



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons: http://cat.creativecommons.org/?page_id=184



ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons: <http://es.creativecommons.org/blog/licencias/>



WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license: <https://creativecommons.org/licenses/?lang=en>



UNIVERSITAT AUTÒNOMA DE BARCELONA

**Departament de Traducció i d'Interpretació
i d'Estudis de l'Àsia Oriental**

Doctorat en Traducció i Estudis Interculturals

**Audiodescripción y propuesta de estandarización de
volumen de audio:
hacia una producción sistematizada**

**Tesis doctoral presentada por:
Alicia Rodríguez**

**Dirigida por:
Dr. Pilar Orero**

2017

AGRADECIMIENTOS

A todos los que durante estos tres años me ayudado a pavimentar un camino que de otra forma me hubiera sido intransitable. Gracias por haber estado ahí y por haberme acompañado. Especialmente les agradezco la confianza y el ánimo a mi directora de tesis, Pilar Orero, y a Jose Luis y Concha, mis “mentores” industriales, por apoyarme siempre y facilitarme el trabajo, sin vosotros esto nunca se hubiera puesto en marcha.

Gracias a mi familia, por estar y por ser. A mis padres por no desesperar ni dudar de que podía conseguirlo; a mi hermano por ponerme ejemplos de los maravillosos absurdos de la vida y, especialmente, gracias a mi hermana Cristina. Sin tu ciencia, que a mí me parece magia, no hubiera sido posible.

RESUMEN

Para que la audiodescripción empiece a producirse a gran escala y forme parte de las parrillas de emisión en televisión de manera sistemática, los aspectos técnicos formales tienen que estar estandarizados. Uno de los aspectos clave para que el servicio funcione como servicio de accesibilidad es el volumen de audio. El volumen de la audiodescripción debe ser suficiente para que los comentarios se escuchen con claridad en todo momento, pero no existe hasta la fecha indicación concreta al respecto. Esta tesis se centra en el volumen de audio como elemento de evaluación de la calidad en la audiodescripción.

La primera parte de la tesis hace una revisión bibliográfica centrada en el tratamiento que ha recibido el sonido como elemento de la audiodescripción en diferentes documentos: artículos científicos, guías de buenas prácticas y estándares de regulación. También se hace una revisión de actuales legislaciones sobre accesibilidad audiovisual en distintos países europeos.

La segunda parte de la tesis presenta dos tests de usuario realizados en 2015 y 2017 respectivamente. El primero de ellos se llevó a cabo exclusivamente con usuarios con discapacidad visual y evalúa la influencia de la mezcla de audio sobre diferentes elementos sonoros de una audiodescripción. El segundo, llevado a cabo tanto con usuarios con discapacidad visual como con usuarios sin problemas de visión, se centra en la diferencia de volumen mínima recomendable entre una locución y un ruido de fondo para asegurar que la inteligibilidad de la voz no se ve afectada.

El objetivo es establecer un rango de volumen de audio testado con usuarios que se pueda aplicar de manera sistemática en el proceso de producción de audiodescripción.

ABSTRACT

In order that audiodescription starts to be produced on a large scale basis and becomes part of television broadcast grids its formal and technical aspects must be standardized.

One of the key aspects that must be regulated in order that the service works effectively as access tool is volume. Audio volume of the audiodescription track must be loud enough to allow the comments be clearly heard in all scenes. Currently there is no specific standard on audio volume level for audiodescription. This thesis focuses on volume as key quality evaluation element in audiodescription.

The first part of the thesis makes a bibliographical review of the approach taken by research papers, regulatory standards and good practices manuals, towards the role of sound in audiodescription. Also, a review of current legislation on audiovisual accessibility laws in different European countries is made, as well as an introduction to basic concepts of the sound wave and digital audio recording.

The second part of the thesis presents two user tests carried out in 2015 and 2017. The first one was carried out exclusively with visually impaired people and evaluates the influence of the audio mixing technique over different audio elements of an audiodescription. The second test was carried out with both visually and non visually impaired users and focuses on minimum audio volume difference that should exist between the audiodescription track and background noise to keep the intelligibility of the narration. The objective is to establish a range of volume in decibels that can be systematically applied in the production and broadcasting process of audio description.

RESUM

Perquè l'audiodescripció comenci a produir-se a gran escala i formi part de les graelles d'emissió en televisió de manera sistemàtica, els aspectes tècnics formals han d'estar estandarditzats. Un dels aspectes clau perquè el servei funcioni com a servei d'accessibilitat és el volum d'àudio. El volum de l'audiodescripció ha de ser suficient perquè els comentaris s'escoltin amb claredat en tot moment, però no hi ha fins a la data indicació concreta al respecte. Aquesta tesi es centra en el volum d'àudio com a element d'avaluació de la qualitat en l'audiodescripció.

La primera part de la tesi fa una revisió bibliogràfica centrada en el tractament que ha rebut el so com a element de l'audiodescripció en diferents documents: articles científics, guies de bones pràctiques i estàndards de regulació. També es fa una revisió d'actuals legislacions sobre accessibilitat audiovisual en diferents països europeus.

La segona part i de la tesi presenta dos tests d'usuari realitzats en 2015 i 2017 respectivament. El primer d'ells es va dur a terme exclusivament amb usuaris amb discapacitat visual i avalua la influència de la barreja d'àudio sobre diferents elements sonors d'una audiodescripció. El segon test, dut a terme tant amb usuaris amb discapacitat visual com amb usuaris sense problemes de visió, se centra la diferència de volum mínima recomanable entre una locució i un soroll de fons per assegurar que la intel·ligibilitat de la veu no es veu afectada. L'objectiu és establir un rang de volum d'àudio testat amb usuaris que es pugui aplicar de manera sistemàtica en el procés de producció d'audiodescripció.

ÍNDICE

AGRADECIMIENTOS.....	5
RESUMEN	7
ABSTRACT	9
RESUM.....	11
ÍNDICE DE TABLAS Y FIGURAS.....	15
ÍNDICE DE ACRÓNIMOS Y ABREVIATURAS.....	17
CAPÍTULO 1. INTRODUCCIÓN.....	19
1.1. Introducción	21
CAPÍTULO 2. MARCO TEÓRICO Y METODOLÓGICO	23
2.1. Introducción	27
2.2. Marco teórico.....	29
2.2.1. AD como objeto de estudio.....	29
2.2.2. AD como servicio de accesibilidad.....	33
2.2.3. AD como objeto sonoro	37
2.3. Marco metodológico	43
2.4. Descripción de los siguientes capítulos y cómo aplica la metodología.....	45
2.5. Conclusiones.....	47
CAPITULO 3. SONIDO, PERCEPCIÓN Y ACCESIBILIDAD.....	49
3.1. Introducción	53
3.2. ¿Física, psiocaústica o accesibilidad?.....	55
3.2.1. Características de la onda sonora e implicaciones en la AD.....	56
3.2.2. Sonido, sistema auditivo y audio digital	58
3.2.3. Audio analógico y audio digital	61
3.3. El sonido en TV: retrasmisión AD	65
3.3.1 Sonido y enmascaramiento; enmascaramiento y AD.....	66
3.4. Conclusiones.....	71
CAPITULO 4. AD COMO AUDIO SOBRE AUDIO: EL PROBLEMA DEL ENMASCARAMIENO	73
4.1. Introducción	77
4.2. Hipótesis y objetivos.....	81

4.3. Metodología	83
4.3.1 Participantes	83
4.3.2. Estímulos.....	84
4.3.3 Criterios de selección de la película y descripción de escenas	85
4.3.4. Creación de guión y grabación de la AD	86
4.3.5. Diseño del experimento, cuestionario y desarrollo del experimento	87
4.3.6. Análisis estadístico.....	88
4.4. Resultados	91
4.5. Discusión	101
4.6. Conclusiones	105
CAPITULO 5. VOZ FRENTE A RUIDO: BÚSQUEDA DE UN MARGEN DE SEGURIDAD	107
5.1. Introducción	111
5.2. Hipótesis y objetivos.....	113
5.3. Metodología	115
5.3.1. Participantes.	115
5.3.2. Creación de estímulos: ruido.....	116
5.3.3. Creación de estímulos: voz	119
5.3.4. Diseño del experimento.....	120
5.3.5. Desarrollo del experimento y cuestionario asociado.....	123
5.3.6. Análisis estadístico.....	124
5.4. Resultados	127
5.5. Discusión	131
5.6. Conclusiones	133
CAPÍTULO 6. CONCLUSIONES Y FUTURA INVESTIGACIÓN.....	135
6.1. Introducción	139
6.2. Mirando hacia el futuro.....	141
BIBLIOGRAFÍA	143
ANEXOS	151
Anexo 1. Tabla de distribución de escenas y mezclas de sonido.....	153
Anexo 2. Cuestionario del experimento de mezclas de sonido.	155

ÍNDICE DE TABLAS Y FIGURAS

Figura 3.1. Bandas de frecuencia de instrumentos musicales y la voz.....	57
Tabla 3.1. Niveles de presión sonora de ruidos y sonidos típicos y sonoridad subjetiva asociada.....	58
Figura 3.2. Umbral de audición en función de la frecuencia.....	60
Figura 3.3. Umbral de audición junto con las zonas correspondientes a música y palabra.....	60
Figura 3.4. Modificación del umbral de audición en presencia de un tono puro de diferentes intensidades.	67
Tabla 4.1. Escenas obtenidas tras la aplicación de los tres tratamientos de sonido.	88
Tabla 4.2. Respuestas de los usuarios a las 9 preguntas.....	93
Tabla 4.3. Valores de P referentes a la comparación entre las mezclas bajar volumen y dinámica con respecto a la mezcla 1 emisión.	95
Figura 4.1. Gráfica de los resultados obtenidos para la pregunta 1.....	96
Figura 4.2. Gráfica de los resultados obtenidos para la pregunta 3.....	97
Figura 4.3. Gráfica de los resultados obtenidos para la pregunta 5.....	97
Figura 4.4. Gráfica de los resultados obtenidos para la pregunta 7.....	98
Figura 4.5. Gráfica de los resultados obtenidos para la pregunta 8.....	99
Tabla 5.1. Características de los participantes.....	116
Figura 5.1. Espectro de frecuencias del ruido 1.	118
Figura 5.2. Espectro de frecuencias del ruido 2.	118
Figura 5.3. Espectro de frecuencias de la voz del locutor.	119
Tabla 5.2. Correspondencia entre tiempo y aumento de volumen del ruido en el modelo A.....	121
Tabla 5.3. Correspondencia entre tiempo y aumento de volumen del ruido en el modelo B.....	122
Figura 5.4. Esquema del experimento realizado para determinar el volumen de AD más adecuado	123
Tabla 5.5. Tabla de cotejo	124
Figura 5.5. Volumen del ruido de fondo que hace incómoda la escucha de la AD a los participantes	127
Tabla 5.6. Volumen del ruido de fondo que los participantes estimaron incómodo para seguir la AD	128

Figura 5.6. Volumen del ruido de fondo al que los usuarios se sienten incapaces de seguir la narrativa de la AD.	129
Tabla 5.7. Volumen del ruido de fondo que los participantes las hacía imposible seguir la AD.....	130

ÍNDICE DE ACRÓNIMOS Y ABREVIATURAS

.mp3	MPEG-2 Audio Layer III
.wav	Wave
AC3	Acoustic Coding 3
ACC	Advanced Audio Coding
AD	Audiodescripción
AENOR	Asociación Española de Normalización y Certificación
BAI	Broadcasting Authority of Ireland
BBC	British Broadcasting Corporation
BSO	Banda sonora original
Codec	codificador-decodificador
CSA	Conseil supérieur de l'audiovisuel
dB	Decibelio
DVB-T	Difusión de Video Digital - Terrestre
EBU	European Broadcasting Union
FFG	Film Funding Law
FM	Frecuencia Modulada
HD	Alta definición
Hz	Hercio
ITC	Independent Television Commission
ITU	International Communication Union
kHz	Kilohercio
LUFs	Loudness Unit Full Scale
Mbps	Megabits por segundo
MHP	Multimedia Home Platform
MPEG-2	Moving Picture Experts Group
NICAM	Near Instantaneous Companded Audio Multiplex
NPS	Nivel de presión sonora
OFCOM	Office of Communications
ONCE	Organización Nacional de Ciegos de España
RMS	root-mean-square
S/N	Señal/Ruido
SPL	Sound Pressure Level
TDT	Televisión Digital Terrestre
UN	United Nations
W/m ²	Vatios metro cuadrado
W3C	World Wide Web Consortium

CAPÍTULO 1. INTRODUCCIÓN

1.1. Introducción

Cuando en 2014 surgió la oportunidad de hacer este doctorado yo ya no contemplaba la posibilidad de seguir estudiando. Con la licenciatura y un master terminados, quería desarrollar una carrera profesional y no pensaba en volver a la universidad. Entonces surgió la posibilidad del doctorado industrial. La empresa Sonidos, tras años produciendo audiolibros accesibles, empezaba en el mundo de la audiodescripción (AD) y yo podía participar en el proyecto: una investigación conjunta universidad/empresa que duraría tres años. La presente tesis recoge los resultados de esta experiencia. Durante tres años he tenido la posibilidad de participar en congresos y acudir a conferencias pero también de enfrentarme a fechas de entrega y peticiones de cliente. He leído artículos de investigación y documentos de buenas prácticas y he visto realizar grabaciones, mezclas y correcciones de audio. He recordado criterios académicos olvidados y aprendido conceptos de sonido nuevos y, en definitiva, he desarrollado una investigación ligada a la industria. El resultado es una tesis de carácter muy concreto cuyo objetivo es aportar datos que ayuden a alcanzar un estándar de volumen de audio de AD que sea aplicable en TV, asumible por empresas y emisoras y que vaya en favor del usuario. Un estándar de este tipo es necesario no sólo para mejorar la calidad del servicio, sino para poder producirlo a gran escala. Desde el principio la idea conductora de la investigación fue la búsqueda de un estándar que ayudara a sistematizar, acelerar y abaratar el proceso de producción de AD y uno de los grandes temas pendientes en este sentido era la normalización de volumen de audio. El estándar de volumen de audio para AD, a diferencia del volumen de audio general permitido en TV, no está regulado, y esa falta de regulación va en detrimento de la calidad del servicio. Desde diferentes asociaciones y agencias de estandarización (AENOR, ITU, ISO) se han publicado normas y códigos de buenas prácticas que en teoría funcionan como regulación para el sector pero cuyas recomendaciones no tienen demasiada aplicación práctica.

Parte del problema surge en parte por la falta de un concepto global de AD como servicio de accesibilidad que implique conjuntamente a legislación, industria, ámbito investigador y sociedad. La mayoría de los documentos de investigación consultados enfocan la AD como concepto y exploran los aspectos más teóricos del servicio,

relacionados con la traducción imagen-palabra, y el guion. La legislación se limita a garantizar que el servicio se presta pero no regula su calidad, y los proveedores consideran el producto como un elemento más de una cadena de producción que tiene que ajustarse a los requerimientos de cliente.

Esta investigación atravesó varias fases, la primera de las cuales se centró en la revisión bibliográfica del tratamiento del tema del sonido en los diferentes documentos disponibles. Esta fase queda reflejada en el capítulo 2 de esta tesis. La revisión confirmó que había una brecha en cuanto a documentación sobre tratamiento técnico de sonido en el proceso de producción de AD. La idea de enfocar la investigación en el tema específico del volumen de audio surgió de la experiencia personal y profesional al comprobar cómo la AD en televisión presenta diferencias de volumen de audio muy acusadas entre cadenas televisivas e incluso entre programas de la misma cadena. Además desde la experiencia como proveedores, se comprobó que archivos de audio con la misma calidad y formato podían sonar de manera muy diferente dependiendo de los parámetros de emisión, lo que denota una doble falta de estándar: en producción y en emisión. El capítulo 3 de esta tesis expone conceptos básicos de sonido que tienen que ver con la calidad de audio digital, cómo se almacena, cómo se transmite y las consecuencias que ello tiene para al AD televisiva. Estos dos primeros capítulos conforman la base teórica de la segunda parte de la tesis que presenta y analiza los resultados de dos experimentos llevados a cabo para evaluar aspectos concretos de la mezcla de sonido y el volumen de audio en AD. El último capítulo presenta las conclusiones alcanzadas y propone futuras vías de investigación.

CAPÍTULO 2. MARCO TEÓRICO Y METODOLÓGICO

2.1. Introducción

2.2. Marco teórico

2.2.1. AD como objeto de estudio

2.2.2. AD como servicio de accesibilidad

2.2.3. AD como objeto sonoro

2.3. Marco metodológico

2.4. Descripción de los siguientes capítulos y cómo aplica la metodología

2.5. Conclusiones

2.1. Introducción

Este capítulo presenta una descripción del marco teórico y metodológico en el que se ha desarrollado la investigación. Dentro del marco teórico, se distinguen tres enfoques distintos de AD ligados a tres sectores diferentes: AD como objeto de estudio cuando lo aborda la investigación, AD como herramienta de accesibilidad y servicio al ciudadano cuando lo abordan las leyes, y AD como objeto de estudio cuando se toma el punto de vista de la producción audiovisual. El punto 2.3. encuadra la investigación en un la línea de estudios los estudios de recepción iniciada ya por otros autores como Matamala (2005) o Rai (2009) y expone brevemente el objetivo de los experimentos presentados en esta tesis. El punto 2.4. da una visión global de la estructura de la tesis. Las conclusiones introducen una cuestión que se desarrollará más en profundidad a lo largo de la tesis y que es la falta de un estándar de calidad y un concepto de AD global, que abarque tanto los aspectos teóricos como los de producción y provisión del servicio

2.2. Marco teórico

Durante los últimos años, diferentes factores han contribuido al aumento de la audiodescripción (AD) para cine y televisión (Salway, 2007) y esto se ha reflejado en un aumento de documentación disponible al respecto. Desde leyes y documentos legales a códigos de buenas prácticas y *white papers* técnicos, pasando por artículos académicos y tesis doctorales, los documentos que conforman el marco teórico de esta tesis provienen de ámbitos dispares. Dado el diferente tratamiento que recibe el servicio de AD según el sector que lo enfoque, la documentación se ha dividido en tres grupos: documentos que consideran la AD como objeto de estudio (artículos académicos, tesis), documentos que consideran la AD un servicio de accesibilidad (leyes, normativas y códigos de buenas prácticas) y, mucho más escasos, los que abordan la AD como actividad audiovisual (*white papers*, páginas web).

2.2.1. AD como objeto de estudio

“Audio description is a description of visual information delivered via an audio channel” (Salway, 2007 p. 151). La definición propuesta por Salway presenta la multiplicidad de canales y códigos que toman parte en el proceso de la AD y encaja con la visión teórica presente en muchos artículos académicos. Es una concepción de la AD que la enfoca sobre todo como discurso (descripción), que traduce un referente que traduce (imagen) a través de un canal (el sonido).

En el ámbito académico, la AD se estudia dentro de la especialidad de la traducción audiovisual y la mayor parte de la investigación se centra en los aspectos lingüísticos de la AD (Gambier, 2008), más que en los sonoros. El primer registro académico del concepto que hoy se tiene de la AD es la tesis de Frazier titulada *The Autobiography of Miss Jane Pitman: An All-audio Adaptation of the Teleplay for the Blind and Visually Handicapped* (Piety, 2004). Piety hace mención a esta tesis de Frazier en un artículo que, ya desde su título (“The language system of AD: An Investigation as a Discursive process”) ilustra la concepción esencialmente discursiva de la AD que prevalece en el entorno académico.

Piety (2004) habla de hacer accesible la imagen a las personas con discapacidad visual pero ya desde el título, subraya que esto se hace desde el discurso. En el artículo no sólo la AD se aborda como texto, sino que también se alude a la información visual que transmite como un tipo de texto tanto cultural y educacional. El concepto de AD como texto más que como impulso sonoro es el dominante en la mayoría de artículos de investigación, tanto desde el punto de vista de traducción (Kruger y Orero, 2010; Igareda y Matamala, 2012) como desde el de la lingüística (Salway, 2007).

Gambier (2006) cuestiona la visión de la AD como proceso de traducción tradicional y critica la paradoja existente entre la acentuación de la relación entre lo verbal y no verbal como base de la AD por un lado, y el hecho de que la mayor parte de la investigación se centre en los aspectos lingüísticos, por otro. De acuerdo con Gambier (2006), aunque ningún texto es estrictamente monomodal, la multimodalidad está especialmente presente en la traducción audiovisual y la investigación debería reflejarlo. El concepto de multimodalidad lo explica en términos de redacción de guion y distingue entre tres tipos de características que ilustran el carácter multimodal de la traducción audiovisual. Uno de estos grupos es lo que él llama “atmósfera” y es importante que el traductor audiovisual la tenga en cuenta. Gambier (ibid) incluye en la atmosfera el sonido, caracterizaciones, vestuario, movimientos de cámara y operaciones de edición y post producción. Estos son elementos en muchos casos visuales pero en todos no verbales y en el guion de AD han de ser especialmente tenidos en cuenta.

Maszerowska (2014) destaca en su tesis doctoral la escasez de investigación centrada en los aspectos no lingüísticos de la AD reafirmando la paradoja a la que aludía Gambier. Ambos reivindican la importancia los lenguajes no verbales del producto audiovisual para la AD y traducción audiovisual respectivamente pero ambos se centran en cómo verbalizarlos. La verbalización está en la base de la mayoría de definiciones académicas de la AD. Snyder (2005 p. 935) define la AD como “the visual made verbal”, Braun (2008) enfatiza la naturaleza de traducción intersemiótica o intermodal de la AD ya afirmada por otros autores (Benecke, 2007; Bourne y Jiménez Hurtado, 2007; Braun, 2007; Orero, 2005) pero coincide con Snyder (ibid) en que su característica más distintiva es que implica “traducir” imágenes (texto visual) a palabras (texto verbal). Para Jimenez Hurtado (Hurtado y Siebel, 2008) la AD para personas con discapacidad visual es una técnica de traducción destinada a aumentar la

accesibilidad de los medios audiovisuales que partiendo de un texto audiovisual crea un nuevo texto, que es el guion audiodescrito.

No se puede negar que verbalización de lo visual es el rasgo más distintivo de la AD pero centrar el concepto de verbalización exclusivamente en el texto escrito sería algo reduccionista. La AD de espectáculos audiovisuales, lo que Braun (2008) llama “AD dinámica” depende del sonido tanto como de la imagen. Primero, porque la posibilidad de insertar texto y crear un guion de AD viene determinada por el sonido y segundo, porque ese texto se transmite por canal sonoro. Desde el punto de vista conceptual, una AD es el resultado de traducir los elementos visuales a palabras pero desde un punto puramente funcional (hacer lo accesible lo inaccesible), no existe texto en AD hasta que no se sonoriza. Si se tiene que leer, no es AD.

La mayor parte de normas de AD y códigos de buenas prácticas mencionan el sonido desde dos puntos de vista: el funcional el narrativo. El punto de vista funcional es el que más frecuentemente se encuentra y al que se le da mayor importancia: el sonido delimita los huecos donde se insertarán los comentarios descriptivos. En segundo lugar y con menos frecuencia, se alude al sonido como elemento narrativo a tener en cuenta en el guion.

Las ITC *Guidance On Standards for Audio Description* (ITC, 2000) establecen como una de las tres reglas doradas de AD no insertar nunca descripción sobre diálogo o comentario. En el paso 5 de los 7 en que esta guía divide el proceso de preparación de un guion de AD, se aconseja bajar el volumen de audio del fondo allí donde se inserte un comentario para asegurar que se escuche y entienda claramente. Más adelante, el documento dedica una sección a la descripción de sonidos importantes y cómo deberían ser tenidos en cuenta en el guion, anticipándolos, si es posible, y nunca describiendo sobre ellos. La norma alemana (Benecke y Dosch, 2004) enfatiza que la idea fundamental de la AD es que esta se produzca entre diálogos y, en sentido estricto, sólo cuando hay completo silencio, aunque más tarde reconoce lo poco realista de la idea y deja a criterio del guionista si describir o no sobre la banda sonora.

El sonido ha recibido tradicionalmente mucha menos atención que la imagen en los artículos académicos sobre AD (Szarkowska y Orero, 2014) y más cuando se trata no de los sonidos que el guion de AD deben tener en cuenta, sino del que la propia

narración aporta. No obstante hay una serie de artículos que se centran específicamente en el sonido como objeto de estudio.

