neutras y alteración tonal. *Círculo de Lingüística Aplicada a la Comunicación* 97 (2024), 195-205. https://dx.doi.org/10.5209/clac.84236

Introducción

En el artículo 27 de la Declaración Universal de Derechos Humanos (ONU, 1948) se hace referencia al derecho de las personas a acceder a la cultura: "toda persona tiene derecho a tomar parte libremente en la vida cultural de la comunidad, a gozar de las artes y a participar en el progreso científico y en los beneficios que de él resulten". Desde entonces se ha avanzado hacia una accesibilidad mayor, pero todavía queda mucho camino por recorrer. En este sentido, los distintos servicios de accesibilidad han permitido acceder a contenidos culturales a aquellas personas que no pueden ver, oír o comprender los elementos auditivos o visuales de dichos contenidos. La subtitulación convierte el texto hablado en texto escrito y, en el caso del subtitulado para sordos, ofrece una presentación visual de los sonidos, música y elementos paralingüísticos. También incluye sistemas para identificar a los personajes que hablan (Matamala & Orero,

2010). La audiodescripción (Fryer, 2016; Maszerowska, Matamala & Orero, 2014; Matamala, 2021, entre otros) ofrece una traducción de las imágenes en palabras que se insertan en los espacios sin diálogos ni música significativa de los contenidos. El audiosubtitulado (Matamala, 2022) convierte los subtítulos en textos orales y dan respuesta a las necesidades de aquellos que no ven los subtítulos o no pueden leerlos adecuadamente. Otra modalidad es la interpretación en lengua de signos, que da respuesta a los espectadores que necesitan el uso de la lengua de signos.

En el caso de la subtitulación, hablamos de lengua escrita, pero en el de la audiodescripción, hablamos de lengua oral, es decir, la voz de una persona o de una máquina (voz sintética) nos describe con palabras qué sucede en las imágenes (Maszerowska, Matamala & Orero, 2014), por lo que la voz es el instrumento mediante el que la persona accede a esos contenidos culturales, bien sea una pieza teatral, una ópera o la proyección de una película; por esta razón, es importante considerar cómo la voz de los audiodescriptores afecta a la experiencia del usuario y cuáles son las preferencias de este respecto a la narración auditiva.

Las recomendaciones que se incluyen en la norma española *UNE:153030. Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías* (AENOR, 2005) sobre las voces de los audiodescriptores consideran que debe evitarse la transmisión de emociones para no influir sobre la interpretación que el usuario pueda tener sobre una determinada obra. No obstante, en algunas normas o recomendaciones se valora positivamente que esa voz, supuestamente neutra, aparezca intercalada en algunos fragmentos con una voz que muestre emoción. Es lo que sucede, por ejemplo, con las recomendaciones irlandesas (BAI, 2012) o francesas (Morisset & Gonant, 2008). Las recomendaciones británicas indican que la audiodescripción tiene que ser "unobtrusive and neutral but not lifeless or monotonous and the delivery should be in keeping with the nature of the programme" (ITC 2000, p. 10). Por otro lado, las recomendaciones alemanas no incluyen información sobre este aspecto (Dosch & Benecke, 2004, traducido al inglés en Greening, Ray & Leen, 2010). Con todo, no debemos olvidar la dificultad de establecer la definición de una voz neutra. Machuca, Matamala & Ríos (2020a) señalan, a propósito de los resultados obtenidos en un análisis cuyo objetivo era caracterizar la neutralidad en las voces de los audiodescriptores, que los sujetos que percibían estas voces sabían fácilmente identificar qué era una voz "no neutra", al contrario de lo que ocurría con las voces neutras. Las voces no neutras se definían a partir de algún rasgo prosódico marcado, determinado por un elevado valor tonal (alta vibración de cuerdas vocales, F0), una mayor velocidad de elocución o un volumen de voz más alto.

Las voces no impostadas se diferencian de otras voces por su tono, por la velocidad que presenta y por el volumen que tiene; estas son las características que identifican a un determinado tipo de voz si no consideramos los modos de fonación que puede incluso realizar un mismo hablante en función de cómo modifique la tensión o la vibración de las cuerdas vocales: voz susurrada, voz rota, voz de hálito, entre otros (para una descripción más detallada, véase Gordon & Ladefoged, 2001; Gil, 2007). El tono está relacionado acústicamente con la frecuencia fundamental (F0) y viene determinado articulatoriamente por la frecuencia en la que los pliegues vocálicos se ponen en contacto, por lo que suele denominarse *vibración de cuerdas vocales*. Cuanto mayor sea esa frecuencia de vibración, más agudo será el tono de voz del hablante, cuanto más baja sea la frecuencia, más grave será el tono. Una voz definida como radiogénica en español (Rodero, 2001, 2002), o lo que es lo mismo, agradable de escuchar, tiende a registros más graves (Rodríguez Bravo, 1989). La velocidad del hablante es un rasgo característico del individuo. A pesar de que los sonidos tienen una duración intrínseca, la velocidad del habla viene determinada no solo por el tiempo que el hablante tarda en emitir un sonido sino también por la duración de las pausas cuando habla. Según Rodero (2003), la voz radiogénica no puede tener una velocidad de elocución rápida, porque las noticias no deben ser solo percibidas sino también comprendidas. Lo mismo sucede en los fragmentos de audiodescripción, el contenido también debe ser comprendido para que los usuarios puedan interpretar correctamente el contenido de la pieza audiodescrita. La velocidad lenta del hablante tampoco produce una eficacia comunicativa (Cebrián Herreros, 1994). En español, el número de sílabas, según Quilis, Cantarero y Esgueva (1993) es de 9,5 en porciones de discurso comprendidos entre dos pausas (grupo fónico). Siguiendo a estos mismos autores, la sílaba muestra una duración media de 158 ms, el discurso comprendido entre dos pausas posee un valor medio de 1305,5 ms, y las pausas tienen una duración de 505 ms. Un incremento de estos valores se podría interpretar como una elocución rápida, mientras que una disminución podría ser considerada como una ralentización en el habla. El volumen de voz viene determinado por la fuerza espiratoria relacionada acústicamente con la amplitud de la onda sonora de la señal de habla. Si el aire sale con mayor presión a través de la cavidad oral, la voz de la persona se percibirá con una intensidad mayor, por el contrario, la voz será percibida con una intensidad más baja. Un nivel de 65 dB permite una buena inteligibilidad de la palabra si el ruido ambiente no es mayor de 40 dB (Jackson-Menaldi, 1992). Se considera que la intensidad media de la voz de un narrador es de 75 dB y un valor neutro de la intensidad se sitúa en 56 dB (Verma, Sarkar & Rao, 2015).

