

Tesis Doctoral

---

# Nuevas métricas de evaluación de la calidad acústica de salas. La incorporación de los modelos perceptivos.

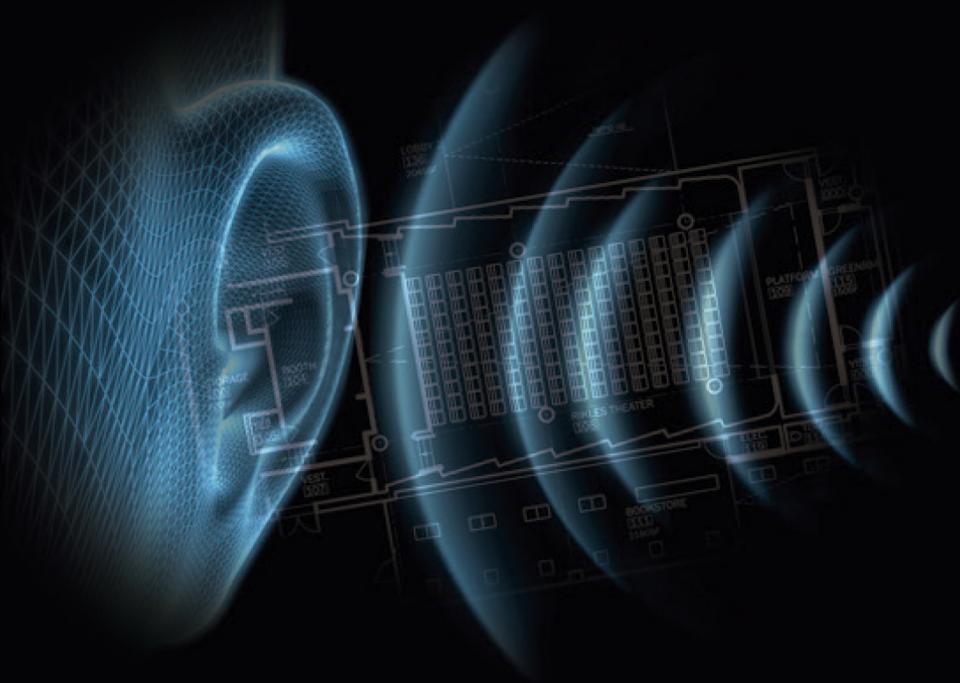
Joan Cruaños Català

---

Directores:

Dr. Jesús Alba Fernández

Dra. Romina del Rey Tormos



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

Enero 2016



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

Nuevas métricas de evaluación de la calidad  
acústica de salas. La incorporación de los modelos  
perceptivos.

Joan Cruaños Català

---

Tesis Doctoral

Enero 2016

Directores:

Dr. Jesús Alba Fernández

Dra. Romina Del Rey Tormos



# Agraïments

Gràcies a Ismael, per esperar-me molts dies abans d'anar-se'n a dormir per dir-me bona nit. També a Irene, que està de camí.

Gràcies a Susana, pel temps i per tindre cura d'ell.

Gràcies a M<sup>a</sup> Carmen i a Santi, per facilitar-me la feina i per fer-me sentir a casa.

Gràcies a Celia Solano i a Javier Redondo, pel seu temps i per les seues veus.

Gràcies a Fernando Hernández i a Laura Bertó, l'un per facilitar-me l'ús de materials i espais, l'altra per les seues lliçons exprés dels equips i pel seu suport.

Gràcies a Carles Madramany, per fer-me més fàcil el treball en alguns espais.

Gràcies a Jaime Ramis, per la seua ajuda i suport incondicionals.

Gràcies a Romina del Rey, per la seua ajuda, per les seues gestions i per les seues paraules terapèutiques.

Gràcies a Jesús Alba, pel seu suport, els seus ànims i per encoratjar-me a finalitzar aquest projecte.

Gràcies a tots aquells que en alguna mesura han fet més fàcil dur a terme aquest treball.



# Resumen

El presente trabajo de Tesis se ubica dentro del programa de Doctorado de Promoción del Conocimiento siendo actualmente la entidad responsable el Departamento de Física Aplicada de la Universidad Politécnica de Valencia.

El análisis cualitativo de sistemas ha cobrado relevancia en los últimos años, incorporando nuevas perspectivas de aproximación a los problemas de evaluación, que incluyen en muchas ocasiones el punto de vista del usuario. Tradicionalmente, en el ámbito de la acústica y el sonido, la evaluación cualitativa de eventos sonoros ha sido tarea encomendada a las técnicas psicoacústicas clásicas. El análisis estadístico de los resultados arrojados por estas técnicas ha ofrecido una respuesta estándar o de referencia a un determinado problema.

Por otra parte, también en el ámbito acústico, el estudio de sistemas mediante el establecimiento de modelos físicos orientados a la descripción de la fenomenología que tiene lugar en éstos, ha sido fuente de una gran cantidad de métricas. Estos modelos y técnicas de medición, históricamente han sido originados por las demandas de unas industrias específicas y sus procedimientos se han ido adaptando posteriormente a diferentes ámbitos de aplicación, en ocasiones, muy diferentes del original para el que fueron concebidos. Las magnitudes ofrecidas por estas métricas, además de cuantificar una determinada variable del modelo, se han relacionado frecuentemente con aspectos cualitativos de los propios sistemas. El establecimiento de una correlación entre los resultados ofrecidos por estas métricas y las valoraciones cualitativas realizadas mediante experimentación psicoacústica no siempre ha sido todo lo satisfactoria que cabría esperar, dando lugar en ocasiones a una dicotomía entre lo objetivo y lo subjetivo, referente a la descripción fenomenológica y a la cualitativa respectivamente.

En las últimas décadas, mucha de la información generada mediante experimentación psicoacústica ha sido utilizada, junto con el desarrollo de técnicas de procesado, con la finalidad de establecer modelados del sistema auditivo humano, dando lugar a los conocidos como modelos perceptuales. Los principales demandantes de técnicas de evaluación cualitativa desde el punto de vista perceptivo han sido las industrias de la telefonía móvil en un primer momento y posteriormente las de difusión de contenidos principalmente audiovisuales a través de los nuevos canales de transmisión. La necesidad de reducir la tasa de transmisión binaria con el menor perjuicio cualitativo posible condujo en primer lugar al desarrollo de *códecs* de compresión de la información basados en cualidades perceptivas – aspecto éste en constante evolución - y posteriormente al establecimiento de técnicas de medición de la degradación de la señal y de monitorización de servicios basadas también en modelos perceptivos. De esta manera,

se produce un cambio del foco de atención sobre el objeto de estudio: el modelado del canal se sustituye por el modelado perceptivo, con la finalidad de establecer una evaluación, en principio, con mayor consideración de los aspectos relevantes en el proceso de percepción – de los eventos sonoros en nuestro caso – y, consecuentemente, con la valoración cualitativa del sistema.

Veremos que históricamente, el intercambio de técnicas entre diferentes ámbitos de aplicación ha sido un denominador común de los modelos, métodos y técnicas de evaluación de la calidad. Con ello, este trabajo pretende evaluar la viabilidad de un modelo perceptual en un ámbito de aplicación distinto del inicialmente previsto para estos modelos. Generalmente, los modelos perceptuales se aplican a señales que han pasado por un canal de transmisión digital, con la finalidad de establecer una comparación entre la señal original o de referencia y la señal transmitida. Las diferencias entre ambas señales son evaluadas por el modelo perceptual, cuantificándose esta evaluación en la salida del modelo. En el caso de este trabajo intentaremos extender el ámbito de aplicación de un modelo perceptual a sistemas continuos, como son un transductor electroacústico o un recinto en el que se propaga un evento sonoro. Tanto un transductor como una sala pueden considerarse como canales de transmisión, si bien las degradaciones de la señal introducidas por éstos serán de naturaleza bien diferenciada y de hecho, su estudio se ha abordado desde perspectivas muy diferentes a lo largo del tiempo. El objetivo de este trabajo es indagar en qué medida puede responder un modelo perceptual a la degradación de señal introducida por canales de transmisión tan diferenciados a los inicialmente abordados con técnicas perceptivas.

El contenido de esta memoria se centrará en primer lugar, en establecer una aproximación al concepto de calidad en el ámbito de la acústica, aspecto que se desarrollará en el capítulo 2. Los capítulos 3 y 4 están dedicados a las métricas clásicas de evaluación de la calidad en los canales de transmisión propuestos, los dispositivos de transducción y los recintos para audición respectivamente. El capítulo 5 contiene los conceptos relativos al sistema auditivo humano sobre los que se fundamenta la métrica perceptual implementada. Los detalles del desarrollo de la métrica se explican en el capítulo 6. Los capítulos 7 y 8 describen el experimento de implementación de la métrica perceptual desarrollada sobre canales de transmisión reales, el primero relativo a dispositivos de transducción y el segundo respecto de recintos, realizándose un análisis de los resultados obtenidos en ambos casos. El último apartado está dedicado al establecimiento de unas conclusiones generales sobre los resultados obtenidos así como a futuras líneas de investigación y mejora sobre la métrica implementada.

# Resum

Aquest treball de Tesi s'ubica dins el programa de Doctorat de Promoció del Coneixement, del qual és actualment l'entitat responsable el Departament de Física Aplicada de la Universitat Politècnica de València.

L'anàlisi qualitativa de sistemes ha augmentat la seua rellevància en els darrers anys amb la incorporació de noves perspectives d'aproximació als problemes d'avaluació, que en moltes ocasions inclouen el punt de vista de l'usuari. Tradicionalment, en l'àmbit de l'acústica i el so, l'avaluació qualitativa d'esdeveniments sonors ha estat una tasca encomanada a les tècniques psicoacústiques clàssiques.

D'altra banda, també en l'àmbit de l'acústica, l'estudi de sistemes mitjançant models físics, establerts amb la finalitat de descriure la fenomenologia que hi ocorre, ha estat una font inesgotable de mètriques. Aquests models i tècniques de mesura han estat, històricament, originats per les demandes d'unes indústries específiques i els seus procediments han anat adaptant-se posteriorment a diferents àmbits d'aplicació, moltes vegades diferents d'aquells per als quals van ser dissenyats originàriament. Les magnituds resultants d'aquestes mètriques, a més de quantificar una determinada variable del model, s'han relacionat sovint amb aspectes qualitatius dels propis sistemes. L'establiment d'una correlació entre els resultats oferts per aquestes mètriques i les valoracions qualitatives realitzades mitjançant experimentació psicoacústica no sempre han resultat ser tan satisfactòries com cabria esperar, de forma que de vegades es produeix una dicotomia entre mètriques objectives i subjectives, on les primeres fan referència a la descripció fenomenològica mentre que les segones a la descripció qualitativa.

En les darreres dècades, molta de la informació generada a partir de l'experimentació psicoacústica ha estat emprada, junt amb el desenvolupament de tècniques de processat del senyal, amb la finalitat d'establir models del sistema auditiu humà, els anomenats models perceptius. Els principals demandants de tècniques d'avaluació qualitativa des del punt de vista perceptiu han estat les indústries de la telefonia mòbil en un primer moment i posteriorment les de difusió de continguts majoritàriament audiovisuals mitjançant els nous canals de transmissió. La necessitat de reduir l'ample de banda de transmissió amb el menor perjudici de qualitat possible dugué en primer lloc al desenvolupament de còdecs de compressió de la informació fonamentats en aspectes perceptius. Aquestes tècniques de compressió continuen en constant evolució però una altra derivada d'aquest enfocament des del punt de vista perceptiu n'és l'establiment de tècniques de mesura de la degradació del senyal i de monitorització de serveis també fonamentades en models perceptius. D'aquesta manera es produeix un canvi de perspectiva: el modelat del canal de transmissió es substitueix pel modelat perceptiu, amb la finalitat d'establir una avaluació que, en principi, estarà més directament

relacionada amb els aspectes més rellevants del procés de percepció – dels esdeveniments sonors en el nostre cas – i consegüentment, amb la valoració qualitativa del sistema.

Veurem que històricament, l'intercanvi de tècniques entre diferents àmbits d'aplicació ha estat una constant dels models, mètodes i tècniques d'avaluació de la qualitat. Amb això, aquest treball pretén avaluar l'aplicació d'un model perceptiu a un àmbit diferent d'aquells inicialment previstos per aquests models. Normalment, els models perceptius s'apliquen a senyals transmèses mitjançant un canal digital, amb la finalitat d'establir una comparació entre el senyal original o de referència i el senyal transmès. Les diferències entre ambdós senyals són avaluades pel model perceptiu, de tal manera que a l'eixida del model es quantifica aquesta avaluació. En el cas d'aquest treball hom intentarà ampliar l'àmbit d'aplicació d'un model perceptiu a sistemes continus, com ho són un transductor electroacústic o un recinte on s'hi propaga un esdeveniment acústic. Un dispositiu de transducció o una sala poden considerar-se canals de transmissió, encara que les degradacions del senyal introduïdes seran de naturalesa diferenciada. De fet, el seu estudi s'ha plantejat al llarg del temps des de perspectives ben diferents. L'objectiu d'aquest treball és indagar en quina mesura pot respondre un model perceptiu a la degradació del senyal introduïda per canals de transmissió tan diferents als inicialment estudiats amb tècniques perceptives.

El contingut d'aquesta memòria se centrarà en primer lloc, en establir una aproximació al concepte de qualitat en l'àmbit de l'acústica, tasca que es desenvoluparà al capítol 2. Els capítols 3 i 4 estan dedicats a les mètriques clàssiques d'avaluació de la qualitat als canals de transmissió proposats, els dispositius de transducció i els recintes d'audició respectivament. Al capítol 5 s'hi presenten els conceptes relatius al sistema auditiu humà sobre els quals es fonamenta la mètrica perceptiva implementada. Els detalls del desenvolupament de la mètrica s'expliquen al capítol 6. Els capítols 7 i 8 descriuen l'experiment d'implementació de la mètrica perceptiva desenvolupada sobre canals de transmissió reals, el primer dels capítols pel que fa als dispositius de transducció i el segon d'ells en relació als recintes. En ambdós casos es realitza una anàlisi dels resultats obtinguts. El darrer apartat està dedicat a l'establiment d'unes conclusions generals així com a futures línies d'investigació i millora sobre la mètrica implementada.

# Summary

The present dissertation is placed within the Doctoral Program for the Knowledge Promotion. This Program currently responsible entity is the Applied Physics Department of the Polytechnic University of Valencia.

Qualitative analysis of systems has gained importance in recent years, by incorporating new approaching perspectives to assessment problems, which often include the user's point of view. Traditionally, in the field of acoustics and sound, qualitative evaluation of sound events has been a task entrusted to the classical psychoacoustic techniques. Statistical analysis of the results produced by these techniques has offered a standard response or reference to a particular problem.

Moreover, also in the acoustic science, the study of systems by establishing physical models designed to describe the phenomenology that takes place in them, has been the source of a lot of metrics. Historically, these models and measurement techniques have been originated by the demands of specific industries and these procedures have been subsequently adapted to scopes of application sometimes very different from the original for which they were conceived. The magnitudes offered by these metrics, in addition to quantify a given variable of the model, are often related to qualitative aspects of the systems themselves. The establishment of a correlation between the results offered by these metrics and qualitative assessments made by psychoacoustics experimentation has not always been as satisfactory than could be expected, leading sometimes to a dichotomy between the objective and the subjective metrics, regarding phenomenological and qualitative description respectively.

In the last decades, the knowledge generated through psychoacoustic experimentation has been used, along with the improvement of signal processing techniques, in order to model the human auditory system. Such work has led to the development of perceptual models. The main applicant of these qualitative assessment techniques was, firstly, the mobile phone industry and subsequently those related to audio-visual content broadcasting through new digital transmission channels. The need to reduce the transmission bit rate with the least possible qualitative degradation, first led to the development of compression codecs, based on perceptual aspects. These compression techniques are constantly evolving. But another consequence of this perceptive approach is the establishment of signal degradation and service monitoring measurement techniques also based on perceptual models. Thus, a change of perspective is produced: modelling of transmission channel is replaced by perceptual modelling, in order to establish an assessment that, in principle, could be more directly related to the most relevant aspects of the perception process, and consequently, to the system's qualitative assessment.

We will see that historically, exchanges of techniques between different application areas has been a usual practice in the context of models and methods for quality assessing. With this, this study aims to assess the application of a perceptual model to a different scope of those initially foreseen for these models. Typically, perceptual models are applied to signals transmitted through a digital channel, in order to establish a comparison between the original signal -or the reference- and the transmitted one. The differences between the two signals are evaluated by the model, in order to obtain a quantification of the assessment. In the case of this work we try to expand the scope of a perceptual model to continuous systems, as are an electroacoustic transducer or a room where an acoustic event propagates.

Both, transducers and rooms, can be considered as transmission channels, although the signal degradations introduced by them will be well differentiated. In fact, their study has been approached from very different perspectives over time. The aim of this study is to investigate to what extent a perceptual model can be used to assess signal degradation introduced by transmission channels so different to those initially dealt with perceptual techniques.