Remael (2012) analiza el papel narrativo del sonido utilizando la escena de desembarco del film de Spielberg “*Salvar al soldado Ryan*” (Spielberg,1998) para introducir la cuestión más básica a tener en cuenta en la AD: dónde están los huecos. Aunque las normas establecen que las descripciones se insertan en los huecos entre diálogo, como se ha mencionado, Benecke y Dosch (2004) ya admiten que esta es una idea poco realista. Establecer qué se puede considerar silencio y su posible función narrativa cuando se produce, es una cuestión que abordan algunos autores planteando varias preguntas.

Szarkowska y Orero (2014) hacen un interesante estudio de los diferentes papeles que el sonido juega el film de Tarantino (2009) *Inglorious Basterds* y su importancia para el discapacitado visual en la AD. El artículo enfatiza la el papel narrativo del sonido en el film y cómo representa en sí mismo un lenguaje que interactúa con la imagen y debe ser tenido en cuenta en el guion descriptivo. Es particularmente interesante la inclusión en este artículo del silencio como parte del lenguaje fílmico y su discusión no sólo desde el punto de vista narrativo, sino también técnico. Más allá de la cuestión, casi filosófica, de hasta qué punto el silencio no es también un sonido, las autoras se centran en cómo abordarlo desde el guion de AD. El silencio en una producción audiovisual no es una casualidad sino una decisión tomada por el director que puede cumplir una función dramática y narrativa y como tal debe entenderlo el guionista de AD. Es el guionista quien decide cómo utilizar este silencio pero las autoras advierten de la falta de datos sobre la impresión que la ausencia de descripción podría causar en el público. Teniendo en cuenta que el silencio es precisamente el lugar en el que se insertan las descripciones, mantenerlo demasiado tiempo sin narración podría llevar al usuario a pensar que ha habido algún error o que la AD se ha desconectado. Esta reflexión es muy interesante porque vincula una investigación académica con la realidad del servicio. Las autoras advierten que más allá de las decisiones sobre el papel, está la experiencia del usuario y que este es un punto que no debe olvidarse. Esta es una visión poco común en un entorno académico en el que Gambier (2006) ya criticó el exceso de especialización como un problema.

Fryer (2010) llama la atención sobre la tendencia de la investigación a centrarse en la AD como texto escrito y pasar por alto el hecho de que ese texto siempre llega a audiencia por canal sonoro. Para Fryer (2010) la AD es una experiencia auditiva en la que la narración de la AD interactúa con la banda sonora y diálogos originales para crear un “*audio drama*” (p. 205). En esta autora, lo que Braun (2008) llama AD dinámica es una traducción de audio e imagen a exclusivamente audio. Este concepto es muy interesante porque plantea la AD como experiencia y desplaza así el foco de estudio del discurso al usuario, acercando el concepto de AD al de servicio al ciudadano y contribuyendo a tender puentes entre esos espacios excesivamente especializados a los que aludía Gambier (2006).

Uno de los factores que hacen que ciertos aspectos de la AD hayan recibido menos atención investigadora es la gran cantidad de interventores que toman parte en ella y su alto grado de especialización (Gambier, 2004). El hecho de que el servicio de AD sea relativamente joven en comparación con la idea de AD como forma de traducción, hace que algunos aspectos importantes para el usuario sigan sin regular. Temas como el volumen de audio o la mezcla de sonido se han convertido en un problema desde que las cuotas de AD en TV han aumentado. Además, el cambio a la televisión digital ha planteado un nuevo escenario en cuanto a requerimientos técnicos (retransmisión, volumen, consumo de banda) que afecta a la AD y que tiene que regularse para asegurar la consistencia del servicio.

Es importante aquí distinguir entre sonido como impulso sonoro, como canal de audio en la producción audiovisual y como fuente de información en lo narrativo. En relación con el servicio televisivo, los dos niveles más relevantes son el audio como impulso sonoro y como canal de audio.

2.2.2. AD como servicio de accesibilidad

La inclusión de la AD entre los servicios de accesibilidad y su regulación por ley es relativamente reciente. Los derechos humanos de personas con discapacidad (Convention on the Rights of Persons with Disabilities, UN) se redactaron entre 2002 y 2006 (Schulze, 2010) y las diferentes leyes de accesibilidad surgieron también a partir de esas fechas.

La legislación ha sido sin duda uno de los motivos del aumento de la AD ya que establece cuotas mínimas de prestación del servicio en los distintos ámbitos culturales. A nivel europeo el acceso a los medios se contempla en las leyes de cada país y aunque las políticas varían, la tendencia es a aumentar progresivamente las cuotas de los servicios de AD, subtítulo y lengua de signos¹. Sin embargo, las diferentes leyes regulan cantidad no calidad. En Reino Unido, la Equity Act (2010) regula la accesibilidad de los medios de comunicación públicos, aunque no los privados.

En el documento Telling our Story: 2011 Equality Information Report² refleja que los porcentajes de AD para los diferentes canales, excepto BBC News, oscilan entre un 13,3% de BBC Two y el 31,5% de BBC Four. El documento también presenta los objetivos y compromisos de la BBC con la accesibilidad, no sólo televisiva sino de todos sus servicios.

En Francia, de acuerdo con la ley No. 86-1067³ sobre libertad de comunicación, la otorgación de licencias televisivas a compañías privadas está subordinada a la firma de un contrato entre la autoridad que regula el sector audiovisual (Conseil supérieur de l'audiovisuel, CSA) y la compañía (art. 28). En 2011 una recomendación con motivo de la campaña de elecciones establece que los canales de televisión con una audiencia media anual de más del 2,5% de audiencia total eran responsables de ofrecer acceso a los programas de contenido electoral (CSA, Deliberation about political pluralism in radio and TV broadcasts during the electoral period (Délibération relative au principe de pluralisme politique dans les services de radio et de télévision en période électorale), 4 January 2011.⁴

En Irlanda, el artículo 38 de la Broadcasting Act de 2009⁵ establece que la Broadcasting Authority of Ireland (BAI) debe redactor un informe anual que incluya una revisión de los progresos hechos para aumentar la accesibilidad televisiva y en particular, los progresos hechos para cumplir los objetivos que se marquen por ley.

¹ Recuperado de <http://fra.europa.eu/en/publications-and-resources/data-and-maps/comparative-data/political-participation/audiovisual-standards> [fecha de consulta: octubre 2016].

² Recuperado de www.bbc.co.uk/diversity/strategy/equality-information-report

³ Recuperado de www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006068930.

⁴ Recuperado de www.csa.fr/Espace-juridique/Deliberations-et-recommandations-du-CSA/Recommandations-du-CSA-en-vue-de-consultations-electorales-ou-referendaires/Deliberation-du-4-janvier-2011-relative-au-principe-de-pluralisme-politique-dans-les-services-de-radio-et-de-telivision-en-periode-electorale.

⁵ Recuperado de www.irishstatutebook.ie/eli/2009/act/18/section/38/enacted/en/html#sec38.

La BAI establece en sus Access Rules⁶ unas tasas del 2,5% y 5% del total de la programación audiodescrita para los RTÉ One y Two, y RTÉ jr respectivamente, para 2018.

En Italia no hay obligación legal que regule la AD o la accesibilidad televisiva, aunque la ley determina que los criterios de accesibilidad para operadores de televisión públicos que se establecen en el contrato entre la RAI⁷ y el ministerio de comunicación, se renueven cada tres años (Testo unico dei servizi audiovisivi e radiofonici)⁸. Actualmente la RAI se compromete a aumentar progresivamente el porcentaje de AD en televisión y a asegurarse de que ésta puede ser recibida en todo el territorio nacional.

En Alemania el tratado de broadcasting estatal (Rundfunkstaatsvertrag, RStV)⁹ estipula que los operadores televisivos, tanto públicos como privados, deberían aumentar su oferta de contenidos accesibles (aunque no hay definición de “accesible” ni marco temporal establecido) dentro de sus posibilidades económicas.

Además la Film Funding Law (Filmförderungsgesetz, FFG)¹⁰ incluye la producción de una versión accesible de los films como uno de los criterios de elegibilidad para recibir financiación pública. Aplica también a documentales.

En Grecia la ley para la nueva radio griega, internet y televisión 4173/2013 (OG A 169/26.7.2013 Νέα Ελληνική Ραδιοφωνία, Ιντερνετ και Τηλεόραση)¹¹ establece que los programas de radio y televisión deberían tener en consideración las necesidades de los colectivos socialmente vulnerables y discapacitados.

Parece claro que los gobiernos apuestan por aumentar progresivamente la accesibilidad a los medios pero la complejidad técnica de la AD hace que sólo una regulación de cuotas no sea suficiente para garantizar la efectividad del servicio. El problema es que el sonido es mucho más subjetivo que la imagen y al contrario que con el lenguaje de signos o los subtítulos, la AD tiene la particularidad de que puede estar emitiéndose sin que el usuario sea capaz de acceder a la información que ofrece. Sería el mismo caso que si, por ejemplo, un edificio instalase rampas en todos los tramos de

⁶ Recuperado de www.bai.ie/en/codes-standards [fecha de consulta: octubre 2016].

⁷ Recuperado de www.segretariatosociale.rai.it/dl/sociale/website/ContentItem-51423263-15e4-4b02-ad05-95da13a172bf.html [fecha de consulta: octubre 2016].

⁸ Recuperado de www.normattiva.it/atto/caricaDetttaglioAtto?atto.dataPubblicazioneGazzetta=2005-09-07&atto.codiceRedazionale=005G0206 [fecha de consulta: octubre 2016].

⁹ Recuperado de www.die-medienanstalten.de/en/legal-basis.html [fecha de consulta: octubre 2016].

¹⁰ Recuperado de <http://www.ffa.de/ffg.html> [fecha de consulta: octubre 2016].

¹¹ Recuperado de <https://nomoi.info/%CE%A6%CE%95%CE%9A-%CE%91-169-2013.html> [fecha de consulta: octubre 2016].

escaleras pero con una inclinación tal, que no fueran útiles para el usuario en silla de ruedas. Este puede ser el caso si la AD se emite como canal de audio aislado y su mínimo de volumen de audio no se regula adecuadamente.

La modalidad de AD en la que no se realiza mezcla de sonido antes de la emisión y es el aparato del usuario el que recibe y mezcla el audio de AD con el general, se llama Receiver Mix (Media Access Australia, 2010). En esta modalidad, la narración de AD se transmite como canal de audio en mono que el usuario activa o no en su TV. El volumen recomendado para este canal de audio extra, no está estipulado.

El volumen de audio es un elemento clave en los medios audiovisuales, tanto para evaluar su calidad como para facilitar su accesibilidad. Prueba de ello son los estándares existentes al respecto, tanto para la programación televisiva general (EBU R128, 2014) como para la accesibilidad de las páginas web en internet (W3C G56, 2016).

La regulación EBU R128 (2014), a pesar de ser un documento regulatorio de obligado cumplimiento para los radiodifusores en países como Reino Unido, es también una estándar de calidad, ya que nace como respuesta a las quejas repetidas de los televidentes por cambios bruscos de volumen en la programación, especialmente en los cortes para publicidad. La regulación supuso un cambio importante para las emisoras ya que se pasó de la limitación de volumen por picos a limitación por medias. Además, esta regulación introdujo una nueva unidad de medida de volumen, el LUFs (*Loudness Unit Full Scale*), que tiene en cuenta la percepción de volumen y no sólo la presión, como el decibelio. A grandes rasgos, el EBU R128 (2014) establece una media de “sonoridad” (*loudness*) de -23 LUFs para todos los programas incluida publicidad. Al basar la medición en la media de volumen y no en picos máximos, se consigue una mayor homogeneidad de la sonoridad y esto se traduce en una menor percepción de aumentos bruscos de volumen. Esta es la regulación que se aplica en Europa actualmente y que, en algunos países como Reino Unido, se ha convertido en ley.

El W3C G56 (2016) es un estándar de accesibilidad a contenidos web que establece la diferencia de volumen que debe existir entre un fondo (musical, ruidos) y una voz para asegurar que la voz se escuche con claridad. La diferencia que propone es de 20 decibelios. Esta es una diferencia de volumen lo suficientemente importante como para asegurar que la voz se oiga claramente frente al fondo.

En el caso de la AD, no hay una referencia de volumen aplicable para realizar la mezcla de sonido, por lo que la diferencia de decibelios entre voz y banda sonora queda a decisión del técnico que realiza la mezcla en cada caso. Traducido al panorama televisivo, esta falta de regulación provoca que el volumen de la pista de AD no sea homogéneo entre canales, lo que puede llevar a problemas de calidad del servicio. La recomendación general es que la AD debe escucharse siempre con claridad para lo que, en caso necesario, se puede bajar de volumen la banda sonora. Seguir esta recomendación implicaría para las emisoras transmitir dos mezclas de sonido completas, una con AD para los espectadores que la solicitasen y una sin ella para el resto. Las emisoras son muy reacias a seguir esta práctica ya que ello implica un consumo de ancho de banda muy elevado y, existiendo la posibilidad de emitir la AD como pista de audio aislada, es esta la opción que usan. La emisión de la AD en televisión como pista aislada que el usuario añade al audio general se conoce como *receiver mix* (mezcla en receptor) y es la utilizada en España por la televisión pública. Este tipo de mezcla es la más frecuente en televisión pero no hay establecidos criterios de calidad mínimos que garanticen su calidad y esta ausencia de regulación puede poner en riesgo la eficacia del servicio de AD.

2.2.3. AD como objeto sonoro

Desde el punto de vista narrativo, la AD podría considerarse un caso especial de sonido extradiegético y como tal debería ser tratada al realizar la mezcla de sonido. Sin embargo, la falta de experiencia en AD como objeto sonoro hace que no haya unos parámetros claros sobre de cómo realizar su mezcla (van der Heijden, 2010). En términos de edición y mezcla de sonido, la AD es una pista de voz que debería escucharse con la misma claridad que la pista de diálogo. Cuando en una escena de acción hay diálogo no suele suponer un problema para el espectador oír las voces de los personajes con claridad pero cuando se trata de AD, la narración en ocasiones se pierde. Tanto en un caso como en el otro, la mezcla de sonido es la responsable de este efecto. La mezcla de sonido en AD es crucial para garantizar la calidad del servicio, pero son muy escasos los autores que han abordado específicamente el tema.

Desde los artículos académicos, van der Heijden (2010) aborda el proceso de mezcla de audio para AD y divide el sonido fílmico en cinco pistas de audio: Voz/diálogos, ambiente, *foley*, efectos de sonido y música. Estas cinco pistas, cuando se visualizan en un programa de edición de audio son ondas sonoras independientes cuyas características se pueden ajustar por separado, desde la ecualización de frecuencias al volumen. Esto permite por ejemplo, atenuar las bajas frecuencias de un ruido de explosión, sin bajar el volumen de la pista dedicada a los efectos de sonido. Durante un proceso de edición lo que se busca alcanzar es un equilibrio que mantenga la pista de voz sobre las demás, sin que por ello se pierda riqueza de sonido. Van der Heijden (2010) defiende que la AD debería tenerse en cuenta como otra pista más en el proceso de mezcla al mismo nivel que la pista de diálogo para obtener un producto audiovisual de calidad. Para ello la AD tendría su propio proceso de remezcla en el que se ajustarían los niveles de cada una de las pistas de audio para garantizar que la AD se escuchase con claridad. Este método sin duda asegura una mezcla de sonido de AD de, al menos, la misma calidad que la de la película original pero resulta poco viable para televisión. El primer problema es que los radiodifusores trabajan con contenidos ya mezclados y en segundo lugar, como ya se ha expuesto, son reacios a emitir dos mezclas de sonido completas por motivos técnicos.

Desde el salto de la televisión analógica a la digital, las televisiones trabajan con estándares de emisión que proporcionan una tasa de transmisión de datos concreta y las mezclas de audio completas consumen hasta el doble que una pista en mono. En el caso de la TDT española, el estándar es la DVB-T, diseñada especialmente para las transmisiones terrestres a través del aire. Este estándar proporciona un total de 19,91 Mbps (Megabits por segundo) a través de cada canal, que puede ser utilizado libremente por los operadores para la difusión simultánea de varios canales de vídeo (TV), varios de audio (radio), aplicaciones interactivas MHP u otros servicios (EPG o EPG extendida). Los canales de televisión se envían codificados en la norma MPEG-2. Este formato necesita alrededor de 3,5 Mbps para cada flujo de vídeo estándar que se envía para ofrecer una calidad de imagen aceptable. Si ese flujo de vídeo contiene imágenes con gran cantidad de variaciones y movimientos (caso de retransmisiones deportivas o películas de acción), el ancho de banda que se recomienda consumir asciende a entre 4,5

y 6 Mbps.¹² En el caso del audio, una mezcla estéreo, dependiendo de su codificación consume entre 126 y 256 Mbps, frente a los 64 a 96 que consume una pista de AD en mono.

Dada la cantidad de información que emiten los canales, el ahorro de *bitrate* es una preocupación constante para las cadenas y, en lo referente a la AD, son reacios a emitir una mezcla de sonido completa premezclada en emisora. Esta circunstancia no es abordada por los códigos de buenas prácticas o normativas existentes. Sin estándar de regulación, varios informes técnicos de ITU (2013), ofrecen detalles técnicos específicos sobre volumen, mezcla y calidad de sonido recomendados en AD. En concreto, el *FG AVA Technical Report Part 5* denuncia la falta de estandarización de la AD tanto en terminología como en práctica y pone de manifiesto la falta de documentos de referencia basados en datos experimentales para garantizar la calidad del servicio. Este es uno de los pocos documentos que propone medidas concretas con respecto al audio en AD y audio subtítulos. En su parte 13, el mismo documento recomienda que la AD suene más alta que el resto de la banda sonora pero intentando mantener esta lo más intacta posible. Para conseguirlo, este documento propone hacer una serie de ajustes en la propia pista de AD para optimizar su claridad. Los ajustes concretos que propone el documento se resumen a continuación:

Sonoridad y picos: la AD debería aproximarse o exceder ligeramente el volumen de la BSO, ajustando dinámicamente la AD si fuera necesario. La AD no debería sobrepasar en más de 6 decibelios los picos de la BSO para evitar saturación o compresión excesiva por limitación de volumen. La AD debería adaptarse al volumen de la BSO bien haciendo ajustes durante la grabación de AD o bien ajustando la pista durante la mezcla.

Ecualización: El documento sugiere ecualizar la pista de AD para enfatizar consonantes potenciando la franja de 1 a 6 kilohercios con una reducción acorde en la franja de los 200 kilohercios. Se aconseja ecualizar lo más parecido posible al audio original pero dando prioridad a la claridad de la narración.

Sincronización: la AD debe sincronizarse con el audio original con una desviación menor del fotograma por segundo (entre 1,50 y 1,60 décimas de segundo dependiendo del formato de retransmisión que utilice el radiodifusor).

¹² Mundoplus.tv. *Datos técnicos: parámetros configurables*. Recuperado de www.mundoplus.tv/zonatdt/datos_tecnicos_tdt_canales.php [fecha de consulta: octubre 2016].

Calidad de grabación (de audio): la calidad de la grabación debe ser, como mínimo la misma que la del audio original. La relación señal-ruido debería ser de calidad profesional y no debería ser necesario tener que limpiar ruido externo o reverberación.

El documento *El FG AVA Technical Report Part 5*, plantea medidas específicas de AD concretas y compatibles con una modalidad de mezcla en receptor pero más estudios son necesarios para llegar a conclusiones claras. Uno de los datos que siguen pendientes de estandarizar es el mínimo de decibelios de diferencia entre narración y banda sonora/ruido de fondo, para garantizar una escucha clara de la AD. Hay que destacar que este es el primer parámetro a definir para garantizar la calidad del servicio ya que la base de cualquier AD es que el público la oiga con claridad y dentro de unos rangos de volumen aceptables. De acuerdo con Chion (1993 p.13) “...si el sonido en el cine es verbo y vococentrista [...] es porque el ser humano también lo es” y “...buscará primero el sentido de las palabras sin pasar a la interpretación de los demás elementos hasta que esté saturado su interés por el sentido”. Que el público pueda acceder a este “sentido” necesariamente pasa por una mezcla de audio que asegure que la AD se escucha con claridad. Tal vez por lo obvio de la cuestión no se han llevado a cabo estudios en esta dirección, a pesar de que un estándar de volumen sería de aplicación global e iría en beneficio de todos los usuarios. En entornos adyacentes a la AD como el de la grabación y producción de sonido o la accesibilidad auditiva, sí se han llevado a cabo estudios destinados a asegurar la inteligibilidad de la voz en entornos de ruido, ya sea para optimizar una mezcla de audio o una grabación, o en relación con dispositivos de ayuda auditiva. Si bien cae fuera del ámbito de esta tesis entrar en esos campos, sí es interesante rescatar algunos datos de los que la AD podría beneficiarse. Hay estudios (Franklin *et al.*, 2006) que establecen una diferencia de 15 dB entre ruido de fondo y voz para que la voz se entienda cómodamente por un oyente medio.

Este dato es algo más bajo que lo que establece el World Wide Web Consortium (W3C, 2016), de 20Db, pero es lógico ya que el estándar de W3C está pensado para asegurar que incluso los usuarios con cierto déficit de audición entienden la voz sobre el fondo sin problema. Esta tesis mantiene que tanto 20 como 15 dB es una diferencia de volumen excesiva para audiodescripción, ya que daría a la voz de la narración una

preponderancia excesiva y además presentaría el problema de que, en escenas de ruido, esta mezcla muy probablemente saturaría cualquier altavoz medio.

Partiendo de la hipótesis de que el estándar de AD para televisión debería regular las características de la pista de AD antes de su inserción en el audio general del programa, esta tesis plantea la necesidad de realizar experimentos para establecer parámetros mínimos de volumen y ecualización de sonido que mejoren la recepción de AD. Para ello, un primer experimento se llevó a cabo en 2015 recreando tres escenarios de mezcla audio diferentes para tratar de establecer si determinadas mezclas de sonido pueden causar dificultades de escucha en la pista de AD. En 2017, y partiendo de las conclusiones extraídas se plantea un segundo estudio, esta vez mucho más específico, centrado en averiguar la diferencia de volumen mínima que se debería mantener entre la AD y el audio general del programa audio descrito para garantizar la inteligibilidad de la narración de AD.

2.3. Marco metodológico

El marco metodológico de la presente tesis se enmarca en los estudios de recepción por el diseño y desarrollo, si bien no sería puramente un estudio de la recepción en cuanto a contenido a evaluar. Diferentes autores han llevado a cabo investigaciones sobre AD a través de estudios de recepción, bien con usuarios sin problemas de visión (Igareda y Matamala, 2012; Kruger, 2012; Vilaró *et al.*, 2012) bien con personas con discapacidad visual (Matamala, 2005; Rai, 2009; Maczynska y Szarkowska, 2011; Walzak y Swarkowska, 2012; Cabeza-Cáceres, 2013), para comprobar diferentes aspectos de AD. Estos estudios evalúan aspectos relacionados con la recepción de diferentes aspectos de AD. Los estudios de Cabeza-Cáceres (2013) Szarkowska (2011) o Fernández-Torné y Matamala (2015) se centran en aspectos sonoros de la AD, pero el enfoque está puesto bien en la comprensión del texto, bien en el grado de satisfacción del público con su formato sonoro. El estudio que se presenta en esta tesis pretende evaluar, en primera instancia, a partir de qué volumen se percibe el pista de AD y se puede entender con comodidad ya que esta es la primera premisa necesaria que debería regularse para asegurar la efectividad del servicio.

En diciembre de 2015 se lleva a cabo un estudio con un grupo de 32 personas con discapacidad visual dirigido a evaluar diferentes aspectos sonoros de tres mezclas de audio diferentes de AD. La evaluación de los distintos estilos de mezcla de audio se realiza a través de cuestionario personal en el que el espectador valora el grado de acuerdo con diferentes afirmaciones de 1 a 3, siendo 1 total acuerdo y 3 total desacuerdo.

Tomando en consideración:

- a) El estándar de sonido más utilizado en televisión, EBU-R 128: todos los contenidos deben estar normalizados a un volumen de -23LUFS
- b) La norma actual sobre mezcla de sonido en AD: bajar el fondo cuando se inserta un comentario (sin datos numéricos)
- c) La práctica común en AD televisiva: añadir el pista de AD a la banda sonora original sin hacer modificaciones de ninguno de los dos

Se eligieron tres clips de una película, se grabó una sola AD para cada uno de ellos y se diseñaron tres escenarios de sonido diferentes de acuerdo con la manera de mezclar el

sonido. Cada escena se mezcló conforme a los tres escenarios diseñados de modo que se obtuvieron nueve clips en total. Cada uno de los participantes en el estudio escuchó, aleatorizadas, cada una de las mezclas de sonido y respondió al cuestionario, cuya primera afirmación era “He podido oír la narración de AD con claridad en todo momento”. La hipótesis en la que se basa el estudio es la de que una pista de AD ajustado dinámicamente a los diferentes momentos sonoros de una escena, permitiría una escucha clara de la AD en todo momento sin comprometer la mezcla de audio original. Un estándar como este sería, además, compatible con la mezcla en receptor y por lo tanto de fácil aplicación para los radiodifusores. Partiendo de esta idea, el diseño del segundo experimento se centró en qué diferencia de volumen sería la óptima entre el audio de AD y el audio general del programa audio descrito para garantizar la inteligibilidad de las descripciones incluso en escenas de ruido. Este segundo experimento se llevó a cabo en Julio de 2017 con un grupo de 4 personas con discapacidad visual y 22 personas sin problemas de visión, que evaluaron la inteligibilidad de una narración frente a un ruido de fondo que subía progresivamente de volumen.