A pesar de que la velocidad de elocución propia de un individuo y el volumen de voz con el que habla, como ya se ha mencionado, son características de su voz, la frecuencia fundamental es un parámetro que funciona como marcador individual, incluso aunque se hayan modificado sus propiedades alterando la voz con gritos o con rugidos (Pisanski, Raine & Reby, 2020). De ahí que el objeto de este estudio sea manipular solo el tono de voz de un hablante que ha sido considerado en español como voz neutra para ver si se altera la percepción de neutralidad o sigue considerándose una voz neutra. Asimismo, queremos estudiar cómo definen una voz neutra usuarios con y sin conocimientos en audiodescripción e identificar los puntos en común y las divergencias entre ambos tipos de participantes.

1. Marco teórico

La audiodescripción tiene dos elementos centrales: por un lado, qué contenidos se seleccionan y cómo se trasladan en palabras; por el otro, cómo se locutan dichas palabras (Maszerowska, Matamala & Orero, 2014). Mientras que la selección de contenidos y la reformulación en palabras se ha estudiado ampliamente (véase, por ejemplo, Mazur, 2020 o Remael, Reviers & Vercauteren, 2015), las características prosódicas de la audiodescripción no han merecido tanta atención. La comparación de voces sintéticas con voces humanas fue objeto de estudio por parte de Szarkowska (2011) en polaco y Fernández-Torné & Matamala (2015), en catalán, así como por parte de Walczak y Fryer (2018) en inglés. Se observó que de modo general los participantes aceptaban las voces sintéticas para la audiodescripción, aunque había cierta preferencia por las voces humanas. Iglesias-Fernández, Martínez Martínez y Chica Núñez (2015) relacionaron las características de las voces de los audiodescriptores con la interpretación y llevaron a cabo un estudio con un total de 15 participantes que evaluaba la carga emocional que perciben los participantes en la voz de los audiodescriptores y su impacto en la evaluación de la calidad. En los dos experimentos de su estudio compararon una audiodescripción cuya locución coincidía con la emoción transmitida por el contenido y una audiodescripción en la que no se producía esta relación. Los participantes parecen considerar más positivamente las primeras, aunque los resultados muestran un interés mayor por el contenido lingüístico que por la prosodia propiamente dicha. De Araújo Carvalho, Aves Leao y Teixeira Palmeira (2017) presentan tres estudios sobre prosodia en AD llevados a cabo en Brasil, en los que analizaron distintos aspectos relacionados con la cualidad del habla (fonación, resonancia, articulación, velocidad, tono, volumen, coordinación fonoarticulatoria y ataque vocal, es decir, la forma con la que se emite el sonido). Los resultados de estos estudios indicaron que los participantes mostraron preferencia por voces expresivas que van más allá de las locuciones tradicionales. Por otro lado, Villoslada (2018) analiza auditivamente las características prosódicas y fonéticas de una audiodescripción en español. En ningún caso se han respaldado las apreciaciones perceptivas observadas en su estudio con un análisis acústico de los datos.

En el marco de los proyectos estatales NEA (Nuevos Enfoques sobre Accesibilidad, https://ddd.uab.cat/pub/pagweb/neaproject/) y RAD (Researching Audio Description: Translation, Delivery and New Scenarios, https://webs.uab.cat/rad), Machuca, Matamala & Ríos (2020b) estudiaron los rasgos prosódicos de un corpus de audiodescripciones en español y llevaron a cabo un estudio de percepción para determinar las preferencias entre usuarios con y sin pérdida de visión. Machuca, Matamala & Ríos (2020a) usan el mismo corpus, pero ahondan en el concepto de neutralidad. Machuca & Matamala (2022) amplían el estudio con resultados para el inglés y el catalán. En el contexto de estos proyectos, Jankowska, Pilarcyk, Woloszyn & Kucniecki (2021) presentan los resultados de un experimento en el que se evaluaron dos tipos de entonaciones (AD adaptada/enfática), según la definición de Cabeza-Cáceres (2013). Cábeza-Cáceres considera que la AD adaptada es la que se encuentra habitualmente en las audiodescripciones de Reino Unido, en las que la audiodescripción se adapta al contenido fílmico, mientras que la enfática correspondería a la propia de Estados Unidos. Mediante el análisis del ritmo cardíaco, la respuesta emocional, un cuestionario de presencia y una entrevista, Jankowska, Pilarcyk, Woloszyn & Kucniecki (2021) demostraron la preferencia por una entonación adaptada. Un estudio parecido dentro también del proyecto NEA comparó el audiosubtitulado con efecto de doblaje y con efecto de voz superpuesta usando medidas biométricas como la respuesta galvánica y el ritmo cardíaco junto a un cuestionario (Iturregui-Gallardo & Matamala, 2021; Iturregui-Gallardo & Soler-Vilageliu, 2021). En el audiosubtitulado con efecto de doblaje el diálogo original desaparecía y se sustituía por una locución actuada de los subtítulos. Sin embargo, a diferencia del doblaje tradicional, no había sincronía labial. En el audiosubtitulado con efecto de voz superpuesta, el diálogo original se oía por debajo del audiosubtitulado, tal y como ocurre con los documentales sonorizados mediante una voz superpuesta o *voice-over*. Los resultados no mostraron diferencias significativas entre las dos modalidades en cuanto a impacto emocional.

A pesar de estas investigaciones, todavía faltan estudios específicos sobre características prosódicas concretas como el que aquí presentamos –centrado en el tono– que no solo pueden hacer una aportación al ámbito de la audiodescripción sino que pueden tener un impacto en el ámbito de la prosodia de modo general.