The content of this dissertation will focus, firstly, on establishing an approach to the concept of quality in the field of acoustics, something that will be developed in chapter 2. Chapters 3 and 4 are dedicated to the classical quality assessment metrics in the proposed transmission channels, transduction devices in chapter 3, and rooms or halls in chapter 4. Chapter 5 contains the concepts relating the human auditory system on which the implemented perceptual metric is based. Details of the metric development are discussed in chapter 6. Chapters 7 and 8 describe the implementation experiment on real transmission channels, transduction devices in chapter 7, and rooms in chapter 8. An analysis of the obtained results is performed in both cases. The last section is devoted to establishing general conclusions on the results as well as future research and improvement over the implemented metric.

# Índice

<b>Agraïments</b>	<b>iii</b>
<b>Resumen</b>	<b>v</b>
<b>Resum</b>	<b>vii</b>
<b>Summary</b>	<b>ix</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Objeto	5
1.2 Estructura	8
<b>2 Calidad acústica</b>	<b>11</b>
2.1 El problema de la definición de calidad	11
2.2 Ámbitos de aplicación. La dificultad para una teoría general.	13
2.2.1 Medida subjetiva de la calidad sonora con encuestas psicoacústicas	13
2.2.2 Medida de la calidad de audio	14
2.2.3 Calidad del habla	18
2.2.4 Calidad sonora de salas, recintos y auditorios	23
2.2.5 Ruido y calidad	28
2.2.6 Calidad sonora de un producto	29

<b>2.3</b>	<b>Algunas consideraciones sobre la calidad</b>	<b>30</b>
<b>2.4</b>	<b>Psicoacústica</b>	<b>32</b>
2.4.1	Una visión histórica	33
2.4.2	Un modelo para describir la experimentación psicoacústica	36
<b>3</b>	<b>Calidad en transductores</b>	<b>39</b>
<b>3.1</b>	<b>El concepto de calidad en los transductores acústicos</b>	<b>39</b>
<b>3.2</b>	<b>Paradigmas de aproximación a la evaluación de la distorsión</b>	<b>41</b>
3.2.1	Medidas de distorsión	41
3.2.2	Métodos de identificación	42
3.2.3	Métodos perceptuales.	42
<b>3.3</b>	<b>Sistemas lineales</b>	<b>43</b>
<b>3.4</b>	<b>Sistemas no lineales</b>	<b>45</b>
<b>3.5</b>	<b>Ejemplos de distorsión lineal y no lineal</b>	<b>48</b>
3.5.1	Distorsión lineal	48
3.5.2	Distorsión no lineal asimétrica	49
3.5.3	Distorsión no lineal simétrica	51
3.5.4	Productos de intermodulación	52
<b>3.6</b>	<b>Evolución histórica de las medidas de distorsión</b>	<b>53</b>
3.6.1	Distorsión armónica total	53
3.6.2	Distorsión armónica total más ruido	55
3.6.3	Distorsión de intermodulación	55
3.6.4	Ponderación de armónicos	56
3.6.5	Excitación multitono	58
3.6.6	Umbrales de audibilidad. Detectabilidad a partir de modelos	60
3.6.7	Uso de la función de coherencia	60
3.6.8	La métrica Geddes-Lee	61
3.6.9	Los modelos perceptuales	66
<b>3.7</b>	<b>Sistemas de radiación indirecta</b>	<b>67</b>
3.7.1	El motor de compresión	68
3.7.2	La bocina	69
3.7.3	Fuentes de no linealidad en el motor de compresión	70
3.7.4	El tubo de onda plana	72
<b>3.8</b>	<b>Sistemas de radiación directa</b>	<b>73</b>
3.8.1	El motor	73
3.8.2	El diafragma	75
3.8.3	La suspensión	75

3.8.4	Fuentes de no linealidad en un transductor de radiación directa _____	76
<b>4</b>	<b><i>Calidad en salas</i></b> _____	<b>79</b>
<b>4.1</b>	<b>La acústica de salas como objeto de estudio</b> _____	<b>79</b>
<b>4.2</b>	<b>El problema de la correlación entre parámetros</b> _____	<b>82</b>
4.2.1	Una perspectiva global _____	83
4.2.2	Nuevas perspectivas de aproximación _____	85
<b>4.3</b>	<b>La respuesta al impulso</b> _____	<b>86</b>
<b>4.4</b>	<b>Parámetros temporales</b> _____	<b>90</b>
4.4.1	Tiempos de reverberación _____	90
4.4.2	Initial Time Delay Gap (ITDG) _____	94
<b>4.5</b>	<b>Parámetros energéticos</b> _____	<b>94</b>
4.5.1	Definición _____	94
4.5.2	Claridad $C_{80}$ _____	94
4.5.3	Claridad $C_{50}$ _____	95
4.5.4	Tiempo central $T_s$ _____	95
4.5.5	Fuerza sonora $G$ _____	96
4.5.6	Curvas de energía reflejada acumulada _____	96
<b>4.6</b>	<b>Parámetros espaciales</b> _____	<b>97</b>
4.6.1	Fracción de energía lateral _____	97
4.6.2	Eficiencia lateral _____	98
4.6.3	Nivel sonoro lateral final _____	98
4.6.4	Correlación cruzada interaural _____	99
<b>4.7</b>	<b>El caso de la evaluación de la palabra</b> _____	<b>100</b>
4.7.1	Parámetros temporales _____	101
4.7.2	Parámetros energéticos _____	101
4.7.3	Medidas de inteligibilidad _____	101
<b>5</b>	<b><i>Percepción auditiva</i></b> _____	<b>107</b>
<b>5.1</b>	<b>El sistema auditivo humano</b> _____	<b>107</b>
5.1.1	Oído externo _____	108
5.1.2	Oído medio _____	109
5.1.3	Oído interno _____	111
<b>5.2</b>	<b>Umbral de audición absolutos</b> _____	<b>118</b>
<b>5.3</b>	<b>Enmascaramiento y banda crítica</b> _____	<b>119</b>
<b>5.4</b>	<b>Estimaciones de la forma de los filtros auditivos</b> _____	<b>124</b>

<b>5.5</b>	<b>Modelado de los filtros auditivos</b>	<b>128</b>
5.5.1	Filtros roex	128
5.5.2	Filtros gammatono	130
<b>6</b>	<b><i>Un modelo perceptual</i></b>	<b>133</b>
<b>6.1</b>	<b>La métrica <math>R_{\text{nonlin}}</math></b>	<b>133</b>
<b>6.2</b>	<b>El modelo</b>	<b>134</b>
<b>6.3</b>	<b>Alineado de señales</b>	<b>137</b>
<b>6.4</b>	<b>Emulación del oído externo y medio</b>	<b>139</b>
<b>6.5</b>	<b>Simulación de los filtros auditivos. Frecuencias centrales, ERB y factor de solapamiento</b>	<b>139</b>
<b>6.6</b>	<b>Simulación de los filtros auditivos. Implementación de los filtros gammatono</b>	<b>143</b>
<b>6.7</b>	<b>Correlación cruzada normalizada</b>	<b>149</b>
<b>6.8</b>	<b>Matriz de ponderación</b>	<b>150</b>
<b>6.9</b>	<b>Predicción</b>	<b>152</b>
<b>6.10</b>	<b>Desarrollo del software</b>	<b>153</b>
<b>7</b>	<b><i>Evaluación en transductores</i></b>	<b>157</b>
<b>7.1</b>	<b>Evaluación en transductores de radiación indirecta</b>	<b>157</b>
7.1.1	Procedimiento	157
7.1.2	Resultados	163
7.1.3	Análisis	167
<b>7.2</b>	<b>Evaluación en transductores de radiación directa</b>	<b>171</b>
7.2.1	Procedimiento	171
7.2.2	Resultados	173
7.2.3	Análisis	178
<b>8</b>	<b><i>Evaluación en salas</i></b>	<b>183</b>
<b>8.1</b>	<b>Procedimiento</b>	<b>184</b>
<b>8.2</b>	<b>Señales utilizadas</b>	<b>185</b>
8.2.1	Señales musicales	185
8.2.2	Señales para palabra	188
<b>8.3</b>	<b>Elección del transductor</b>	<b>190</b>

<b>8.4</b>	<b>Salas</b>	<b>194</b>
8.4.1	Sala 1. Cámara Reverberante Campus de Gandia UPV	194
8.4.2	Sala 2. Aula Docente Campus de Gandia	196
8.4.3	Sala 3. Sala de actos Casa de la Cultura de L'Alqueria de la Comtessa	197
8.4.4	Sala 4. Aula Magna Campus de Gandia UPV	198
8.4.5	Sala 5. Sala de Conferencias Campus de Gandia	200
8.4.6	Sala 6. Locutorio Estudio de radio EPSG	201
8.4.7	Sala 7. Locutorio de doblaje EPSG	202
<b>8.5</b>	<b>Resultados</b>	<b>203</b>
8.5.1	Señales musicales	203
8.5.2	Señales de voz	225
<b>9</b>	<b>Conclusiones y futuras líneas de trabajo</b>	<b>241</b>
9.1	Implementación de la métrica	243
9.2	Resultados en transductores	243
9.3	Resultados en recintos	245
9.4	Futuras líneas de trabajo	248
<b>10</b>	<b>Referencias bibliográficas</b>	<b>251</b>
<b>Anexo I</b>		<b><i>i</i></b>
<b>Anexo II</b>		<b><i>xi</i></b>
<b>Anexo III</b>		<b><i>xvii</i></b>



*esse est percipi*

George Berkeley (1685-1753)



# 1 Introducción

A la hora de afrontar una definición del término calidad aplicado a un ámbito determinado, el primer obstáculo que surge es la dificultad para la delimitación y acotamiento del concepto, inicialmente disperso entre multitud de definiciones, técnicas de medición, ámbitos de aplicación y postulados teóricos. En lo referente a la acústica y el sonido, esta visión difusa del concepto ocurre de igual manera que en otros ámbitos, y puede deducirse, haciendo una revisión histórica, que el término se ha ido utilizando constantemente en ámbitos de aplicación muy específicos, dejando a un lado modelos o teorías generales sobre la calidad. Si bien es cierto que en la actualidad una teoría general bien establecida es la referente a la calidad del producto, esta unificación teórica es de reciente aplicación en el ámbito del sonido y la acústica y se debe fundamentalmente a los trabajos de Jekosh (Blauert & Jekosh, 1997), (Jekosh, 2004), (Jekosh, 2005), (Blauert & Jekosh, 2012) y en su definición de calidad sonora del producto – *Product Sound Quality* -. Este enfoque se centra principalmente en establecer la dimensión acústica de un producto como un elemento comunicativo más del mismo, de forma que el sonido del producto es fruto de un proceso de diseño de éste orientado a satisfacer unas necesidades

y a cubrir unas determinadas expectativas. En todo este proceso cobra de forma creciente cada vez más importancia la valoración subjetiva del usuario como un instrumento de evaluación de la calidad (Le Callet, Moller, & Perkins, 2012), en lo que se ha venido a definir como calidad de la experiencia del usuario.

Este marco teórico encaja perfectamente en el ámbito de aplicación de la industria de fabricación de productos de consumo, en los que la dimensión sonora de éstos se diseña y evalúa como un elemento semiótico que influye de forma decisiva en la comunicación entre el producto y el usuario, de manera que un diseño adecuado del sonido del producto será un elemento más del conjunto de características del producto que contribuyen a la percepción cualitativa de éste. Tenemos muchos ejemplos de aplicación de este modelo en la industria de la automoción, de la fabricación de electrodomésticos y otros dispositivos de uso común en la vida cotidiana. Sin embargo, fuera del ámbito de aplicación de la industria de productos de consumo, el marco de referencia del *Product Sound Quality*, aunque igualmente de aplicación directa al menos de forma conceptual, se encontrará fuertemente condicionado por la ingente cantidad de técnicas, métodos de evaluación y modelos que le han precedido en cada campo de aplicación. A nivel estrictamente formal, es cierto que desde el punto de vista acústico, cualquier elemento puede considerarse como producto: un instrumento musical, un transductor acústico, un teatro, una sala de audiciones, etc.

También es cierto que las cualidades sonoras de estos “productos” estarán directamente relacionadas con las valoraciones cualitativas que de éstas realicen sus usuarios. Pero a la hora de evaluar, como será el objeto principal de este trabajo, un transductor acústico o una sala para la audición musical o de palabra hay otros elementos a considerar. Desde el punto de vista de la calidad sonora del producto, el sonido es un atributo comunicativo de éste. La peculiaridad en el caso de los transductores acústicos así como también de las salas es que son, en realidad, canales de transmisión, con lo cual, no poseen un atributo sonoro intrínseco – como lo puede tener un vehículo o una aspiradora – sino que aportarán sus propiedades al mensaje o evento sonoro que se transmita a través de ellos.

Así pues, en el momento del diseño de una sala o un transductor no podemos pensar directamente en un sonido que le atribuya un valor semántico a éstos para cumplir unas determinadas expectativas sino más bien en un conjunto de cualidades que necesariamente modificarán los mensajes sonoros que en ellos tengan lugar dando lugar a diferentes percepciones cualitativas por parte de los usuarios. Es cierto que, desde el punto de vista del paradigma de *Product Sound Quality*, este conjunto de cualidades acústicas o sonoras otorgará un determinado significado al “producto” – canal de transmisión – que será decisivo en la valoración cualitativa que de él realicen los usuarios, pero también lo es que el resultado de la valoración cualitativa estará fuertemente condicionado por las características del evento o mensaje sonoro que tenga lugar en el canal de transmisión, así como por la idoneidad de éste – el mensaje sonoro – para ser transmitido por un determinado canal. A este aspecto diferencial hay que añadir todos los esfuerzos que históricamente se han hecho para realizar evaluaciones

cualitativas de canales de transmisión y que, lógicamente, no se realizaron bajo el paraguas de un marco conceptual predeterminado, sino más bien ajustando el término calidad a las condiciones específicas del problema bajo estudio.

A lo largo de esta memoria veremos que históricamente los problemas relativos a la evaluación de canales de transmisión se han abordado en función de las necesidades concretas de cada momento, con una fuerte dependencia de los recursos tecnológicos disponibles en cada época y teniendo casi exclusivamente como impulsoras a unas industrias que en su momento eran las demandantes de estas evaluaciones. En este último aspecto, veremos que la incipiente industria telefónica desde principios del siglo XX será una fuente inagotable de creación de técnicas de evaluación de canales de transmisión, así como de producción de conocimiento respecto de las características de los mensajes sonoros – principalmente la voz humana - (Fletcher, 1922). El relevo a la industria de la telefonía lo tomarán las radiocomunicaciones desde el primer tercio del siglo XX y la industria de la electrónica de consumo ya en la segunda mitad del siglo pasado. A finales de éste y durante el principio del siglo XXI, la industria demandante pasará a ser la de la telefonía móvil así como los sistemas de transmisión digital de contenidos.

Cada una de estas industrias ha aportado su propia perspectiva al problema, así como una serie de técnicas que después se han extrapolado con mayor o menor éxito a otros ámbitos de aplicación. Por ejemplo, los métodos incipientes de medida de distorsión en canales de transmisión telefónica fueron adoptados directamente para la evaluación de transductores acústicos – medida de distorsión armónica total, medidas de productos de intermodulación, etc... - convirtiéndose en un estándar *de facto* para la evaluación de estos dispositivos aun en plena vigencia (Hilliard, 1941), (Shorter, 1950). Técnicas pensadas a priori para la evaluación de la inteligibilidad de la palabra en canales eléctricos de transmisión (Steeneken & Houtgast, 1980) han sido adaptadas con éxito, siguiendo los mismos fundamentos, a la evaluación de la inteligibilidad en canales puramente acústicos (Houtgast & Steeneken, 1984). Algunos procedimientos concebidos a priori para el estudio de dispositivos electrónicos como por ejemplo es la caracterización de éstos por su respuesta al impulso han constituido una auténtica revolución en otros ámbitos como la acústica de salas, disciplina en la que actualmente se obtienen la mayor parte de parámetros de evaluación mediante técnicas de respuesta al impulso (Berkhout, de Vries, & Boone, 1980). Más recientemente, la necesidad surgida de mejorar las ratios de compresión de los sistemas de transmisión digitales conllevará a la búsqueda de soluciones desde el punto de vista perceptivo: la investigación psicoacústica ha proporcionado un conjunto de conocimientos mediante el cual poder reducir la información transmitida sin perjuicio de la calidad percibida. Todo este conocimiento cristalizó en un primer momento en el desarrollo de técnicas de compresión de datos, tal como demandaba la telefonía móvil o los sistemas de difusión de contenidos, pero posteriormente ha posibilitado el desarrollo de técnicas de modelado perceptual del sistema auditivo humano que son de aplicabilidad en otros ámbitos,

incluida la evaluación cualitativa de los propios sistemas de transmisión (Beaton, Beerends, Keyhl, & Treurniet, 1996).