2.4. Descripción de los siguientes capítulos y cómo aplica la metodología.

El capítulo 3 de esta tesis se centra en los aspectos técnicos del sonido. Es un capítulo teórico que describe conceptos fundamentales de la percepción, grabación y reproducción de la onda de sonido. La primera parte del capítulo aborda los conceptos más generales de la onda de sonido y la segunda se centra en el salto de analógico a digital y lo que ha supuesto en la producción audiovisual, posibilitando, entre otros, un aumento de la presencia de AD en TV.

El capítulo 4 presenta el primero de los experimentos que recoge esta tesis. Se trata de un test de usuario llevado a cabo en Diciembre de 2015, en Madrid, con un grupo de 32 personas con discapacidad visual. El experimento se lleva a cabo para comprobar si el tipo de mezcla que se hace en AD afecta, y cómo lo hace, a los diferentes elementos sonoros que la componen: la banda sonora original y la voz del locutor de AD. Para este experimento se recrearon tres escenarios de mezcla de sonido: ajuste de volumen de la banda sonora original, banda sonora original y locución de AD a un volumen similar, y ajuste de volumen de la locución de AD.

El capítulo 5 presenta un segundo experimento, realizado en Julio de 2017 que se centra específicamente en el grado de tolerancia de ruido antes de que la inteligibilidad de una narración superpuesta se vea afectada. Dada la dificultad de encontrar un grupo numeroso de usuarios con discapacidad visual, este experimento se realizó con un grupo de 4 usuarios con discapacidad visual y de 22 sin problemas de visión. El objetivo es medir ese grado de tolerancia de ruido, en decibelios, y usarlo para sugerir un estándar de diferencia de volumen que debería mantenerse entre banda sonora original y AD para asegurar la inteligibilidad de la AD incluso en escenas de ruido.

En el último capítulo se presentan las conclusiones alcanzadas a lo largo del doctorado y de la realización y se apuntan vías de futura investigación que podrían ser útiles en el proceso de alcanzar un estándar de AD aplicable a nivel universal.

2.5. Conclusiones

El concepto de AD varía según el ámbito desde el que se aborde y este desencuentro dificulta el alcance de un estándar global de calidad del servicio. Académicos, legislación y empresas de producción actúan de forma unilateral en su campo y esto revierte negativamente en la calidad del servicio de accesibilidad. Mientras que la AD como objeto de estudio es esencialmente un texto lingüístico (Braun, 2007; Piety, 2004), a efectos legales es un servicio que tiene que cumplir unos porcentajes mínimos (Ley General de la Comunicación Audiovisual, 2010; Equity Act, 2010) en cuanto a cuotas pero no en cuanto a calidad. Finalmente, para los estudios de grabación y productoras, es un producto audiovisual todavía con poca tradición y sin estándar técnico aplicable (Media Access Australia)¹³.

Esta desconexión entre los conceptos de AD que se manejan en los tres ámbitos dificulta el alcance de un estándar de calidad común que aporte homogeneidad al servicio y permita una producción sistemática de calidad.

¹³ Recuperado de www.mediaaccess.org.au/practical-web-accessibility/media/audio-description-guidelines [fecha de consulta: octubre 2016].

CAPITULO 3. SONIDO, PERCEPCIÓN Y ACCESIBILIDAD

3.1. Introducción

3.2. ¿Física, psicoacústica o accesibilidad?

3.2.1. Características de la onda sonora e implicaciones en la AD

3.2.1.1. Sonido, sistema auditivo y audio digital

3.2.1.2. Audio analógico y audio digital

3.3. El sonido en TV: retrasmisión y AD

3.3.1. Sonido y enmascaramiento; enmascaramiento y AD

3.4. Conclusiones

3.1. Introducción

“Quality measurement methods must be designed so that the parameter really quantified is the user’s perception” (Cotê, 2011 p. 38). El sonido se puede definir como una perturbación que se propaga a través de un medio elástico a la velocidad característica de dicho medio causando una alteración de la presión o un desplazamiento de las partículas (López, 1993) y, de acuerdo con Cotê (2011), la calidad puede definirse en función de la percepción del usuario. La AD es un servicio de accesibilidad que depende del sonido para cumplir con su función y por eso, garantizar su calidad es clave para que el servicio funcione. Cotê (2011) habla de percepción del usuario como parámetro a cuantificar para medir la calidad, pero, en el caso del sonido, su realidad física es la que determina esa percepción.

La importancia del sonido en el mundo audiovisual es evidente en el propio término “audiovisual”. Es uno de los elementos integrantes de la ecuación junto a la imagen y puede cumplir funciones narrativas y artísticas varias. Desde la caracterización de un personaje (a través de un leimotiv, por ejemplo) hasta la contextualización de lugares, momentos o estados de ánimo (Szarkowska y Orero, 2014), el sonido es parte fundamental de la obra audiovisual. En el caso de la AD, además, es la única parte de la obra audiovisual que percibe el usuario. Como se ha expuesto en el capítulo 2, no hay un estándar de sonido de AD que marque a los radiodifusores o proveedores conforme a qué medidas (volumen, formato, tratamiento) se debe insertar una pista de AD en un programa y esto puede causar problemas al usuario. La falta de homogeneidad en las características acústicas de la AD que se recibe en TV entre programa y programa, o entre canal y canal, es uno de esos problemas; el riesgo de que la inteligibilidad de los comentarios del locutor de AD se vea disminuida a causa de una mezcla de sonido deficiente, es el peor. Un estándar de que regulase los aspectos más básicos del sonido de AD: volumen, formato, compresión y ecualización, podría solucionar estos problemas pero la complejidad del proceso de producción de la AD, así como el gran número de actores implicados en él, hace que sea muy difícil alcanzarlo. El viejo acertijo de si un árbol hace ruido al caer en un bosque cuando no hay nadie cerca para oírlo, ilustra la complejidad que entraña intentar regular algo que depende tanto de la percepción humana como el sonido. Sin

embargo, si el acertijo se plantease a un científico la respuesta sería clara: sí, lo hace, y además se puede grabar. Los aspectos técnicos del sonido son objetivos y como tales deberían estar regulados. Ningún estándar puede garantizar homogeneidad en cómo se va a percibir una AD en cada uno de los casos particular, pero sí en cómo se emite. Los procesos de grabación, codificación y emisión de AD influyen en cómo se percibe y cuestiones como el volumen de normalización, la profundidad de bits del archivo o el códec de audio que se aplique durante la emisión, no pueden permanecer sin regular si se pretende alcanzar una homogeneidad en el servicio.

Este capítulo trata las cuestiones físicas y técnicas del sonido que sí son objetivas y regulables, y cómo su tratamiento en el mundo audiovisual puede afectar a la AD como servicio de accesibilidad.

3.2. ¿Física, psiocaústica o accesibilidad?

Uno de los problemas de AD como servicio de accesibilidad es el gran número de actores que intervienen en su proceso, no sólo de creación (guionista, locutores, estudio de sonido, montadores) sino también de distribución (emisoras de televisión, productoras de cine, fabricantes de televisiones y monitores), y la dificultad que implica establecer baremos de evaluación de calidad para cada uno de ellos. El guion ha sido abordado con frecuencia por la investigación (Maszerowska *et al.*, 2014; Vera, 2006; Beneke, 2004; Snyder, 2005) y recientemente la locución también a raíz de la irrupción en el mundo audiovisual de las voces sintéticas como alternativa a la voz humana (Szarkowska, 2011; Szarkowska y Jankowska, 2012; Fernández i Torné y Matamala, 2012). Sin embargo, el sonido como elemento técnico cuyas características físicas (frecuencia, intensidad) determinan cómo lo percibe el oído, no es abordado en ninguno de los manuales de buenas prácticas consultados en esta investigación.

Los aspectos técnicos del sonido como elemento de AD se encuentran en una tierra de nadie que se sitúa, en lo teórico, entre los campos de la física y lo psicoacústica y en lo práctico, entre los estudios de sonido, las emisoras de televisión, los organismos que regulan los servicios de accesibilidad y los fabricantes de reproductores de TV y audio. La actual legislación de accesibilidad audiovisual establece cuotas mínimas de AD a los canales de TV en la mayoría de países europeos pero el único criterio que se contempla en las leyes es el cuantitativo. Los criterios de calidad de AD se encuentran diseminados en diferentes documentos no vinculantes ni para proveedores ni para canales de radiodifusión que, además, no presentan un criterio unificado. Con respecto al tema que ocupa esta tesis, el sonido de la AD, sí hay un consenso en cuanto a que los comentarios descriptivos tienen que oírse con claridad. Esta indicación es una obviedad que sin embargo no ha recibido la atención suficiente ya que, hasta la fecha, no hay estándar ni de proceso de producción, ni de formato de archivo o emisión, ni de volumen, que establezca cómo asegurar esa claridad. Para alcanzar este estándar uno de los pasos necesarios es la regulación de los aspectos técnicos del proceso de grabación, mezcla y emisión de AD. Estos tres procesos, al igual que el resto de elementos relacionados con la producción audiovisual, han experimentado una transformación total con el salto de analógico a digital y es necesario establecer criterios de evaluación

de calidad nuevos adaptados al nuevo panorama. A continuación se describen aspectos básicos de sonido, como elemento físico y como archivo digital, que influyen directamente sobre la calidad y claridad de la AD como producto audiovisual.

3.2.1. Características de la onda sonora e implicaciones en la AD

Las características de la onda sonora son determinadas por el modo en que vibra la fuente que la origina y determinan cómo percibe el sonido el oído humano. La frecuencia, el periodo, la longitud y la intensidad de onda son los cuatro parámetros que determinan cómo el oído humano va a percibir un sonido. Estos cuatro parámetros son estrictamente físicos, susceptibles de ser medidos, evaluados y expresados en cifras. De la misma manera, la calidad de audio en el proceso de creación audiovisual se puede medir desde parámetros objetivos.

Frecuencia, periodo, longitud de onda e intensidad son las principales características de la onda de sonido que tienen que registrarse en un proceso de grabación para que ese sonido pueda ser reproducido en otro momento de forma fidedigna. La frecuencia se expresa en ciclos/segundo o Hertzios (Hz) y representa el número de vibraciones completas de la fuente emisora de la onda de sonido. El periodo es lo inverso a la frecuencia, es decir, representa el lapso de tiempo necesario para que una vibración complete un ciclo completo. La longitud de onda representa la distancia en metros entre dos puntos consecutivos del campo sonoro que se encuentran en el mismo estado de vibración. La intensidad evalúa el flujo de energía que atraviesa en un segundo una superficie de un m^2 colocada perpendicularmente a la dirección de propagación del mismo y es lo que determina el “volumen” al que escuchamos el sonido. Esto en cuanto a física; pero para esta tesis y en general para el mundo audiovisual, lo importante es cómo el ser humano (en este caso el espectador) percibe e interpreta esas características. Casi todos los autores (Carrión, 1998) añaden este matiz final a la definición de sonido anterior, y por tanto resulta inevitable si se pretende establecer un estándar objetivo de calidad de sonido, referirse a cómo el cerebro procesa e interpreta (ya por tanto de una forma subjetiva) las características físicas de una onda sonora. Documentos como el EBU R128, recurrieron a la psicoacústica para elaborar un estándar aplicable a nivel industrial y audiovisual que de hecho es actualmente la

norma de regulación de sonoridad en televisión a nivel europeo y ley en países como Reino Unido.

La psicoacústica estudia la relación entre las propiedades físicas del sonido y la interpretación que hace de ellas el cerebro. Gracias a ella se pueden establecer conceptos fundamentales como la banda de audiofrecuencia o rango de trabajo del oído humano (que se establece de forma aproximada entre los 20Hz y los 20.000Hz) o el umbral de audición que sería la intensidad acústica mínima que somos capaces de detectar a una determinada frecuencia (20μPascales a para una onda sonora de 1.000Hz) (Recuero y Gil, 1993). Los conceptos de timbre, grave o agudo vienen determinados por la interpretación que el cerebro hace de las características físicas del sonido (Figura 3.1)

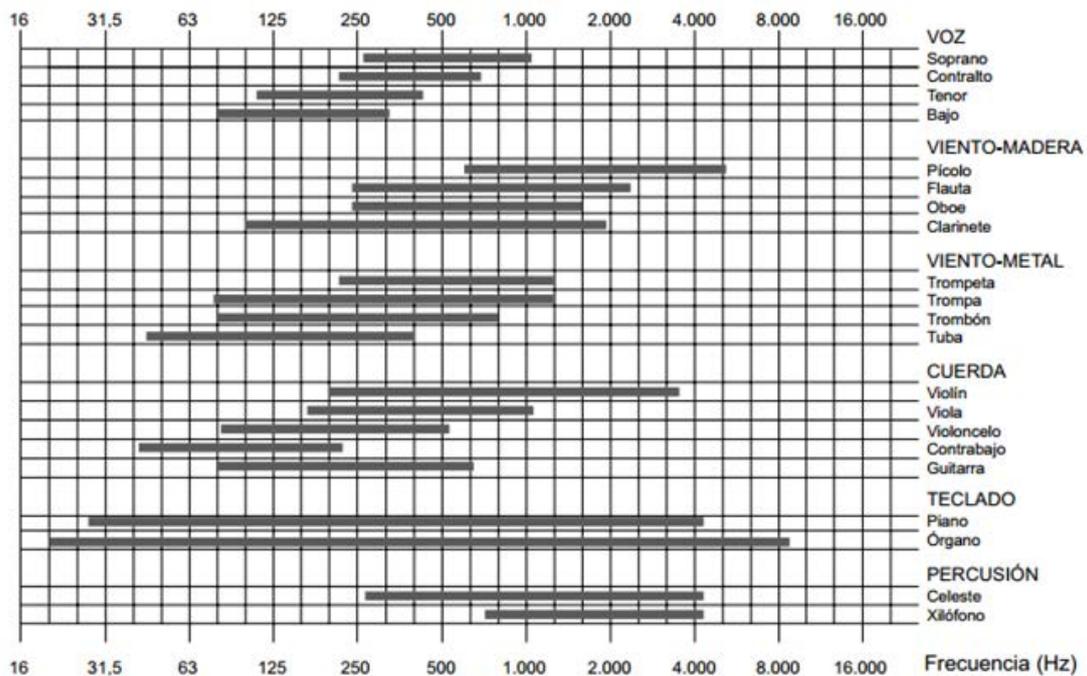


Figura 3.1. Bandas de frecuencia de instrumentos musicales y la voz.

La figura 3.1 muestra en el eje horizontal, el rango de frecuencia entre los 16 y los 16.000 Hercios. En la parte vertical derecha aparecen la voz humana y sus distintos registros tonales y los diferentes grupos de instrumentos musicales. El rango de frecuencias que cubre cada uno de ellos aparece resaltado en gris oscuro.

Estos son conceptos básicos que los estudios de grabación y técnicos de sonido usan continuamente en procesos fundamentales de la producción de audio como la normalización o la ecualización. Estos procesos también están en la base de la

producción de AD y son muy importantes para garantizar su calidad. Merece la pena, por lo tanto, analizar con algo más de profundidad las implicaciones que tiene el modo en el que se produce la interpretación del sonido por parte del ser humano antes de realizar una hipótesis sobre qué estándar de grabación, tratamiento de sonido y emisión garantizaría mejor la calidad de sonido de una AD.

3.2.2. Sonido, sistema auditivo y audio digital

El sistema auditivo no responde de una manera lineal a la intensidad de los estímulos, esto, unido al margen enorme que existe desde el sonido más leve que el oído es capaz de detectar (10^{-12} W/m^2) hasta aquel que produciría dolor y dificultades en la audición (1 W/m^2) (Jackson-Menaldi, 1992), hace que habitualmente las variaciones de presión provocadas por una onda sonora se expresen mediante unidades logarítmicas. Para cuantificar la intensidad de un sonido se hace referencia al nivel de presión sonora (NPS ó SPL en sus siglas en inglés) cuya unidad de medida es el decibelio SPL (dB SPL), (Tabla 3.1.)

FUENTE SONORA	NIVEL DE PRESIÓN SONORA SPL (dB)	VALORACIÓN SUBJETIVA DEL NIVEL
Despegue avión (a 60 m)	120	Muy elevado
Edificio en construcción	110	
Martillo neumático	100	
Camión pesado (a 15 m)	90	Elevado
Calle (ciudad)	80	
Interior automóvil	70	
Conversación normal (a 1 m)	60	Moderado
Oficina, aula	50	
Sala de estar	40	
Dormitorio (noche)	30	Bajo
Estudio de radiodifusión	20	

Tabla 3.1. Niveles de presión sonora de ruidos y sonidos típicos y sonoridad subjetiva asociada.

La tabla 3.1. muestra medidas de decibelios SPL en la columna central y los relaciona, a la izquierda con distintas fuentes sonoras, y a la derecha, con la sonoridad subjetiva asociada.

De acuerdo con esta escala, 0 dB SPL es el nivel mínimo dentro del umbral de audición para 1 KHz y el nivel máximo o umbral de dolor, a partir del cual se producen daños irreversibles en el oído, se sitúa en torno a los 120dB. La diferencia en dB entre los valores mínimo y máximo de sonido que el oído es capaz de percibir se denomina rango dinámico. En producción audiovisual el rango dinámico es la diferencia entre el pico máximo y mínimo de intensidad de la onda en un archivo de sonido.

En relación con este concepto está uno de los parámetros que en el mundo audiovisual se utiliza para evaluar la calidad de audio: La relación señal ruido (o S/N en inglés de *Signal/Noise*). La relación señal ruido representa en dB la diferencia entre el ruido de fondo inherente a toda grabación y el valor máximo que puede alcanzar una señal. Valores iguales o inferiores a ese nivel de ruido no serían interpretables para nuestro oído y se confundirían con él.

Es importante destacar que el oído humano no posee la misma sensibilidad para todo el umbral de audición y por tanto la amplitud del rango dinámico que somos capaces de percibir varía en función de la frecuencia de la onda sonora. Con valores bajos de presión sonora el oído es muy insensible a las bajas y altas frecuencias mientras que con frecuencias medias, puede captar sonidos a menos decibelios. Por ejemplo un sonido con 20Hz de frecuencia a 70dB SPL produce la misma sonoridad que 1.000Hz un nivel de 5dB. A medida que los niveles de presión sonora aumentan el oído tiende a responder de forma más homogénea en toda la banda de frecuencias audibles hasta el punto de que cuando son muy elevados la sonoridad asociada a tonos puros de diferente frecuencia es muy similar. En las figuras 3.2. y 3.3., se representa el umbral de audición humano (Figura 3.2.) con los decibelios en el eje vertical y los Hercios en el horizontal y el campo de audición humano resaltado en gris. En la figura 3.3. se representan, dentro del umbral de audición humano, los campos que ocupan la palabra y la música.

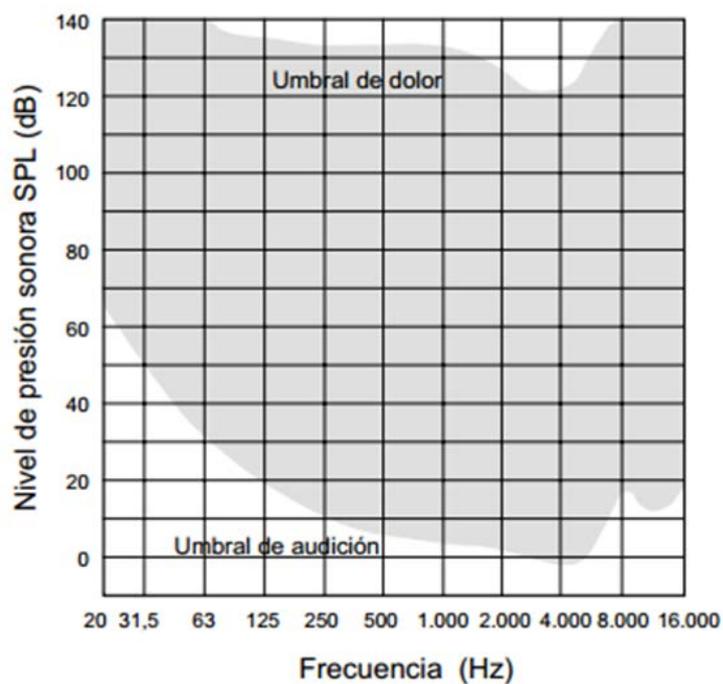


Figura 3.2. Umbral de audição em função de la frecuencia.

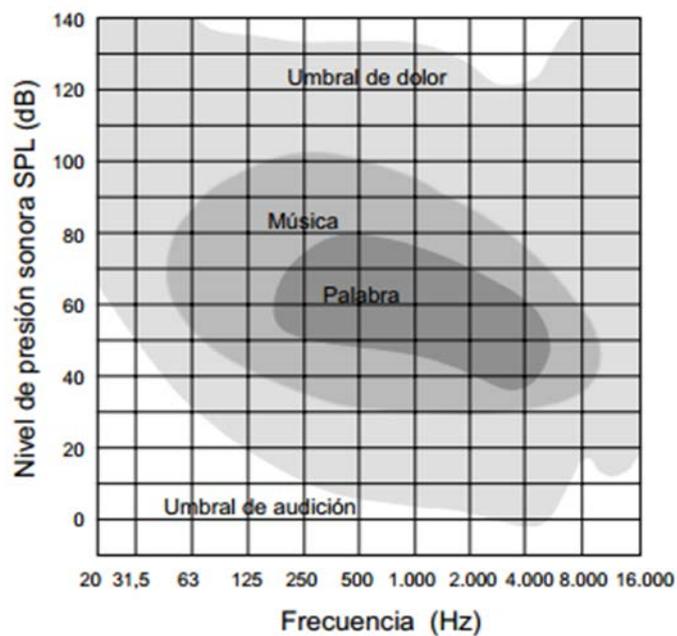


Figura 3.3. Umbral de audição junto con las zonas correspondientes a música y palabra.

Esto es lo que lo que hace que el proceso de ecualización de audio sea tan importante en la grabación y producción de audio, ya que posibilita enfatizar o atenuar

sólo ciertas frecuencias dentro de una grabación para dar más o menos presencia a determinados sonidos. Esto puede ser muy útil en AD para optimizar la claridad de las descripciones en escenas ruidosas y de hecho es uno de los recursos habituales en el montaje de audio al introducir diálogos en escenas con banda sonora. El documento técnico de características de audio para audiodescripción y audiosubtítulos (ITU, 2013) habla precisamente de la importancia de la ecualización para mejorar la claridad de la voz en AD y audio subtítulos y recomienda enfatizar frecuencias entre 1-6 kHz y atenuar frecuencias por debajo de los 200 Hz.

El proceso de grabación y posproducción son responsables por tanto de que algo “suene bien” o “suene mal” a oídos del consumidor y ambos han sufrido una transformación radical con el paso de analógico a digital. Este paso es fundamental también para la AD ya que abre posibilidades tanto de posproducción como de emisión imposibles con el audio analógico.

3.2.3. Audio analógico y audio digital

Durante el proceso de grabación, las ondas sonoras se transforman en variaciones de corriente o voltaje análogas a las variaciones de presión asociadas a la onda mediante transductores electroacústicos como los micrófonos. Para convertir esas variaciones de corriente de nuevo en ondas sonoras, basta con emplear el transductor electroacústico inverso, como un altavoz. El altavoz interpretará estas variaciones de corriente para vibrar de forma proporcional a la corriente variable que atraviesa su bobina y esa vibración desplazará el aire generando la misma onda sonora que se grabó. Entre estos dos extremos de grabación y reproducción están el almacenamiento, transmisión y procesado o no, de las señales sonoras.