2. Marco metodológico

Con el fin de manipular una voz neutra, se partió de las voces que habían sido elegidas como neutras en el estudio de Machuca, Matamala & Ríos (2020b), en el que se utilizó un subcorpus compuesto por 10 audiodescriptores profesionales del español, extraído del corpus Visuals Into Words (VIW) (Matamala, 2019), para analizar los valores medios de F0 (Hz), de amplitud (dB) y de duración (ms) de las voces de los audiodescriptores, parámetros acústicos que han permitido a los autores caracterizar tales voces, cinco femeninas y cinco masculinas. Después, se seleccionaron las tres voces, tanto del grupo de los hombres como del de las mujeres, que tuvieran más diferencia entre ellos a partir de los valores de los rasgos prosódicos (tono, volumen y velocidad de habla) considerando el valor medio de todos los hablantes de cada sexo. Una vez seleccionadas las voces de estos audiodescriptores, se elaboró una prueba de percepción con el objetivo de obtener las preferencias de los oyentes en términos de neutralidad y de agradabilidad. La prueba fue completada por 60 participantes, divididos en dos grupos en función de su discapacidad visual: 29 sin pérdida de visión y 31 con pérdida de visión. Tanto en las voces femeninas como en las masculinas, los sujetos descartaron aquellas voces que presentaban valores máximos en cualquiera de los parámetros

analizados. Esas voces, una masculina y una femenina, se han seleccionado como neutras, entre otras voces del mismo sexo, y se han empleado en este estudio para obtener los estímulos base sobre los que se van a llevar a cabo las modificaciones tonales.

2.1. Test de percepción

Para la manipulación de los estímulos, se segmentó una unidad de audiodescripción que contenía 11 palabras y tenía una duración de 2,5 s, lo cual nos permitía realizar un experimento de percepción que no fuera demasiado largo y evitar así el cansancio propio de los sujetos que deben valorar voces. El tono del audiodescriptor se alteró mediante el programa Audacity 2.4.2, un editor de audio que permite la modificación del tono en semitonos. La escala en semitonos es más adecuada para la percepción de las voces que en hercios (Rietveld & Gussenhoven, 1985), de esta forma, se normalizan las diferencias tonales que pueden existir entre los dos segmentos manipulados, pues la percepción de la altura musical es la misma. Un tono equivale a dos semitonos, es decir, el intervalo mínimo que existe entre dos notas consecutivas de una escala diatónica. Se crearon estímulos modificando, como máximo, un semitono hacia un registro tonal agudo y hacia un registro tonal grave. Aunque Rietveld & Gussenhoven (1985) indicaron que el umbral en la F0 es de 1,5 semitonos, se manipuló la señal hasta 2 semitonos (1 tono), a partir de esta escala el tono de voz del audiodescriptor se apreciaba tan alterado que su manipulación era obvia. Se obtuvieron cinco estímulos para cada una de las voces, la voz neutra femenina y la masculina: el estímulo base, los dos estímulos modificados un semitono y un tono hacia el registro tonal agudo, y los dos estímulos modificados un semitono y un tono hacia el registro tonal grave. A continuación, para estar seguros de la modificación realizada se analizaron los valores de la F0 para cada una de las vocales de la unidad de audiodescripción seleccionada mediante el programa Praat. La Tabla 1 muestra los valores en hercios de cada uno de los segmentos tomados como estímulos base y de los manipulados 1 tono (t) o 1 semitono (st).

Tabla 1. Valores medios en hercios (Hz) del tono en los estímulos base y en los manipulados

Voz	Estímulo base	+1 st	+1 t	-1 st	-1 t
Femenina	144.67	153.42	161.97	136.76	129.21
Masculina	103.55	109.39	115.56	97.26	91.67

Los sujetos percibían los estímulos en una combinación de dos segmentos. Para las combinaciones se incluyeron siempre el estímulo base, es decir, el audio sin modificar, y uno de los audios modificados. Con el fin de que la posición del estímulo base no influyera en las respuestas de los participantes, se crearon dos estímulos para cada combinación: uno con el audio sin modificar en primera posición y otro con el audio sin modificar en segunda posición. Esto dio lugar a un total de dieciséis estímulos, ocho para cada una de las voces seleccionadas (*estímulo base*+1st*, *estímulo base*+1t*, *estímulo base*-1st*, *estímulo base*-1t*, *+1st* estímulo base*, *+1t*estímulo base*, *-1st*estímulo base* y *-1t*estímulo base*).

La prueba de percepción se llevó a cabo mediante un formulario de Google debido a su facilidad de distribución (véase Anexo), a pesar de que presentaba alguna limitación técnica, como el hecho de que los audios debían incluirse en un archivo de vídeo y que no era posible establecer una presentación aleatoria de los estímulos. Para minimizar este efecto, se decidió ordenar los estímulos por secciones diferentes, de tal modo que en primer lugar se presentaron los estímulos correspondientes a la voz femenina y, en segundo lugar, los estímulos de la voz masculina. Dentro de cada sección se presentaban primero los cuatro audios con locución original, estímulos sin modificar, en primer lugar y luego los cuatro audios con la locución original en segundo lugar. La duración de la prueba era de aproximadamente cinco minutos.

El cuestionario incluía el consentimiento informado, una pregunta sobre la edad del participante, una pregunta acerca de sus conocimientos previos sobre audiodescripción, una pregunta abierta sobre el concepto de "neutralidad" ("Defina qué es para usted una voz neutra") y 16 preguntas cerradas que constituían la prueba de percepción propiamente dicha en la que los participantes tenían que seleccionar una de las tres opciones siguientes, después de haber oído dos estímulos de la misma voz: "La primera es más neutra", "La segunda es más neutra", "Las dos son igual de neutras".

Se recogieron un total de 65 respuestas: 4 tenían menos de 20 años, 55 tenían entre 20 y 30 años, y 6 tenían más de 30 años. Para el análisis se eliminó a los participantes menores de 20 años y a los mayores de 30 años para evitar que este factor influyera en los resultados. Así pues, se analizaron las respuestas de 55 participantes, 26 de ellos tenían conocimientos de audiodescripción y eran fundamentalmente estudiantes del Máster Universitario en Traducción Audiovisual de la UAB. 29 no poseían ningún conocimiento sobre este tema y estaba compuesto por personas ajenas a la traducción y accesibilidad audiovisual. Nos pareció interesante tener en cuenta también esta variable para evaluar las voces de los audiodescriptores.

3. Resultados

Presentamos en apartados diferentes los resultados de las respuestas a la pregunta abierta sobre la definición de una voz neutra (apartado 3.1) y los obtenidos para la prueba de percepción (apartado 3.2).