Toda esta evolución de la evaluación de los canales de transmisión ha dado lugar al establecimiento de modelos teóricos para explicar el comportamiento de éstos, modelados como sistemas. En el caso de los recintos o salas, los modelos teóricos principalmente establecidos son la teoría estadística, la teoría geométrica y la teoría ondulatoria. En lo concerniente a los transductores, la electroacústica modela el comportamiento de estos sistemas mediante el establecimiento de circuitos equivalentes considerando el mayor número posible de variables que actúan en el sistema, pudiendo éste ser caracterizado como un sistema lineal o no lineal. El establecimiento de los modelos teóricos con un determinado nivel de abstracción tiene como finalidad explicar los fenómenos físicos que ocurren en el sistema y los resultados de estos modelos teóricos se denominan generalmente parámetros objetivos. Ocurre con cierta frecuencia que los resultados ofrecidos por un modelo teórico no ofrecen una correlación suficientemente satisfactoria con los resultados arrojados por evaluaciones cualitativas abordadas mediante procedimientos psicoacústicos, por ejemplo. Con lo cual, históricamente se ha establecido una diferenciación entre parámetros o medidas objetivas y subjetivas, siendo las primeras las obtenidas mediante la abstracción del modelo teórico y las segundas las conseguidas mediante procedimientos evaluativos de calidad directamente sobre el usuario, generalmente el oyente o espectador en nuestro caso. Así, se ha establecido una especie de dualismo entre lo objetivo y lo subjetivo que sienta sus bases en un modelo epistemológico cartesiano, como veremos en el siguiente capítulo.

Las explicaciones a las no satisfactorias correlaciones entre lo objetivo y lo subjetivo frecuentemente se abordan desde la explicación de las imperfecciones de nuestros sistemas perceptivos, pero esta perspectiva encuentra también sus detractores argumentando que, en esencia, cualquier modelo teórico está fundamentado también en la percepción de un fenómeno, difiriendo únicamente de una valoración cualitativa en el grado de abstracción (Dominicus, 2009), (Blauert & Dominicus, 2013). Sea como fuere, parece ser que en la valoración cualitativa influyen muchos más elementos de los tenidos en cuenta por un modelo con mayor o menor grado de abstracción, tales como factores ambientales y cognitivos del usuario. Por ejemplo, ningún modelo que explique el comportamiento del campo acústico dentro de un recinto considerará, evidentemente, elementos arquitectónicos desde una perspectiva puramente estética para describir el comportamiento de éste. Sin embargo, según el perfil de usuario que valore el recinto, estos elementos podrían influir en la valoración cualitativa de la acústica de la sala (Galiana, Llinares, & Llinares, 2014). En otras ocasiones ocurre que los descriptores cualitativos utilizados no están consensuados de forma homogénea por todos los individuos que son objeto de un experimento psicoacústico, dando lugar a errores conceptuales que se traducen en dispersión e indefinición de resultados (Lokki, 2014), una de las posibles causas de una deficiente correlación con las métricas denominadas objetivas. Estos factores, entre otros, pueden llevarnos a la conclusión, desde un punto

de vista estrictamente dualista, de que la evaluación cualitativa es inabordable con procedimientos científicos, opinión que tradicionalmente ha sido predominante en el ámbito científico-tecnológico. Sin embargo, contra este postulado se podría argumentar que todas las métricas, técnicas y procedimientos considerados a priori objetivos tienen como principal finalidad la evaluación del sistema bajo estudio y su clasificación, en referencia a otros sistemas, en un rango cualitativo, que, en última instancia, sólo tiene sentido si hay un observador o usuario del evento o del sistema.

### **1.1 Objeto**

Los modelos y métodos perceptuales, surgidos, como se ha comentado más arriba, a partir de la necesidad de las industrias de telefonía móvil y de difusión digital de contenidos, de reducir la cantidad de información transmitida sin perjuicio de un determinado nivel de inteligibilidad o de calidad, pueden considerarse como una aproximación que intenta aunar esos dos mundos de lo objetivo y lo subjetivo vistos como separados en tantas ocasiones. En realidad, la diferencia estriba en que en lugar de modelar el sistema estudiado – un canal de transmisión, por ejemplo – se modela la audición humana, es decir, la herramienta con que se evalúa cualitativamente el sistema. El foco se centra pues en la fenomenología que ocurre en el propio receptor, sin considerar los fenómenos del sistema, ya que estos tendrán sus consecuencias en el modelo auditivo utilizado, de manera que se podrá dar cuenta de ellos desde la perspectiva del receptor.

Los métodos perceptuales también son herederos de una serie de conocimientos previos y del aprovechamiento de técnicas pensadas originalmente con otras finalidades, tal como ha ocurrido en otros ámbitos de aplicación en tantas ocasiones, como ya se ha comentado. Por ejemplo, de los resultados obtenidos en experimentos psicoacústicos a lo largo de los poco más de 150 años de existencia de la psicoacústica se han podido derivar, mediante procedimientos de encuesta, escalado, obtención de umbral diferencial, etc., los umbrales absolutos de audición, la respuesta en frecuencia del oído humano, se han podido establecer estimaciones de las resoluciones en frecuencia de la cóclea o determinar patrones de enmascaramiento en función de la frecuencia, de la intensidad del sonido o del tiempo entre diferentes eventos sonoros. También se ha investigado acerca de las formas en que se procesa esa información. Todo este conocimiento, combinado con técnicas de procesado de señal adaptadas para el modelado auditivo, ha posibilitado el establecimiento de modelos de procesado auditivo que pueden utilizarse como indicadores de calidad de una señal de audio o como descriptores de una determinada fenomenología.

Los métodos perceptuales pueden considerarse pues como un puente entre el modelado abstracto de un sistema y la percepción de éste, si bien, un puente aún incipiente y con sus limitaciones. En primer lugar, el propio modelado del sistema auditivo humano es, *per se*, un modelo abstracto y, como tal, deberá definir claramente las variables que lo conforman, con lo cual, algunos elementos de posible importancia que no se consideren en el modelo podrán no tener su correspondiente peso en el modelado. En segundo lugar,

el modelado es incipiente en cuanto a que realiza una representación de la imagen auditiva interna llevada a cabo por el sistema auditivo. Lo que ocurre después con esa información es inaccesible, es decir, no existe un modelo ni una aproximación de las funciones cognitivas que procesan la imagen auditiva generada (ver sección 2.4.2). Con lo cual, la aproximación perceptual llega un poco más allá de la puerta de entrada del sistema perceptivo, quedando muchos rasgos de la percepción, principalmente todos los asociados a las funciones cognitivas, aún por explorar. En este sentido, recientes planteamientos sobre el objeto de estudio de la psicoacústica han puesto de manifiesto la necesidad de ir más allá de las funciones estrictamente auditivas, considerando una perspectiva ecológica, que aborde la integridad de las relaciones establecidas entre el individuo y su entorno acústico (Neuhoff, 2004).

Aun a pesar de estas limitaciones, los modelos perceptuales han demostrado ser eficaces en la evaluación cualitativa de determinados tipos de canales de transmisión, hasta el punto de que algunos modelos se utilizan como sistemas de monitorización de calidad de algunos servicios de transmisión de contenidos y otros han obtenidos excelentes correlaciones con resultados de evaluación cualitativa realizados mediante procedimientos psicoacústicos. Generalmente aplicados a canales de transmisión basados en codificación digital, con éstos se han obtenido mejores resultados que con algunas de las métricas clásicas de evaluación de la distorsión, por ejemplo, lo cual ha llevado a la estandarización de dos modelos principales PESQ y PEAQ, aplicados a la evaluación de palabra con ancho de banda telefónico y audio con ancho de banda musical respectivamente (ITU BS.1387-1, 2001) (ITU-R BS.1534-1, 2003) (ITU-T P.862, 2001).

El principal objetivo de este trabajo es evaluar la posibilidad de aplicar una métrica basada en un modelo perceptual en unos canales de transmisión diferentes a los utilizados inicialmente con estos modelos. Normalmente estos modelos se aplican a señales sometidas a un canal de transmisión digital, de manera que se establece una comparación entre la señal original o de referencia y la señal transmitida. Las diferencias entre ambas señales son evaluadas por el modelo perceptual, obteniéndose a su salida una cuantificación de esta evaluación. Las validaciones de las correlaciones con evaluaciones subjetivas mediante experimentos psicoacústicos suelen realizarse degradando la señal matemáticamente mediante procesado y comparando después los resultados de la métrica con los de la encuesta cualitativa. Es decir, en todos estos procesos, se actúa sobre canales de transmisión o sistemas de degradación de la señal que son intrínsecamente discretos, como lo son un canal de transmisión digital o una degradación de señal por procesado.

En el caso de este trabajo intentaremos extender el ámbito de aplicación de un modelo perceptual a sistemas continuos, como son un transductor electroacústico o un recinto en el que se propaga un evento sonoro. Tanto un transductor como una sala pueden considerarse como canales de transmisión, si bien las degradaciones de la señal introducidas por éstos serán de naturaleza bien diferenciada y de hecho, históricamente su estudio se ha abordado desde perspectivas muy diferentes. Así, en el caso de los

transductores el problema se ha focalizado en la fenomenología de la distorsión, mientras que en los recintos se han estudiado los fenómenos relativos a la propagación del campo acústico en su interior. Pero si cambiamos el foco de atención de nuevo del sistema al receptor, después de estos canales de transmisión tendremos igualmente una señal degradada. El objetivo de este trabajo es indagar en qué medida puede responder un modelo perceptual a la degradación de señal introducida por este tipo de canales de transmisión.

Esto implica, por una parte, definir la métrica a utilizar e implementarla y por otra, evaluar su ámbito de aplicación relacionándola o estableciendo paralelismos con las técnicas conocidas de estimación de la calidad en esos ámbitos. Hemos visto que los modelos perceptuales se fundamentan en el análisis de la señal de audio, realizando una comparación respecto de una señal patrón o de referencia. Se necesitan, por lo tanto, una referencia, un canal de transmisión de esa referencia y la señal transmitida, a la que posteriormente aplicaremos el modelo.

Abordemos ahora el problema del canal de transmisión. Para poder establecer una comparación entre la señal de referencia y la registrada, la primera necesita ser reproducida en el canal. En el caso de las salas, el canal es a priori puramente acústico, pero si queremos reproducir una señal registrada en una sala necesitaremos un elemento adicional: un transductor electroacústico que actúe como fuente de señal o también podría ser el propio sistema de refuerzo sonoro de la sala si lo tuviera. De esta manera, el canal de transmisión lo formarán la suma de uno o varios transductores, en primer lugar, y la propia sala, en segundo. La degradación de la señal introducida por los transductores se estudia en términos de distorsión lineal y no lineal introducida por estos dispositivos. Por otra parte, la degradación de la señal introducida por la sala vendrá dada por el ruido de fondo presente en ésta y por las características de propagación del sonido en su interior, atendiendo a parámetros dependientes del tiempo y de relaciones energéticas, entre otros. La señal registrada estará sometida pues a toda este serie de transformaciones. Es por ello que en primer lugar se estudiará la aplicación de la métrica estrictamente sobre transductores, para analizar el comportamiento de ésta por separado únicamente con dispositivos de transducción. Para realizar esta tarea, se han elegido dos tipos de transductores muy diferentes en cuanto a fundamentos de funcionamiento, características y aplicaciones, de forma que se puedan analizar los resultados que ofrece la métrica ante distintos tipos de canales de transmisión – ahora estrictamente transductores -. Concretamente, se utilizarán transductores de radiación indirecta – caracterizados por ser muy eficientes y ofrecer altos niveles de presión sonora a costa de ofrecer una mayor distorsión, y transductores de radiación directa autoamplificados en forma de monitores de campo cercano. En segundo lugar, una vez determinada la respuesta de la métrica a la primera parte del canal, se estudiará cómo responde ésta ante la degradación de la señal producida por el conjunto transductor más sala, realizando para ello mediciones en 7 salas diferentes en cuanto a volumen y características acústicas. Esta segunda aproximación puede observarse desde dos perspectivas. Por un

lado cuando se realiza el análisis de una sala siempre es necesario disponer de una fuente de excitación acústica de ésta, que generalmente será un transductor omnidireccional emitiendo algún tipo de ruido, barrido en frecuencia o señal determinista. En nuestro caso, dada la necesidad de disponer de una señal de referencia que sea introducida en el canal, la presencia de una fuente de excitación electroacústica es igualmente necesaria, si bien la métrica ya considerará la degradación introducida por el propio transductor.

Por otro lado, si consideramos que actualmente, en la mayoría de los casos, el diseño de salas está orientado a un uso para múltiples aplicaciones, es de esperar que muchos de los recintos cuenten con sus propios sistemas de refuerzo sonoro, con lo cual, una métrica perceptual podría ser también de aplicación directa en la evaluación del conjunto refuerzo sonoro – sala. El conjunto de sistemas refuerzo sonoro-sala también puede ser evaluado utilizando técnicas clásicas (Cunha, I.B., Smiderle, & Bertoli, 2013). En nuestro caso es más correcto plantear que se ha realizado la primera aproximación – el transductor como fuente de excitación de la sala – si bien la implementación de la métrica desde el segundo punto de vista sería realizable siguiendo exactamente el mismo procedimiento.

La métrica que se utilizará es el parámetro  $R_{\text{nonlin}}$  propuesto en (Tan, Moore, Zacharov, & Ville-Veikko, 2004), basada enteramente en un modelo perceptual. Ésta ha sido derivada a partir de otras anteriores (Moore, Tan, & Zacharov, 2003) y validada mediante procedimientos psicoacústicos utilizando degradaciones de la señal realizadas artificialmente (Moore, Tan, Zacharov, & Ville-Veikko, 2004). Se implementará el modelo perceptivo propuesto y se aplicará en los canales de transmisión reales ya mencionados: transductores en primer lugar y el conjunto transductor más sala a continuación.

## 1.2 Estructura

La memoria de este trabajo se estructura de la siguiente manera:

- a) El segundo capítulo pretende dar una visión general del concepto de calidad y sus diferentes definiciones y objetos de estudio dentro de la acústica. Se verá que frecuentemente estos objetos de estudio están interrelacionados, de manera que el estudio o implementación de una métrica de calidad frecuentemente se aborda desde un punto de vista multidisciplinar, como será también nuestro caso.
- b) En el tercer capítulo se analizará el concepto de calidad aplicado a los transductores acústicos. Se estudiará cómo han evolucionado las medidas de distorsión históricamente para intentar establecer mejores correlaciones con los resultados de los test cualitativos de audición y cómo se han introducido los modelos perceptuales en este terreno. Se introducirán los conceptos de distorsión lineal y no lineal. Finalmente, se explicarán algunos conceptos fundamentales de los dispositivos sobre los que se aplicará la métrica experimentalmente: los transductores de radiación indirecta y los de radiación

- directa, haciendo una breve exposición de sus principios de funcionamiento y sus principales fuentes de distorsión o degradación de señal.
- c) En el cuarto capítulo se describirán en mayor detalle algunos de los parámetros de evaluación de calidad de salas que serán introducidos en la sección 2.2.4. Se profundizará más en aquellos que se han utilizado en el procedimiento experimental de este trabajo con la finalidad de poder establecer alguna relación con los resultados de la métrica implementada. También se realizará una introducción al concepto de respuesta al impulso y las posibilidades que ésta ofrece en la obtención de parámetros de calidad en las salas.
  - d) En el apartado quinto se hará una descripción del sistema auditivo humano, centrándonos básicamente en los aspectos que son relevantes para la implementación de la métrica perceptual. Se realizará una descripción de cómo mediante aproximaciones psicoacústicas se obtienen los umbrales de audición, los patrones de enmascaramiento y el ancho de banda crítico. Se describirá también la técnica utilizada para el modelado de los filtros auditivos del oído interno.
  - e) El capítulo sexto explica el trabajo llevado a cabo para la implementación del modelo perceptual. Se explicarán todos los pasos implicados en el procesado de la señal, la metodología utilizada para su implementación y algunos detalles del desarrollo del software que se ha utilizado para el análisis de las medidas experimentales.
  - f) El séptimo capítulo se dedicará a la presentación de los procedimientos y técnicas llevadas a cabo para la aplicación de la métrica a los transductores acústicos, tanto de radiación indirecta como directa, se presentarán los resultados obtenidos y se realizará una discusión de éstos.
  - g) El capítulo octavo expondrá la metodología seguida para la realización de las medidas *in situ* en las diferentes salas, el equipo utilizado y los parámetros medidos además del propio registro de la señal para su análisis mediante la métrica perceptual. Se expondrán los resultados obtenidos para las diferentes salas y las varias señales de excitación utilizadas intentando establecer relaciones entre las métricas clásicas y los valores obtenidos mediante la métrica implementada. Se realizará un análisis de los resultados conseguidos.
  - h) Finalmente, el último capítulo se dedicará al establecimiento de unas conclusiones a partir de los análisis de los resultados obtenidos, a evaluar la viabilidad de uso del sistema perceptual en el canal/es de transmisión propuesto y a enumerar posibles futuras líneas de mejora del modelo.