Antes de la revolución digital todos los aspectos del proceso eran analógicos, las grabaciones se imprimían físicamente en cintas o discos que se almacenaban y reproducían con procesos que poco a poco iban deteriorando la copia física usada para su reproducción y que, además, ofrecían relativamente poca capacidad de manipulación, más allá de reproducción o no. El cambio de analógico a digital cambió esto por completo. La señal de audio digital no es ya la impresión física del impulso en un soporte sino la transformación de la señal eléctrica registrada en información que se

representa con una serie finita de valores en código binario. Esta transformación hace que el almacenamiento, transmisión y procesado de la señal de audio sea infinitamente más sencillo, eficiente, y añade numerosas ventajas como una mayor capacidad de manipulación del sonido. Esta es la principal de que en casi todos los campos de la producción audiovisual se trabaje con audio digital, si no en el proceso completo, al menos en una gran parte del mismo. Los dos grandes factores que influyen en la calidad del audio digital hacen referencia al proceso de transformación de los impulsos eléctricos a información en código binario y son la frecuencia de muestreo y la profundidad de bits.

El proceso de transformación de la onda sonora a valores en código binario se hace por muestreo. Esto quiere decir que se toman muestras de la onda de sonido con una frecuencia X , se registran sus características exactas en ese momento y se almacenan en forma de código binario. A priori pudiera parecer que renunciar a una señal continua (analógica) para pasar a una secuencia de muestras entre las cuales necesariamente hay lapsos, aunque sean mínimos, implica una irremediable pérdida de información pero no es así. Queda fuera del alcance de esta tesis entrar a valorar las teorías y leyes físicas que respaldan este dato, baste decir que el Teorema de Nyquist-Shannon (Nyquist, 1928) confirma que si el número de muestras que se toma (frecuencia de muestreo) se toma al menos al doble de la velocidad máxima a la que varía la señal original, no existirá tal pérdida de información. Esto quiere decir que si el oído humano no capta vibraciones a una frecuencia mayor de los 20.000Hz, y es posible tomar muestras con una frecuencia de 40.000Hz, no se producirá pérdida de información.

En los comienzos del audio digital las limitaciones tecnológicas del momento establecían la frecuencia de muestreo en 32.000Hz lo que limitaba el ancho de banda de la señal a 16.000Hz, lejos del límite de 20.000Hz de la audición humana, por lo que se producía una ligera pérdida de información y por lo tanto de calidad. Sin embargo años más tarde se alcanzaron los 44.100Hz (Frecuencia de muestreo empleada en el Compact Disc) y actualmente el estándar de trabajo en el mundo audiovisual son los 48.000Hz, por lo que si se cumple este estándar, no hay pérdida de información ni de calidad para el oído entre una señal sonora y su representación digital. Este valor, expresado en Hz, es la frecuencia de muestreo.

La segunda característica determinante es la profundidad de bits, y hace referencia al número de valores disponibles para asignar a cada muestra de audio que se registra. Este número de estados de cuantificación entre los que se puede “elegir” para asociar a la muestra, determina el margen de error que existe entre el valor real de la muestra y el valor que se le asigna. Este margen de error a efectos de percepción, se manifiesta como ruido. Trabajar con muestras de 2 bits, por ejemplo, significaría que independientemente del valor que tenga la onda en el momento que se analiza sólo se le podrían asignar cuatro valores: 00; 01; 10; 11 con el error (ruido) que eso implica. Una profundidad de 8 bits permite 256 valores para cada muestra y profundidades de 16 ó 24 bits, como las que se emplean en el ámbito profesional, proporcionan 65.536 y 16.777.216 valores respectivamente. Conceptualmente la codificación implica siempre un determinado error, pues por elevado que sea el número de estados de cuantificación disponibles siempre se reduce un valor con precisión infinita a uno finito como el que se le otorgaría tras la cuantificación. Ese error, en el audio digital, se transforma en ruido. Este ruido acompaña a la señal cuando esta se devuelve al entorno analógico al ser reproducida y varía de una manera proporcional a esa señal, lo que lo hace que sea especialmente perceptible y molesto para el oído humano. John Watkinson (2001) establece la fórmula para calcular la relación señal ruido y determina que $S/N = 6,02 * N^{\circ} \text{Bits} + 1,76$ expresado en dB. Muestras de 16 bits harían factible trabajar con una relación señal ruido de 96dB (inferior al rango dinámico con el que es capaz de funcionar el oído) y empleando 24 bits alcanzaríamos los 144dB S/N, ahora sí por encima del margen de trabajo del oído humano.

El estándar de trabajo en estudios de grabación y productoras de audio ha sido hasta hace poco de 48.000Hz (frecuencia de muestreo), 16 bits (profundidad de bits) aunque ya están empezando a ser frecuentes los 96.000Hz 24 bits. Esta tesis tomará por defecto como estándar de calidad 48.000Hz, 16 bits, que es el estándar de trabajo en la TV pública española para hablar de sonido en AD. Esta es una calidad muy alta pero tiene el factor negativo de que la cantidad de información generada por estos procesos es enorme y esto puede causar problemas a la hora de retransmitirla, por ejemplo, en televisión.

3.3. El sonido en TV: retransmisión AD

Al margen de la calidad del sonido con el que trabaje una emisora, los parámetros de emisión influyen en la calidad del audio que recibe el usuario pudiendo rebajarlo por ejemplo, si se aplican tasas de compresión muy altas. El sonido es parte de prácticamente cualquier retransmisión televisiva, pero en general, recibe menos atención por parte de las emisoras que la imagen, tal y como puede verse en los datos de las principales emisoras españolas¹⁴ cuyo formato de imagen es en alta definición en muchos casos pero cuyo formato de audio nunca es en 5.1, por ejemplo. Durante muchos años el sonido en “mono” (un único canal monofónico formado por la suma de diálogos música y efectos) era la única señal de audio que acompañaba al video, empleando técnicas de modulación similares a las utilizadas en la radio FM. En los años 90 el sistema Nicam desarrollado por la BBC, permitió que aquellos espectadores con TVs equipadas con la electrónica necesaria pudieran decodificar una señal digital de sonido que se añadía a la señal original de TV. Para mantener la compatibilidad con los receptores convencionales, se mantenía intacta la señal monoaural analógica original. Aunque el público en los años 90 llevaba décadas disfrutando del sonido estereofónico en equipos de alta fidelidad domésticos, no fue hasta ese momento cuando pudieron experimentar una sensación similar cuando veía la televisión.

Con el audio multicanal ocurrió algo parecido, aunque estaba presente en el cine desde los presente en el cine desde los años 80 y completamente extendido desde los 90 con unas cotas de calidad muy alta (el espectador recibe información sonora desde 6 fuentes independientes que reproducen señales distintas: canales izquierdo, derecho, central trasero izquierdo, trasero derecho y bajas frecuencias) el multicanal no irrumpió en la radiodifusión hasta la llegada de la TDT. Al igual que en el caso de la imagen, el paso de analógico a digital supuso una revolución en cuanto a las posibilidades de emisión de sonido. Aunque el audio sigue disponiendo de un espacio muy limitado dentro del ancho de banda que ocupa cada canal de TV frente a la señal de video, la señal digital permite gracias a técnicas de compresión de datos que el espectador reciba audio multicanal y, esta vez sí, pueda disfrutar de la versión original del programa y

¹⁴ Mundoplus.tv. *Datos técnicos: parámetros configurables*. Recuperado de www.mundoplus.tv/zonatdt/datos_tecnicos_tdt_parametros.php [fecha de consulta: septiembre 2017].

servicios añadidos como la AD. Sin embargo todo servicio añadido representa consumo de ancho de banda y la AD en concreto no comenzó a emitirse de forma sistemática hasta que la Ley Audiovisual de 2010 -y sus posteriores modificaciones- impuso cuotas mínimas de AD a las diferentes cadenas televisivas. Para cumplir con la legislación haciendo el menor consumo posible de ancho de banda, lo que hicieron (y hacen actualmente) las cadenas de TV es emitir la AD como canal de audio aislado, en mono, que el espectador activa o no en su receptor sumándolo al audio general del programa. Este sistema consume sensiblemente menos ancho de banda que la emisión de un audio remezclado completo con la AD por un canal de alternativo, pero conlleva limitaciones importante como que el audio original no se puede adaptar en ningún momento para facilitar la escucha del pista de AD (escenas ruidosas, por ejemplo). Dado que cada canal de TV tiene asignado un ancho de banda fijo (número de bits que se manejan o transmiten en un segundo asociados a dicho canal), y teniendo en cuenta que la imagen de alta definición ya consume gran parte de él, las emisoras persiguen manejar la mayor cantidad de información posible en su ancho de banda, y para ello utilizan algoritmos de compresión de datos.

Los algoritmos de compresión de audio más utilizados son el MPEG 1 LayerIII, MPEG 2 Layer III (conocidos popularmente como mp3), MPEG4 ó AC3 (propiedad de Laboratorios Dolby) y son algoritmos de compresión perceptuales. Este tipo de algoritmos elimina cierta cantidad de información de audio para reducir la cantidad de datos a transmitir basándose en un fenómeno clave en el mundo del sonido y determinante para la AD que es el enmascaramiento.

3.3.1 Sonido y enmascaramiento; enmascaramiento y AD

El enmascaramiento es un fenómeno acústico ligado a la forma en que el oído humano procesa el sonido cuando dos o más tonos puros de frecuencias diferentes se perciben a la vez. Cuando esto ocurre, y dependiendo de la intensidad de cada uno de ellos, existe el riesgo de que el oído sólo perciba uno de estos tonos quedando los otros enmascarados de forma total o parcial. El enmascaramiento es un fenómeno objetivo derivado de la naturaleza interna del oído humano y es la base de los algoritmos de compresión de audio. Queda fuera del alcance de esta tesis entrar en los detalles físicos

del oído que causan el fenómeno (Isbert, 1998), pero la consecuencia es que en la concurrencia de dos o más tonos puros un tono de baja frecuencia puede enmascarar a otro de frecuencia más elevada, tanto más cuanto mayor sea su nivel de presión sonora (Figura 3.4.). En la figura 3.4. se aprecia cómo la cola de enmascaramiento de un tono puro a un Kilohercio enmascara más hacia los tonos agudos que hacia los graves y cómo a medida que aumenta de intensidad, aumenta el campo enmascarado.

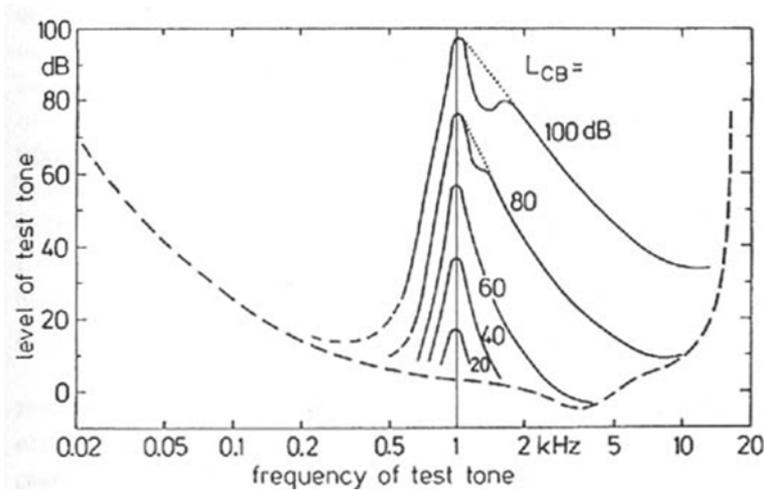


Figura 3.4. Modificación del umbral de audición en presencia de un tono puro de diferentes intensidades.

Este fenómeno se tiene muy en cuenta durante las mezclas y montaje de sonido ya que es el causante de, por ejemplo, la dificultad para entender un diálogo en presencia de ruido de fondo de baja frecuencia típico en explosiones o ruido de motor.

La Figura 3.4 muestra la modificación del umbral de audición para un tono puro a diferentes frecuencias en presencia de un tono puro a un 1 kilohercio a intensidades de 20, 40, 60, 80 y 100 decibelios. Como puede observar, en dicha figura el potencial de enmascaramiento de un tono no es lineal, sino que la cola de enmascaramiento afecta más a las frecuencias más agudas que a las más graves. A efectos prácticos, esta circunstancia es la responsable de que, a un mismo volumen, el sonido de una sirena de policía, por ejemplo, y un extractor de humos no ejerzan el mismo grado de enmascaramiento sobre una voz, siendo el ruido del extractor más entorpecedor para entender la voz, que el de la sirena.

Los algoritmos de compresión MPEG 1 LayerIII MPEG 2 Layer III, MPEG4 ó AC3 usados comúnmente en radiodifusión basan su funcionamiento en eliminar o

disminuir de calidad aquella información sonora que, debido al enmascaramiento, no sería percibida por el oído humano o sería percibida con menor claridad. Si algo no va a ser percibido por el espectador, no tiene demasiado sentido dedicar esfuerzo y bits en incluirlo en la señal de audio. En aquellos casos en que sí va a ser percibido, pero con menor claridad, se puede rebajar la calidad de audio rebajando bits, ya que, aunque el ruido asociado a la pérdida de audio será mayor, el espectador no lo escuchará porque quedará enmascarado por las bajas frecuencias enmascarantes. Siguiendo esta estrategia se puede reducir la cantidad de bits necesarios para representar una señal en márgenes de hasta 13:1. (Davis y Davis, 1984).

Los diferentes algoritmos de compresión (también llamados codecs) utilizan fórmulas muy complejas y se distinguen fundamentalmente en el modo en que analizan la señal y los modelos psicoacústicos en los que se basan. Cuanto más sofisticados y complejos son, mayor es la capacidad de compresión sin que el oyente perciba que no está recibiendo el audio “entero”. El grado de compresión es un parámetro que se puede variar y es el técnico de sonido quien lo establece durante el proceso de producción y elaboración de la señal.

Como puede observarse en el resumen de datos técnicos de canales de televisión y radio nacionales españolas de Mundoplus.tv¹⁵, las distintas cadenas españolas emplean audio digital con diferentes frecuencias de muestreo, profundidad de bit, algoritmos y grados de compresión para cada una de las señales de audio que transmiten asociadas a un canal (Audio Español, QAA, AD). Por ejemplo La1 HD trabaja con audio muestreado a 48KHz codificación AC3 y un flujo de bits de 256kbs (kilobits/segundo) para el audio en español. Los valores empleados para la audiodescripción son muy similares con la única diferencia de que se trata de un único canal (señal monofónica) y el grado de compresión es mayor ya que se está transmitiendo una sola voz. Gracias a esto el número de bits dedicados a la AD es menor (64kbs) que si se emitiera una mezcla de audio alternativa específica para AD. La calidad por tanto del pista de audio de AD emitido por TV es alta, pero la calidad que recibe el espectador puede no serlo debido precisamente al fenómeno del enmascaramiento.

¹⁵ Mundoplus.tv. *Datos técnicos: parámetros configurables*. Recuperado de www.mundoplus.tv/zonatdt/datos_tecnicos_tdt_canales.php [fecha de consulta: octubre 2016].

El fenómeno de enmascaramiento es el responsable de que ante ciertas circunstancias sonoras como bombas, explosiones y otros ruidos de baja frecuencia sea difícil entender un diálogo. Esto es algo que los montadores de sonido tienen muy en cuenta durante el proceso de montaje de una banda sonora y solucionan mediante la equalización o la disminución de intensidad de ciertas frecuencias de la banda sonora para facilitar la escucha de las voces de los personajes. Tanto la voz como la banda sonora son impulsos de audio complejos, en los que se dan múltiples frecuencias a la vez, por lo que un enmascaramiento total y continuado una narración o un diálogo es improbable. El problema es que cuando hay un enmascaramiento parcial de una banda sonora como música o explosiones, no le damos importancia, pero si enmascaramiento parcial se produce sobre una voz, la inteligibilidad se ve muy disminuida, ya que basta con perder o malentender una consonante por otra, la consecuencia puede ser que la voz se oye pero no se entiende.

En el caso de la AD emitida como canal de audio independiente, no hay un técnico encargado de hacer una mezcla de audio o atenuar las frecuencias que más problemas causen a la voz, la situación es similar a la de una película en la que no se realizara mezcla de sonido, algo impensable en cualquier producción audiovisual. Si se tiene en cuenta que, como norma, la regulación de volumen aplicada al audio general se aplica también a la AD, el resultado es una voz aislada que en momentos de ruido suena al mismo volumen que todos los tonos potencialmente enmascarantes presentes en una banda sonora. En estas circunstancias, las probabilidades de que la inteligibilidad de los comentarios se vea afectada en algún momento, es muy alta.

Cadenas como la BBC asocian al archivo de audio de AD una serie de metadatos que hacen que el audio del programa audiodescrito baje de volumen automáticamente cuando los detecta, de manera que el volumen de la AD quede siempre por encima pero esta técnica no se aplica en España y tampoco se establece en ningún documento el número concreto de decibelios de bajada de volumen que se aplica a la banda sonora.

3.4. Conclusiones

El efecto de enmascaramiento es un fenómeno crucial en el mundo del sonido y se utiliza en beneficio de radiodifusores y usuarios en general al posibilitar la retransmisión y el almacenaje de audio en menos espacio (mp3 frente a .wav, por ejemplo) sin perjuicio del espectador pero puede representar un problema cuando voces y ruido se dan en una misma escena. Durante lo que se conoce como proceso de mezcla de sonido los técnicos de sonido y montadores de audio ecualizan y modifican el sonido de una banda sonora para garantizar que los diálogos de los personajes se entiendan con claridad en todas las escenas, pero cuando un pista de AD se añade sin más a el audio general de un programa, es como si esta figura del técnico de sonido desapareciera.

La modalidad de emisión de AD en España no permite hacer ningún tipo de ajuste sobre el audio general donde se insertan los comentarios y esto causa problemas al usuario en ciertos momentos ya que, en ausencia de una mezcla de sonido adecuada, la voz puede quedar enmascarada por la banda sonora subyacente. Ciertas cadenas como la BBC asocian metadatos al audio para que este se atenúe cuando se insertan comentarios de AD, pero esto no es un estándar industrial sino una decisión de la cadena y en otros países este procedimiento no se lleva a cabo. En un escenario como el de España en donde la banda sonora original de los programas no se modifica para adaptarla a la AD, preajustar la AD antes de su emisión podría paliar el efecto de enmascaramiento y mejorar la calidad de la AD que se emite. Este preajuste con unos valores determinados, junto con la exigencia de un formato y calidad de audio mínimos podría servir de base a un estándar industrial de AD que mejorase la experiencia del consumidor sin ir en perjuicio de los intereses de las cadenas.

En el siguiente capítulo se exponen los resultados de un test llevado a cabo con el fin de comparar tres diferentes escenarios sonoros de AD y comprobar si adecuar la grabación de AD a las circunstancias de la banda sonora antes de su inserción mejora la escucha de los comentarios narrativos.

CAPITULO 4. AD COMO AUDIO SOBRE AUDIO: EL PROBLEMA DEL ENMASCARAMIENTO

4.1 Introducción

4.2 Hipótesis y objetivos

4.3 Metodología

4.3.1. Participantes

4.3.2. Mezclas de sonido

4.3.3. Selección, elaboración y mezcla de los clips

4.3.4. Diseño de experimento, cuestionario y desarrollo del experimento

4.4. Análisis estadístico

4.5. Conclusiones

4.1. Introducción

Se ha hecho un repaso en el capítulo 3 de las principales características de la onda de sonido y cómo los métodos de grabación, compresión y retransmisión de audio actuales pueden afectar a su calidad. Este capítulo trata específicamente el problema del enmascaramiento, uno de los más serios que pueden aparecer en la AD y presenta los resultados de un estudio llevado a cabo para comprobar qué repercusión tiene la mezcla de audio en este problema.

A efectos de montaje de sonido, la AD consiste en insertar un conjunto de frecuencias sonoras (la narración), a otro, normalmente mucho más complejo, compuesto por la banda sonora. Aunque idealmente la AD sólo debe producirse en momentos de completo silencio, es muy común que escenas con una banda sonora de fondo, necesiten descripciones para que la acción pueda seguirse correctamente (escenas de acción, por ejemplo). En estas circunstancias, el riesgo de que la voz quede enmascarada por el espectro de frecuencias de la banda sonora es muy alto. Esta es una circunstancia bastante común en España ya que el estándar de sonido que se aplica en TV se aplica a todo lo que emite y como consecuencia el volumen medio de la AD y el de la banda sonora son muy similares (-23 LUFs). La forma de que una frecuencia enmascarada se vuelva a oír, es subirla de volumen hasta que sobrepase el umbral de enmascaramiento.

En términos audiovisuales esto se solventa bien subiendo el volumen de la frecuencia enmascarada, bien bajando el de la enmascarante. En términos de mezcla y emisión de AD, esto significa que la pista de AD tiene que mantener un margen de diferencia de decibelios con respecto a la pista de la BSO. De cuánto ha de ser ese margen es algo que está por determinar, y queda fuera del alcance de esta tesis intentar establecer un principio general respecto a potenciales de enmascaramiento, pero el método de conseguirlo si puede regularse.

Cadenas como la BBC asocian una serie de metadatos a la pista de AD que, cuando son detectados, hacen que la BSO baje de volumen automáticamente, pero las cadenas de televisión en España, no se acogen a esta práctica. En el caso de las cadenas españolas, por ejemplo, esta práctica no se lleva a cabo y el resultado en muchos casos

es un a AD con sonido deficiente que es difícil de entender en escenas con banda sonora de fondo.

Durante esta investigación, la doctoranda participó en el proceso de suministro de audiodescripción a TVE. La empresa Sonidos S2, donde se desarrolló la estancia doctoral y en papel de proveedor cubría todo el proceso de realización de AD, desde elaboración de guion hasta la grabación y ajuste a código de tiempo del audio de AD. No había mezcla con el programa ya que la emisión de la AD en TVE se hace como pista aislada. El flujo de trabajo era: Recepción de pedido de AD, normalmente semanal, con los episodios de serie, programas, o películas que se iban a emitir y fecha de entrega del pista de AD. Asignación de guionista y locutor para cada uno de los pedidos, descarga del archivo vía ftp de cliente y envío a guionista. Los plazos habituales eran de unos 4-5 días para largometrajes y alrededor de 3 para series de duración media. El flujo de trabajo una vez recibido el guión de AD con códigos de tiempo se realizaba la grabación en un estudio profesional equipado con micrófono Rode NT1000, programa de grabación Protools versión 11, tarjeta de sonido RME Fireface UC y *plugins* de sonido de *Wave* e *Izotope*. La grabación se realizaba ajustando en tiempos, esto es, con el soporte visual del programa y el código de tiempo del guion, de modo que las posibles desviaciones de tiempo o errores de guion, eran detectados en esta fase. Una vez realizada la grabación, el técnico de sonido hacía un ajuste manual de la locución para asegurar que no hay solapamiento entre descripciones y diálogos y trataba el audio para optimizar la sonoridad de la voz de cada locutor, después el audio se normalizaba a -16 RMS (petición de cliente) y guardaba en formato ACC. Este formato de archivo es poco frecuente en el ámbito audiovisual profesional por el porcentaje de pérdida de datos que implica frente a un .wav, pero es un estándar de emisión común en radiodifusión, ya que es mucho más ligero.

Aunque el audio entregado siempre era de calidad formato .wav sin pérdidas de alta calidad (48.000 Hz), los testimonios de locutores, técnicos y otros implicados en el proceso eran muy diferentes en cuanto a cómo era la recepción de la AD en sus televisores. Dejando al margen los problemas de ciertos monitores de TV antiguos para recibir el canal de audio de AD, la queja más frecuente de los implicados al convertirse en televidentes de su propio producto era la dificultad para oír claramente la voz del locutor de AD por encima de la banda sonora de fondo de los programas. Hasta el punto

de que en ciertas escenas (películas de acción, por ejemplo), simplemente no se oían. Este problema para entender claramente las palabras de una narración en presencia de ruido de fondo se debe al efecto de enmascaramiento expuesto en el capítulo 3. En un escenario en el que el volumen del programa audiodescrito y la AD es el mismo, el riesgo de enmascaramiento de la voz en escenas de ruido, es muy alto.

4.2. Hipótesis y objetivos

En una AD la narración se tiene que oír claramente y para ello, la guía ITC (2000) recomienda bajar el volumen del ruido de fondo, si lo hubiera, cuando sea necesario. La hipótesis de este experimento es que subir el volumen de audio de una AD en las escenas donde no se oye con claridad, debería tener el mismo efecto que bajar la banda sonora.

El primer objetivo de este estudio fue determinar si una mezcla de audio en la que el volumen de la AD se ajusta manualmente para aumentarlo en escenas ruidosas soluciona el problema del enmascaramiento y mejora la escucha de los comentarios de AD. El segundo objetivo fue comparar varios tipos de mezcla de sonido de AD para establecer preferencias de usuario. Para ello se diseñó un experimento en el que un grupo de usuarios evaluó diferentes aspectos sonoros de tres mezclas de sonido distintas.