3.1. Definiciones de "voz neutra"

A partir del análisis de las respuestas recibidas, se categorizaron las respuestas en 10 ítems que podían incidir en las definiciones de una voz neutra: emociones, acento dialectal, tono, entonación, objetividad, timbre, calma/tranquilidad, claridad/buena pronunciación, ritmo/velocidad, musicalidad/melodía, voz no destacable. No se tuvieron en cuenta respuestas aisladas que no se podían incluir en ninguna de las categorías mencionadas por la falta de precisión en la respuesta. A modo de ejemplo, respuestas como "moldeable y con la que se puede trabajar" o "no perturba demasiado al oído" eran definiciones aisladas que no se han considerado para el análisis de esta respuesta abierta y que se han incluido en el recuento dentro de la categoría de "Otros". Al tratarse de una pregunta de respuesta abierta, los resultados obtenidos son bastante variados. Uno de los participantes introdujo como respuesta en el formulario «no lo sé», por lo que se excluirá de esta sección del análisis quedando para ella un total de 26 participantes con conocimientos sobre audiodescripción y 28 sin conocimientos sobre audiodescripción.

Un 82,14 % de los participantes sin conocimientos de audiodescripción ofrecieron una única característica como respuesta, mientras que solo el 34,86 % de los participantes con conocimientos de audiodescripción definieron la voz neutra mediante una única característica. Las definiciones de estos participantes, los que sabían qué era y para qué servía una audiodescripción, eran mucho más precisas, como lo demuestra el hecho de que su respuesta generalmente iba asociada a más de una característica. Así, un 38,46 % de los participantes con conocimientos de audiodescripción aportaron dos características (10,71 % en el caso de los que no tenían conocimientos); 15,38 % aportaron tres características (contra un 3,57 % del otro grupo), y 7,69 % aportaron cuatro (contra un 3,57 %). A pesar de estas diferencias, hay una gran coincidencia en las características escogidas, con pequeñas variaciones, tal y como se ilustra en la Figura 1.

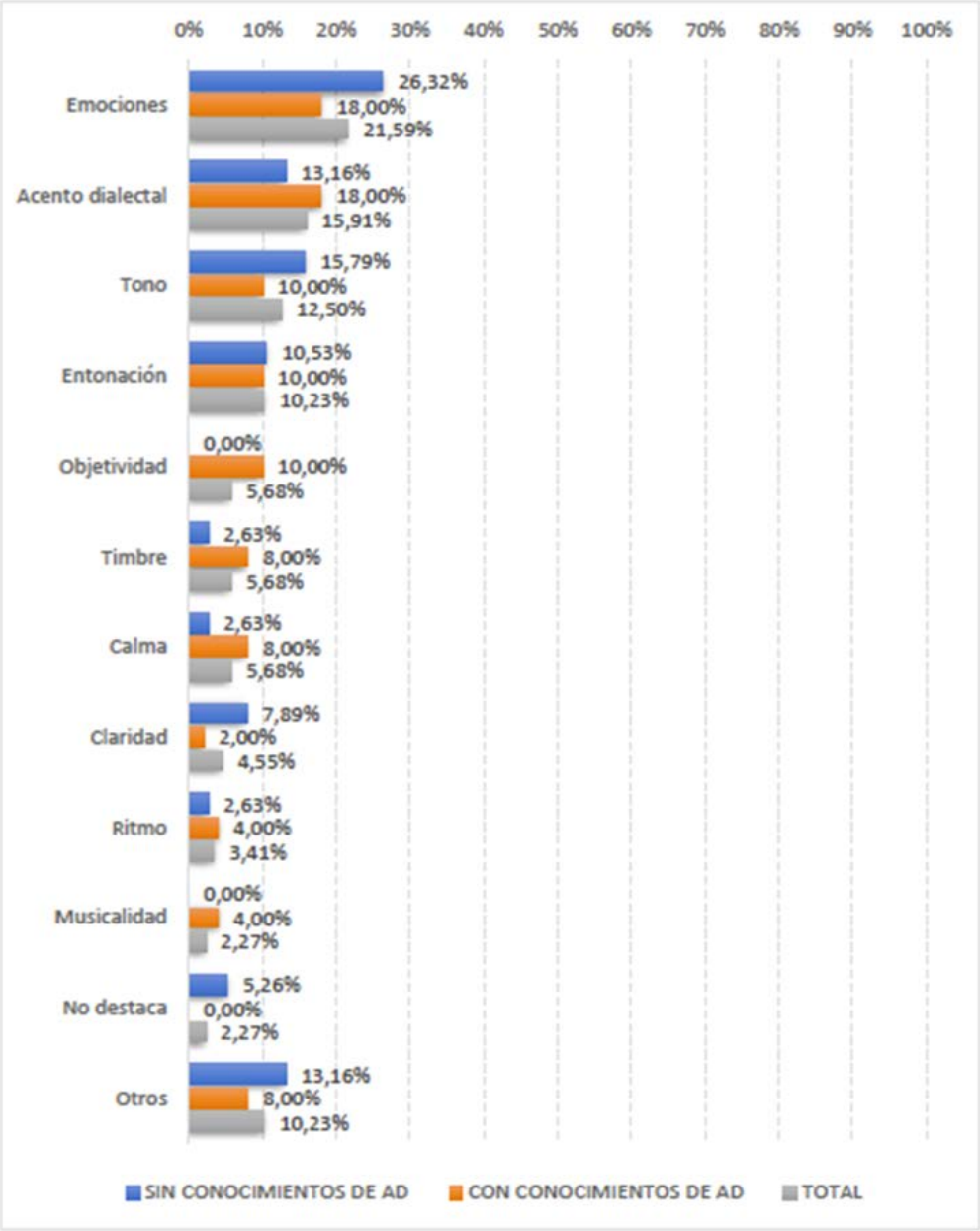

Figura 1. Porcentaje de aparición de la categorización de las respuestas

Como se puede observar en el gráfico (Fig. 1), el rasgo que se menciona de modo más frecuente por parte de ambos grupos de participantes tiene relación con las emociones o sentimientos: "aquella voz que no genera sentimientos", "aquella que no transmite emoción", "una voz que no suscita ningún tipo de emoción (alegría, tristeza, intriga, desconfianza...)". En segundo lugar aparecen respuestas con mención al acento dialectal que los participantes consideran que debe ser estándar y no marcado: "alguien que habla un idioma con acento neutro o estándar", "sin emociones ni acentos", "sin acentos concretos", "una voz sin acento marcado y que a ningún hablante de esa lengua le suene extraña". Esta característica ocupa la tercera posición en los participantes sin conocimientos sobre audiodescripción, que indican el tono como segunda característica importante. La mayoría de participantes indican que este no debe ser ni demasiado grave ni demasiado agudo, aunque de modo aislado un participante considera que una voz neutra es "una voz con una frecuencia baja". Tanto en el total de respuestas como en el grupo con conocimientos de audiodescripción el tono ocupa la tercera posición. En cuarto lugar aparece en ambos grupos la entonación: una voz neutra se define como "plana", "monótona" o "sin variaciones", pero también como "buena" o "controlada". Estas cuatro primeras categorías corresponden al 60,23 % de las respuestas. En el 40 % restante es donde se aprecian más diferencias entre los participantes.