## 2 Calidad acústica

### **2.1 El problema de la definición de calidad**

El término calidad lleva implícito un profundo debate a la hora de definirlo y contextualizarlo en diferentes ámbitos. En general, se piensa en el término calidad como un sinónimo de “excelencia” a la hora de evaluar un determinado objeto, proceso o evento en una escala de preferencias subjetiva como “bueno – malo”, “excelente – inaceptable”. Sin entrar en consideraciones generales y únicamente en el ámbito que nos ocupa, el término se ha adaptado a la evolución que hayan seguido las diferentes disciplinas. Así, al hablar de calidad acústica, puede uno referirse a la calidad de un transductor, a la calidad de un canal o un sistema de transmisión, a la impresión general que nos causa la audición en una sala o a la perceptibilidad o no de artefactos en la codificación de una señal de audio. El ámbito de aplicación del término es tan extenso, que los intentos por definirlo, a pesar de cuantiosos, difícilmente consiguen un atisbo de unanimidad y mínimo consenso.

Uno de estos casos de cuasi-unanimidad es la generalización del concepto *Product Sound Quality* (calidad sonora del producto) establecida en (Blauert & Jekosh, 1997) (Blauert & Jekosh, 2012). En su consideración más general, este modelo contempla también los aspectos más clásicos de lo que tradicionalmente se ha entendido por calidad sonora, desde una sala de audiciones hasta un instrumento musical, pasando por un dispositivo de reproducción de audio o un electrodoméstico que genera un determinado ruido. En todos estos casos, el objetivo es que el sonido del producto (la sala, el instrumento, el dispositivo o el electrodoméstico) cumpla las necesidades o expectativas del usuario, espectador, consumidor. Como vemos, se trata de una definición intrínsecamente relacionada con las disciplinas denominadas de diseño del producto. En este caso, el sonido es un elemento más del proceso de diseño de un determinado producto. En el último estadio, el diseñador de sonido de un determinado producto es un ingeniero de comunicación (Jekosh, 2005). El sonido transmitirá al consumidor-receptor una serie de cualidades definitorias del producto. Jekosh considera la semiótica de los eventos sonoros en este proceso: en las consideraciones del diseño deben tenerse en cuenta como los individuos perciben el sonido y qué tipo de comunicación se establece durante los eventos sonoros. Esto otorgará al producto unas determinadas cualidades que contribuirán a la percepción de éste. Ésta última perspectiva es aplicable únicamente a productos de fabricación industrial, desde un vehículo hasta un electrodoméstico.

Tradicionalmente, el sonido generado por estos productos no era tenido en consideración. En el mejor de los casos, se trataba como un ruido a minimizar. Si bien este aspecto es importante, el enfoque aportado por Jekosh y Blauert le otorga al sonido la misma categoría que puedan tener los aspectos visuales del producto, en tanto que forman parte de los elementos comunicativos de éste.

En la medida que entramos en el término de diseño del sonido de un producto, nos alejamos del ámbito de aplicación del término calidad que desarrollaremos en este trabajo. Sin embargo, como ya se ha mencionado, el paradigma del *Product Sound Quality* permite definir el término calidad en escenarios más tradicionales, como son los transductores acústicos o la acústica de salas. Al fin y al cabo, un transductor o una sala pueden concebirse como producto. Así podemos definir el término calidad en este contexto siguiendo las definiciones propuestas en (Jekosh, 2004):

**Calidad:** naturaleza física de una entidad en relación con su capacidad para cumplir unos requisitos predefinidos y predeterminados.

**Naturaleza Física:** el conjunto de características de una entidad y los valores asignados a esas características.

**Entidad:** El ítem bajo investigación, ya sea éste material o inmaterial.

Como se ha dicho, estas definiciones están relacionadas con el paradigma del *Product Sound Quality*, pero sin modificaciones son de aplicación directa en otros ámbitos como

los que nos ocuparán: los transductores y las salas. Siguiendo las definiciones de Jekosh, la calidad puede descomponerse en características o cualidades.

**Cualidades:** propiedades que permiten reconocer o distinguir entre entidades.

En definitiva, para ser capaces de evaluar la calidad acústica de una sala o la calidad de un transductor (entidades) han de conocerse las múltiples características (naturaleza física) que contribuyen a la calidad acústica (atributos perceptuales) de manera que puedan buscarse parámetros objetivos que puedan correlar con esas características (cualidades).

Otra perspectiva muy en boga en la actualidad es la denominada Calidad de la experiencia del usuario, que hace referencia a la aceptabilidad general de un producto, aplicación o servicio tal y como se percibe subjetivamente por parte del usuario final. (Le Callet, Moller, & Perkins, 2012) Este enfoque prioriza sobre cualquier otro aspecto la valoración subjetiva del usuario y aplica técnicas de marketing al desarrollo del producto. Aunque a priori fuera de nuestro ámbito de aplicación, se menciona aquí dada la relevancia que tiene actualmente especialmente en el caso de evaluación de la calidad de servicios. En el caso del audio, este enfoque se utiliza en la evaluación de plataformas o servicios de *streaming* musical u otros contenidos de audio.

## **2.2 Ámbitos de aplicación. La dificultad para una teoría general.**

La posibilidad de un modelo universal para la evaluación de la calidad sonora es objeto de discusión sobre su posible existencia o no. Por un lado, una corriente de opinión postula que la experiencia de la calidad es un factor tan altamente subjetivo e individual que no da margen a una posible generalización bajo un mismo paradigma. De otra parte, en el lado opuesto, se busca una teoría general para la medida de la calidad sonora.

Ciertamente, viendo con perspectiva cómo se han desarrollado a lo largo del último siglo las diferentes técnicas de evaluación de los diferentes dispositivos, procesos, recintos y otras entidades relacionadas con la acústica y el sonido, resulta difícil vislumbrar una única teoría que pueda dar cobertura a todos ellos. Pero por otro lado, la mera existencia de todos estos métodos científicos y procedimientos de ingeniería sugiere que la calidad es un atributo medible. Parece ser pues que los diferentes dominios en los que se puede aplicar el concepto llegan a ser tan diferentes que no parece posible alcanzar un modelo generalizado. Estos dominios incluyen ámbitos tan dispares las que veremos a continuación.

### *2.2.1 Medida subjetiva de la calidad sonora con encuestas psicoacústicas*

En este ámbito, se trata de conseguir una opinión general de una cantidad de individuos suficientemente elevada. Generalmente se pide a los participantes que valoren diferentes eventos sonoros en una escala subjetiva o que comparen entre sí estos eventos. Existen dos técnicas predominantes. Por un lado, la denominada *mean opinion score* (MOS) se basa en la asignación a cada evento sonoro de la muestra de un valor de una escala subjetiva discreta verbalizada, es decir, a cada valor de la escala (por ejemplo entre 1-5)

le asocia un descriptor verbal, que ayuda a la toma de decisión. De estos métodos, aunque formalizados e incluso estandarizados hace tan sólo una o dos décadas (Bech & Zacharov, 2006) (ITU-T P.800, 1996), como se verá más adelante, hay referencias en la literatura de su uso desde mediados del siglo pasado (Shorter, 1950)<sup>1</sup>. Por otra parte, una evolución de las técnicas MOS consiste en las denominadas técnicas *multiple stimulus hidden reference with anchors* (MUSHRA). En una encuesta tipo MUSHRA, varios eventos sonoros se valoran simultáneamente en una escala MOS. Los eventos se suelen presentar en conjunto, el individuo que realiza la encuesta dispone de un evento sonoro de referencia y una serie de eventos sonoros que debe valorar respecto al de referencia. Entre estos últimos se incluye una referencia oculta – como indica el propio nombre del método –, es decir, uno de los eventos a evaluar es idéntico al que se le presenta como referencia a la persona que realiza el test. Desarrollada originariamente para la evaluación subjetiva de *codecs* de compresión de audio (ITU-R BS.1534-1, 2003), la metodología se ha extendido a otros ámbitos de aplicación (Sporer, Liebetrau, & Schneider, 2009) y se encuentra actualmente en proceso de estandarización.

### 2.2.2 Medida de la calidad de audio

La evolución que históricamente han desarrollado los dispositivos de captación, transmisión y reproducción del audio ha ido generando, a medida que ésta avanzaba, diferentes técnicas de evaluación de la calidad. Un sistema ideal de reproducción de audio requeriría las mismas condiciones de reproducción que aquellas que se tenían en el proceso de realización del máster o programa sonoro a transmitir. Esto obviamente, es imposible, pero la capacidad de adaptación del sistema auditivo humano a diferentes condiciones de escucha es capaz de mitigar estas diferencias.

Históricamente, las principales limitaciones tecnológicas se han focalizado sobre los sistemas de almacenamiento, transporte o transmisión, que eran los que introducían una mayor degradación sobre el evento sonoro. Estas técnicas se centraban en el concepto de distorsión y en la medida de éste. La distorsión de la señal original introducida por el sistema de almacenamiento o por el canal de transmisión era el principal objeto de estudio. Se consideran los casos de la distorsión lineal y de la distorsión no lineal, que se estudiarán con más detalle más adelante.

Desde la perspectiva de la distorsión lineal, los principales indicadores de calidad se centran en la respuesta en frecuencia del sistema y el retardo de fase. Estas distorsiones se obtienen a partir de la medida de la respuesta al impulso, que también se definirá posteriormente. Estas mismas técnicas han tenido sus homólogas en la acústica de salas, en la que se ha utilizado también la respuesta al impulso para la caracterización de éstas (Berkhout, de Vries, & Boone, 1980). En lo que respecta a la distorsión no lineal, la métrica pionera fue la Distorsión Armónica Total (THD), aún en plena vigencia en los catálogos comerciales de transductores. A partir de ésta se sucedieron una serie de métricas que la mejoraban centrándose en un estudio más detallado de los armónicos, la

---

<sup>1</sup> En este caso para la evaluación de la percepción de armónicos

ponderación de éstos (Shorter, 1950) o el uso de señales más complejas, como las medidas de intermodulación (Hilliard, 1941) o las señales de excitación multitono (Belcher, 1978). Cabe mencionar también las técnicas y procedimientos encaminados a evaluar el efecto del ruido introducido en estos sistemas, como la relación señal-ruido, o la medida de THD+N (distorsión armónica total más ruido). Aun así, el caso del ruido merece una discusión aparte ya que es un atributo común a muchas de las aproximaciones a las medidas de calidad y en otras ocasiones es el propio objeto de estudio.

En la actualidad, con el pleno establecimiento de los sistemas de almacenamiento y transmisión digital de audio, los problemas vinculados a estas etapas de la cadena se han minimizado considerablemente. Se abre un nuevo ámbito de aplicación en el caso del estudio de los sistemas de transmisión basados en *códecs* con pérdidas, pero éstos tienen también un alto componente subjetivo que a menudo se evalúa con los procedimientos vistos en la sección 2.2.1. Así pues, los elementos analógicos que inevitablemente quedan en la cadena, es decir, los micrófonos y los diferentes tipos de altavoces se han convertido en los principales objetos de estudio abordados por estas técnicas. Dado que la producción de voz y los mecanismos humanos de escucha son intrínsecamente “analógicos”, los transductores electroacústicos se han convertido en los principales elementos estudiados bajo la perspectiva de la distorsión – tanto lineal como no lineal -, relevando así el puesto que tradicionalmente habían ocupado los sistemas de almacenamiento y transmisión.

En este mismo ámbito se desarrollaron entre la última década del siglo anterior y la primera de éste los denominados métodos perceptuales, que se fundamentan en el uso de modelados artificiales del sistema auditivo humano para la evaluación de la calidad. Teóricamente, se trata de implementar un modelo computacional del sistema de escucha humano que pueda explicar los procesos perceptivos que tienen lugar al escuchar un determinado evento sonoro. Aunque en la actualidad el modelado auditivo se encuentra bastante lejos de su objetivo ideal, los métodos basados en modelos perceptuales han demostrado ser más eficaces para evaluar la calidad sonora de *códecs* para la compresión y transmisión de audio que los métodos más convencionales de medida de la distorsión (Meddis & López Poveda, 2010).

Existen una gran cantidad de métodos desarrollados para la evaluación de la calidad sonora en sistemas de compresión de baja tasa binaria, los archiconocidos *codecs* de audio (mp3, aac, wma, etc...). Estos métodos se desarrollan alrededor de dos estándares, uno para la evaluación de señales de audio musicales comprimidas –PEAQ, *perceptual evaluation of audio quality* - (ITU BS.1387-1, 2001) y el otro para señales de voz con menor ancho de banda denominado PESQ - *perceptual evaluation of speech quality*- (ITU-T P.862, 2001). En modelos perceptuales existen dos técnicas básicas de aproximación: una de ellas se denomina *masked threshold concept* o también *noise to mask ratio* y funciona con una señal de error (distorsión) obtenida a partir de la diferencia entre la señal distorsionada y la señal de referencia. Esta diferencia entre la señal

distorsionada y la referencia puede calcularse tanto en el dominio temporal como en el espectro en frecuencia. La obtención de esta diferencia en el dominio temporal suele ser demasiado sensible a las distorsiones de fase por lo que se prefieren los cálculos en el dominio de la frecuencia. Esta señal de error se compara para todos los filtros paso-banda que simulan los filtros auditivos con los patrones de enmascaramiento obtenidos del modelo auditivo. Si el nivel de la señal de error o señal de distorsión es menor que el patrón de enmascaramiento producido por la señal original en la misma banda de frecuencia y el mismo *frame* temporal, esa muestra de señal de distorsión se considera inaudible. Las distancias psicoacústicas obtenidas para cada filtro auditivo y *frame* temporal se superponen y/o ponderan en un único número final que caracterizará la calidad sonora.

Esta técnica se utiliza mayoritariamente en la evaluación de *codecs* de compresión y su diagrama de bloques simplificado puede verse en la Figura 2.1. La segunda metodología de aproximación definida en (ITU BS.1387-1, 2001) se denomina comparación de representaciones internas y está más cerca de los procesos psicológicos que tienen lugar en el sistema auditivo humano. Se realiza una representación interna de la señal de referencia y de la señal bajo test. Esta representación interna es una estimación de la información de que dispondrá el sistema auditivo humano para realizar comparaciones entre ambas señales. La información indicativa de la calidad percibida se derivará a través de la simulación de los patrones de excitación de la membrana basilar provocados tanto por las señales de referencia como por la distorsionada. Las representaciones internas asignan a la señal de audio inicial la reacción fisiológica de la membrana basilar causada por esa misma señal.

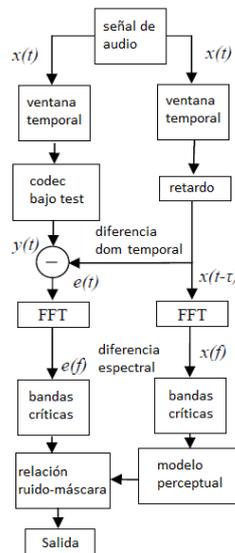


Figura 2.1. Diagrama de bloques de una representación *noise to mask ratio*.

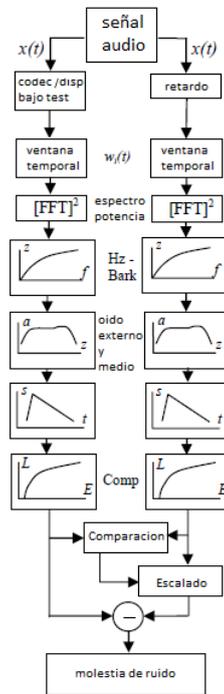


Figura 2.2. Diagrama de bloques de una metodología de representaciones internas.

En el caso de señales estacionarias la representación interna está mejor descrita en el dominio de la frecuencia. Para señales no estacionarias la representación se realiza en el dominio temporal. La señal original  $x(t)$  y la distorsionada  $x_d(t)$  son enventanadas en el dominio temporal con una ventana  $w(t)$  y transformadas al dominio de la frecuencia mediante FFT. Posteriormente se calcula el espectro de potencia y se pasa a una escala basada en Barks. La señal se ecualiza para simular la transformación del espectro producida por el oído externo y medio. A continuación la señal se propaga de acuerdo con la reacción fisiológica de la membrana basilar y se comprime debido a la relación no lineal entre la sonoridad percibida y la intensidad del sonido. Finalmente las señales se comparan para obtener la diferencia en todos los filtros auditivos y ventanas de tiempo. El procesado descrito sería el más idóneo para señales estacionarias. Como se verá en la implementación de la métrica  $R_{nonlin}$ , que es la que se analizará en la parte experimental de este trabajo, el procesado se realizará en el dominio del tiempo, al tratarse de señales complejas.

De la misma forma que las técnicas de medida de distorsión han sido aplicadas a la evaluación de sistemas de compresión de audio, las técnicas perceptuales han extendido su dominio de aplicación a otros ámbitos, como el caso de los transductores acústicos (Voishvillo A. , 2006) , con lo cual existen un gran número de variaciones sobre estas

dos metodologías, adecuándose en cada caso al objeto concreto de estudio. Por ejemplo, la distorsión introducida por un *codec* de compresión difiere significativamente de la de un sistema radiante, de forma que el método de evaluación deberá ajustarse a las particularidades del sistema bajo test.