4.3. Metodología

A lo largo de este capítulo, las medidas de sonido se expresarán en dB (decibelios) para expresar el volumen en un momento concreto, y RMS (*Root Mean Square*) para hablar de nivel de normalización (media de volumen de una muestra de audio de más duración). Aunque el estándar de regulación en radiodifusión es el EBU R128, las unidades de sonoridad que utiliza este estándar, los LUFS (*Loudness Unit Full Scale*), no se usan en el ámbito de la grabación de sonido. En el flujo de trabajo habitual, se graba y postproduce en decibelios y posteriormente el audio se trata con un proceso de edición específico que evalúa si el nivel de sonoridad se ajusta a los criterios de EBU R128. En caso negativo, un técnico realiza los ajustes de volumen necesarios hasta que el audio cumple con la regulación. Este es el proceso que se siguió también para la grabación y mezcla de los clips en este experimento, y los términos dB y RMS serán los que se utilicen en adelante. Si bien en todos los casos, tanto el audio de partida (antes de mezclar) como el que se utilizó en la prueba, se normalizaron conforme a EBU R128. En este capítulo se aludirá a la película escogida, “*Malditos Bastardos*” (2009) por su título en español ya que se trabajó con el audio de su versión doblada al español.

4.3.1 Participantes

El grupo de participantes fue facilitado por la Organización Nacional de Ciegos de España (ONCE) gracias a la colaboración entre la empresa donde se realizó este doctorado industrial dicha institución. Se consiguió una muestra homogénea de 32 participantes, 17 hombres y 15 mujeres. Todos los participantes eran ciegos totales o con deficiencia visual grave. Se entiende aquí como deficiencia visual grave la que exigen los estatutos de la ONCE, definida como disponer en ambos ojos de al menos una agudeza visual igual o inferior a 0,1 obtenida con la mejor corrección óptica posible o campo visual reducido a 10 grados o menos¹⁶. Dada la especificidad de los estatutos de la ONCE no se suministró un cuestionario adicional de evaluación de la visión a los

¹⁶ ONCE (2017). *Requisitos para la afiliación a la ONCE*. Recuperado de www.once.es/new/afiliacion/requisitos [fecha de consulta: septiembre 2017].

participantes ya que se consideró que la pertenencia a la organización acreditaba este punto satisfactoriamente.

4.3.2. Estímulos

Para este experimento, el primer paso fue definir cuántas y con qué características serían las mezclas de sonido a evaluar y se definieron tres por considerarse las más representativas de la realidad de la AD como servicio de accesibilidad audiovisual en TV:

Mezcla 1: se realizó sin hacer ningún ajuste de volumen, ni a la BSO original ni a la grabación de AD. Para esta mezcla simplemente se normalizaron tanto el volumen de la escena como el de la pista de AD a -23 LUFS y la AD se insertó conforme a los códigos de tiempo.

Mezcla 2: se realizó emulando el sistema de emisión utilizado por la BBC y aconsejado en las ITC guidelines (ITC, 2000). Para esta mezcla, la AD se insertó conforme a los códigos de tiempo, normalizada a -23 LUFS y el audio original se redujo entre 5 dB y 7 Dbs cuando había comentarios de AD en presencia de ruido de fondo.

Mezcla 3: se realizó adaptando manualmente el volumen de la AD a la BSO original. En esta mezcla el audio original de la BSO se mantuvo intacto y la AD se adecuó manualmente a cada pasaje de la escena, tanto subiendo volumen en presencia de ruido de fondo, como bajándolo en momentos de silencio. En ambos casos, el ajuste de decibelios fue de entre 5 y 7.

Se decidió hacer estas tres mezclas porque se consideró que son las más representativas de la realidad de AD. La mezcla 1 es la manera en que se emite en los canales de la televisión pública española (TVE), la mezcla 2 es la recomendación de ITC (2000) y la forma de emisión de la BBC, y la mezcla 3 la que esta tesis propone como estándar. En el apartado de análisis estadístico la mezcla 1, o método de emisión en España será la que se tome como referente.

4.3.3 Criterios de selección de la película y descripción de escenas

Para la realización de los clips se consideró que la manera más fiable de controlar todas las variables sonoras desde grabación hasta mezcla, era generar los estímulos en lugar de intentar diferentes muestras de ADs emitidas. Una vez tomada esta decisión había total libertad de elección de la película. El único criterio fijo era que se tratara de una película con escenas de banda sonora ruidosa en las que poder probar las diferentes mezclas de audio.

Se escogió la película de Quentin Tarantino (2009) "*Malditos bastardos*". Esta película ya había sido objeto de estudio en relación con diferentes aspectos de la AD en el libro Audio description: New perspectives illustrated (Maszerowska, Matamala y Orero, 2014) y esta tesis deseaba que las posibles conclusiones alcanzadas en este estudio pudieran contribuir, aunque fuera acumulativamente, a aquellas investigaciones. Además esta película es particularmente rica en elementos y referencias sonoras cuyo papel va mucho más allá del simple acompañamiento a las imágenes (Szarkowska y Orero, 2014) y se consideró que este podía ser un factor que hiciese el test más atractivo para los participantes.

Dentro de la película se seleccionaron tres escenas siguiendo principalmente un criterio temático y sonoro. El criterio temático consistía en escoger escenas que presentasen una unidad narrativa comprensible aun desligadas del marco general de la película, y que no durasen más de 2 a 3 minutos. El criterio sonoro era encontrar de amplitud de rango dinámico, para asegurar que cada escena tenía momentos de diálogo sin apenas ruido de fondo y momentos de banda sonora potente que mostrase fuentes de ruido de diferente naturaleza para cada una de las escenas. En ese sentido se seleccionó: ruido de disparos/explosiones (escena 1), ruido de banda musical con instrumentos de viento predominantes (escena 2) y ruido metálico/agudo (guitarra eléctrica, escena 3). A continuación se describen las características de cada uno de las escenas seleccionadas:

Escena 1: 2:24 minutos de duración.

Entre el minuto 00:00 y el 00:40 la banda sonora es prácticamente nula y cuando los diálogos se detienen el silencio es casi absoluto por lo que en la primera parte de este lo único que se oye es el dialogo original de la película y las descripciones de AD. Los diálogos hasta el minuto 00:40 muestran un RMS de -38,4 dBs y picos de -19 dBs. A

partir del 00:40 se inicia un crescendo de la banda sonora de fondo que culmina en el minuto 1:45 en el que la confluencia de banda musical y fuertes ruidos ambiente llega a alcanzar picos de -2,1 dBs. El RMS entre el minuto 1:15 y 1:45 es de -23,8 dBs.

Escena 2: 2:22 minutos de duración.

Hasta el minuto 00:35 no hay apenas banda sonora. Los comentarios se insertan en un ambiente de silencio. Las voces de los personajes hasta el minuto 00:35 muestran un RMS de -39 dBs con picos de -20. En el minuto 00:35 suena un ruido de golpeo con picos de -18,8 dBs, clave para entender la escena. Este ruido se mantiene intermitente mientras la banda sonora inicia una subida de volumen progresiva que culmina entre el minuto 1:20 y 1:25, periodo en el que el RMS es de -23,5 dBs con picos de -5 dBs.

Escena 3: 1:02 minutos de duración.

Entre el minuto 00:03 y el 00:44 de este clip hay una banda sonora de guitarra eléctrica continuada, con RMS de -23,4 dBs y picos de -5,8 dBs. En la parte final del clip, la guitarra desaparece y el volumen general de la escena baja considerablemente, no hay banda sonora y el RMS de la voz del diálogo es de -37,8 dBs con picos de -19,4 dBs.

4.3.4. Creación de guión y grabación de la AD

Dado que el elemento a evaluar era la mezcla de sonido, se descartó la posibilidad de reutilizar una AD existente de la película y tanto el guión como la grabación de las narraciones utilizadas se crearon ex profeso para este test. El criterio principal de elaboración de los guiones fue sonoro, no temático. La intención era mantener un ritmo y tono de narración continuo en todos los momentos que no hubiera diálogo para exponer la voz al mayor número de circunstancias sonoras de fondo posible.

No se encontró un criterio justificado en el que basar la elección del género de la voz del locutor, por lo que la elección se hizo en base a criterios sonoros. Como se expone en el capítulo 3 (cfr. Figura 3.5), el efecto de enmascaramiento no es lineal y se ejerce más hacia los agudos que hacia los graves. Siguiendo este criterio se eligió una voz grave, con mayor potencial enmascarador.

El locutor presenta en su voz unas frecuencias medias que se sitúan entre 47 y 890 Hz, pero llega a cubrir frecuencias entre 20 y 1590 Hz. Esta amplitud de rango optimiza la claridad de su discurso incluso en entornos ruidosos. La grabación de los tres guiones (escenas 1, 2 y 3) se realizó en una sola sesión para que las características sonoras de la pista de AD fueran idénticas para las tres escenas. La sesión de grabación se realizó en estudio, con un micrófono Neumann BCM 705 dinámico. Los micrófonos dinámicos se diferencian de los de condensador en el mecanismo que utilizan para convertir la presión acústica provocada por la onda sonora en impulsos eléctricos. Actualmente hay micrófonos de alta calidad de uno y otro tipo y la elección se basa en preferencias personales en cuanto a los matices que se quiere enfatizar o el entorno donde van a utilizarse. El modelo de micrófono utilizado para este experimento realiza un registro de más calidad en las bajas frecuencias características de la voz humana que los micrófonos de condensador. Cuenta con una sensibilidad de 20 Hz a 20000 Hz, una relación señal ruido muy baja (68 dB) que asegura que aparezca ruido de estática de fondo. También cuenta con un realce en las frecuencias de 2 kHz a 9 kHz para facilitar la inteligibilidad del habla, y de esta forma compensar la mayor sensibilidad a frecuencias graves. La tarjeta de sonido utilizada fue una Fireface UC con software de grabación un Protools 11. Este es un equipamiento profesional estandarizado en el sector de la grabación de audio. No se realizó ecualización de la voz ni se aplicaron filtros durante la grabación ni en la posproducción para evitar alteraciones de las características de la voz. El proceso de posproducción se limitó a normalizar el audio para cumplir con los parámetros del EBU R 128 y a ajustar la grabación a los códigos de tiempo. La locución resultante tiene un volumen muy estable, con un RMS de -23,6 dBs y pico máximo de -2,0 dBs.

Una vez grabadas, normalizadas y ajustadas las tres audiodescripciones, se insertaron en cada una de las escenas conforme a los parámetros de mezcla 1, 2 y 3 descritos en el apartado 3.2 Mezclas de sonido.

4.3.5. Diseño del experimento, cuestionario y desarrollo del experimento

Se decidió utilizar un diseño de cuadrado latino intrasujeto, que permite maximizar el número de registros a partir del número de participantes en un estudio. Una vez obtenidos los nueve clips de audio conforme al esquema siguiente:

	Escena 1	Escena 2	Escena 3
Mezcla 1	escena 1.1	escena 1.2	escena 1.3
Mezcla 2	escena 2.1	escena 2.2	escena 2.3
Mezcla 3	escena 3.1	escena 3.2	escena 3.3

Tabla 4.1. Escenas obtenidas tras la aplicación de los tres tratamientos de sonido.

Se diseñó una distribución aleatorizada (anexo 1), para evitar el efecto aprendizaje, en la que cada participante podría evaluar todas las muestras de sonido sin repetir nunca escena. Así, con una muestra de 32 participantes, se obtuvieron 96 registros.

La toma de datos se hizo mediante cuestionario personal (anexo 2) que cada participante respondía después de cada uno de los clips. El cuestionario constaba de 12 preguntas en las que se pedía a los participantes que evaluaran diferentes aspectos sonoros de las escenas, del 1 al 3, considerando el uno total desacuerdo con la afirmación y el tres total acuerdo. El experimento se llevó a cabo en dos jornadas consecutivas en una de las salas de las instalaciones de la ONCE en Madrid y la reproducción se realizó con una Tablet Samsung Note 3 conectada a altavoces. No se utilizaron auriculares porque se consideró más cercano a las condiciones habituales de un domicilio la reproducción con altavoces. El orden de reproducción de escenas con cada uno de los participantes se aleatorizó de manera que cada participante escuchase las tres mezclas sin repetir ninguna y las tres escenas sin repetir ninguna. El anexo 1 presenta el patrón de aleatorización utilizado.

4.3.6. Análisis estadístico

Se utilizó la prueba no-paramétrica U de Mann-Whitney para comparar los distintos escenarios. Para ello se consideraron las distintas valoraciones como determinaciones independientes.

El nivel de significación fijado para el estudio fue de 0.05. De este modo para P valores inferiores a 0.05, la hipótesis nula es rechazada, y el resultado se definió como

estadísticamente significativo'. El software utilizado para realizar los tests estadísticos y generar los gráficos fue el GraphPad Prism 7¹⁷.

¹⁷ Disponible en www.graphpad.com.

4.4. Resultados

Los resultados obtenidos en el experimento fueron estadísticamente significativos para las preguntas 1, 3, 5 y 7. La pregunta 8 no llega a obtener un valor de P suficiente pero queda muy cerca y se comentará su caso en particular. En la Tabla 4.2. se muestran las valoraciones expresadas en porcentaje de la valoración de los usuarios (1 en total desacuerdo, 2 no sabe y 3 total acuerdo) de cada una de las preguntas para cada una de las mezclas. En la tabla 4.3. se presentan los P valores referentes a la comparación entre las mezclas 2 (bajar volumen) y 3 (mezcla dinámica) dinámica con respecto a la 1 (emisión).

Se comentan en esta sección las preguntas con resultado estadísticamente significativo.

Preguntas	Valoración de usuarios (%)*								
	Emisión			Bajar vol.			Dinámica		
	1	2	3	1	2	3	1	2	3
1. En general, la locución de AD era clara y comprensible.	28	28	44	0	6	94	9	31	59
2. He escuchado con claridad todos los sonidos ambiente en todo momento	6	9	84	6	28	66	13	9	78
3. El volumen de la narración me parece adecuado en todo momento a lo largo de la escena.	28	53	19	6	47	47	28	34	38
4. En algunos momentos de la escena el volumen de la AD me parece excesivamente alto.	59	16	25	38	28	34	44	19	38
5. En algunos momentos de la escena el volumen de AD me parece demasiado bajo.	41	22	38	81	13	6	50	16	34
6. En algunos momentos, la AD dificulta la escucha de los sonidos ambiente de la escena.	72	19	9	59	25	16	69	19	13
7. En algunos momentos, la banda sonora dificulta la escucha clara de la AD	25	38	38	81	6	13	50	19	31
8. He notado cambios de volumen en la narración de AD a lo largo de la escena.	50	22	28	63	13	25	34	6	59
9. En general me parece que el volumen de la AD y de la escena original están a un nivel similar.	28	25	47	31	22	47	31	28	41

*La valoración indica el % de usuarios que respondieron: 1 en total desacuerdo, 2 no sabe y 3 totalmente de acuerdo.

Tabla 4.2. Respuestas de los usuarios a las 9 preguntas.

Preguntas	Valores de P	
	Mezcla 2. Bajar vol.	Mezcla 3. Dinámica
1. En general, la locución de AD era clara y comprensible.	<0,0001	0,050
2. He escuchado con claridad todos los sonidos ambiente en todo momento	0,141	0,544
3. El volumen de la narración me parece adecuado en todo momento a lo largo de la escena.	0,006	0,122
4. En algunos momentos de la escena el volumen de la AD me parece excesivamente alto.	0,097	0,232
5. En algunos momentos de la escena el volumen de AD me parece demasiado bajo.	0,001	0,413
6. En algunos momentos, la AD dificulta la escucha de los sonidos ambiente de la escena.	0,329	0,651
7. En algunos momentos, la banda sonora dificulta la escucha clara de la AD	<0,0001	0,062
8. He notado cambios de volumen en la narración de AD a lo largo de la escena.	0,573	0,035
9. En general me parece que el volumen de la AD y de la escena original están a un nivel similar.	0,961	0,772

Tabla 4.3. Valores de P referentes a la comparación entre las mezclas bajar volumen y dinámica con respecto a la mezcla 1 emisión.

Cuando se analizaron las diferencias entre las valoraciones a cada una de las preguntas, tomando como referencia la mezcla 1 (emisión), se encontraron diferencias significativas en las preguntas 1, 3 5 y 7 de la mezcla 2 (bajar volumen) y la 8 de la mezcla 3 (dinámica).

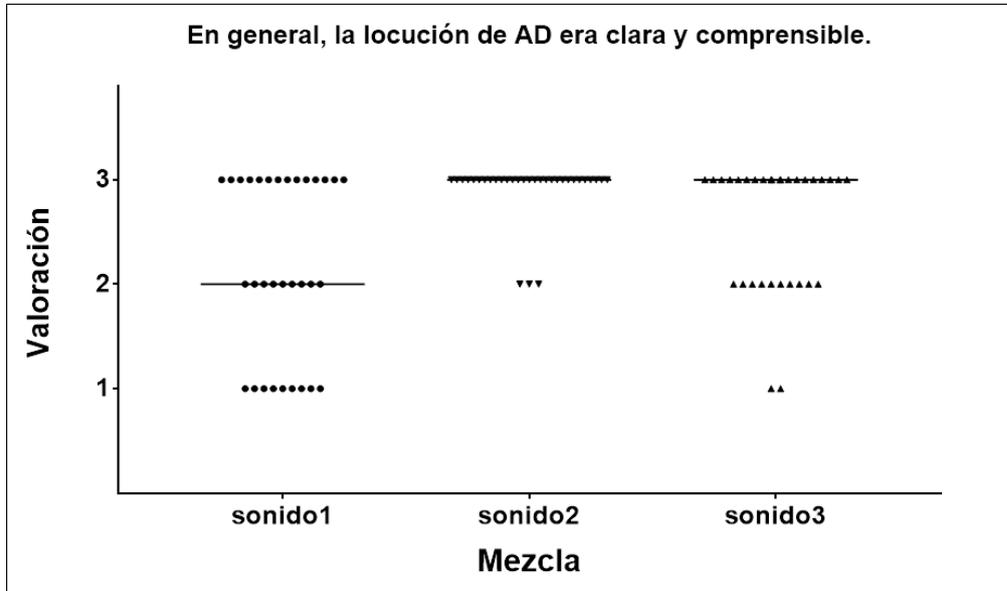


Figura 4.1. Gráfica de los resultados obtenidos para la pregunta 1.

En relación con la pregunta 1, un cero 0% de los participantes tuvo problemas para la escucha clara de los comentarios de AD en la mezcla 2 frente al 28% de la mezcla 1 (emisión) y el 9% de la mezcla 3 (ajuste dinámico). Si comparamos los resultados, la diferencia entre las mezclas 1 y 2 es claramente significativa. La comparación entre las mezclas 2 y 3 da un p valor de 0,050, justo en el límite de ser significativa.

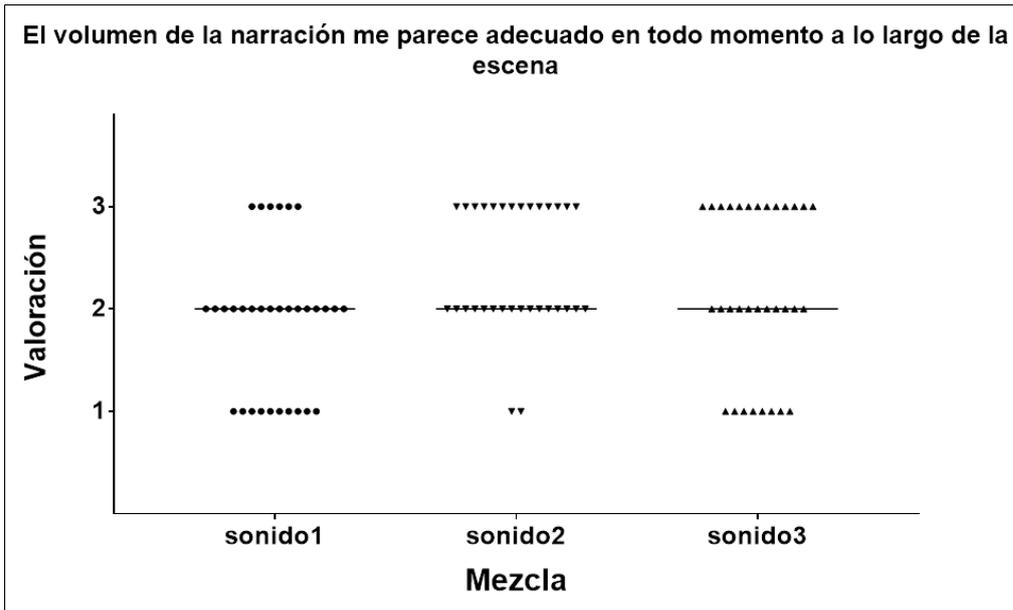


Figura 4.2. Gráfica de los resultados obtenidos para la pregunta 3.

En la pregunta 3 hay una diferencia significativa entre las mezclas 1 y 2 pero no entre 1 y 3. El porcentaje de participantes que estuvo en total acuerdo con la afirmación fue de un 47% para la mezcla 2 frente al 19% de la mezcla 1, el porcentaje de participantes en total desacuerdo fue del 6% para la mezcla 2 frente al 28% de la mezcla

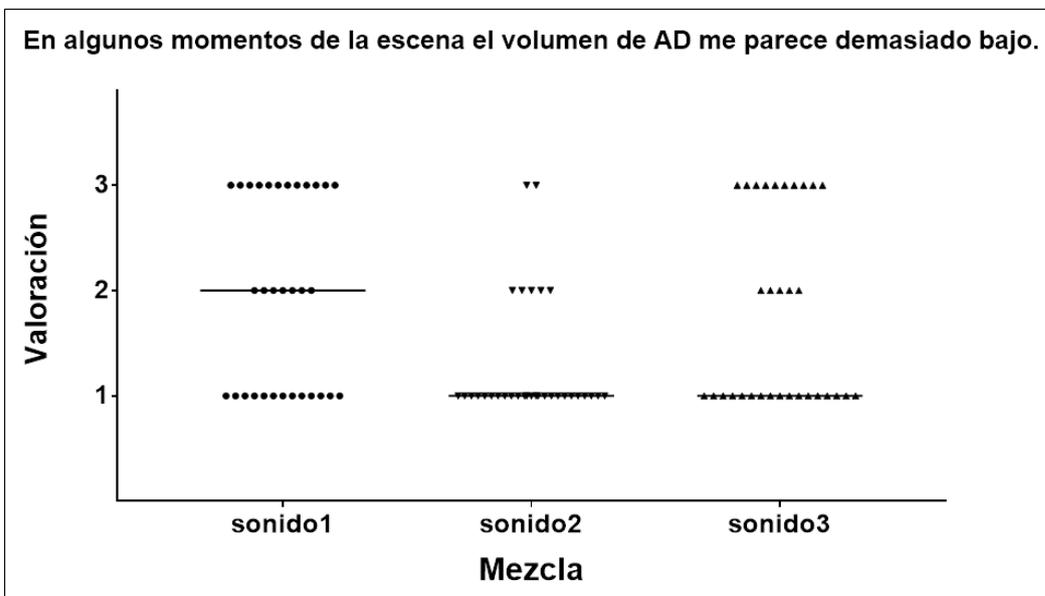


Figura 4.3. Gráfica de los resultados obtenidos para la pregunta 5.

La pregunta 5 muestra diferencias significativas entre las mezclas 1 y 2 pero no entre las 1 y 3. Para esta afirmación un 6 % estuvo en total acuerdo para la mezcla 2 y un 38% para la mezcla 1.