Es relevante comentar que el concepto de objetividad solo lo mencionan participantes con conocimientos de audiodescripción. Esto se debe probablemente al hecho de que un debate recurrente en esta área es el de la objetividad o subjetividad de las audiodescripciones, tal como exponen Mazur y Chmiel (2012) y Limbach (2013). Cabe señalar que Limbach usa el término "neutralidad" para referirse al hecho de no añadir valoraciones adicionales o subjetivas en la audiodescripción, no para referirse a la locución neutra. Si bien el debate se ha centrado en la selección de contenidos y su traducción en palabras, los participantes lo trasladan al ámbito de la voz: "voz que transmita el mensaje con el tono más objetivo posible", "una voz que transmita objetividad".

En cuanto a timbre o calma, las cifras globales son idénticas, con diferencias entre ambos grupos: mientras que el porcentaje de aparición es del 8 % en aquellos con conocimientos de audiodescripción, el porcentaje se reduce a un 2,63 % entre los que no tienen dichos conocimientos. Los comentarios indican que una voz neutra es una "voz pausada", "calmada", "con un timbre y tono agradable". El resto de características no cuentan con suficientes respuestas para comparar ambos grupos.

Las respuestas obtenidas nos ayudan a definir las características de una voz neutra como un tipo de voz que no transmite emociones, plana, en la que no deben apreciarse cambios tonales y en la que tampoco haya rasgos que nos indiquen la procedencia geográfica del hablante.

3.2. Prueba de percepción

Los resultados de percepción muestran que, a pesar de haber separado las voces masculinas y femeninas en dos sesiones diferentes, la alteración de esas voces ha producido un efecto similar, de manera que no se obtienen resultados diferentes en relación con el sexo del audiodescriptor. La prueba no paramétrica de chi cuadrado de Pearson muestra que no hay diferencias significativas en las respuestas de los participantes en función del sexo del hablante ($p = 0,92$), por lo que hemos tratado las alteraciones hacia una voz más grave o más aguda, independientemente de que estas procedan de una voz femenina o de una voz masculina. Los resultados que se muestran en la Figura 2 son muy claros: las voces agudas se rechazan como neutras, mientras que las voces graves son más aceptables. Los participantes siguen identificando, en general, como voz neutra la voz original en la que no ha habido modificaciones de tono. No obstante, cuando la voz original se modifica un semitono hacia una escala más grave, los participantes la identifican como voz neutra de igual forma (46,82 %) que la voz original (31,82 %), incluso un 21,36 % de los participantes la prefieren a la voz original, hecho que no ocurre con las otras manipulaciones de la voz. Si la voz se manipula un semitono hacia la escala de los agudos, también se identifica de igual manera que la voz original en un 35 % de los casos, no sucede lo mismo con las voces manipuladas un tono hacia cualquiera de las dos tonalidades. Las respuestas en función de que los participantes hayan escogido la original, la manipulada o les sirva cualquiera de las dos (igual) presentan un nivel de significación de $p < 0,05$, por lo que el rango tonal influye en la percepción de una voz neutra en los sujetos.

Por último, el hecho de que los participantes tuvieran o no conocimientos sobre audiodescripción influye sobre las respuestas en cuanto a cómo perciben una voz neutra. La prueba de chi2 arroja un valor de $p < 0,05$ considerando este factor. No obstante, en la Figura 3 puede observarse que las preferencias entre los estímulos con la voz modificada y la original presentan las mismas tendencias, aunque el porcentaje es diferente si comparamos los participantes con conocimiento de audiodescripción y los que no lo poseen. Así, los estímulos con una voz de un semitono más grave son considerados neutros de igual forma que la voz original en un 57,69 % de los casos para los sujetos con conocimiento de AD y solo en un 37,07 % para los que no disponen de él. Esto significa que los dos grupos consideran que la voz grave es una voz neutra, pero con proporciones diferentes, lo que hace que la variable con o sin conocimientos sobre audiodescripción resulte significativa.

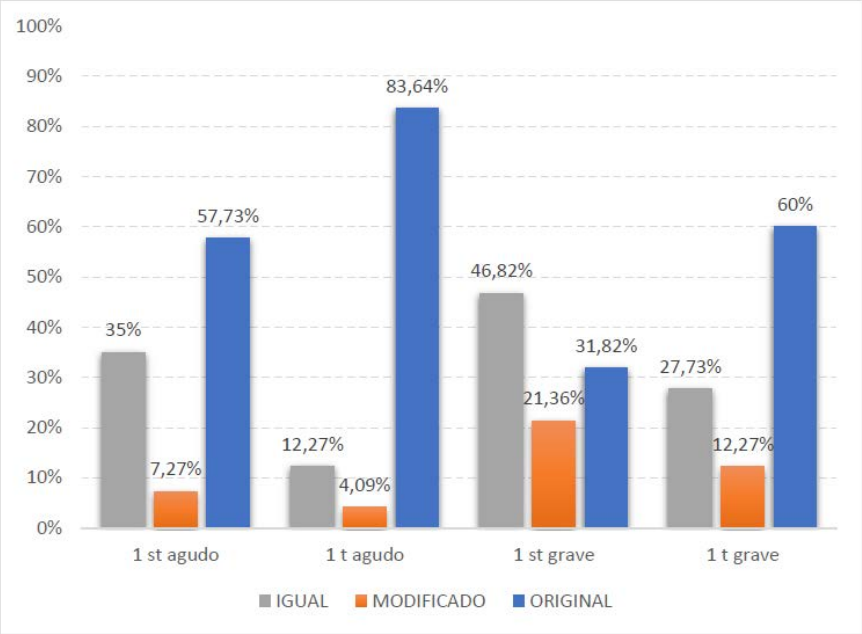

Figura 2. Porcentaje de voces identificadas como neutras en función de la alteración tonal