Siguiendo en el territorio de la evaluación de la calidad de señales de audio y en lo referente a los sistemas desarrollados para reproducir o sintetizar las características espaciales del sonido – los sistemas de reproducción de audio conocidos como sistemas multicanal, desde una simple imagen estéreo a una más compleja 5.1 o 7.1 – también se han desarrollado técnicas para la evaluación de éstos (Rumsey, Zielinski, Kassier, & Bech, 2005) haciéndose estimaciones de la contribución de la espacialidad al conjunto de la valoración cualitativa. El principal hándicap de esta técnica reside en que las condiciones de escucha de la sala donde se realiza el experimento serán significativamente diferentes de las de cualquier sala en la que se instale un sistema multicanal de manera que se han realizado algunas aproximaciones mediante grabaciones multicanal anecoicas en las que posteriormente se reproducen diferentes condiciones de escucha (Vilkamo, Lokki, & Pulkki, 2009).

Los modelos perceptuales binaurales son los que a priori están llamados a resolver la problemática de la evaluación de la espacialidad del sonido y su aportación a la valoración cualitativa. Sin embargo, de momento, sólo en un número limitado de casos son de aplicabilidad para la evaluación de la calidad espacial (Blauert, 2013). En términos generales se hace necesaria una extensión del modelo PEAQ a la medida de calidad sobre señales multicanal.

### 2.2.3 *Calidad del habla*

Este ámbito de la evaluación de la calidad ha estado históricamente vinculado de forma muy estrecha con la telefonía y la evolución tecnológica experimentada por ésta. Desde principios del siglo XX y a lo largo de éste la aportación de los laboratorios Bell contribuyó a la elaboración de una potente base de conocimiento sobre la voz humana (Fletcher, 1922), el estudio de los mecanismos o canales de transmisión, así como sentó también la base de los estudios psicoacústicos sobre la percepción de la voz humana a través de un canal de transmisión (French & Steinberg, 1947). La reciente irrupción de la telefonía a través de IP, así como los sistemas de intercomunicación hombre máquina o los sistemas de reconocimiento de voz han abierto un nuevo campo de aplicación para la evaluación de la calidad de éstos. Una versión de los modelos de evaluación de calidad en los sistemas de telefonía actuales puede verse en (Möller, 2005).

En una conversación humana en la que no media sistema de comunicación eléctrico o electrónico alguno, el canal de transmisión es el camino acústico entre el órgano fonador del emisor y el sistema auditivo del receptor. Este canal de transmisión, puramente acústico también ha sido evaluado de diferentes maneras a lo largo de la historia. Un ejemplo son las métricas y procedimientos de evaluación de la inteligibilidad de la voz en salas. Lo que resulta más curioso es que, a la hora de evaluar la efectividad de un

canal estrictamente acústico para la transmisión de la voz, frecuentemente, se han utilizado técnicas o métodos derivados o adaptados de aquellos empleados para la evaluación y estudio de un canal de transmisión eléctrico, como es el caso de la telefonía, la radiodifusión o el PA (*public address* – sistemas de refuerzo sonoro tanto al aire libre como en salas). Concretamente, el caso más extendido y estandarizado es el de la métrica STI (*speech transmission index*) propuesto en (Hougast & Steeneken, 1984) para su uso en recintos, salas y auditorios a partir de (Steeneken & Houtgast, 1980) en el que formulaban la misma métrica de forma genérica para un canal de transmisión retomando el trabajo iniciado décadas antes por French y Steinberg en la formulación de una métrica objetiva para la medida de calidad de un canal de transmisión eléctrico, el índice de articulación (AI) (French & Steinberg, 1947). En el siguiente epígrafe se introducirá más detalladamente el procedimiento de medida STI, basado en la función MTF (*modulation transfer function*), que se explicará en el capítulo 4.

En lo que respecta a la evaluación de los canales de transmisión, desde un primer momento se acuñó el término inteligibilidad. Frecuentemente, en el caso de la voz humana los términos inteligibilidad y calidad se usan indistintamente, como si la única cualidad valorable subjetiva u objetivamente del canal de transmisión fuese la inteligibilidad de los mensajes a través de éste. De nuevo, la perspectiva histórica aporta una explicación a esta incongruencia: en los tiempos de la incipiente telefonía el objetivo primordial era asegurar la inteligibilidad del mensaje, es decir, encontrar un sistema de transmisión, que con el mínimo coste – de infraestructura y de ancho de banda – asegurase que el mensaje se percibía de forma inteligible en el otro extremo de la línea telefónica. Esto, sin duda, es un indicador de calidad, aunque muy básico. Pero no resulta muy incongruente pensar que en la primera mitad del siglo pasado el concepto de calidad se restringiera a la inteligibilidad, especialmente debido a las limitaciones tecnológicas de los sistemas de telefonía de la época.

El término ha continuado siendo preponderante a la hora de hablar de la evaluación de sistemas de transmisión del habla, si bien, hoy en día, también se consideran otros aspectos que contribuyen a la calidad global de un sistema, más allá de la inteligibilidad y, obviamente, una vez asegurada ésta. Entre estos últimos destaca el reconocimiento del hablante por parte de interlocutores conocidos y otros como la naturalidad del habla. Esta última es especialmente relevante en el caso del habla sintetizada. Durante décadas, se han desarrollado numerosos métodos tanto objetivos como subjetivos para medir la calidad del habla. Se enumeran a continuación algunos de los más utilizados de entre los subjetivos.

- a) Test de inteligibilidad mediante el uso de logatomos. La tarea de los receptores es escuchar e identificar una secuencia de palabras generalmente monosilábicas y sin significado alguno formadas por la concatenación de una consonante más una vocal (CV) o, usado más frecuentemente debido a su mayor dificultad de identificación, una sílaba del tipo CVC (consonante-vocal-consonante). El porcentaje de respuestas correctas constituye el índice de inteligibilidad.

- b) Además del uso de logatomos de la técnica anterior también se utiliza esta misma técnica de identificación del mensaje mediante el uso de frases o palabras reales. Al igual que en el caso anterior, el porcentaje de respuestas correctas constituye el índice de inteligibilidad. Generalmente, el hecho de que el material de test (las frases o las palabras) sean reales otorga a los mensajes una contextualización en la que los elementos cognitivos del receptor pueden ayudar a mejorar la inteligibilidad del mensaje, con lo cual, normalmente los índices obtenidos con esta técnica son mayores que los obtenidos en el caso de los logatomos monosilábicos.
- c) Test de rimas. Esta técnica utiliza palabras en las que el cambio de un único fonema altera el significado de la palabra. Existen multitud de modalidades de test de rimas en función de su aplicación, que puede extenderse desde la inteligibilidad de una línea de comunicación hasta el ámbito de la logopedia, por ejemplo.
- d) Métodos de comparación de calidad, mediante encuesta psicoacústica, en la que al encuestado se le pide que ordene por orden de preferencia varias muestras de audio con contenido hablado.
- e) Métodos de isopreferencia. La idea básica consiste en registrar la señal de habla a través del canal a evaluar a diferentes niveles y añadiendo también diferentes niveles de ruido de fondo. En un plano de coordenadas determinado por el nivel de la señal hablada y el nivel del ruido de fondo el encuestado ubica cada una de las muestras que se le presentan
- f) Encuestas de tipo MOS (ver sección 2.2.1).
- g) Test de comunicabilidad. Se utilizan dos sujetos que se comunican a través de un canal. En esta comunicación uno de los sujetos dirige al otro en la realización de una determinada tarea. Al finalizar ésta se le pregunta a los sujetos por la facilidad de comprensión de la comunicación y el esfuerzo que hayan debido realizar para transmitir y/o comprender las instrucciones.
- h) Test de supresión de ruido. Utilizado en el ámbito de la telefonía móvil, para evaluar los algoritmos de supresión del ruido de fondo que utilizan estos dispositivos (ITU-T P.835, 2003).

Como se verá en la parte experimental de este trabajo, el material fuente de cualquier procedimiento destinado a la evaluación de la calidad del habla debe estar fonéticamente equilibrado. Esto significa que el contenido fonético del material que se utiliza como habla de prueba debe ser proporcional a la probabilidad de aparición de dichos fonemas en una determinada lengua. Esto introduce la variable del idioma en los resultados obtenidos y materiales a utilizar. Por ejemplo, el primer test de rimas propuesto en (Fairbanks, 1958) y modificado y estandarizado por (House, Williams, Hecker, & Kryter, 1965) no tuvo su homólogo en español hasta la última década del siglo XX (Aguilar, 1991) y su versión en catalán en (Alías & Triviño, 2007). De la misma manera, las conocidas como frases psicoacústicas de Harvard, introducidas por (Ergan, 1948) desarrolladas inicialmente para realizar audiometrías y posteriormente utilizadas para la

evaluación de la calidad del habla, se adaptaron al español en (Valero, 1991). Se trata de un corpus de frases fonéticamente equilibradas, con significado – a diferencia de los logatomos – y sintácticamente bien formadas (Llisterri & Poch, 1991).

Generalmente, los materiales lingüísticos son originariamente en lengua inglesa y con el tiempo aparecen las adaptaciones a otras lenguas. En el caso de los logatomos, existe una norma que correla un corpus concreto de logatomos fonéticamente equilibrado en inglés con el parámetro STI, de forma que establece una equivalencia entre el valor de articulación obtenido – porcentaje logatomos bien percibidos – y la métrica objetiva *Sound Transmission Index* (STI) (ISO 9921, 2003). En lengua española, existe un corpus de logatomos propuesto en (Pérez Miñana, 1969), cuya correlación con STI ha sido establecida en (Sommerhoff & Rosas, 2007) observándose diferencias significativas respecto de la correlación estandarizada en lengua inglesa. Trabajos posteriores como (Rosas & Sommerhoff, 2008) (Hidalgo, Sommerhoff, & Rosas, 2011) han tenido como objetivo la elaboración de un corpus de logatomos adaptado para el español de Hispanoamérica con la finalidad de que sea equiparable al corpus normalizado del inglés. Así, pues, la variable del idioma resulta determinante no sólo en lo que respecta a la disponibilidad y preparación de los materiales de test. Además, las características fonéticas y fonológicas de un determinado idioma pueden ser determinantes en cuanto a los resultados de la evaluación se refiere. Sirva como ejemplo la siguiente tabla, en la que se comparan los pesos determinados en la inteligibilidad de la lengua de cada banda de octava para las lenguas inglesa y húngara (Tarnóczy, 1986).

**Tabla 2.1. % de inteligibilidad asociado a cada banda de octava**

<b>Banda (Hz)</b>	<b>125</b>	<b>250</b>	<b>500</b>	<b>1000</b>	<b>2000</b>	<b>4000</b>	<b>8000</b>
<b>Inglés</b>	3%	15%	29%	28%	17%	8%	0%
<b>Húngaro</b>	2%	13%	18%	22%	22%	20%	3%

Fuente: (Tarnóczy, 1986)

A la vista de la tabla anterior, resulta inmediato pensar que el comportamiento de un determinado canal de transmisión de habla en frecuencia puede ser valorado cualitativamente de forma muy diferente si la evaluación se efectúa en inglés o en húngaro. Un canal con deficiencias en alta frecuencia podría llegar a ofrecer resultados aceptables en lengua inglesa, pero inadmisibles en húngaro, por ejemplo.

En lo que respecta a los denominados métodos objetivos para la evaluación de la calidad del habla cabría destacar los siguientes procedimientos.

- a) Índice de articulación (AI). Propuesto inicialmente por (French & Steinberg, 1947) y desarrollado para medir la inteligibilidad del habla a través de un canal considerado cuasi lineal pero con la presencia de ruido. La pérdida de articulación se obtiene sumando los valores de AI obtenidos en diferentes bandas de frecuencia, considerando el espectro en escala Bark (ver sección 4.3).
- b) Porcentaje de pérdida en la articulación de consonantes (% $AL_{cons}$ ). Introducido inicialmente en (Peutz, 1971) y reformulado posteriormente en (Davis &

Patronis, 2006) es un parámetro utilizado con relativa frecuencia para evaluar la inteligibilidad del habla en salas, auditorios y grandes espacios. Se verá con más detalle en el capítulo 4.

- c) Índice de transmisión del habla (STI). Es el más estandarizado de entre los denominados métodos objetivos. Fue propuesto por (Steeneken & Houtgast, 1980) para la evaluación de canales de transmisión, siguiendo la línea iniciada por el AI. Posteriormente, los mismos autores propusieron la utilización del mismo índice para la evaluación en salas (Houtgast & Steeneken, 1984). El cálculo del índice se basa en la función de transferencia de modulación (MTF). Como material del test se utilizan señales moduladas que imitan el habla humana en diferentes bandas de frecuencia. Dado que la profundidad de modulación de la señal de test es conocida, se compara con la profundidad de modulación de las señales obtenidas una vez han pasado por el canal. En el caso de las salas, en presencia de ruido y de reverberación, se producirá una reducción de la profundidad de modulación. La métrica se obtiene a partir del cálculo de estas reducciones para cada banda de frecuencia. Esta métrica se verá con detalle en la sección 4.7.
- d) Relación señal/ruido. La relación señal ruido cuantifica en qué medida una señal se diferencia del ruido de fondo. En su definición más tradicional se define como SNR, en la que  $x(n)$  es la señal original y  $x_d(n)$  es la señal distorsionada una vez ha pasado por el canal de comunicación. La relación señal ruido es muy sensible a cualquier tipo de diferencias entre la señal original y la distorsionada, incluyendo aquellas que son imperceptibles, lo cual delimita su ámbito de aplicación a la hora de evaluar aspectos cualitativos.
- e) Medidas de distancia espectral. Al igual que las dos aproximaciones anteriores, se basan en la comparación entre una señal original y una alterada tras su paso por el canal de comunicación. La particularidad radica esta vez en que las diferencias entre ambas se analizan en el dominio frecuencial, en lugar de analizar las señales temporales. Las señales se representan en un mapa de intensidades tiempo-frecuencia, comparándose la original y la comunicada. Se verá algún ejemplo en la parte experimental de este trabajo.
- f) Medidas de calidad sonora basadas en modelados del sistema auditivo humano. Hacen referencia a los dos estándares expuestos en el apartado anterior (PEAQ y PESQ). Como se ha adelantado, se trata de modelos computacionales que realizan una representación interna de las señales a partir de un modelado del sistema auditivo humano. Una vez realizado ese procesado existen diferentes técnicas para evaluar la diferencia entre las señales. Aunque el estándar PESQ hace referencia explícita al habla (*Perceptual Evaluation of Speech Quality*), éste está más vinculado a canales de transmisión de banda estrecha y se utiliza frecuentemente en telefonía móvil. El estándar PEAQ (*Perceptual Evaluation of Audio Quality*) es igualmente aplicable al habla en sistemas con más ancho

de banda. Esto constituye una de las bases de los procedimientos experimentales que se desarrollarán en este trabajo, como se explicará en el capítulo 6.

#### 2.2.4 *Calidad sonora de salas, recintos y auditorios*

Los espacios para representación teatral, interpretación musical, intervención de oradores, etc... se han venido utilizando desde tiempos ancestrales en diferentes culturas. Es muy difícil hacer abstracción de lo que para un teatro griego o romano podía considerarse calidad acústica, pero parece ser que la calidad del sonido percibido por la audiencia de estos recintos ha sido históricamente de interés (Blauert, 2013), estableciéndose comparaciones entre diferentes recintos cuyas características se podían adaptar o modificar por ensayo y error. Hasta casi los inicios del siglo XX no se realizó la primera aproximación científica en el ámbito de la acústica de salas. Los resultados de estas primeras investigaciones concluyeron, entre otros, en el establecimiento de la formulación del tiempo de reverberación (RT).

$$RT = T_{60} = 0.161 \frac{V}{\bar{\alpha}S} \quad (2.1)$$

La ecuación anterior es el principal resultado de estos trabajos pioneros en la acústica de salas. En 1895 Wallace Sabine recibió el encargo de la Universidad de Harvard de mejorar la inteligibilidad de una sala para clases magistrales y empleando un tubo de órgano accionado por una electroválvula neumática que emitía a 512 Hz y un cronógrafo controlado por la misma señal eléctrica que la electroválvula determinó la relación existente entre el volumen de la sala, la absorción acústica de las superficies y el tiempo de reverberación (Sabine, 1922). Suponiendo una sala en la que una fuente acústica está emitiendo un sonido constante y continuo, éste produce un determinado nivel de presión sonora en el interior de la sala,  $L_0$  (dB). Si la fuente cesa la emisión, se producirá un descenso gradual de este nivel de presión sonora. La manera en que se producirá este descenso vendrá determinada por las características de la sala, tanto geométricas, como de absorción de los materiales que revisten o conforman sus superficies.

Sabine definió RT como el tiempo que transcurre desde el momento en que la fuente cesa la emisión hasta que el nivel de presión sonora en la sala es de  $L_0-60$  (dB), es decir, el tiempo en que el nivel de presión tarda en descender 60 decibelios. En la ecuación 1,  $V$  es el volumen de la sala,  $S$  la superficie total y  $\alpha$  el coeficiente de absorción medio de los materiales que confinan el recinto. La formulación de Sabine es todavía hoy de uso común para una estimación rápida del RT. Así, en un primer momento y a lo largo de las décadas siguientes se abordó la problemática de la acústica de salas mediante parámetros temporales. A la propuesta de cálculo de Sabine le siguieron varias modificaciones respecto al establecimiento del tiempo de reverberación, las más conocidas de las cuales son (Eyring, 1930) y (Millington, 1932). Aún bien avanzado el siglo XX parecía que el único dato disponible para abordar la clasificación cualitativa de las salas era el tiempo de reverberación.