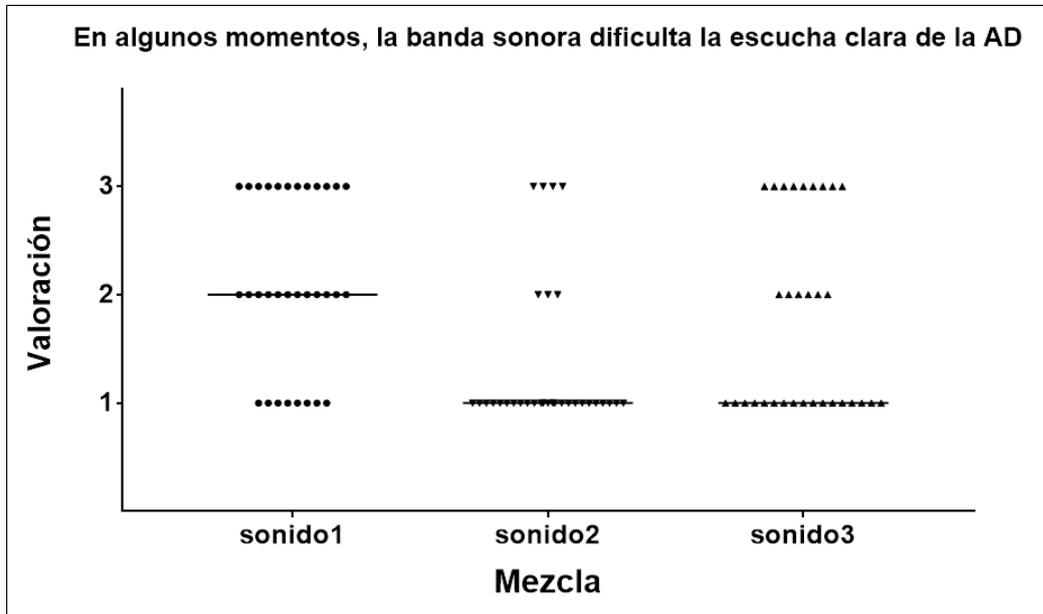


Figura 4.4. Gráfica de los resultados obtenidos para la pregunta 7.

De nuevo en esta pregunta las diferencias son significativas al comparar las mezclas 1 y 2, pero no las 1 y 3. Un 38% de participantes encontró que la banda sonora dificultaba la escucha de los comentarios en la mezcla 1, frente a un 13% de la mezcla 2.

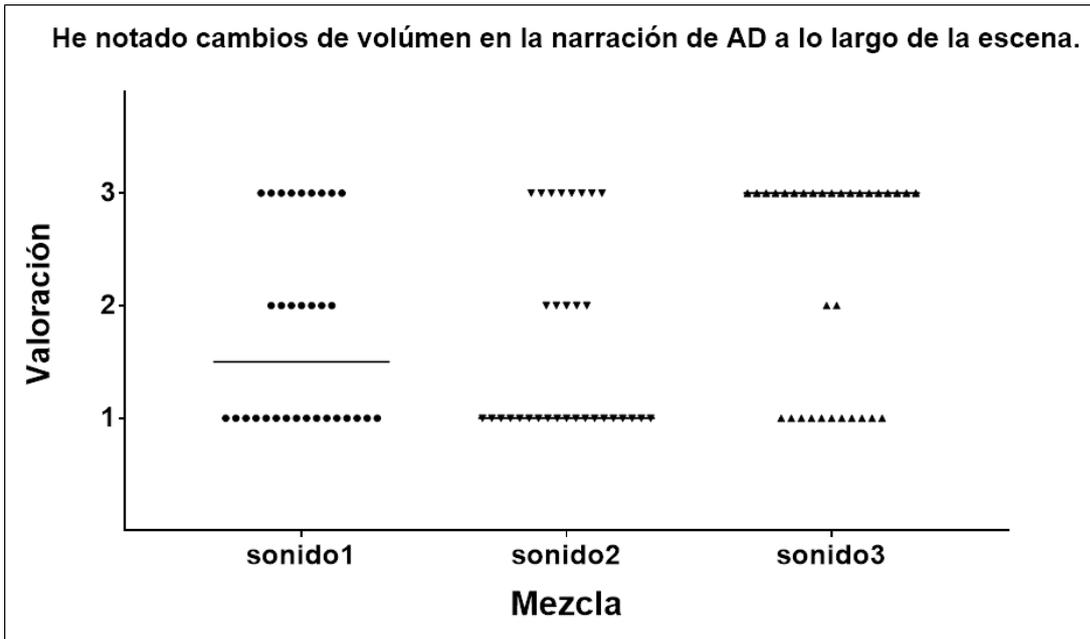


Figura 4.5. Gráfica de los resultados obtenidos para la pregunta 8.

Para la pregunta 8 se encontró significativo el resultado de la comparación entre las mezclas 1 y 3. En la mezcla 3, un 59% de los participantes notó cambios de volumen en la narración de la audiodescripción frente al 28% de la mezcla 1.

4.5. Discusión

Como consideración general previa, es interesante observar que las preguntas que obtienen un resultado estadístico más significativo (p valor más bajo), son precisamente las que hacen referencia a la tener dificultad para escuchar la narración de AD en algún momento. Aunque todas las afirmaciones se refieren a la percepción de elementos sonoros, sólo la 1 (En general, la locución de AD era clara y comprensible.); la 5 (En algunos momentos de la escena el volumen de AD me parece demasiado bajo); y la 7 (En algunos momentos, la banda sonora dificulta la escucha clara de la AD) implican un peligro real de pérdida de información de la narración de AD, con los problemas de accesibilidad que ello conlleva.

La pregunta 1 es la más relevante desde el punto de vista de la accesibilidad, dado que lo que se evalúa es la posibilidad del usuario de acceder al mensaje y los datos muestran que con el referente de emisión actual en España, habría un 28% de usuarios que no lo consiguen. De todos los datos éste es el más alarmante ya que no compromete la calidad del servicio, sino el servicio en sí. Durante el análisis de resultados se detectó el problema de la poca especificidad de la pregunta, en cuanto que no se apunta el motivo por el que no se entiende bien la locución de AD. Si tenemos en cuenta que la locución es la misma para las tres mezclas de audio y que en la mezcla 2 ningún usuario tuvo problemas para oírla claramente, se descarta que el problema pueda venir de la grabación de AD. Si el problema no es la grabación en sí las dos posibilidades restantes son enmascaramiento o saturación por exceso de volumen. La saturación de volumen queda descartada porque la postproducción y normalización de los audios a -23 LUFS garantiza que el audio no sature y la reproducción se hizo a una potencia media de los altavoces (volumen de 8 sobre 15) lo que también descarta problemas de saturación. El factor restante es el enmascaramiento. En el clip 1, el audio original de la BSO entre el minuto 1:15 y 1:45 tiene un RMS -23,8 dBs, en el clip 2 entre el minuto 01:20 y 01:25 el RMS de la BSO es de -23,5 dBs y en el clip 3, entre el minuto 00:03 y el 00:44 la BSO un RMS de -23,4 dBs. Los comentarios de AD en estos momentos estaban a un RMS de -23,6 dBs, de lo que se deduce que a un volumen similar, los comentarios de AD quedan enmascarados por la BSO. En la mezcla 2 el volumen de la banda sonora se redujo cuando era necesario para que nunca superara los -18 a 18,5 dBs (5 dBs menos

que la RMS de la AD) y ningún usuario reportó problemas de escucha de la AD. Los resultados para esta pregunta con la mezcla 3 están en el límite de ser estadísticamente significativos (p valor 0,050) por lo que también se comentarán también. En la mezcla 3 la banda sonora original se dejó intacta y se subió el volumen de la AD para que siempre estuviera unos 5dBs por encima. Un 9% de usuarios tuvo problemas para escuchar claramente la locución. Aunque con este tipo de mezcla se mejora el porcentaje de usuarios que no escucha con claridad la AD frente a la mezcla referencia de emisión, el problema no se elimina.

Los resultados de la pregunta 3, podrían tener varias lecturas ya que la posible no adecuación del volumen de la narración de AD no se especifica. Aun así, en términos generales, la mezcla 2 puntúa significativamente mejor que la 1, con un 47% de participantes que encontraban que el volumen era adecuado en todo momento en la mezcla 2 frente al 6% de la mezcla 1. Es difícil saber el porqué del alto número de usuarios que se acogieron al “no sabe” en esta pregunta (un 53%, 47% y 34% respectivamente, los porcentajes más altos en todo el estudio) una posible interpretación es que el término “adecuado” como algo al margen de la percepción del usuario ya que en preguntas sobre lo que se oye o no se oye, las respuestas son más claras. La redacción de las preguntas puede haber sido un problema de diseño del cuestionario.

Las diferencias entre las mezclas 1 y 2 para las preguntas 5 y 7 vienen a corroborar los resultados obtenidos para la pregunta 1. Con 38% de usuarios en la mezcla 1 frente a un 6% en la 2 que encontraron la AD estaba demasiado baja; y un 38% de participantes en la mezcla 1 frente a un 13% en la dos que tuvieron problemas para escuchar la AD sobre la BSO, se puede afirmar que las mezclas de AD en las que el volumen de la BSO original y el de la AD están a niveles similares provoca problemas de accesibilidad a parte de los usuarios.

En la pregunta 8, referente a si se perciben cambios de volumen de la voz del locutor de AD, se corrobora que los participantes notan que el volumen de la AD varía a lo largo de las escenas.

Para las preguntas 2, 4, 6 y 9, no se obtuvieron resultados significativos. Esto puede deberse a que no hay diferencias en la percepción de los elementos de la banda sonora entre las tres mezclas, pero dado que sí las hay en la percepción de la grabación de AD una posibilidad sería que los participantes en el estudio focalizasen su atención

en la narración de AD, ya que es su herramienta de acceso a los elementos que no perciben que lo que la rodea.

4.6. Conclusiones

Los resultados confirman que la mezcla de sonido tiene un impacto directo sobre el grado de accesibilidad real del producto audiodescrito. Cuando el volumen medio de la AD y el de la BSO están a un mismo nivel y hay ruido de fondo, escuchar las descripciones con claridad resulta difícil para un 28% de los usuarios con el detrimento del servicio que ello supone.

Mantener una diferencia de volumen de alrededor de 5 dBs a favor de la AD mejora el problema pero sólo lo elimina si el la BSO la que se baja de volumen. Cuando es la voz de la AD la que se aumenta por encima del ruido de fondo, el resultado mejora frente a no hacer ajustes pero sigue causando problemas a un número de usuarios. Esto podría deberse a un efecto de saturación por exceso de volumen en los altavoces pero harían falta más pruebas para establecer la causa del problema.

Aunque hacen más estudios que ahonden en el tema, la regulación del volumen debe formar parte de cualquier estándar de AD si se quiere garantizar la efectividad del servicio y, en caso de que una manipulación del audio general de las emisiones no sea posible por parte de los radiodifusores, el ajuste manual del volumen de la AD previo a su inserción representaría una mejora frente a la práctica habitual de mantener el volumen de AD estable.

CAPITULO 5. VOZ FRENTE A RUIDO: BÚSQUEDA DE UN MARGEN DE SEGURIDAD

5.1. Introducción

5.2. Hipótesis y objetivos

5.3. Metodología

5.3.1. Participantes

5.3.2. Estímulos

5.3.3. Diseño y desarrollo del experimento y cuestionario asociado

5.4. Análisis estadístico

5.4.1. Resultados

5.4.2. Discusión

5.5. Conclusiones

5.1. Introducción

Una de las principales amenazas para la calidad de la AD es la realización de una mezcla de sonido de mala calidad. Dada la naturaleza sonora de la audiodescripción, una mezcla que no tenga en cuenta el ruido de fondo en los momentos en que se inserta un comentario puede llegar a poner en peligro la inteligibilidad de la narración del locutor. El capítulo 4 de esta tesis presenta resultados que demuestran que la mezcla de sonido tiene un impacto directo sobre la AD y la experiencia de usuario.

En el capítulo 2 se expone la forma en que el sonido en la AD ha sido tratado en artículos de investigación, códigos de buenas prácticas y estándares de regulación y no se haya una conclusión en cuanto a cuál es el volumen de audio recomendado para una pista de AD. El volumen de audio en AD se aborda de manera fugaz en *guidelines* y códigos de buenas prácticas (ITC, 2000; ITU, 2013), que se limitan a señalar que la AD debe oírse siempre con claridad sin concluir qué volumen con respecto a la BSO es el adecuado. En ITU (2013) curiosamente no se establecen medidas mínimas de diferencia de decibelios entre AD y banda sonora para garantizar la escucha clara de la AD, pero sí un límite máximo para evitar saturación (6 dB) que el volumen de audio de la AD no debe rebasar con respecto al volumen del audio general. Este documento también aconseja intensificar ligeramente las frecuencias entre los 1 y 6 kHz y bajar las frecuencias por debajo de los 200 Hz de la grabación de AD durante el proceso de mezcla para conseguir una mayor claridad del habla. Esta medida resalta el sonido de las consonantes mientras que palia las posibles distorsiones y reverberaciones características de las frecuencias más bajas y contribuye a una mayor claridad de la voz. Estas medidas, aunque útiles para mejorar la calidad de sonido de la voz, no sirven de nada si la normalización de volumen de audio de AD se realiza a un volumen demasiado bajo. Se han realizado estudios con usuarios para evaluar la percepción de diversos parámetros de AD tales como la velocidad de la narración (Cabeza-Cáceres, 2013) o la utilización de voz sintética frente a voz humana (Szarkowska, 2011). El estilo de las descripciones y su grado de objetividad también han sido estudiados con tests de usuario (Walczak y Fryer, 2017) pero se dispone de estudios que aborden específicamente la diferencia de volumen aconsejable entre narración de AD y posible ruido de fondo.

Como se exponía en el capítulo 3, el fenómeno del enmascaramiento se produce cuando un sonido impide la percepción de otro y es un riesgo importante para la AD ya que puede comprometer la efectividad del servicio al dificultar la escucha de los comentarios descriptivos en escenas ruidosas. Para revertir un enmascaramiento y hacer que la señal enmascarada vuelva ser percibida hay que aumentar su intensidad hasta superar el umbral de enmascaramiento. En el terreno audiovisual esto se traduce en subir el volumen pero la ausencia de estándar actual de mezcla de sonido en AD hace que la emisión de AD en TV se haga al mismo volumen que el audio general. En estas circunstancias el problema del enmascaramiento resulta frecuente en ciertos programas o géneros en los que la BSO tiene más protagonismo. El capítulo 4 presenta los resultados de un test llevado a cabo para determinar si la mezcla de audio afecta y cómo al usuario de AD. Los resultados muestran que la mezcla afecta a la facilidad del usuario de entender los comentarios de AD en las escenas de ruido y que los mejores resultados se obtienen bajando el ruido de fondo cuando se insertan comentarios.

Averiguar qué diferencia de volumen es aconsejable mantener, como mínimo, entre la banda sonora y la voz de la AD sería un primer paso para alcanzar un estándar de mezcla de sonido aplicable para TV y productoras de AD.

Este capítulo presenta un estudio enfocado específicamente a determinar el margen de volumen mínimo que se debería mantener entre la voz de AD y el posible ruido de fondo para garantizar la inteligibilidad de la narración de AD.

El estudio fue llevado a cabo con 4 usuarios con discapacidad visual y 22 sin discapacidad visual, que evaluaron el volumen de ruido máximo deseable y aceptable, para entender una narración sobre un fondo de ruido.

5.2. Hipótesis y objetivos

La hipótesis de este estudio es que es posible establecer una diferencia de volumen de audio concreta en dB que garantice que los comentarios de AD se oyen con claridad incluso si se insertan en escenas ruidosas. Para comprobar esta hipótesis se diseña un experimento en el que los usuarios evalúan la inteligibilidad de una narración de AD frente a un ruido de fondo que va aumentando de volumen.

El objetivo principal de este experimento es obtener un dato en dB que exprese la diferencia de volumen que debe mantenerse entre el audio general de un programa y la pista de AD para garantizar que AD se escucha con claridad incluso en escenas de ruido. El objetivo secundario es determinar si la práctica común de emisión en la que la pista de AD y el audio general se emiten al mismo volumen, resulta cómoda para el espectador en momentos de ruido de fondo.

5.3. Metodología

Esta sección recoge las características de los participantes en el estudio y la creación de los estímulos de ruido y voz utilizados para determinar la diferencia de volumen aconsejable entre voz y ruido de fondo.

5.3.1. Participantes.

Debido a la dificultad de localizar y coordinar a un número elevado de invidentes, se realizó un pre-test con 4 usuarios con discapacidad visual facilitados y pertenecientes a la ONCE y posteriormente se extendió el estudio a 22 usuarios sin discapacidad visual. Dado que ésta no era una prueba médica ni una evaluación de nivel de audición sino un test para evaluar la satisfacción de usuario, se consideró que el parámetro de la discapacidad visual no era fundamental para la obtención de resultados válidos. Los resultados obtenidos por los dos grupos se analizaron por separado y se compararon para determinar si existían diferencias significativas entre ambos, antes de hacer extensivos al colectivo de discapacitados visuales los resultados obtenidos por el grupo de personas sin problemas de visión. Por tanto, los participantes del test se dividieron en dos grupos: un grupo formado por 4 usuarios con discapacidad visual y grupo más amplio de 22 personas sin problemas de visión. Todos ellos tenían el español como lengua materna

La ausencia de problemas de audición se determinó a través de una prueba auditiva que se realizó a todos los participantes (discapacitados y no discapacitados visuales) antes de realizar el test. La prueba consistía en la reproducción de una serie de tonos puros a 80 Hz, 1kHz, 5kHz, 10 kHz, 13 kHz y 15 kHz, en intervalos de 4 segundos de tono seguidos de 2 de silencio. Los participantes levantaban la mano tras cada uno de los tonos que escuchaban. En estas condiciones, y sin que este sea un dato con validez médica, una incapacidad para escuchar los tonos a 1 kHz, 5 kHz o 10 kHz, denotaría una deficiencia auditiva clara.

Tanto la prueba de audición como el test de AD/ruido se llevaron a cabo a un volumen de reproducción de 50 dB medidos con un sonómetro, con los participantes sentados frente a los altavoces a una distancia aproximada de 1 metro. De acuerdo con

la Organización mundial de la salud (OMS, 1999) 50 dB es el volumen de una conversación normal, en un ambiente tranquilo, medida a un metro de distancia. El equipo de reproducción utilizado fue un ordenador portátil conectado altavoces.

El grupo de usuarios con discapacidad visual estaba formado por 2 hombres y 2 mujeres con edades comprendidas entre los 52 y 75 años, sin problemas de audición (Tabla 5.1.). El grupo de grupo de usuarios sin discapacidad visual estaba formado por 22 personas, 8 hombres y 16 mujeres de edades comprendidas entre los 26 y los 76 años, sin problemas de audición. Todos tenían el español como lengua materna. Ninguno los participantes incluidos en el este estudio presentó deficiencia auditiva de acuerdo a la prueba realizada, anteriormente descrita.

Las características del grupo de discapacitados visuales y de participantes sin discapacidad visual fueron similares, tal y como indican los P valores de la tabla 5 (P>0.05).

Características	Discapacitados visuales	No discapacitados visuales	Valor de P
Edad (años)	Mediana= 57 Min= 53 Max= 75	Mediana= 50 Min= 18 Max= 76	0.38
Género (%)	Hombres= 50% Mujeres= 50%	Hombres=36% Mujeres=64%	0.52
Prueba de audición satisfactoria (%)*	100%	100%	0.99

* Se definió test satisfactorio cuando no había incapacidad para escuchar los tonos a 1kHz, 5kHz o 10kHz.

Tabla 5.1. Características de los participantes.

5.3.2. Creación de estímulos: ruido

No hay en AD estudios con usuarios previos en los que basarse para realizar el test, por lo que no se disponía de un modelo de ruido “estándar” para utilizar como

fuente de interferencia con la voz. El uso de los ruidos rosa o blanco que comúnmente se usan para distintas mediciones acústicas y calibración de equipos de sonido se descartó, ya que son ruidos generados por ordenador con unas características muy concretas que no se dan en el mundo natural. Para este estudio se deseaba crear un entorno realista con ruidos lo más parecidos posible a los que se dan en una BSO de cine o TV y, a ser posible, extraídos de una BSO real.

Dado que la película “*Malditos bastardos*” (Tarantino, 2009) tiene una banda sonora particularmente rica, y que ya se había utilizado para evaluar elementos sonoros en el experimento anterior, se decidió buscar los ruidos enmascarantes entre las tres escenas utilizadas para evaluar las mezclas de sonido. La reutilización de estímulos ofrecía la ventaja que los resultados obtenidos podrían ser complementarios a los del experimento anterior. Cada una de las escenas utilizadas presentaba unas características sonoras distintas: la escena 1 acaba con un ruido de tiroteo característico de las películas/escenas de acción o guerra; la escena 2, presenta una banda sonora musical con predominancia de instrumentos de viento mezclada con ruido ambiente y griterío también representativa de escenas de lucha/pánico/persecución; la banda sonora de la escena 3 es muy llamativa pero menos común, ya que consiste casi exclusivamente en una melodía de guitarra eléctrica con un marcado sonido metálico. Se decidió utilizar los ruidos de las escenas 1 y 2 por considerarse más representativos de una escena de acción estándar. Se seleccionaron sólo unos segundos de ruido de cada una de las escenas, para mantener la homogeneidad del impulso. Los fragmentos de ruido seleccionados muestran un espectro de frecuencias más amplio que el de la voz humana que cubre entre los 100 y los 9.000 Hz aproximadamente (López, 1999), y un rango dinámico reducido, con ausencia de picos bajos.

Para el ruido 1 se escogieron 2,3 segundos (minuto 1:41-1:43 de la escena 1) correspondientes a la secuencia del tiroteo y cuyo espectro de frecuencias se muestra en la Figura 5.1.

La Figura 5.1 muestra el rango de frecuencias cubierta por el ruido 1, expresado en Hz en el eje horizontal y su intensidad expresada en dB en el eje vertical. Como se puede observar, los mayores picos se dan en la franja de los 100 y 400-500 Hz, y toda la franja correspondiente a las frecuencias predominantes del habla está cubierta

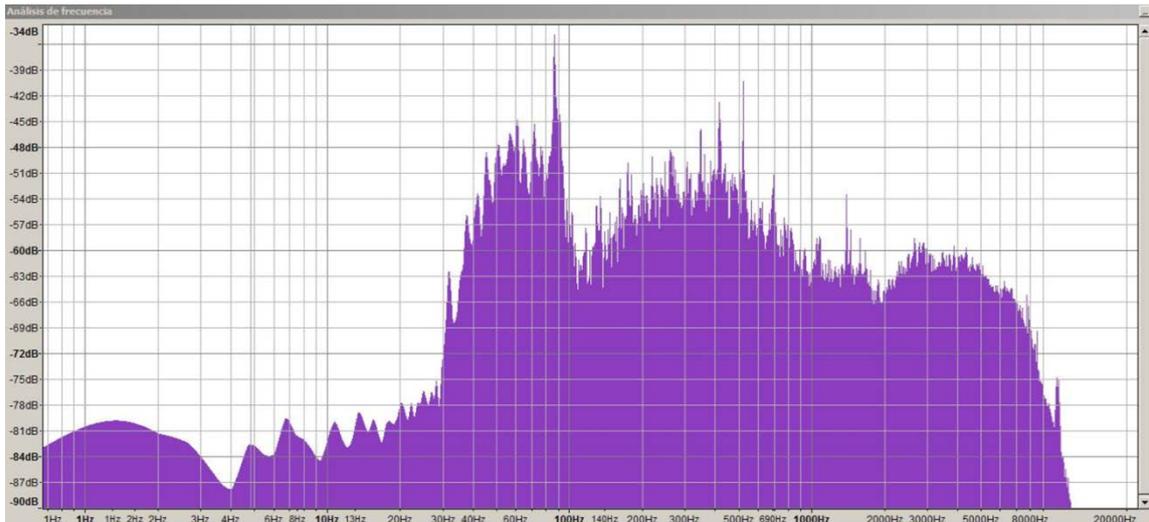


Figura 5.1. Espectro de frecuencias del ruido 1.

Para el ruido 2 se escogieron y cortaron 4,6 segundos (minuto 1:20 - 1:25 de la escena 2) con predominancia de banda musical mezclada con gritos y aplausos. El espectro de frecuencias se muestra en la figura 5.2

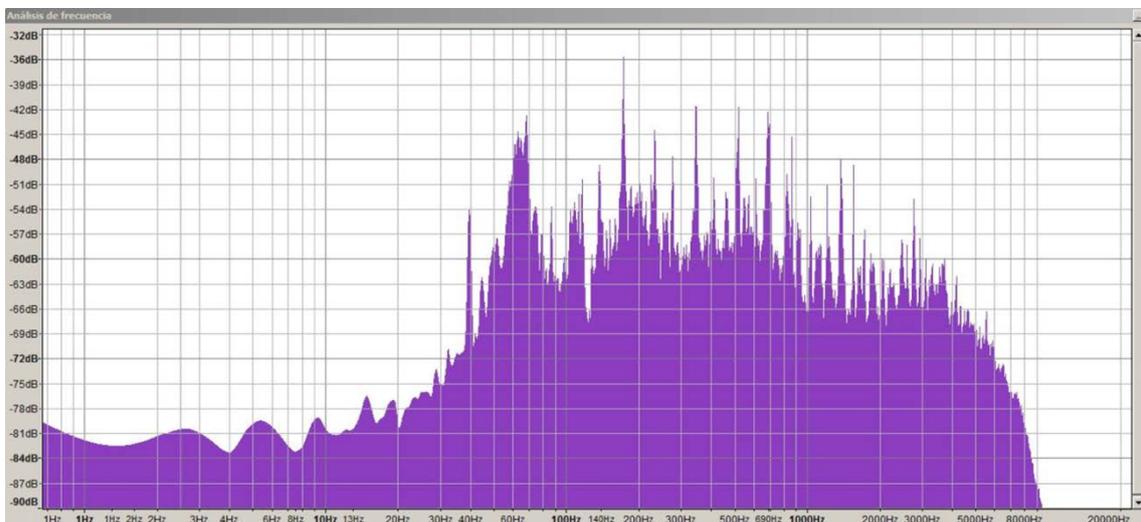


Figura 5.2. Espectro de frecuencias del ruido 2.