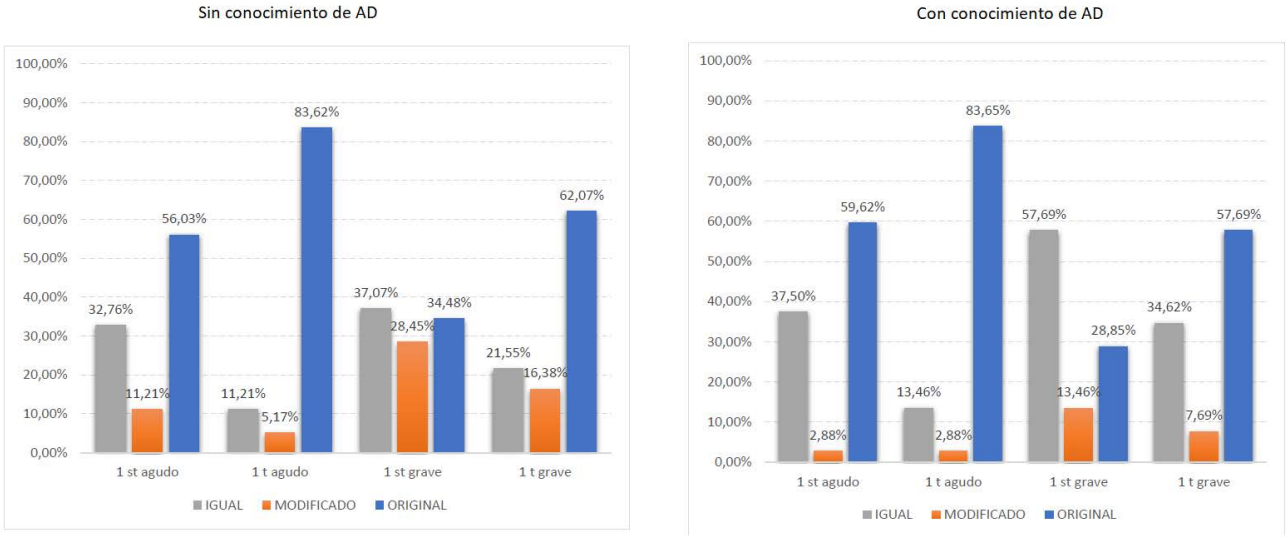

Figura 3. Percepción de la neutralidad considerando el conocimiento sobre audiodescripción (AD)

4. Discusión

Los resultados obtenidos muestran, por un lado, que las definiciones relacionadas con la voz, y sobre todo con la neutralidad, están enlazadas con aspectos que pueden transmitir diferentes emociones, que nos pueden proporcionar información geográfica del hablante o que pueden indicarnos un tono determinado del hablante. La ausencia de este tipo de informaciones en la voz se relaciona con una voz neutra, desde el punto de vista del usuario.

Cabe señalar también que es imprescindible establecer criterios objetivos medibles para la caracterización de las voces. A propósito de esto, Iglesias Fernández (2013) ya menciona la subjetividad que hay a la hora de definir las voces. Por ejemplo, la definición de una voz agradable, en la que se han manejado, según esta autora, aspectos tan subjetivos como pueden ser la entonación, la fluidez, el tempo y la dicción. No obstante, aunque trabajos como los de Machuca, Matamala & Ríos (2020a) han demostrado que la neutralidad y la agradabilidad no van de la mano, los autores definen la "voz neutra" a partir del análisis acústico de lo que los sujetos han percibido como una voz no neutra en español: las voces no neutras poseen un elevado valor de F0, es decir un tono agudo, una mayor velocidad de elocución y un volumen de voz más alto.

Otro de los objetivos de este trabajo ha sido observar cómo la manipulación del tono hacia frecuencias más altas (voces agudas) o más bajas (voces graves) puede alterar la percepción de la neutralidad en las voces. Es curioso observar que tanto para voces masculinas como para femeninas los registros graves, modificados en un semitono, son considerados tan neutros como la muestra original, mientras que los registros agudos se rechazan como voces neutras. Rodero (2003: 6) ya había señalado la preferencia por las voces graves porque transmiten "credibilidad, veracidad, seguridad y tranquilidad", además de que

una voz grave está relacionada con una voz más intensa, por lo que no hay necesidad de incrementar el volumen de la voz. En la misma línea, Rodríguez Bravo (1989), en la búsqueda de parámetros para definir una voz radiogénica, manifestaba que las voces graves en español eran más aceptadas que las agudas, aunque, como señala Rodero (2003), no es la única condición para obtener una buena locución radiofónica, la intensidad de la voz y la velocidad de habla son importantes también para que los oyentes asimilen el mensaje que se transmite. No obstante, algunos trabajos como el de Re et al. (2012) muestran que en inglés las preferencias por un tono más grave o agudo dependen del sexo, así para los hombres se prefieren voces con tonos más graves y para las mujeres se prefieren voces con tonos agudos. Quizá habría que analizar también hasta qué punto las preferencias sobre el tono de la voz o el hecho de que un tono determinado sea considerado neutro depende de la lengua de los audiodescriptores. Este no parece ser el caso en español, las preferencias siempre tienden hacia voces más graves.

Conclusiones

Las conclusiones extraídas del análisis de los resultados se centran en dos aspectos claves: uno, que cualquier hablante se siente con capacidad de definir términos que, en principio, son difíciles de describir y, dos, que la percepción de la voz grave tiende más hacia la neutralidad que la de la de una voz aguda.

Las definiciones aportadas por los participantes sobre una voz neutra sugieren que este término está ligado a las emociones que puede percibir el hablante, al acento que indica la zona geográfica de un determinado individuo y a la tonalidad de la voz. Los hablantes que tienen formación en audiodescripción son capaces de dar diferentes acepciones de este término en contraposición a los sujetos que no saben nada sobre la audiodescripción. Un aspecto que vale la pena destacar es cómo el debate sobre la objetividad de la audiodescripción –que en principio está circunscrito al ámbito de la selección de qué es lo que se audiodescribe y con qué palabras se audiodescribe– surge también en relación con la voz. Cabe señalar que algunos autores como Limbach (2015: 150) prefieren hablar de neutralidad, en vez de objetividad, ya que "it covers precisely the idea of hot the AD should be a neutral description of the film which does not reflect the audio describer's personal opinion or preferences and which transmits the film's communicative function and value without changing the director's intended effect in any way". Sin embargo, en este caso estaríamos hablando de una audiodescripción neutra, no de una voz neutra, que es lo que se pedía a los participantes que definieran.