Sin embargo, como es lógico suponer, no existe un tiempo de reverberación ideal, de forma que el RT no es *per se* un indicador de calidad. En la famosa sala del *Fogg Arts Museum* en que Sabine pasó días y noches en busca de la ecuación 2.1 el problema era la inteligibilidad de la palabra debida a un tiempo de reverberación excesivo. En general, un tiempo de reverberación largo tiende a calificarse intuitivamente como perjudicial desde el punto de vista acústico, pero ésta es una simplificación con muchas objeciones. El establecimiento de un tiempo de reverberación adecuado vendrá determinado por el uso que se le vaya a dar al recinto. Así, en el caso de un teatro, por ejemplo, hay que priorizar la inteligibilidad de la palabra, con lo cual, los tiempos de reverberación deberán ser reducidos. El caso opuesto puede darse en un recinto para la audición de música sinfónica. Una sala excesivamente apagada, i. e., con muy poca reverberación, puede desnaturalizar el sonido y ser percibida como inadecuada por la audiencia. Los dos casos anteriores se refieren a ejemplos en que el sonido presente en el recinto – excluyendo, obviamente, el ruido de fondo - lo generan las fuentes acústicas únicamente, es decir, los actores, oradores, instrumentos musicales.

Frecuentemente, se ha focalizado la evaluación de la calidad de recintos basándose en supuestos donde las fuentes son estrictamente acústicas, pero lo cierto, quizá exceptuando el caso de la música clásica, es que generalmente se recurre a sistemas de refuerzo sonoro y en la gran mayoría de recintos, auditorios, teatros, etc... tienen lugar eventos con características acústicas bien diferenciadas (teatro, conciertos electroacústicos, conciertos acústicos con muy diferentes de formaciones –algunas de las cuales requieren refuerzo sonoro- , intervención de oradores...). El caso de las salas destinadas a un uso concreto queda reservado para unos pocos recintos de ámbito y prestigio mundial. Pero la inmensa mayoría de recintos albergan eventos de muy diversa índole. Esta característica “multiuso” de los recintos hace acaso más difícil considerar un tiempo de reverberación ideal, dado por una parte las características acústicas diferenciadas de las fuentes sonoras como el uso o no de sistemas de refuerzo sonoro.

Durante el transcurso del siglo XX, se introducirán numerosas variaciones para la obtención del RT, ya sea en su cálculo o en los procedimientos para su medición *in situ* en los recintos. De los datos obtenidos en estas estimaciones o mediciones, se derivarán otros criterios para la estimación de la calidad relacionados con aspectos perceptivos, como pueden ser la calidez (*Bass Ratio* - BR), el brillo (*Brilliance* – Br) o la intimidad (*Intimacy*) (Beranek L. L., 1962). Beranek elaboró una lista de hasta 18 parámetros para la evaluación global de una sala como se verá en el capítulo 4. El valor de cada uno de estos parámetros era después ponderado según la finalidad de la sala para obtener una valoración general. Muchos de estos parámetros, como los mencionados anteriormente, están estrechamente relacionados con el tiempo de reverberación o con una medición de aspectos temporales de la propagación acústica en el recinto. Así, la calidez y el brillo son una relación entre los tiempos de reverberación en las octavas de baja y alta frecuencia respectivamente.

La intimidad viene determinada por el tiempo que transcurre entre la llegada del sonido directo y las primeras reflexiones. Siguiendo con el análisis desde el punto de vista temporal, se derivaron otros parámetros de uso frecuente como el *Early Decay Time* (EDT) (Jordan V. , 1970) que establece el tiempo de reverberación extrapolando la pendiente de la caída energética desde 0 a -10 dB para el caso de -60 dB. EDT está considerado mayoritariamente como el más representativo, o el *Initial Time Delay GAP* (ITDG) (Davis, 1979), que estima el tiempo entre la llegada del sonido directo y la primera reflexión al oyente. En (Jordan V. L., 1982) se definirá también el tiempo de subida, o *rise time*, definido también como el tiempo de subida a -3 dB, que corresponderá al tiempo necesario para que el nivel del campo reverberado transitorio alcance la mitad de su nivel estacionario.

A partir de la década de los 50 empiezan a adoptarse otros puntos de vista, más allá del estrictamente temporal, aunque ésta, el tiempo, continuará siendo una variable fundamental, como veremos. Se trata de métricas que calculan relaciones energéticas delimitadas por una constante de tiempo. A éstas, se las suele denominar parámetros energéticos. Así, en (Thiele, 1953) se introduce la Definición, que posteriormente dará lugar a una serie de parámetros basados también en relaciones energéticas. La Definición compara la energía recibida en los primeros 50ms con la energía total recibida por el oyente incluidos esos primeros 50ms.

Otros parámetros que seguirán un planteamiento similar y que son ampliamente utilizados en la acústica de salas son los de Claridad  $C_{50}$  y  $C_{80}$ , que establecen la relación entre la energía recibida en los primeros 50 u 80 ms y el resto de energía recibida a partir de esos instantes. (Reichardt, Abdel Alim, & Schmidt, 1974) (Reichardt, Abdel Alim, & Schmidt, 1975). Se suele utilizar la constante de 50 ms para la palabra y la de 80 para la música. En (Kürer, 1971) se introducirá el concepto de tiempo central  $T_s$ , o centro de gravedad de la curva de caída como otra de las métricas denominadas objetivas dentro de las consideradas energéticas.

En (Lehmann, 1976) se presentó el *Strength* o factor de fuerza introducido por la sala al realizar la comparación de los niveles de presión sonora producidos por una misma fuente omnidireccional en el interior de una sala y en campo libre a una distancia de 10m. El refuerzo sonoro introducido por la sala está estrechamente vinculado a la sonoridad percibida por el espectador. Otros parámetros en esta línea pueden ser las curvas de energía reflejada acumulada, propuestas en (Toyota, Oguchi, & Nagata, 1988). Las curvas RECC (*Reflection Energy cumulative Curve excepting direct sound*) grafían la energía reflejada acumulada hasta un instante  $t$ , generalmente 80ms, sin considerar el sonido directo. RECC, hasta aproximadamente los 80ms muestran diferentes formas en función de las características del recinto y de los puntos de medición dentro de éste. A partir de los 80ms, la forma de las curvas tiende a converger para todos los recintos. Establecen que casi la totalidad de la energía reflejada viene determinada por el parámetro ARE80 (energía reflejada hasta los 80ms). Este parámetro mostrará importantes variaciones en función del punto de medida dentro de la sala,

independientemente del tiempo de reverberación, que puede ser similar o incluso idéntico para dos puntos donde el parámetro ARE80 varíe significativamente. A partir de la consideración energética, las curvas RECC ponen de manifiesto la importancia de la forma y geometría de la sala en la contribución a la creación del campo acústico en un determinado punto y como éste puede variar ostensiblemente en diferentes puntos de la sala.

Otro criterio de aproximación a la evaluación de la calidad de las salas lo constituyen los denominados parámetros espaciales. Estos criterios están orientados a establecer la impresión espacial percibida por el espectador y están asociados a aspectos perceptivos que pueden describirse como la sensación de envolvente del sonido, la impresión del tamaño de la sala o de proximidad o no a la fuente. Hemos visto que estas descripciones cualitativas pueden asociarse también a mediciones basadas en aspectos temporales. Los espaciales introducirán un nuevo enfoque para describirlas. Parámetros como LF - *Lateral Fraction* (fracción de energía lateral) - (Barron, 1981) se obtiene a partir de la respuesta al impulso de la sala realizada con un micrófono bidireccional comparándola con la respuesta al impulso medida con un micrófono omnidireccional, de manera que refleja la ratio de la energía lateral en el conjunto de energía recibida en un determinado punto de la sala. En la misma línea (Jordan V. , 1980) propone LE - (*Lateral Efficiency* - eficiencia lateral -) de forma muy similar a Barron pero están vez considerando la relación entre la energía lateral (respuesta al impulso bidireccional) entre 25 y 80ms y la energía total (respuesta al impulso omnidireccional) considerando para ésta unos límites de integración entre 0 y 80ms. Algunas variaciones para la medición del factor de energía lateral teniendo en consideración el ángulo de incidencia fueron propuestas en (Kleiner, 1989).

Las técnicas basadas en mediciones o grabaciones binaurales supusieron un importante avance para la definición de nuevos parámetros que analizan las características espaciales (Schroeder, Gottlob, & Siebrasse, 1974) centrándose en la comparación de las señales recibidas en los dos oídos. Para estas grabaciones se utilizan torsos artificiales o *dummies*, cuyas cabezas llevan acopladas unos micrófonos que ejercen la función de "oídos". La técnica utilizada es la correlación cruzada, que, como puede verse en el apartado 6.7 es una evaluación del grado de similitud entre dos señales. Así, se introduce el parámetro IACC (*InterAural Cross correlation*) que medirá esta similitud para un ancho de banda de entre 100 Hz y 8 kHz y para un tiempo entre 0 y 1s como límites de integración. Posteriormente se definirán los parámetros IACC<sub>E</sub> (*early*) e IACC<sub>L</sub> (*late*). IACC<sub>E</sub> hace referencia a la correlación entre las señales binaurales en un período de 0 a 80ms, mientras que IACC<sub>L</sub> considera los tiempos entre 80ms y 500ms. En lo que respecta al ancho de banda utilizado para su medición, normalmente, ambos valores se determinan promediando las correlaciones obtenidas en las octavas de 500 Hz, 1000 Hz y 2000 Hz. Estos procedimientos de medida se estandarizaron en (ISO 3382, 1997). En (Beranek L. , 1996) se estableció una correlación entre IACC<sub>E</sub> y la que denominó ASW (*Aparent Sound Width*), que podría traducirse como amplitud de sonido aparente, de

manera que ésta venía determinada por  $\rho = 1 - IACC_E$ . Según el autor este parámetro correlaba muy bien con la aparente sensación de amplitud de la fuente sonora percibida por el espectador. Igualmente definió LEV (*Listener Envelopment*) como  $\varepsilon = 1 - IACC_L$ , que correlaba con la sensación subjetiva por parte del oyente de sentirse “envuelto” por el sonido. Con esto, las diferencias binaurales correspondientes a los primeros instantes (80ms), es decir, las asociadas a la recepción de las primeras reflexiones y la reverberación inicial resultarán determinantes en la formación de una imagen de amplitud de la fuente sonora por parte del receptor. Las diferencias de recepción entre el oído derecho y el izquierdo respecto a la reverberación tardía constituyen la base del mecanismo que genera la sensación cualitativa de “sentirse envuelto” por el sonido. La validez de ASW y LE ha sido corroborada también en entornos acústicos generados artificialmente mediante síntesis del campo acústico (Nowak, Liebetrau, & Sporer, 2013).

Además del espacial, otro criterio específico para la evaluación de la calidad de una sala es el que engloba a todas las métricas asociadas a la calidad de la palabra hablada. Acerca de esto, hemos visto la mayor parte de conceptos en el epígrafe anterior. Además de los métodos expuestos allí (test subjetivos, el índice de articulación, el %AL<sub>cons</sub> y el STI como los más destacados) cabría mencionar que tanto la Definición (Thiele, 1953) como la Claridad con una constante de tiempo de 50ms también se han utilizado para la evaluación de la calidad del habla en salas. Adicionalmente, el criterio de eco, (Dietsch & Kraak, 1986) se ha utilizado en ocasiones preferentemente en el ámbito de la palabra.

Finalmente, cabría hacer una mención a las técnicas de síntesis del campo acústico, en un uso cada vez más frecuente a la hora de evaluar y predecir comportamientos de diferentes parámetros. Como se ha visto, en la evaluación del audio los estándares PASQ y PESQ una de las finalidades es la reducción de costes económicos y de tiempo en realizaciones de encuestas extensísimas, reduciendo éstas a las estrictamente necesarias para la validación del modelo. Algo similar puede entenderse respecto a las virtualizaciones del campo acústico. Las mediciones *in situ* en salas pueden minimizarse, al menos en un estadio inicial, realizando simulaciones de campos acústicos y obteniendo las respuestas al impulso en estos entornos virtuales. De esta manera, puede realizarse una validación a priori de los parámetros que deseen evaluarse.

Existen diferentes técnicas de simulación del campo acústico. Entre las más simples destaca la inicialmente propuesta en (Allen & Berkley, 1979) basada en geometrías básicas, generalmente en forma de prismas rectangulares. Las reflexiones se generan a partir de imágenes-espejo de la ubicación de la fuente circunscrita entre las superficies del recinto. A partir de las fuentes especulares se generan las sucesivas reflexiones permitiendo obtener una recreación del campo acústico. Otras técnicas están basadas en el trazado de rayos (Krokstad, Strom, & Sorsdal, 1968), en el trazado de haces (Heckbert & Hanrahan, 1984) o en modelos de propagación ondulatoria (Berkhout A. J., de Vries, Baan, & van den Oetelaar, 1999).

### 2.2.5 Ruido y calidad

Como se ha visto hasta ahora, el término ruido está presente de forma transversal en cualquier aproximación que se realice para la medida de la calidad. Siguiendo una definición clásica se entiende por ruido todo aquel artefacto que estando presente en el canal de comunicación, interfiere o modifica en cierta manera el contenido del mensaje original pudiendo influir de forma determinante en el proceso de recepción y/o decodificación del mensaje. Esta sería una definición cercana al modelo teórico de comunicación propuesto por (Shannon, 1948). Esta perspectiva es la abordada en la mayoría de procedimientos en los que se compara una señal consigo misma una vez ha sido transmitida por un canal de comunicación. El modelo ha demostrado sobradamente su validez desde el punto de vista de la comunicación entre máquinas o de la comunicación entre humanos en la que media algún sistema tecnológico. Es también de aplicabilidad en la comunicación directa entre humanos en muchos casos: en una conversación normal, el nivel de presión sonora ambiental se considera ruido de fondo y este nivel afectará a la inteligibilidad o no del mensaje pudiéndolo hacer incluso imperceptible o ininteligible (son numerosos los ejemplos de la vida cotidiana que se pueden ajustar a esta descripción: una conversación a escasos metros de un martillo neumático puede resultar imposible de mantener).

En el ámbito de la comunicación no hablada, el modelo también es aplicable, por ejemplo, la mala calidad de una impresión o una serie de borrones en un texto manuscrito puede contribuir a una degradación cualitativa del mensaje original e interferir en su recepción e inteligibilidad. Desde este punto de vista, el término ruido se ha asociado tradicionalmente – ahora en el ámbito de la acústica y el sonido – a eventos sonoros a priori no deseados, que producen una perturbación en el proceso de comunicación o constituyen una molestia para los actores de la comunicación: el transmisor y el receptor. Así, de una forma más simple, haciendo una definición subjetiva, podríamos decir que acústicamente el ruido es cualquier sonido molesto. Esta definición, aunque podría generar a priori bastante consenso, es muy subjetiva, dado que la molestia no es una cualidad del sonido *per se* o de un evento sonoro concreto. Un mismo evento sonoro podrá resultar molesto o no según las circunstancias y los actores de la comunicación.

Al intentar establecer un vínculo entre los conceptos de calidad y de ruido, abordándolo desde la perspectiva explicada en el párrafo anterior, lo primero que parece lógico pensar es que la mayor calidad se tendrá en la ausencia total de ruido y, a medida que este incrementa se producirá una degradación progresiva de la calidad. Esto es cierto para las metodologías y métricas centradas en el proceso de evaluación del canal de comunicación estrictamente. Pero en el ámbito de la comunicación humana existen muchos más factores que influyen en el proceso comunicativo, además del canal. Se trata de elementos cognitivos, psicológicos, culturales y fisiológicos que contribuirán a determinar en qué medida un evento sonoro es considerado como ruido o no, y en el caso de ser considerado como ruido, qué grado de molestia causa. Como todos estos factores son variables culturalmente y entre individuos, parece ser que una forma de aproximarse

a la evaluación del ruido es mediante valoraciones estadísticas utilizando procedimientos propios de la psicoacústica más clásica. Así, se pueden evaluar diferentes eventos sonoros y establecer una comparación con sus atributos psicoacústicos clásicos como la sonoridad, la fluctuación, la agudeza, la rugosidad o la tonalidad – para una definición en profundidad de estos parámetros psicoacústicos ver (Zwicker & Fastl, 1990) -.

Una técnica de aproximación es crear un modelo a partir de estos atributos mediante el cual se pueda establecer el grado de molestia o perturbación (según su sonoridad, rugosidad, agudeza, etc...). Posteriormente, se han de correlar las descripciones realizadas por el modelo con encuestas psicoacústicas para determinar su validez (Marquis-Favre, Premat, Aubree, & Vallet, 2005a) (Marquis-Favre, Premat, & Aubree, 2005b). En resumen, el concepto de molestia es el que toma mayor preponderancia al hablar de ruido en el contexto de calidad sonora (Guski, Felscher-Suhr, & Schuemer, 1999).

Otro contexto en el que se pueden vincular los conceptos de calidad y ruido es en el ámbito de la exposición al ruido en entornos laborales y los riesgos derivados de la exposición prolongada a altos niveles de presión sonora. En este contexto, la calidad hace referencia a las técnicas y procedimientos encaminados a la disminución de la exposición y del riesgo.