La figura 5.2 muestra el rango de frecuencias cubierta por el ruido 2 expresado en Hz en el eje horizontal, y su intensidad, expresada en dB en el eje vertical. Como se

puede observar, los mayores picos de intensidad se dan en la franja de los 60 y 200 Hz, aunque hay picos altos intermitentes en toda la franja entre los 140 y los 3000 Hz.

Con ayuda del programa de edición de audio Cool Edit Pro 2.1., cada una de las muestras de ruido se pegó en bucle hasta alargar su duración a 2 minutos y se normalizó a -23 dB. Una vez obtenidas las dos muestras, se hicieron los ajustes de volumen que se describen el diseño de experimento en el punto 5.4 de este capítulo.

5.3.3. Creación de estímulos: voz

Para la voz también se recurrió a la grabación realizada para el experimento previo. Como se necesitaba una narración constante los silencios durante la narración se redujeron a 1,5 segundos como máximo. No se disponía de un valor de referencia para aplicar en cuanto a la duración de las pausas durante la narración pero se consideró que acortando a menos de 1,5 segundos se producía un efecto poco natural en la narración y las pausas naturales del discurso, correspondientes a la puntuación de un texto, quedaban distorsionadas. Tras el acortamiento de silencios se obtuvo una pista de audio de 5:23 minutos de narración continua que se normalizó a un volumen de -23 RMS. El espectro de frecuencia de la voz se muestra en la Figura 5.3.

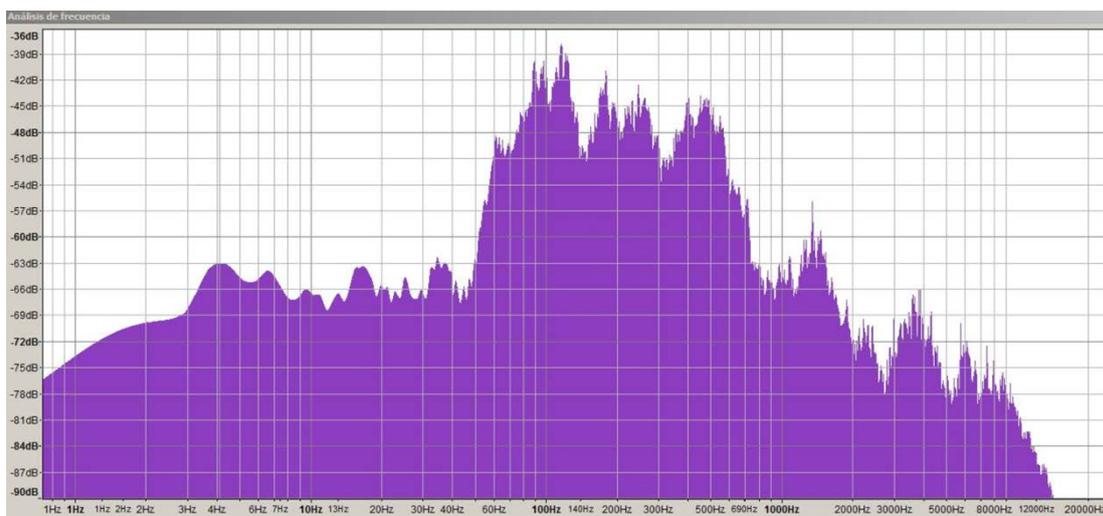


Figura 5.3. Espectro de frecuencias de la voz del locutor.

La Figura 5.3. muestra el rango de frecuencias cubierto por la voz expresado en Hz en el eje horizontal y su intensidad expresada en dB en el eje vertical. Como puede observarse, los picos de intensidad se dan de los 100 a 200 Hz, y de nuevo en torno a los 500 Hz.

5.3.4. Diseño del experimento

En este experimento se decidió mantener la voz un volumen constante de -23 dB RMS y ajustar el volumen del ruido de fondo para que fuera ascendiendo progresivamente. Se hizo de esta manera porque se consideró que era lo que reflejaba con más realismo las circunstancias reales de una AD, en las que la banda sonora de fondo a veces es silenciosa y a veces ruidosa. Esta narración se mezcló con dos ruidos de fondo que ascendían progresivamente de volumen y se pidió a los participantes que evaluaran cuándo el ruido les resultaba molesto para seguir la narración, y cuando el ruido les hacía imposible seguir la narración.

De acuerdo con los datos obtenidos en el experimento previo una diferencia de 5dB a favor de la AD garantizaría una escucha clara de la narración para la mayoría de los usuarios aun con ruido de fondo, mientras que a igualdad de volumen entre la pista de AD y el audio general, un 28% tendría dificultades para escuchar la AD con claridad. Teniendo estos datos en cuenta, la franja de peligro en la que los usuarios empiezan a tener dificultad para entender la AD empezaría cuando la diferencia de volumen entre la voz y el ruido es menor de 5dB. En un ejercicio de escucha sencillo, se aprecia que los problemas se producen cuando los niveles de volumen de audio entre ambas pistas son parecidos y que con la BSO 5dB más alta que la voz, comprender la narración es muy difícil. Para tener un margen de seguridad en ambos extremos se estableció el volumen de partida del ruido en -33 dB RMS (10 dB por debajo de la voz) y el volumen máximo en -12 dB (9 dB más alto que la voz).

Se diseñaron dos modelos de aumento de volumen de ruido: modelo A y modelos B. Para ambos los escalones de volumen fueron iguales; Los primeros tres aumentos de volumen son de 3 dB cada uno y a partir del tercero, que está a -24 dB (un dB más bajo que la voz), se aumenta dB a dB, para afinar la medición. Modelo A: en este modelo el ruido aumentaba de manera progresiva y continua desde -33 dBs hasta -

12 dBs, a razón de 3 dB cada 4 segundos para los primeros 3 tramos (hasta los -24 dB), y 1 dB cada cuatro segundos para el resto. La duración, el volumen y la diferencia de volumen con la voz de cada uno de los tramos se muestra en la tabla 5.2.

Escalones	Intervalo de tiempo (s)	Volumen del ruido en dB RMS	Diferencia de volumen entre ruido y voz en dB RMS
1	0:00.000- 0:04.000	-33	-10
2	0:04.100- 0:08.000	-30	-7
3	0:08.100- 0:12.000	-27	-4
4	0:12.100- 0:16.000	-24	-1
5	0:20.100- 0:24.000	-23	0
6	0:24.100- 0:28.000	-22	1
7	0:28.100- 0:32.000	-21	2
8	0:32.100- 0:36.000	-20	3
9	0:36.100- 0:40.000	-19	4
10	0:40.100- 0:44.000	-18	5
11	0:48.100- 0:52.000	-17	6
12	0:56.100- 1:00.000	-16	7
13	1:04.100- 1:08.000	-15	8
14	1:08.100- 1:12.000	-14	9
15	1:14.100- 1:16.000	-13	10
16	1:16.100- 1:20.000	-12	11

Tabla 5.2. Correspondencia entre tiempo y aumento de volumen del ruido en el modelo A.

Modelo B: los escalones de volumen del ruido en el modelo 2 son los mismos, pero este modelo intercala un segundo de descanso de ruido entre escalón y escalón. La narración se mantiene constante de forma que durante los descansos de ruido el usuario escucha la narración con un fondo de silencio. La duración, el volumen y la diferencia de volumen con la voz de cada uno de los tramos se muestra en la tabla 5.3.

Escalones	Intervalo tiempo (s)	Volumen del ruido en dB RMS	Diferencia de volumen entre voz y ruido en dB RMS
1	0:00.000- 0:06.000	-33	-10
2	0:07.000- 0:15.000	-30	-7
3	0:16.000- 0:22.000	-27	-4
4	0:23.000- 0:29.000	-24	-1
5	0:30.000- 0:36.000	-23	0
6	0:37.000- 0:43.000	-22	1
7	0:44.000- 0:50.000	-21	2
8	0:51.000- 0:57.000	-20	3
9	0:58.000- 1:04.000	-19	4
10	1:05.000- 1:11.000	-18	5
11	1:13.000- 1:20.000	-17	6
12	1:21.000- 1:27.000	-16	7
13	1:28.000- 1:34.000	-15	8
14	1:35.000- 1:41.000	-14	9
15	1:42.000- 1:48.000	-13	10
16	1:49.000- 1:55.000	-12	11

Tabla 5.3. Correspondencia entre tiempo y aumento de volumen del ruido en el modelo B.

Para el modelo B la duración de los tramos de volumen se alargó a 6 segundos, frente a los 4 del modelo A, para compensar el posible efecto residual de compresión de discurso acarreado del tramo de descanso de ruido.

Una vez obtenidas las muestras de ruido, de 1:20 y 1:55 minutos para el modelo A y B respectivamente, se insertó la voz del locutor a un volumen constante de -23 dB.

Estos dos modelos se probaron con 10 personas con edades comprendidas entre los 37 y los 76 años. Siete de estas personas refirieron que el modelo B, con los descansos de ruido intercalados, les resultaba menos confuso a la hora de identificar las sensaciones sonoras de entender o no la voz sobre el ruido. Refirieron también que los tramos de silencio ayudaban a reevaluar el nivel de ruido con más claridad que el

modelo que no tenía intervalos de descanso. Por tanto el modelo B se seleccionó para la realización del estudio, de acuerdo con el esquema representado en la figura 5.4.

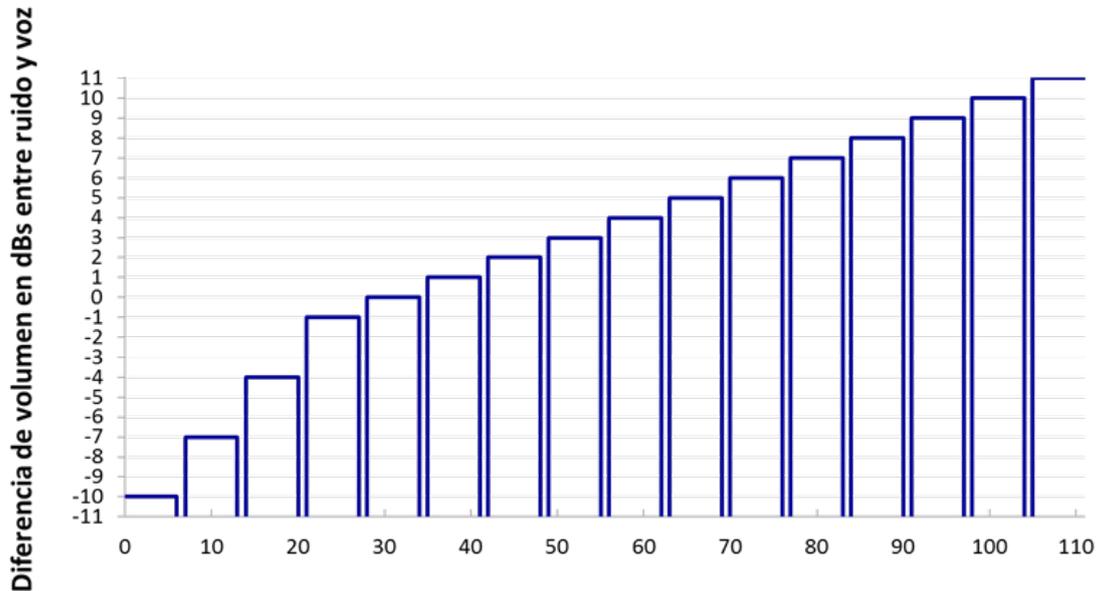


Figura 5.4. Esquema del experimento realizado para determinar el volumen de AD más adecuado.

La figura 5.4 muestra la duración e intensidad de volumen de los tramos de ruido en el modelo B, utilizado en el estudio. El volumen de la AD se mantuvo constante a -23 dB RMS (línea horizontal en la figura) y el ruido de fondo aumentó de volumen de -33 a -12 dB (10 dBs por más bajo y 11 más alto que la voz). Cada escalón de volumen del ruido de fondo se mantuvo constante durante 6 segundos seguidos de 1 segundo de silencio. El eje vertical muestra la diferencia de volumen entre el ruido y la voz.

5.3.5. Desarrollo del experimento y cuestionario asociado

El test se realizó mediante entrevista individual y toma de datos manual y se solicitó a participantes dos datos diferentes, uno, el momento en que el ruido comenzaba a hacerles incómodo seguir la narración de AD y dos, el momento en que el ruido les hacía imposible seguir la narración de AD. Antes de la prueba, se explicó a los participantes el funcionamiento del test y los dos datos que se les solicitaba.

Para la realización del test los participantes iniciaban y paraban la reproducción de los audios ellos mismos presionando la barra espaciadora del teclado del ordenador portátil que se usaba como reproductor. De esta manera se evitaba la posible desviación de tiempos derivada del tiempo de reacción del entrevistador para apuntar el dato. Cuando el participante paraba la reproducción, el dato en minutos y segundos se apuntaba manualmente, luego, con un segundo toque de la barra espaciadora el participante retomaba la reproducción donde se había quedado hasta que paraba por segunda vez y el entrevistador apuntaba el segundo dato. Los datos se introducían en una tabla Excel, que después se cotejaba con los datos de tiempo y decibelios (Tabla 5.4).

Intervalo de tiempo (s)	Volumen del ruido en dB RMS	Diferencia de volumen entre voz y ruido en dB RMS
0:00.000- 0:06.000	-33	-10
0:07.000- 0:15.000	-30	-7
0:16.000- 0:22.000	-27	-4
0:23.000- 0:29.000	-24	-1
0:30.000- 0:36.000	-23	0
0:37.000- 0:43.000	-22	1
0:44.000- 0:50.000	-21	2
0:51.000- 0:57.000	-20	3
0:58.000- 1:04.000	-19	4
1:05.000- 1:11.000	-18	5
1:13.000- 1:20.000	-17	6
1:21.000- 1:27.000	-16	7
1:28.000- 1:34.000	-15	8
1:35.000- 1:41.000	-14	9
1:42.000- 1:48.000	-13	10
1:49.000- 1:55.000	-12	11

Tabla 5.5. Tabla de cotejo.

5.3.6. Análisis estadístico

El análisis de los datos se llevó a cabo utilizando el programa estadístico GraphPad. Para la comparación entre los resultados obtenidos entre el grupo de discapacitados visuales y usuarios sin discapacidad visual se utilizó la prueba U de

Mann-Whitney. Se consideró que un P valor inferior a 0.05 indicaba una diferencia estadísticamente significativa.

parar el ruido 1 y 4 dB para el ruido 2. Pero como se puede apreciar en la Figura 5, el resto de participantes se encontraban incómodos en estos valores. La mediana en el grupo sin problemas de visión se situó en -1 dB, tanto para el ruido 1 como para el ruido 2. Se compararon los resultados obtenidos entre el grupo con y sin discapacidad visual con respecto a los ruidos 1 y 2. En ninguno de los casos se obtuvieron diferencias estadísticamente significativas con un valor de P superior al 0.05 (Tabla 5.6.), sugiriendo que los resultados del grupo sin discapacidad visual pueden ser extensibles a la población con discapacidad visual.

Diferencia de volumen en dB RMS	Discapacitados visuales (n=4)	No-discapacitados visuales (n=22)	Valor de P
Ruido 1 incómodo			
Mediana	0	-1	0.66
Mínimo	-4	-7	
Máximo	0	5	
Ruido 2 incómodo			
Mediana	-1	-1	0.32
Mínimo	-1	-7	
Máximo	1	4	

Tabla 5.6. Volumen del ruido de fondo que los participantes estimaron incómodo para seguir la AD.

En segundo lugar se estudió la diferencia de volumen que hacía imposible seguir la narración. Al igual que en el apartado anterior, cuando se determinó el volumen del ruido que hacía imposible seguir la AD, los resultados para los ruidos 1 y 2 fueron parecidos, aunque de nuevo se apreció una mayor tolerancia al ruido 1. En el grupo de usuarios con discapacidad visual, cuando el volumen del ruido 1 se situaba 2 dB por encima de la AD éste era intolerable para la mayor parte (tres) de los participantes (Figura 5.6).

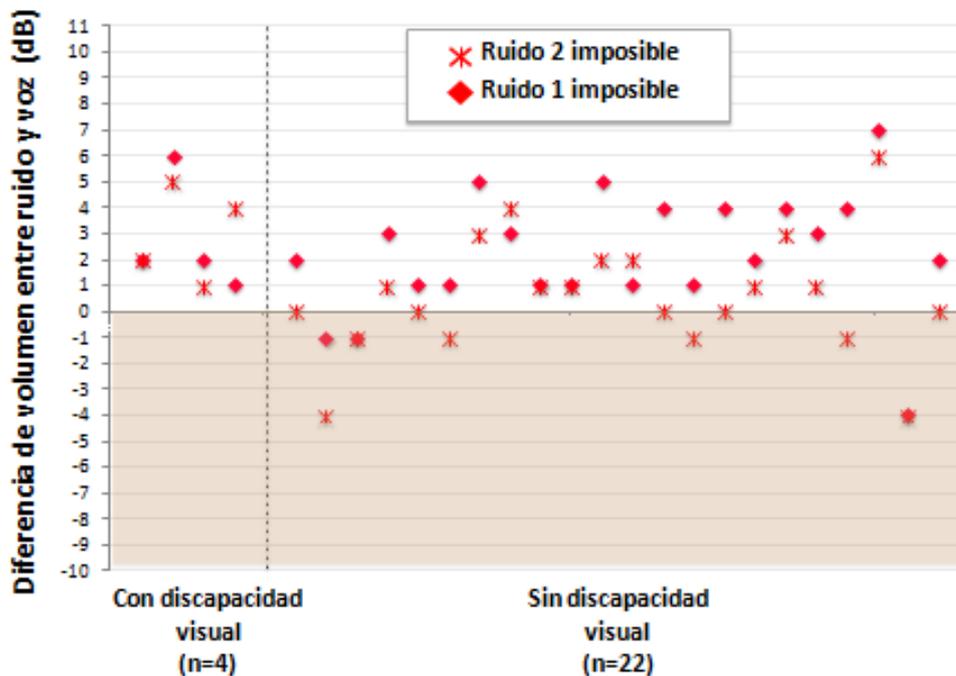


Figura 5.6. Volumen del ruido de fondo al que los usuarios se sienten incapaces de seguir la narrativa de la AD.

Para el ruido 2 en esta situación 2 participantes lo consideraban intolerable. El máximo de ruido tolerable fue de 6 dB de ruido por encima de la voz, diferencia a partir de la cual la narración no era inteligible para ninguno de los participantes. La mediana para los volúmenes de los ruidos 1 y 2, se situaron en 2 y 3 dB más altos que la voz, respectivamente (Tabla 5.7). En el grupo de personas sin discapacidad visual la mediana de diferencia de volumen entre ruido y AD para el ruido 1 fue de 2 dB a favor del ruido y de 0,5 dB para el ruido 2, también a favor del ruido. El participante que primero dejó de ser capaz de entender la voz frente al ruido lo hizo cuando el ruido estaba 4 dB más bajo que la voz, tanto en el ruido 1 como en el 2. A partir de aquí, el número de participantes que no eran capaces de entender la narración fue ascendiendo.

Diferencia de volumen en dB RMS	Discapitados visuales (n=4)	No-d discapitados visuales (n=22)	Valor de P
Ruido 1 imposible			
Mediana	2	2	0.78
Mínimo	1	-4	
Máximo	6	7	
Ruido 2 imposible			
Mediana	3	0,5	0.044
Mínimo	1	-4	
Máximo	5	6	

Tabla 5.7. Volumen del ruido de fondo que los participantes las hacía imposible seguir la AD..

A igualdad de volumen entre ruido y voz, 2 participantes no entendían la narración con el ruido 1 y 10 no entendían la narración con el ruido 2. Con el ruido 1, cuando el volumen se situó 6 dB más alto que la voz, ningún participante era capaz de entender la narración de AD. Con el ruido 2, cuando el volumen se situó 7 dB más alto que la voz, ningún participante era capaz de entender la narración. No se encontraron diferencias significativas entre los resultados del grupo de discapitados y no discapitados visuales para el ruido 1 ($p=0.78$) pero sí para el ruido 2 ($p=0.044$).

5.5. Discusión

Las conclusiones de este capítulo se harán generales abarcando el grupo de discapacitados y no discapacitados visuales dado que no se encontraron diferencias significativas entre el comportamiento de ambos grupos para la gran mayoría de los parámetros estudiados. Sólo en el caso del ruido 2 para el test de comprensión de la AD se apreciaron diferencias. En este caso el grupo de usuarios con discapacidad visual pareció tener mayor umbral de tolerancia que el grupo sin discapacidad visual para seguir la narración frente a ruido de banda musical. Esto podría deberse a una mejor capacidad auditiva de los discapacitados visuales o a una mayor costumbre de escuchar voz sobre un fondo musical.

Los resultados obtenidos en este experimento prueban que cuando una narración y un ruido de fondo están al mismo volumen el riesgo de que parte de los usuarios no pueda entender la narración es alto. Estos datos refrendan los obtenidos en el experimento de evaluación de mezclas de sonido que se presenta en el capítulo 4, en los que, cuando la pista de AD y el volumen de audio general estaban al mismo volumen un 28% de los participantes tenía problemas para entender la narración de AD con claridad.

A pesar de la dificultad de establecer medidas absolutas cuando se trata de sonido, los resultados obtenidos en este capítulo indican que lo aconsejable sería mantener una diferencia de volumen de 5 dB a favor de la AD para garantizar que los comentarios se entienden la narración con claridad incluso en escenas de ruido. Este es un dato que debería tenerse en cuenta para la emisión de AD ya que el gran problema del efecto de enmascaramiento sobre la voz es que aunque el efecto sea sólo parcial, la inteligibilidad se ve muy afectada. Esto ocurre porque basta con que la percepción de los sonidos consonánticos de la voz se vea afectada, para que la narración deje de ser comprensible, ya que son las consonantes las que más estrechamente con la inteligibilidad del habla (Carrión1998). La disminución de la inteligibilidad de la voz en presencia de ruidos de fondo es uno de los mayores riesgos en la edición de sonido de cine y televisión. Cuando una mezcla de audio no se hace correctamente, y el potencial de enmascaramiento de los ruidos ambiente o música de fondo sobre la voz no se tienen en cuenta el resultado es un diálogo o voz en off que se oye pero no se entiende.

En los resultados obtenidos en este experimento la diferencia en la fuente de ruido que se usara como interferencia no tiene un gran impacto sobre la diferencia de volumen que debería mantener la voz para entenderse con claridad aunque la tolerancia a la música parece ser algo peor que al ruido, situándose el umbral de tolerancia máxima 1 dB por debajo del de ruido de explosiones y disparos.

Con respecto a los datos obtenidos cuando la voz y el ruido están al mismo volumen, la mayoría de participantes encontró incómodo (entre un 88% y un 92%, dependiendo del tipo de ruido) o imposible (entre un 11% y un 42%, dependiendo del tipo de ruido) seguir la narración de AD. Este dato implicaría que si la AD se emite al mismo volumen que el audio general el servicio de accesibilidad estaría siendo infectivo para al menos un 11% de los usuarios en las escenas ruidosas.

Con la AD 5 dB más alta que el ruido, el 88% de los usuarios seguía con comodidad la narración sobre el ruido de banda musical, el 96% sobre el ruido de disparos y a ningún usuario le resultaba imposible seguirla. De acuerdo con estos datos, la diferencia de volumen mínima que habría que mantener para asegurar que el servicio de AD funcionara de forma óptima sería de 5 dB a favor de la voz sobre el ruido. Tomando como estándar de emisión los -23 dB RMS, esto supondría normalizar la AD a -19 RMS.