En cuanto a la manipulación del tono de los hablantes, cuya voz se ha seleccionado como neutra en anteriores estudios, se puede observar que si las manipulaciones se hacen hacia una escala tonal grave siguen interpretándose como neutras para la audiodescripción, pero si se modifican hacia una escala tonal aguda, dejan de reconocerse como una voz neutra. Esto puede ponerse en relación con las preferencias de las voces, en español hay una preferencia de las voces graves como voces más estéticas, por lo que también suelen ser las elegidas para voces neutras. En este sentido, podemos afirmar que existe una relación entre voces neutras y voces graves. Sería interesante también indagar el papel de otros parámetros relacionados con la voz, como la intensidad o la velocidad de elocución, en la categorización de una voz neutra. Queda, además, por analizar las voces en contextos más amplios, en géneros diversos (de documentales a películas de distintas temáticas), combinándose con diálogos emitidos por voces con distintas características prosódicas. Las posibilidades son múltiples puesto que la prosodia en la audiodescripción es uno de los temas que menos se ha tratado en comparación con otros aspectos más relacionados con la selección de contenido y su traducción en palabras.

Agradecimientos

Esta investigación forma parte del proyecto RAD (Researching Audio Description: Translation, Delivery and New Scenarios), código de referencia PGC2018-096566-B-I00. Las autoras son miembros de TransMedia Catalonia, un grupo de investigación financiado por el Gobierno catalán bajo el esquema de financiación SGR (2021SGR00077).

Contribución de autoría CREdIT

Las tres autoras han contribuido de igual manera en el desarrollo de este artículo.

Referencias bibliográficas

AENOR. (2005). *Norma UNE:153030. Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. Madrid: AENOR.

BAI (2012). BAI guidelines – audiodescription [en línea]. Disponible en: https://www.bai.ie/en/media/sites/2/dlm_uploads/2018/05/20171220_AccessRules_MinisterReport_vFinal_DMcL.pdf

Cabeza-Cáceres, Cristobal (2013). Audiodescripció i recepció. Efecte de la velocitat de narració, l'entonació i l'explicitació en la comprensió fílmica [en línea]. Barcelona: TDX. Disponible en: https://www.tdx.cat/handle/10803/113556

Cebrián Herreros, Mariano (1994). Información radiofónica. Mediación técnica, tratamiento y programación. Síntesis: Madrid.

De Araújo Carvalho, Wilson Júnior, Alves Leao, Bruna & Teixeria Palmeira, Charleston (2017). Locuçao e audiodescriçao nos estudios de traduçao audiovisual. *Trabalhos em Lingüística Aplicada*, 56(2). https://doi.org/10.1590/010318138649286277551

Dosch, Elmar & Benecke, Bernd (2004). *Wenn aus Bildern Worte werden. Durch Audio- Description zum Hörfilm*. Bayerischer Rundfunk.

Fernández-Torné, Anna & Matamala, Anna (2015). Text-to-speech versus human-voiced audio descriptions: a reception study in films dubbed into Catalan. *JosTrans. Journal of Specialised Translation*, 24, 61-88.

Fryer, Louise (2016). *An Introduction to Audio Description. A practical guide.* Londres: Routledge.

Gil, Juana (2007). *Fonética para profesores de español: de la teoría a la práctica*. Madrid: Arco/Libros.

Gordon, Matthew & Ladefoged, Peter (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29, 383–406.

Greening, Joan, Petré, Leen & Rai, Sonai (2010). *A Comparative Study of Audio Description Guidelines Prevalent in Different Countries*. RNIB. http://audiodescription.co.uk/uploads/general/RNIB._AD_standards.pdf

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

Iglesias Fernández, Emilia, Martínez Martínez, Silvia & Chica Núñez, Antonio (2015). Cross-fertilization between reception studies in audio description and interpreting quality assessment: the role of the describer's voice. En R. Baños Piñero & J. Díaz-Cintas (Eds.), *Audiovisual translation in a global context* (pp. 72-95). London: Palgrave Macmillan.

ITC (2000). *ITC Guidance on Standards for Audio Description*. http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf

Iturregui-Gallardo, Gonzalo & Matamala, Anna (2021). Audio subtitling: dubbing and voice-over effects and their impact on user experience. *Perspectives. Studies in Translatology*, 29(1), 64-83.

Iturregui-Gallardo, Gonzalo & Soler-Vilageliu, Olga (2021). Audio subtitling and subtitling: a comparison of their emotional effect on blind/partially sighted and sighted users. *Onomazéin.* Special Issue VIII-Emotions in Translation and Interpreting, 61-82.

Jackson-Menaldi, María Cristina (1992). *La voz normal*. Buenos Aires: Editorial Media Panamericana.

Janckowska, Anna, Pilarcyk, Joanna, Woloszyn, Kinga & Kucniecki, Michal (2021). Enough is enough: how much intonation is needed in the vocal delivery of audio description? *Perspectives. Studies in Translation Theory and Practice*. https://doi.org/10.1080/0907676X.2022.2026423

Limbach, Christiane (2013). El aspecto de la neutralidad en la audiodescripción fílmica. En Ortega Arjonilla (ed.) *Translating Culture, Traduire la Culture, Traducir la Cultura*, Granada: Comares, 793–806

Machuca, María J. & Matamala, Anna (2022). Neutral voices in audio descriptions. What does it mean? *Babel*. *Revue internationale de la traduction.* DOI 10.1075/babel.00287.mac

Machuca, María J., Matamala, Anna & Ríos, Antonio (2020a). Los audiodescriptores: voces neutras y voces agradables. Loquens, 7(2), e076. https://doi.org/10.3989/loquens.2020.076.

Machuca, María J., Matamala, Anna & Ríos, Antonio (2020b). Prosodic features in Spanish audio descriptions of the VIW corpus. *MonTI*, 12, 53-77.

Maszerowska, Anna, Matamala, Anna & Orero, Pilar (2014). (Eds) *Audio description. New perspectives illustrated*. Ámsterdam: Benjamins.

Matamala, Anna & Orero, Pilar (Eds.). (2010). *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing.* Peter Lang. https://doi.org/10.3726/978-3-0351-0147-8

Matamala, Anna (2019). The VIW corpus: multimodal corpus linguistics for audio description analysis. *RESLA. Revista Española de Lingüística Aplicada*, 32(2), 515-542.