#### *2.2.6 Calidad sonora de un producto*

“La idoneidad de un sonido en el contexto de una tarea u objetivo técnico específico” es como define (Blauert & Jekosh, 2012) el concepto de calidad sonora del producto. Ya se ha visto que este concepto puede tener un marco de definición mucho más amplio pero aquí está aplicada en lo que concierne a productos de consumo. Esta perspectiva, como ya se ha adelantado en el apartado 1.1, considera el sonido producido por un producto como un elemento más de los atributos comunicativos del producto. Todos los productos que producen un ruido tienen su propia calidad de sonido, que se evalúa cada vez que son usados. En contraste con la perspectiva anterior, de análisis de la molestia causada por el ruido, este nuevo enfoque permite relacionar el sonido producido por el producto (que bajo otro paradigma en ocasiones podría considerarse ruido) con las expectativas del usuario del producto y la experiencia de su uso.

Se considera el sonido emitido por el producto como una fuente de información sobre las características y/o el correcto funcionamiento de éste. Bajo este punto de vista, el objetivo no es únicamente minimizar la molestia del sonido del producto, sino que éste sea comunicativamente útil, como lo puede ser el diseño visual del producto o la propia imagen de marca. Así, las características psicoacústicas que veíamos en el caso anterior deben extenderse para tener en cuenta factores cognitivos y emocionales del usuario. No basta con definir un sonido como más o menos rugoso, por ejemplo, sino estudiar cómo puede influir la mayor o menor rugosidad del sonido en la percepción del producto por parte del usuario.

Ejemplos de aplicación de estas técnicas los tenemos en el diseño del sonido de producto aplicado a los vehículos, especialmente los de alta gama, así como también está muy extendido su uso a los electrodomésticos. No es suficiente que un electrodoméstico sea poco ruidoso. El ruido debe transmitir al usuario atributos como fiabilidad, buen funcionamiento, idoneidad. En otras ocasiones incluso un bajo nivel de ruido emitido por un producto puede ser contraproducente en términos de calidad sonora del producto. Por ejemplo, cuando un usuario tiene expectativas de atributos como potencia respecto de una motocicleta, por ejemplo, un nivel de ruido emitido por el vehículo excesivamente bajo puede contribuir al incumplimiento de esa expectativa. En definitiva, se trata, a grandes rasgos, de diseñar cuidadosamente el sonido que generará el producto teniendo en cuenta las expectativas de los usuarios potenciales para los que el producto está diseñado, de manera que el sonido contribuya a cumplir estas expectativas y afianzarlas a través de la experiencia de uso.

### **2.3 Algunas consideraciones sobre la calidad**

De todo lo expuesto hasta aquí se puede deducir la complejidad implícita que lleva el término calidad en el ámbito acústico. Se pueden distinguir varios aspectos que son recurrentes a la hora de abordar el concepto.

El primero es el acotamiento y restricción del término según el problema concreto que se esté abordando. Hemos visto que la perspectiva adoptada según el objeto de estudio evaluado es totalmente diferente. Esto está estrechamente vinculado con las necesidades específicas surgidas en los orígenes de cada ámbito. Las necesidades de la telefonía de principios de siglo pasado originaron los métodos de cálculo y medición de la distorsión, que posteriormente se adoptaron a otros canales de transmisión. La irrupción de los sistemas de almacenamiento y reproducción de audio y la necesidad de su estudio a partir de ese momento hizo que estas técnicas se adoptaran para estos sistemas y que se fueran ampliando para adaptarse a la medición de los nuevos dispositivos a medida que estos iban experimentando una mejora tecnológica. El estudio de los recintos, por su parte, empezó abordando el problema desde un punto de vista estrictamente temporal. Esta perspectiva inicial fue prácticamente la única disponible durante casi medio siglo y condiciona fuertemente las posteriores aproximaciones que se realizan para la evaluación de la calidad.

El segundo es el fuerte grado de dependencia que se establece entre las necesidades específicas de un problema muy concreto y la posterior generalización de la solución adoptada para ese problema a un ámbito más amplio. Ya hemos mencionado el caso de la incipiente telefonía. En las décadas de los 70 y 80, surgió el término *Hi-Fi*, alta fidelidad, en referencia a los sistemas de almacenamiento y reproducción de sonido con mayor ancho de banda y una considerablemente menor distorsión que sus predecesores. Esto supuso un nuevo incentivo para la búsqueda e implementación de nuevos sistemas de evaluación, utilizando señales y métodos de análisis mucho más complejos que los iniciales, como veremos en el capítulo 3. La irrupción a finales del XX e inicios del XXI de la telefonía móvil y de los nuevos sistemas de transmisión de contenidos generará un

nuevo marco de necesidades al que se dará respuesta con nuevas metodologías. Como ya se ha mencionado, los sistemas de transmisión digitales minimizan el problema del canal, que había sido el principal quebradero de cabeza de los métodos de evaluación de la calidad, de manera que el foco de atención se traslada a los extremos del sistema de comunicación. Se adoptará una perspectiva desde el punto de vista del receptor, siendo el usuario final del servicio o del producto el que evalúa la calidad. Evidentemente, esto sólo puede hacerse preguntando directamente al usuario, técnica que se realiza en el ámbito de la valoración cualitativa de la experiencia de usuario, pero que resulta poco realista desde un punto de vista de ingeniería. Los recursos disponibles son pues, las metodologías psicoacústicas basadas en amplios test de percepción, como veremos en el siguiente epígrafe o una modelización de ciertos parámetros y resultados obtenidos originalmente mediante técnicas propias de la psicoacústica. Esto conllevará la aparición del modelado perceptual y el desarrollo de modelos que intentan emular el comportamiento del sistema auditivo humano y abordar el tradicional problema de la comparación de señales (original - transmitida) incluyendo ahora un tercer invitado: la percepción (original – transmitida – percibida). Hemos visto que estos sistemas, al menos en su faceta monoaural, están ya plenamente establecidos y estandarizados.

El tercer aspecto es el establecimiento de interrelaciones y dependencias entre los diferentes ámbitos de abordaje del problema. Técnicas surgidas en un principio para la resolución de un problema concreto, se extenderán, no sólo a otros problemas dentro de su mismo ámbito como se menciona anteriormente, sino también a otros ámbitos de aplicación. Por ejemplo, las técnicas de medición de respuesta al impulso, de aplicación en la electroacústica y el procesado electrónico, constituirán una auténtica revolución en la acústica de salas, incorporando procedimientos de medida y evaluación que aumentarán de forma muy significativa las posibilidades de análisis. La propia métrica STI experimentó una migración desde su concepción original basada en inteligibilidad en sistemas de transmisión para convertirse finalmente en un estándar en la evaluación de la inteligibilidad de la palabra en salas. Parece recurrente que al establecerse un método de evaluación de calidad del habla, por ejemplo, en un contexto específico, éste método o alguna variante fundamentada en éste intento aplicarse a otro ámbito donde el habla también tiene lugar. Lo mismo podría decirse del análisis del ruido o de la señal de audio.

Finalmente, como cuarta característica, remarcar que se ha establecido una especie de tradición histórica que diferencia las técnicas y los resultados de éstas entre objetivas y subjetivas. Generalmente se establece un modelo teórico con un mayor o menor nivel de abstracción que es el que intenta explicar los fenómenos físicos que tienen lugar en el objeto de estudio. A estos métodos y procedimientos se les denomina objetivos. Más frecuentemente de lo deseado ocurre que la descripción aportada desde esta perspectiva “objetiva” no puede explicar las descripciones aportadas por los individuos receptores de ese fenómeno, por ejemplo, un evento acústico. Es un aspecto fundamental en todo lo concerniente al término calidad. La calidad, en última instancia, sólo tiene sentido si

hay un observador o usuario del evento o del sistema. Este observador o usuario, mediante su percepción y sus funciones cognitivas elaborará un diagnóstico del fenómeno o evento. Al final, posiblemente, el único denominador común que tienen todas las técnicas, metodologías y procedimientos vistos en las páginas precedentes es el sujeto receptor, llámese oyente, espectador, observador o usuario. De manera que, para intentar establecer un vínculo entre los resultados ofrecidos por los modelos teóricos y los aportados o descritos por los observadores surgen los denominados parámetros o métricas subjetivas, normalmente obtenidos o derivados a partir de procedimientos psicoacústicos.

Se podría decir que en la evaluación de la calidad acústica, como en tantos otros ámbitos científico-tecnológicos ha predominado una visión dualista del mundo, de clara tradición epistemológica cartesiana: el modelo explica el mundo, pero la percepción de éste por nuestra parte es limitada debido a las imperfecciones de nuestros sistemas perceptivos. Bajo este paradigma epistemológico, cualquier fenómeno perceptivo es sólo una imagen o remembranza de un supuesto mundo real que nos es inaccesible a través de los sentidos. La psicoacústica, dado que sólo dispone como objeto de estudio de estas “imágenes” de la realidad ve limitadas sus posibilidades de análisis. Esta visión del mundo ha sido duramente criticada en (Dominicus, 2009) (Blauert & Guski, 2009) (Blauert, 2012) (Blauert & Dominicus, 2013), que bajo un paradigma empirista, o “perceptualista” – como lo define el propio Blauert – defiende que los modelos teóricos en última instancia también están surgidos y generados a partir de la percepción y únicamente difieren de cualquier evento perceptivo en su nivel de abstracción. En realidad, el germen de la psicoacústica, la psicofísica fechneriana de la segunda mitad del siglo XIX, nació como respuesta al dualismo, como han señalado (Tortosa & Civera, 2006):

Para superar el materialismo intentó demostrar la identidad de lo psíquico y lo físico (mente y cuerpo), una demostración que le permitiera argumentar que la conciencia, lo único de lo que tenemos idea directa e inmediata, constituye la verdadera y única realidad, no siendo el resto de los objetos y procesos físicos otra cosa que ficciones. Toda la fuerza del argumento dependía de la posibilidad de demostrar científicamente aquella identidad, y para ello buscó una solución por vía experimental.

## 2.4 Psicoacústica

Desde que Gustav Theodor Fechner sentara las bases de la psicofísica, la psicología de la percepción inició su curso como disciplina científica y, lógicamente, también el de la percepción auditiva. Durante menos de dos siglos de evolución, esta disciplina se ha considerado desde diferentes paradigmas epistemológicos, todos ellos influenciados por los cambios históricos y filosóficos que les fueron contemporáneos.

#### 2.4.1 Una visión histórica

Generalmente, se atribuye el nacimiento de la psicología científica a una especie de emancipación del estudio de la mente respecto de la filosofía. El papel desarrollado por los físicos y fisiólogos desde mediados del siglo XIX resulta fundamental en este proceso de emancipación. Algunos autores consideran que tal emancipación no surge de forma espontánea sino que puede interpretarse como una evolución surgida de la necesidad de dar el relevo al límite impuesto por René Descartes al estudio de la mente, la dicotomía mente-cuerpo, la división entre *res cogitans* y *res extensa* (Morabito & Della Roca, 2010). El rechazo de Fechner al dualismo cartesiano queda evidenciado en *Elemente der Psychophysic*, obra en la que define la interdependencia entre las dimensiones físicas y psíquicas del ser humano (Fechner, 1860). El dualismo cartesiano plantea el desafío más importante a la investigación de Fechner, un desafío cuya solución habría desembocado en la fundación de la psicología como disciplina científica. El problema podría sintetizarse en la siguiente pregunta: ¿Cómo podría una ciencia de la mente poner al alcance de la *res extensa* la *res cogitans*? es decir, ¿cómo podría ser observada y estudiada – la mente – con metodologías experimentales? Para poder encontrar una solución, habría que limitarse a establecer las relaciones funcionales entre los fenómenos físicos y los psíquicos.

En este punto, la formación experimental de Fechner resultó determinante y a ello contribuyeron también notablemente los hallazgos y métodos de Weber respecto a la percepción de las sensaciones de peso. Algunos autores sostienen que los desarrollos experimentales de Fechner eran un aspecto de su trabajo subordinado al que consideran su principal objetivo: sus postulados filosóficos antimaterialistas, con lo cual estos desarrollos pudieran surgir de la necesidad de fundamentar de forma empírica su teoría filosófica (Cagigas, 2001). Sea así o no, lo irrefutable es que la influencia ejercida por Fechner desde el último tercio del siglo XIX ha sido decisiva en el ámbito de la psicofísica y la psicología, posibilitando la incorporación de técnicas experimentales y estadísticas al estudio de fenómenos psíquicos. La trascendencia de su obra en este ámbito ha sido fundamental. Las implicaciones filosóficas de ésta, en cambio, no tuvieron esa repercusión.

Siendo extremadamente sintéticos, se podría decir que el planteamiento de Fechner no propone implementar un método de medición directa de las sensaciones sino de establecer una relación entre éstas y los fenómenos físicos que las desencadenan. Para ello, se centrará en la medición de las variaciones en las sensaciones en función de las variaciones de energía que las provocan. Partirá de tres supuestos (Tortosa & Civera, 2006):

- a) Un incremento en la sensación, dado que es mínimo y equivale al umbral diferencial, tiene valor de unidad y constituye una ‘mínima diferencia perceptible’.
- b) Así, una determinada sensación constituirá una suma de ‘mínimas diferencias perceptibles’.

- c) Los estímulos, esto es, la cantidad de energía para generar una sensación han de incrementarse en aumentos proporcionales.

La ley de Fechner establece que:

$$S = k \log I \quad (2.2)$$

La sensación  $S$  es igual al logaritmo del estímulo (multiplicada por una constante  $k$  que se ajusta experimentalmente según el tipo de sensación objeto de medida). Es decir los estímulos crecen en progresión geométrica mientras que las sensaciones lo hacen siguiendo una progresión aritmética.

Los tres principales fundamentos de la psicofísica fechneriana esbozados anteriormente, esto es, la concepción de la ley logarítmica como una ley psicofísica, la cuestión de la proporcionalidad indirecta así como la propia concepción de la posibilidad de medición de lo mental fueron objeto de muchas y en ocasiones feroces críticas. Algunos filósofos como Julius von Kries o Hermann Cohen directamente descartaban la posibilidad de medición de cualquier magnitud psicológica. Por otra parte, fisiólogos y psicólogos como Ewald Hering – conocido por sus experimentos sobre la percepción del color, entre otros – centraron sus ataques en la cuestión de la proporcionalidad indirecta fundamentándolos principalmente en la defensa de la proporcionalidad directa establecida por Johann Herbart unos años antes del alumbramiento de la ley logarítmica. Una visión intermedia sería la aportada por el filósofo y psicólogo también alemán Franz Brentano. De hecho, algunos autores (Seron, 2011) consideran que las aportaciones de Brentano cristalizarían un siglo más tarde de forma implícita en la propuesta de Stanley Smith Stevens –psicólogo americano fundador del laboratorio de psicoacústica de Harvard - y su conocida ley de potencias. Las objeciones al modelo de Fechner y su propuesta de ley de potencias pueden verse en (Stevens, 1961). Stevens está considerado como el psicofísico que mayor influencia ejerció en este ámbito científico durante la segunda mitad del siglo XX. Su aproximación, aunque fundamentalmente empirista en cuanto basada en la experimentación, se aleja bastante de los ideales monistas que inspiraron a Fechner. Así, Stevens tiene una concepción mecanicista de los sistemas perceptivos humanos, los concibe como una caja negra sobre la que analiza los estímulos de entrada y los resultados de la salida. Su propia definición sobre la psicofísica resulta bastante clarificadora en este sentido (Stevens, 1975):

La psicofísica estudia la respuesta de un organismo, no una suerte de cosa mental no física que, por definición, desafía cualquier prueba objetiva. (...)

(...) La medición es de especial interés en la psicofísica, no sólo en el sentido estricto de término, sino de la psicofísica en el más antiguo y amplio espíritu de ésta, que intenta descubrir reglas que relacionen las

respuestas de los organismos a las configuraciones energéticas del entorno. (...) <sup>2</sup>

Él mismo definió su propio paradigma epistémico, que denominó *schemapiric*, y que se fundamentaba en la definición de un sistema formal creado a través del análisis de los datos empíricos obtenidos a partir de experimentos que abordaban la dimensión neurofisiológica del individuo. Puede pensarse que en el fondo de esta concepción subyace una visión del mundo dualista (Morabito & Della Roca, 2010). Desde este punto de vista, la percepción humana es la respuesta de un organismo a un estímulo exterior. Los datos obtenidos son la respuesta del sistema, que una vez analizados, permitirán modelar el comportamiento del sistema (*schema*) a partir de esas respuestas obtenidas mediante experimentación (*empiric*). Finalmente, para ilustrar hasta qué punto Stevens eludía tener en cuenta cualquier otra consideración de carácter cognitivo en el proceso perceptivo se reproduce otra cita del autor (Stevens, 1966):

(...) Cuando estudiamos las características de entrada/salida de un amperímetro, no nos sentimos llamados a imaginar cómo se siente ser amperímetro, ni tampoco tratamos de relacionar nuestras propias experiencias a las de los amperímetros. En el estudio científico del hombre, y especialmente en el estudio de las características de sus sistemas sensoriales, muchos pseudo-problemas pueden ser obviados si tomamos la misma actitud hacia un participante humano en un experimento que la que tomamos con el amperímetro (...) <sup>3</sup>

Ya en el siglo XXI, (Neuhoff, 2004) propone una psicoacústica fundamentada en una aproximación ecológica, es decir, abordando todas las relaciones que se producen entre el individuo y el medio del que forma parte. El planteamiento teórico de la psicoacústica ecológica pretende variar el foco de atención sobre el objeto de estudio. Así, establece que la mediación entre estímulo y sensación no debe ser el objeto central a analizar sino una puerta de entrada para el análisis de procesos sensoriales y cognitivos de orden más alto. Entiende cada elemento del proceso auditivo como una parte de una cadena más compleja de acciones cognitivas, neurobiológicas y comportamentales. Neuhoff plantea que los métodos habituales de la psicoacústica – identificación, reconocimiento, escalado de similitudes, categorizaciones, etc... – frecuentemente se limitan a focalizar la atención en lo que él mismo denomina “comportamiento auditivo”.