5.6. Conclusiones

Aunque establecer un estándar de calidad de AD es muy complejo, establecer un estándar de volumen que garantice la efectividad del servicio, sí es posible. De acuerdo con los datos obtenidos, ese volumen mínimo sería de 5 dB de diferencia entre voz y ruido de fondo. Esta diferencia garantizaría que los comentarios de AD se oyeran frente al posible ruido de fondo y además proporcionaría además una escucha relajada de los comentarios para la mayoría de usuarios, incluso en escenas de ruido intenso. Por lo tanto, y aunque la variabilidad entre usuarios es alta, una diferencia de 5 a 6 dB a favor de la banda de AD frente al audio general, sería un marco de seguridad razonable siempre que con ello no se incurriese en riesgo de saturación de volumen. En opinión de esta tesis, un estándar de volumen de AD pasaría por involucrar la actual regulación de sonido de TV para permitir aumentos de volumen de más de 6 dB sobre el volumen general, sin entrar en riesgo de saturación.

CAPÍTULO 6. CONCLUSIONES Y FUTURA INVESTIGACIÓN

6.1. Introducción

6.2. Mirando hacia el futuro

6.1. Introducción

La AD es un servicio de accesibilidad joven que aún carece de un estándar aplicable a nivel global para la producción y retransmisión televisiva a gran escala. Uno de los impedimentos para alcanzar este estándar es el número de sectores diferentes que toman parte en el proceso de creación de AD, y la falta de comunicación entre ellos. Esta investigación ha sido testigo de la distancia existente entre industria, ámbito investigador académico, emisoras de TV y usuarios. Esta lejanía entre los diferentes sectores en parte se debe a que la naturaleza de la AD, como servicio, no se enfoca de manera global. Los esfuerzos realizados por la investigación académica, en ocasiones quedan reflejados sólo en papel, sin que la industria llegue a aplicarlos de forma sistemática, mientras que las necesidades de estandarización de las empresas proveedoras no se tienen en cuenta en ámbitos más teóricos. Los proveedores de AD son empresas que lidian con la entrega y satisfacción de cliente y a falta de estándar claro para AD, aplican como criterio los requerimientos de cliente. Las emisoras de TV buscan en primera instancia cumplir con la legislación de accesibilidad vigente y destinan a ello los recursos justos y necesarios, pero no más. El usuario, por su parte, depende de los fabricantes de monitores de TV y la publicitación del servicio de AD por las emisoras para saber cómo, cuándo y con qué programas pueden recibir AD en casa. Muchos de los participantes con discapacidad visual que tomaron parte en los estudios que se presentan confesaron no ser conscientes de que el servicio de AD en TV formaba parte de la parrilla habitual de contenidos. Esto no es extraño porque, de hecho, no hay información en la página general de programación que informe al respecto¹⁸. Estos usuarios conocían el servicio a través de iniciativas específicas promovidas por organizaciones como la ONCE o gracias a APPs¹⁹ que ofrecen los audios de AD para descargar y sincronizar mediante un dispositivo portátil, en el cine o en casa, con el audio general de las películas. Todas estas iniciativas son muy valiosas y permiten que la AD esté disponible para el usuario de manera más o menos sencilla, sin embargo, lo excluyen del panorama general que es la programación televisiva. Esta tesis defiende la

¹⁸ TVE Programación (2017) Recuperado de <http://www.rtve.es/television/programacion/> [Fecha de consulta septiembre 2017]

¹⁹ *AudescMobile*. <https://itunes.apple.com/es/app/audescmobile/id705749595?mt=8> [Fecha de consulta septiembre 2017]

necesidad de establecer un estándar industrial y de producción si lo que se pretende es que la AD pase a ser parte integral de los servicios televisivos. En este proceso la empresa tiene que implicarse pero a menudo carece de la capacidad, el tiempo, o la preparación necesarios para llevar a cabo iniciativas de investigación complejas y costosas en tiempo y recursos. Este doctorado es una excepción y, desde sus limitaciones, ha enfocado la investigación sobre el aspecto más visible de la emisión de AD, que el volumen. La normalización de volumen es un paso necesario de la edición y tratamiento de audio y un estándar claro a este respecto facilitaría la producción a gran escala y garantizaría a los usuarios una homogeneidad del servicio.

6.2. Mirando hacia el futuro

La mezcla de sonido en AD es un aspecto que influye de forma decisiva en cómo le va a “sonar” el producto final al usuario y la propuesta de esta tesis es sólo un primer paso hacia un estándar completo. Aunque el dato propuesto, si se aplicase, podría solucionar el problema de la dificultad para escuchar una narración en una escena de ruido, también presenta riesgos y carencias que deben ser estudiados. ITU (2013) ya advierte claramente que la pista de AD no debe sobrepasar el volumen del audio general en más de 6dB para evitar saturación del audio. La diferencia propuesta de 5 dB se acerca peligrosamente a ese límite y debería estudiarse si una aplicación sistemática de esta medida causa o no problemas durante la emisión. Los resultados del experimento que se presentan en el capítulo 4 sugieren que aumentar el volumen de la AD por encima de la banda sonora no es igual de efectivo para mejorar la escucha de la narración frente a un fondo de ruido que bajar la BSO. Aunque ajustar la pista de AD mejora el problema, los resultados en cuanto a comprensión son peores que los obtenidos bajando la BSO. Además, las respuestas obtenidas a la pregunta 8 del cuestionario del cuestionario utilizado en el experimento (Anexo 1) muestran que los usuarios perciben esos ajustes de volumen y haría falta hacer más estudios para comprobar si esas variaciones son o no de su gusto.

Desde un punto de vista de calidad, tal y como apunta M. van der Heijden (2010), una mezcla de AD específica hecha en estudio y con todas las pistas de audio ajustadas ex profeso para la versión audio descrita del programa o la película sería la mejor opción, pero en el panorama televisivo esta es una posibilidad poco realista. En general, los radiodifusores trabajan con productos ya mezclados y además ese tipo de mezcla supondría un gasto de recursos y tiempo que, a menos que fuera requisito legal, es poco probable que quisieran asumir. Además, el panorama audiovisual seguirá cambiando y protocolos como la Hbb TV ya apuntan hacia que la personalización de los contenidos y la programación a la carta son el futuro. Un panorama en el que televisores híbridos permitan manipular las opciones de audio para ajustarlas al gusto del consumidor no es para nada improbable. Ya hay televisores que incluyen posibilidades de ajuste de audio por canales y la suma o no de pistas adicionales como la de la AD.

Cabe preguntarse si en un panorama así, en el que el grado de personalización sea tan alto, merece la pena el esfuerzo de investigar qué parámetros de volumen de AD son los más efectivos. La opinión de esta tesis es que, precisamente porque cada vez una mayor parte de las mezclas audiovisuales van a recaer en el receptor, la edición del audio de AD como elemento aislado antes de su emisión se presenta como un campo de estudio con muchas posibilidades. Ajustes como la ecualización selectiva o el ajuste dinámico de volumen no deberían recaer en el usuario, ya que aunque es positivo que el usuario puede personalizar opciones según sus necesidades, no sería justo que tuviera que estar haciéndolo constantemente. En el caso concreto de la AD, la posibilidad de ajustar el volumen de manera independiente no debería suponer para el usuario la necesidad de estar haciéndolo constantemente. Esta sería la situación si, por ejemplo, se fijara un volumen de emisión de AD 5 dB por encima del audio general de manera constante. Esto solucionaría los problemas de escucha de los comentarios en escenas ruidosas, pero probablemente supondría tener que bajar el volumen de la pista de AD con respecto al audio general en las escenas más silenciosas para evitar la sensación de que los comentarios están a un volumen atronador. De momento no hay estudios centrados en qué parámetros de la edición de audio de la pista de AD son los más importantes para garantizar la calidad de audio del producto final. El volumen es un elemento muy importante, pero es sólo la punta del iceberg. Aspectos como la ecualización o la adición de efectos para integrar mejor el audio de AD en el ambiente general de la escena son terrenos muy interesantes que ofrecen muchas posibilidades de investigación.

BIBLIOGRAFÍA

- Benecke, B. (2004). Audio-description. *Meta: journal des traducteur/Meta: Translators' Journal*, 49(1), 78-80.
- Benecke, B. (2007). Audio description: phenomena of information sequencing. En Gerhard Budin (ed.), *Proceedings of the Marie Curie Euroconferences MuTra: Audiovisual Translation Scenarios*. Conferencia celebrada en Viena del 4 de abril al 30 de mayo de 2007.
- Benecke, B. y Dosch, E. (2004). *Wenn aus Bildern Worte werden*. Múnich: Bayerischer Rundfunk.
- Berglund, B., Lindvall, T. y Schwela, D. H. (1999). Guidelines for community noise. En B. Berglund, T. Lindvall y D. H. Schwela (eds.), *Guidelines for community noise*. Ginebra: Organización Mundial de la Salud. Recuperado de <http://www.bvsde.paho.org/bvsci/i/fulltext/noise/noise.pdf> [fecha de consulta: julio 2017].
- Bourne, J. y Jiménez Hurtado, C. (2007). From the visual to the verbal in two languages: a contrastive analysis of the audio description of *The Hours* in English and Spanish. En J. Díaz-Cintas, P. Orero y A. Remael (eds), *Media for all. Subtitling for the Deaf, Audio description and Sign Language* (pp. 175-188). Ámsterdam: Rodopi.
- Braun, S. (2007). Audio Description from a discourse perspective: a socially relevant framework for research and training. *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, 6, 357-369.
- Broadcasting Authority of Ireland (2016). *Access Rules*. Recuperado de www.bai.ie/en/media/sites/2/2016/08/20160106_BAI_AccessRules2016_vFinal.pdf [fecha de consulta: octubre 2016].
- Cabeza-Cáceres, C. (2013). *Audiodescripció i recepció. Efecte de la velocitat de narració, l'entonació i l'explicitació en la comprensió fílmica* (tesis doctoral). Universitat Autònoma de Barcelona. Recuperada de www.tdx.cat/bitstream/handle/10803/113556/cc1de1.pdf?sequence=1 [fecha de

consulta: julio 2017].

Carrión, A. (1998). *Diseño acústico de espacios arquitectónicos*. Barcelona: Edicions UPC.

Chalmers, R. (1997). The Broadcast Wave Format. *EBU Technical Review*, 3-6. https://tech.ebu.ch/publications/trev_274-chalmers [fecha de consulta: junio 2017]

Chion, M. (1993). *La audiovisión. Introducción a un análisis de la imagen y el sonido*. Barcelona: Paidós.

Côté, N. (2011). Speech Quality Measurement Methods. En *Integral and Diagnostic Intrusive Prediction of Speech Quality* (pp. 37-85). Berlin y Heidelberg: Springer.

Davis, G. y Davis, G. D. (1989). *The sound reinforcement handbook* (2.^a ed.). Milwaukee: Hal Leonard Corporation.

EBU (2014). *R128. Loudness normalisation and permitted maximum level of audio signals*. Ginebra. Recuperado de <https://tech.ebu.ch/docs/r/r128.pdf> [fecha de consulta: septiembre 2017].

Fernández-Torné, A. y Matamala, A. (2012). Technology for accessibility in multilingual settings. En *The translation and reception of multilingual films*. Conferencia celebrada en Université Paul-Valéry (Montpellier) del 15 al 16 de junio de 2012.

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

Franklin, C. A., Thelin, J. W., Nabelek, A. K. y Burchfield, S. B. (2006). The effect of speech presentation level on acceptance of background noise in listeners with normal hearing. *Journal of the American Academy of Audiology*, 17(2), 141-146.

Fresno Cañada, N. (2014). *La (re) construcción de los personajes fílmicos en la audiodescripción Efectos de la cantidad de información y de su segmentación en el recuerdo de los receptores* (tesis doctoral). Universitat Autònoma de

- Barcelona. Recuperada de https://ddd.uab.cat/pub/tesis/2014/hdl_10803_285420/nfc1de1.pdf [fecha de consulta: junio 2017].
- Fryer, L. (2010). Audio description as audio drama—a practitioner’s point of view. *Perspectives: Studies in translatology*, 18(3), 205-213.
- Gambier, Y. (2004). *Traduction audiovisuelle*. Montreal: Presses de l’Université de Montréal.
- Gambier, Y. (2006). Multimodality and audiovisual translation. En *MuTra 2006-Audiovisual Scenarios: Conference Proceedings* (pp. 1-8). Conferencia celebrada del 1 al 5 de mayo de 2006 en Copenhague.
- Gambier, Y. (2008). Recent developments and challenges in Audiovisual Translation. En Chiaro, D., Heiss, C., & Bucaria, C. (Eds) *International Conference Between Text and Image: Updating Research in Screen Translation Vol 78* (pp. 27-29). Amsterdam: John Benjamins
- Hurtado, C. J. y Seibel, C. (2010). Traducción accesible: narratología y semántica de la audiodescripción. En L. González y P. Hernández (coords.), *IV Congreso El español, lengua de traducción para la cooperación y el diálogo* (pp. 451-468). Comunicación llevada a cabo en el Congreso celebrado del 8 al 10 de mayo de 2008 en Toledo. Madrid: Esletra.
- Igareda, P. y Matamala, A. (2012). Variations on the Pear Tree experiment: different variables, new results? *Perspectives*, 20(1), 103-123.
- Independent Television Commission. (2000). *ITC guidance on standards for audio description*. Recuperado de msradio.huji.ac.il/narration.doc [fecha de consulta: junio 2017].
- ISO (1998). *ISO/IEC 13818-3:1998(en) Information technology. Generic coding of moving pictures and associated audio information. Part 3: Audio*. Recuperado de www.iso.org/obp/ui/#iso:std:iso-iec:13818:-3:ed-2:v1:en [fecha de consulta: julio 2017].
- International Telecommunications Union (2013). *FG AVA TR. Version 1.0. Technical*

Report Part 13. Audio characteristics for audio descriptions and/or spoken subtitles. Recuperado de www.itu.int/dms_pub/itu-t/opb/fg/T-FG-AVA-2013-P13-PDF-E.pdf [fecha de consulta: junio 2017].

International Telecommunications Union (2013). *FG AVA Technical Report Part 5: Final report of activities: Working Group B «Audio/Video description and spoken captions».* Recuperado de www.itu.int/dms_pub/itu-t/opb/fg/T-FG-AVA-2013-P5-PDF-E.pdf [fecha de consulta: junio 2017].

Jackson-Menaldi, M. C. A. (1992). *La voz normal.* Buenos Aires: Médica Panamericana.

Kruger, J. L. (2012). Making meaning in AVT: Eye tracking and viewer construction of narrative. *Perspectives*, 20(1), 67-86.

Kruger, J. L., & Orero, P. (2010). Audio description, audio narration—a new era in AVT. *Perspectives: Studies in Translatology*, 18(3), 141-142.

Lehman, A., O'Rourke, N., Hatcher, L. y Stepanski, E. (2005). *JMP for basic univariate and multivariate statistics.* Cary: SAS Institute.

Maczyńska, M. y Szarkowska, A. (2011). Text-to-Speech Audio Description with Audio Subtitling to a Non-Fiction Film «La Soufriere» by Werner Herzog. En *The 4th Media For All Conference.* Conferencia celebrada en Londres en 2011.

Maszerowska, A. (2014). *The verbalizations of the effects of light and contrast in audio description* (tesis doctoral). Universidad Autónoma de Barcelona. Recuperada de <http://www.tdx.cat/handle/10803/285576> [fecha de consulta: septiembre 2017].

Maszerowska, A., Matamala, A. y Orero, P. (eds.). (2014). *Audio description: New perspectives illustrated* (vol. 112). Amsterdam: John Benjamins.

Matamala, A. (2005). Live audio description in Catalonia. *Translating Today*, 4, 9-11.

Matamala, A. y Orero, P. (2014). Standardising Audio Description. *Italian Journal of Special Education for Inclusion*, 1(1), 149-155.

Miyara, F. (2004). El ruido y la inteligibilidad de la palabra. En la conferencia organizada por ASOLOFAL el 9 de agosto de 2004. Recuperado de

<https://www.fceia.unr.edu.ar/acustica/biblio/inteligibilidad.pdf> [fecha de consulta: 14 de septiembre de 2017].

- Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2), 617-644.
- Orero, P. (2005a). Audio description: Professional recognition, practice and standards in Spain. *Translation Watch Quarterly*, 1(1), 7-18.
- Orero, P. (2005b). La inclusión de la accesibilidad en comunicación audiovisual dentro de los estudios de traducción audiovisual. *Quaderns: revista de traducció*, 12, 173-185.
- Orero, P., Casacuberta, D. y Maszerowska, A. Audio describing Silence: Lost for Words. En A. Jankowska y A. Szarkowska (eds.), *New Points of View in Audiovisual Translation* (pp. 219-236). Berna: Peter Lang.
- Piety, P. (2004). The language system of audio description: an investigation as a discursive process. *Journal of Visual Impairment & Blindness (JVIB)*, 98(08). Recuperado de <http://files.eric.ed.gov/fulltext/EJ683817.pdf> [fecha de consulta: mayo 2017].
- Rai, S. (2009). *Bollywood for all: The demand for audio described Bollywood films*. Londres: Royal National Institute of Blind People.
- Recuero López, M. (1999). *Acústica arquitectónica aplicada*. Madrid: Paraninfo.
- Recuero López, M. (2000). *Ingeniería acústica*. Madrid: Paraninfo.
- Recuero López, M. y Gil González, C. (1993). *Acústica arquitectónica*. Madrid: Izquierdo.
- Remael, A. (2012). For the use of sound. Film sound analysis for audio-description: Some key issues. *MonTI. Monografías de Traducción e Interpretación* (4) 255-276.
- Rowe, B. M. y Levine, D. P. (2016). *A concise introduction to linguistics* (4.^a ed.). Nueva York: Routledge.
- Salway, A. (2007). A corpus-based analysis of audio description. En J. Díaz Cintas, P.

- Orero y A. Remael (Eds), *Media for all: Subtitling for the deaf, audio description and sign language* (pp. 151-175). Amsterdam: Rodopi.
- Schulze, M. (2010). *Understanding the UN Convention on the Rights of Persons with Disabilities* (3.^a ed.). Nueva York: Handicap International.
- Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), 3-55. Recuperado de <http://lanethames.com/dataStore/ECE/InfoTheory/shannon.pdf> [fecha de consulta: julio 2017].
- Snyder, J. (2005). Audio description: The visual made verbal. *International Congress Series*, 1282, 935-939.
- Szarkowska, A. (2011). Text-to-speech audio description: towards wider availability of AD. *Journal of Specialised Translation*, 15(1), 81-98.
- Szarkowska, A. y Jankowska, A. (2012). *Text-to-speech audio description of voiced-over films. A case study of audio described Volver in Polish*. Recuperado de www.openstarts.units.it/bitstream/10077/6362/1/Szarkowska_Jankowska_EmergingTopics.pdf [fecha de consulta: abril 2017].
- Szarkowska, A. y Orero, P. (2014). The importance of sound for audio description. En A. Szerowska, A. Matamala y P. Orero (eds.), *Audio Description: New perspectives illustrated* (pp. 121-140). Amsterdam: John Benjamins.
- Van der Heijden, M. (2007). Making film and television accessible to the blind and visually impaired (tesina de máster). Dirigida por E. Grimm, Utrecht School of the Arts. Recuperada de <https://pdfs.semanticscholar.org/6b7c/507262678e9d01083b8fe09897c2f3081ed7.pdf> [fecha de consulta: noviembre 2016].
- Vera, J. F. (2006). Translating audio description scripts: the way forward? Tentative first stage project results. En *MuTra 2006-Audiovisual Scenarios: Conference Proceedings* (pp. 148-181). Conferencia celebrada del 1 al 5 de mayo de 2006 en Copenhagen. Recuperado de http://translationconcepts.org/pdf/MuTra_2006_Proceedings.pdf#page=152

[fecha de consulta: noviembre 2016].

- Vercauteren, G. (2012). A narratological approach to content selection in audio description. Towards a strategy for the description of narratological time. *MonTI. Monografías de Traducción e Interpretación*, 4, 207-231.
- Vilaró, A., Duchowski, A. T., Orero, P., Grindinger, T., Tetreault, S. y Di Giovanni, E. (2012). How sound is the Pear Tree Story? Testing the effect of varying audio stimuli on visual attention distribution. *Perspectives*, 20(1), 55-65.
- Voss, P., Tabry, V. y Zatorre, R. J. (2015). Trade-off in the sound localization abilities of early blind individuals between the horizontal and vertical planes. *Journal of Neuroscience*, 35(15), 6051-6056.
- Walczak, A. y Fryer, L. (2017). Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment*, 35(1), 6-17.
- Walczak, A. y Fryer, L. (2017). Vocal delivery of audio description by genre: measuring user presence. *Perspectives*, *Published online: 13 Jul 2017* 1-15.
Recuperado
<http://www.tandfonline.com/eprint/rd6TRUUP9qGrpHZQNeb5/full> [fecha de consulta septiembre 2017]
- Walczak, A. y Szarkowska, A. (2012). Text-to-speech audio description of educational materials for visually impaired children. En S. Bruti y E. di Giovanni (Eds), *Audio Visual Translation across Europe: An Ever-Changing Landscape* (pp. 209-234). Berna; Berlín: Peter Lang.
- Watkinson, J. (2001). *The art of digital audio*. Nueva York; Londres: Taylor & Francis.
- Watkinson, J. (2002). *An introduction to digital audio* (2.^a ed.). Nueva York: Taylor & Francis.
- Watkinson, J. (2003). *Introducción al audio digital*. Waltham: Focal Press.
- World Wide Web Consortium (2016). G56: Mixing audio files so that non-speech sounds are at least 20 decibels lower than the speech audio content. *Techniques for WCAG 2.0. Techniques and Failures for Web Content Accessibility*

Guidelines 2.0. Recuperado de <https://www.w3.org/TR/WCAG20-TECHS/G56.html> [fecha consulta septiembre 2017]

ANEXOS

Anexo 1. Tabla de distribución de escenas y mezclas de sonido.

	ESCENA	MEZCLA	ESCENA	MEZCLA	ESCENA	MEZCLA
INDIVIDUO 1	1	1	2	3	3	2
INDIVIDUO 2	2	1	1	2	3	3
INDIVIDUO 3	3	1	2	2	1	3
INDIVIDUO 4	2	3	3	2	1	1
INDIVIDUO 5	3	3	2	1	1	2
INDIVIDUO 6	1	3	3	1	2	2
INDIVIDUO 7	3	2	1	1	2	3
INDIVIDUO 8	1	2	3	3	2	1
INDIVIDUO 9	2	2	1	3	3	1
INDIVIDUO 10	1	2	3	1	2	3
INDIVIDUO 11	1	1	2	3	3	2
INDIVIDUO 12	2	1	1	2	3	3
INDIVIDUO 13	3	1	2	2	1	3
INDIVIDUO 14	2	3	1	2	1	1
INDIVIDUO 15	3	3	2	1	1	2
INDIVIDUO 16	1	3	3	1	2	2
INDIVIDUO 17	3	2	1	1	2	3
INDIVIDUO 18	1	2	3	3	2	1
INDIVIDUO 19	2	2	1	3	3	1
INDIVIDUO 20	1	2	3	3	2	1
INDIVIDUO 21	1	1	2	3	3	2
INDIVIDUO 22	2	1	1	2	3	3
INDIVIDUO 23	3	1	2	2	1	3
INDIVIDUO 24	2	3	3	2	1	1
INDIVIDUO 25	3	3	2	1	1	2
INDIVIDUO 26	1	3	3	1	2	2
INDIVIDUO 27	3	2	1	1	2	3
INDIVIDUO 28	1	2	3	3	2	1
INDIVIDUO 29	2	2	1	3	3	1
INDIVIDUO 30	1	2	3	1	2	1
INDIVIDUO 31	1	1	2	3	3	2
INDIVIDUO 32	2	1	1	2	3	3

Anexo 2. Cuestionario del experimento de mezclas de sonido.

Valore del 1 al 3, siendo 1 en total desacuerdo y 3 totalmente de acuerdo, las siguientes afirmaciones.			
	Escena 1	Escena 2	Escena 3
1. En general, la locución de AD era clara y comprensible.			
2. He escuchado con claridad todos los sonidos ambiente en todo momento			
3. El volumen de la narración me parece adecuado en todo momento a lo largo de la escena.			
4. En algunos momentos de la escena el volumen de la AD me parece excesivamente alto.			
5. En algunos momentos de la escena el volumen de AD me parece demasiado bajo.			
6. En algunos momentos, la AD dificulta la escucha de los sonidos ambiente de la escena.			
7. En algunos momentos, la banda sonora dificulta la escucha clara de la AD			
8. He notado cambios de volúmen en la narración de AD a lo largo de la escena.			
9. En general me parece que el volumen de la AD y de la escena original están a un nivel similar.			