Mazur, Iwona & Chmiel, Agnieszka (2012). Audio Description Made to Measure: Reflections on Interpretation in AD Based on the Pear Tree Project Data. https://pdfs.semanticscholar.org/7d50/5dc4f112633517c9ca8a0d4d2d73e7d3539e.pdf.

Morisset, Laure & Frédéric, Gonant (2008). *La charte de l'audiodescription*. Ministère des Affaires Sociales et de la Santé. Disponible en: https://www.sdicine.fr/wp-content/uploads/2015/05/Charte-de-laudio-description-1008.pdf

NEA (2015) Nuevos Enfoques sobre Accesibilidad: modalidades híbridas, inmersión y tecnología en audiodescripción. Ministerio de Economía y Competitividad (FFI2015-64038-P, MINECO/FEDER, UE). https://ddd.uab.cat/pub/pagweb/neaproject/

ONU. Asamblea General. (1948). *Declaración Universal de Derechos Humanos*, 217 A (III). París. Disponible en: https://www.un.org/es/about-us/universal-declaration-of-human-rights.

Pisanski, Katarzyna, Raine, Jordan & Reby, David (2020). Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *Royal Society Open Science*, 7 (2). https://doi.org/10.1098/rsos.191642

Quilis, Antonio, Cantarero, Margarita & Esgueva, Manuel (1993). El grupo fónico y el grupo de entonación en el español hablado**.** *Revista de Filologia Espanola* 73(1/2), 55-64.

Re Daniel E., O'Connor Jilian, Bennett Patrick J. & Feinberg David R. (2012). Preferences for very low and very high voice pitch in humans. *PLoS* One, 7(3): e32719. doi: 10.1371/journal.pone.0032719.

Remael, Aline, Reviers, Nina & Vercauteren, Gert (2015) *Pictures painted into words: ADLAB Audio Description guidelines*. EUT Edizioni Università di Trieste. http://hdl.handle.net/1077/11838

Rietveld, Antonious C. & Gussenhoven, Carlos (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13(3), 299–308.
Rodero, Enma (2001). *Locución Informativa Radiofónica*. Tesis doctoral. Universidad Pontificia de Salamanca.
Rodero, Enma (2002). El tono de la voz masculina y femenina en los informativos radiofónicos: un análisis comparativo. *Mujeres, Hombres y Medios de Comunicación*. Madrid: Lex Nova, 319-331.
Rodero, Enma (2003). Para hablar de manera correcta y agradable desde la cabina. La voz informativa radiogénica. *Revista mexicana de comunicación*, 25, 36-41.
Rodríguez Bravo, Ángel (1989). *La construcción de una voz radiogénica*. Tesis doctoral. Universitat Autònoma de Barcelona.
Szarkowska, Agnieszka (2011). Text-to-speech audio description: towards wider availability of audio description. *JosTrans*. *Journal of Specialised Translation*, 15, 142-162.
Verma, Rashmi, Sarkar, Parakrant & Rao, K. Sreenivasa (2015). Conversion of Neutral Speech to Storytelling Style Speech. ICAPR 2015 - 2015 8th International Conference on Advances in Pattern Recognition. 10.1109/ICAPR.2015.7050705.
Villoslada, Ana (2018). La lengua de la audiodescripción del guion audiodescrito de Ocho apellidos vascos. *TRANS: revista de traductología*, 22, 61-80.
Walczak, Agnieszka & Fryer, Louise (2018). Vocal delivery of audio description by genre: measuring users' presence. *Perspectives. Studies in Translatology*, 26(1): 69-83. https://doi.org/10.1080/0907676X.2017.1298634

Anexo: Formulario empleado para realizar el test de percepción

Sección 1

El objetivo de esta encuesta es conocer su opinión sobre la neutralidad de las voces que se comparan en cada fragmento.

Al hacer la valoración, debe tener en cuenta únicamente las voces e ignorar el contenido de lo que están diciendo. El contenido se repite, aunque no es el mismo en todos los casos.

En primer lugar, se le hará una pregunta abierta. Responda con total libertad.

En segundo lugar, escuchará 16 pares de voces. Se le hará una pregunta sobre cada par de voces y tendrá que responder seleccionando una de las dos.

Al hacer clic en «Siguiente» se le mostrará un nuevo grupo de voces que tendrá que comparar.

Es importante que responda tanto sobre el primer grupo de voces como sobre el segundo. Recuerde prestar atención a las voces y no a lo que dicen.

Cuando llegue al final de la encuesta, haga clic en el botón «Enviar».

Finalizar esta encuesta le llevará unos 10 minutos. No recopilaremos ningún dato personal. Solo le preguntaremos por su edad aproximada.

Si tiene cualquier pregunta o quiere tener más información sobre la investigación y su desarrollo, ya sea antes de empezar el test o una vez finalizado, puede ponerse en contacto con Inés Fernández por correo electrónico (se indicaba el correo). La responsable de esta encuesta es la profesora del Máster Universitario en Traducción Audiovisual de la UAB Anna Matamala, puede ponerse en contacto con ella por correo electrónico (se indicaba el correo).

Contestar «Sí» a la primera pregunta implica:

- que ha recibido y entendido la información sobre el estudio
- que ha tenido la posibilidad de hacer preguntas sobre el estudio
- que ha entendido que la participación es voluntaria y puede abandonar la encuesta en cualquier momento, sin necesidad de dar ninguna explicación y sin repercusiones negativas
- que da libremente su consentimiento para participar en este estudio

Muchas gracias por su participación.

1. ¿Da su consentimiento para participar en este estudio?
   - o Sí
   - o No
2. ¿Cuál es su edad?
   - o Menos de 20 años
   - o Entre 20 y 30 años
   - o Más de 30 años
3. ¿Tiene conocimientos sobre audiodescripción para ciegos?
   - o Sí
   - o No

Sección 2

4. Defina qué es para usted una voz neutra
5. Al hacer clic en Audio 1 escuchará dos voces describiendo una escena. Indique cuál le parece más neutra.
   - La primera es más neutra
   - La segunda es más neutra
   - Las dos son igual de neutras
6. Al hacer clic en Audio 2 escuchará dos voces describiendo una escena. Indique cuál le parece más neutra.
   - La primera es más neutra
   - La segunda es más neutra
   - Las dos son igual de neutras

(El resto de las preguntas eran iguales hasta un total de 16 audios)