Aun reconociendo la necesidad del conocimiento acerca de este comportamiento auditivo, plantea que el objeto de estudio no debe quedarse ahí, sino que este conocimiento es el primer eslabón para abordar un estudio que permita comprender los procesos de alto orden que se establecen cuando una persona oye un determinado sonido. Entiende la percepción como la relación que se establece entre la información disponible

---

<sup>2</sup> Trad. del original

<sup>3</sup> Trad. del original

en el entorno y las capacidades, sensibilidades e intereses del perceptor. En resumen, limita el modelo Stevens al primer paso que puede darse para abordar una psicoacústica que considere al sujeto perceptor inmerso en su entorno, con lo cual, los factores cognitivos y culturales resultan tan determinantes, si no más, que la estricta percepción fisiológica.

#### 2.4.2 Un modelo para describir la experimentación psicoacústica

Dejando a un lado los diferentes paradigmas epistemológicos que han intentado determinar tanto los métodos como el objeto de estudio de la psicoacústica, la aproximación de la experimentación realizada por ésta puede sintetizarse en un modelo como el propuesto por (Pulkki & Karjalainen, 2015), que se esquematiza en la Figura 2.3 y que explicaremos a continuación brevemente.

La primera distinción que debe hacerse es entre evento sonoro y evento auditivo. El evento sonoro forma parte del entorno. Fechner diría que es un estímulo energético. Stevens lo clasificaría como una configuración energética del entorno del organismo, mientras que Neuhoff podría definirlo como información presente en el ecosistema. El rasgo común es que el evento sonoro es un estímulo que entra en el sistema auditivo del sujeto bajo estudio. Además del evento sonoro, al sistema auditivo también le llega ruido. El ruido procedente del exterior puede entenderse como ruido acústico, bajo un punto de vista estrictamente auditivo.

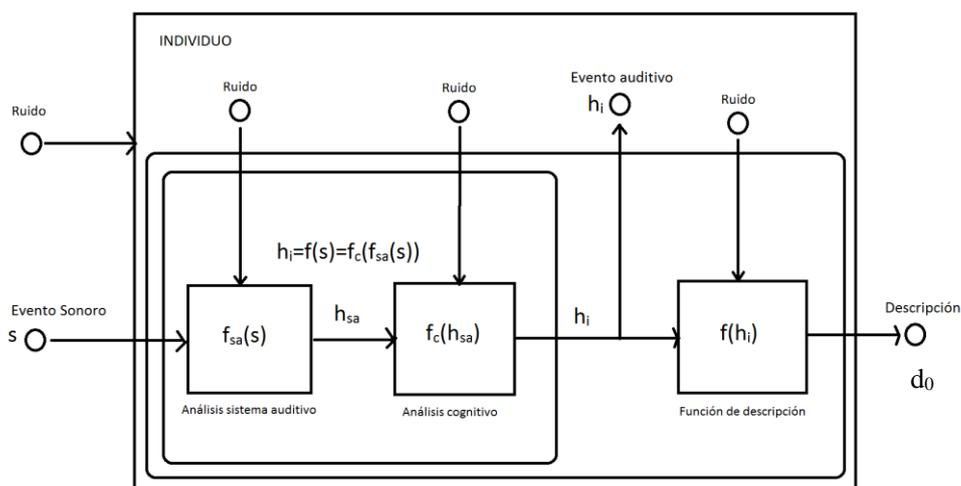


Figura 2.3. Representación gráfica de un proceso experimental psicoacústico. Adaptado de (Pulkki & Karjalainen, 2015)

El evento sonoro  $S$  genera en el individuo un evento auditivo  $h_i$ , que es únicamente observable por éste mediante introspección. Para un observador externo, el evento

auditivo no es observable. La conformación del evento auditivo viene determinada por el análisis realizado por el sistema auditivo fisiológico  $f_{sa}(s)$ , que genera en el sistema auditivo un estímulo  $h_{sa}$  que no es accesible ni para el propio individuo ni, obviamente, para un observador externo. El estímulo  $h_{sa}$  se somete a las funciones cognitivas del individuo  $f_c(h_{sa})$  originando finalmente el evento auditivo  $h_i$ . El objetivo clásico de los procedimientos psicoacústicos ha sido el de establecer la relación entre el evento sonoro y el evento auditivo, es decir, obtener la función  $f_c(f_{sa}(s))$ . Pero existe un obstáculo más que salvar. El evento auditivo no es directamente observable por un observador externo. Para que lo sea debe sufrir una transformación más, dada por la función de descripción  $f(h_i)$ , que es la que determina la caracterización del evento auditivo expresada por el sujeto. Esta descripción  $d_0 = f(h_i)$  será la información que finalmente tendrá disponible el observador externo.

Dado que lo que se busca es la relación  $f_c(f_{sa}(s))$  la manera en que se minimiza el efecto de esta última transformación dada por la función de descripción es mediante la utilización de procedimientos estadísticos sobre experimentos realizados a un número determinado de individuos. En lo que respecta al resto de fuentes de ruido mostradas en la Figura 2.3, la que influye sobre el sistema auditivo hace referencia a posibles problemas fisiológicos que éste pueda tener, la que actúa sobre las funciones cognitivas puede relacionarse con diferentes estados de concentración o actitudes por parte del sujeto bajo experimentación, mientras que la fuente de ruido de la función de descripción hace referencia a las diferentes maneras en que los sujetos pueden expresar el evento auditivo y que, como ya se ha mencionado, se intenta minimizar mediante técnicas estadísticas. Este esquema es válido para los métodos y procedimientos que a lo largo del último siglo han contribuido a determinar parámetros de la audición humana, como diferentes tipos de umbrales, fenómenos de enmascaramiento, caracterizaciones y descriptores de eventos sonoros, etc.



# 3 Calidad en transductores

## **3.1 El concepto de calidad en los transductores acústicos**

Históricamente, al hablar de calidad en un transductor acústico o en un dispositivo de audio se ha hecho referencia a la evaluación de la distorsión tanto lineal como no lineal introducida por el dispositivo, de forma que cuanto menor sea ésta se ha considerado que el equipo analizado tenía una mayor 'calidad'. De esta forma, al hablar de evaluación de calidad se estaba hablando de evaluación de la distorsión. Los indicadores de distorsión (el clásico THD – distorsión armónica total -, la distorsión de intermodulación y las diferentes métricas objetivas conducentes a la obtención de una representación gráfica o un valor numérico) pretenden ser una herramienta que determina de un rápido vistazo el grado de fidelidad del dispositivo con respecto a la señal de entrada. Sin embargo, estos métodos han tenido que enfrentarse continuamente a un duro hándicap: la baja

correlación existente entre las magnitudes obtenidas en estas métricas y la calidad sonora percibida debida a la audibilidad de la distorsión no lineal.

Si bien en un principio sería más propio hablar de distorsión y omitir el término calidad, es cuando se intenta establecer una correspondencia entre la magnitud física obtenida para medir la distorsión de un dispositivo y la valoración perceptiva que se hace de este dispositivo cuando aparece el término calidad. De hecho, los trabajos pioneros en la tarea de encontrar esta correlación grafaban THD versus ‘*perceived sound quality*’ (Shorter, 1950) (Hilliard, 1941). La posterior introducción de los métodos perceptuales ha contribuido a incrementar el uso del término calidad, que les es más propio, dado el objetivo primordial para el que fueron desarrollados (la evaluación perceptiva de la señal de voz o de la señal de audio para sistemas de compresión de baja tasa binaria en transmisiones digitales). Es pues en ese anhelado nexo de unión entre la medida objetiva y la experiencia perceptiva donde cobra sentido hablar de indicadores de calidad.

La primera métrica aplicada para la evaluación de la distorsión no lineal de dispositivos de audio – la medida de la distorsión armónica total o THD - data de los tiempos de la incipiente electroacústica, en los inicios del siglo pasado y se toma prestada de los procedimientos para medir equipos de radiocomunicaciones (Czewrinski & Voishvillo, 2001). Este dato no sería relevante de no ser porque esta métrica sigue siendo hoy en día de uso común y estandarizado en el ámbito de los transductores electroacústicos y no se ha logrado plenamente su sustitución por otra – al menos con un uso tan extendido -. Sin embargo, a lo largo de estos casi 100 años de industria se ha producido todo el conocimiento, herramientas, ingeniería y desarrollo de la electroacústica. En el caso concreto de los altavoces, el grado de conocimiento alcanzado en el comportamiento de cada uno de los componentes del sistema y su aportación a la distorsión no lineal del dispositivo tiene su máximo exponente en los trabajos de Klippel, basados en el modelado de los transductores aplicando la dinámica de sistemas no lineales (Klippel, 1996).

A pesar de todo esto, sigue pendiente la tarea de encontrar y estandarizar una métrica que dé cuenta inequívoca de la calidad sonora que se le supondrá a un dispositivo según el valor obtenido en la métrica. En palabras de A. Voishvillo (Voishvillo A. , 2006):

*El objetivo final de la medida de distorsión no lineal de un altavoz sería la obtención de datos que proporcionen información sobre la no linealidad, de forma que esta información pueda relacionarse de forma inequívoca con la calidad sonora percibida del dispositivo evaluado, de manera que pudieran compararse de forma objetiva diferentes dispositivos. Aún la aparente simplicidad de estos objetivos y de los numerosos esfuerzos realizados por muchos investigadores durante 90 años, estos objetivos nunca han sido alcanzados.*

La búsqueda de métodos objetivos de medida de la distorsión no lineal en equipos de sonido ha existido siempre pero la investigación de la audibilidad de la distorsión no

lineal en sistemas de sonido no ha producido buenas correlaciones entre los parámetros tradicionales de medida de la distorsión y la percepción de ésta.

### **3.2 Paradigmas de aproximación a la evaluación de la distorsión**

La evaluación de la distorsión no lineal en transductores electroacústicos se ha realizado históricamente desde tres perspectivas diferentes (Voishvillo A. , 2006): medidas de distorsión, métodos de identificación y métodos perceptuales.

#### *3.2.1 Medidas de distorsión*

Entenderemos como medidas de distorsión todas aquellas metodologías encaminadas a obtener síntomas de no linealidad del dispositivo bajo análisis utilizando una señal de test conocida y analizando la salida del sistema a ese estímulo. Esta perspectiva es la que se ha adoptado en la mayor parte de propuestas realizadas para la evaluación de la distorsión no lineal. La evolución en esta línea de trabajo se ha producido a medida que se ha ido disponiendo de herramientas para la construcción de señales de test más complejas y también a medida que ha ido aumentando la capacidad de análisis de la respuesta al sistema. De esta forma, desde la simple senoide utilizada para la medida de la THD, pasando por el uso de dos tonos modulados, se llega al uso de señales multitono, señales de *burst* con determinados factores de forma o señales de ruido pseudoaleatorio.

En lo que respecta al análisis de la respuesta del sistema, desde el simple recuento de armónicos (THD), pasando por el análisis de los productos de intermodulación, se llega a la manipulación de la ingente cantidad de productos de modulación generada por excitaciones multitono o el uso de la función de coherencia, por ejemplo. A medida que han ido evolucionando estas técnicas se han ido realizando intentos de establecer una relación entre los resultados obtenidos por estos métodos (que podríamos llamar *medidas objetivas*) y la valoración subjetiva realizada de los dispositivos generalmente mediante test de audición como los mencionados en el apartado 1.2.1. No todos los métodos de medida de distorsión han intentado su correlación cualitativa particular, algunos no han abordado este aspecto, pero en términos generales, si bien las correlaciones han ido mejorando a medida que se introducían nuevos procedimientos, se puede concluir que los métodos convencionales de medida de distorsión no lineal utilizando barridos en frecuencia producen información con una correlación sólo parcial con la calidad sonora percibida. Volviendo a citar a Voishvillo:

*Subestimar la naturaleza compleja y multidimensional de la no linealidad y la intrincación de la respuesta del sistema auditivo humano a una señal distorsionada conduce generalmente a conclusiones erróneas. Un error típico en el análisis y la evaluación de la no linealidad en audio es la confusión entre los síntomas de no linealidad obtenidos a través de la aplicación de determinadas señales de test (generalmente sinusoidales) y la reacción del sistema a señales no estacionarias.*

Es decir, el principal problema estriba en que frecuentemente, para estimar la calidad sonora esperada de un determinado dispositivo al reproducir una señal musical, por ejemplo, se han considerado como datos de referencia los obtenidos mediante señales de test sinusoidales y estacionarias distorsionadas, produciéndose una generalización de esta respuesta concreta a la respuesta del sistema no lineal frente a una señal de entrada arbitraria. Si además añadimos el intento de vincular al dato obtenido una estimación de la calidad sonora percibida, nos encontraremos con otro gran problema: el sistema auditivo humano, es decir, el único instrumento real del que disponemos para la evaluación de la calidad, es también un sistema no lineal. Esto aumenta aún más si cabe el problema pero al mismo tiempo abre una nueva perspectiva para abordarlo: los métodos perceptuales, que, como veremos a más adelante, tratan el problema desde el punto de vista del sujeto evaluador de la calidad, sin entrar demasiado en las consideraciones propias del sistema generador de la distorsión no lineal.

### 3.2.2 *Métodos de identificación*

Aunque los conceptos matemáticos que describen la dinámica de sistemas no lineales se desarrollaron a principios del siglo XX, sus aplicaciones en ingeniería no empezaron hasta mediados de la década de los 40 del mismo siglo (Schetzen, 1989). En las décadas de los 50 y 60 se desarrolló la teoría de los sistemas dinámicos no lineales basada en los trabajos de Volterra y Wiener. Las primeras aplicaciones de estas teorías en el análisis de altavoces datan de mediados de los 80 (Kaizer, 1986), mientras que el progreso más significativo tiene lugar al iniciar la década de los 90, con los trabajos de Klippel. Las aproximaciones desde esta perspectiva han posibilitado la comprensión de los mecanismos implícitos en las no linealidades de los dispositivos (Klippel W. , 2007), estableciendo relaciones entre la distorsión no lineal generada por un dispositivo y las no linealidades de éste, es decir, las causas físicas que producen la distorsión. Desde esta perspectiva se han hecho aún más evidentes las carencias de los métodos convencionales de medida de distorsión: conociendo los mecanismos implicados en la generación de la distorsión resulta más fácil determinar si la distorsión generada al inyectar una señal de test concreta puede generalizarse al sistema o no y en qué condiciones (Klippel W. , 1990). En los siguientes apartados de esta memoria se hará una breve introducción de algunos de estos conceptos, que están fuera del ámbito central de este trabajo, dado que desde esta perspectiva no es aplicable una evaluación de la calidad desde el punto de vista de un modelo perceptual. Por otra parte, desde estas metodologías no se ha abordado el aspecto perceptivo.

### 3.2.3 *Métodos perceptuales.*

Son los métodos mencionados en el apartado 2.2.2 (ver Figura 2.1 y Figura 2.2). Como respuesta a la eclosión de los sistemas de transmisión digitales y la telefonía móvil han emergido durante las últimas décadas una serie de herramientas para la evaluación objetiva de sistemas de compresión de audio y señales de voz (Beaton & Beerends, 1996). Para los sistemas de transmisión de audio digital o de voz por telefonía, en principio, la mejor manera de obtener una evaluación de la calidad sonora sería una

evaluación estrictamente subjetiva. Y la mejor manera de llevar a cabo esta función sería mediante la realización de test de evaluación subjetiva cuidadosamente diseñados (Watanabe, 2008). Lógicamente, este procedimiento no es viable para funciones de monitorización de un servicio – función indispensable en cualquier servicio de transmisión – y tampoco resulta cómodo para funciones de diseño o implementación, dado el elevado coste en tiempo y dinero que suponen los experimentos de audición. Dada esta necesidad surgen los métodos de evaluación de calidad basados en modelos del sistema auditivo humano. Estos métodos perceptuales posibilitan la determinación de cuáles de los componentes de una señal objetivamente distorsionada son psicoacústicamente audibles y cuáles no. En las técnicas de codificación perceptual utilizadas para transmisión digital – vía radio o IP – la señal se comprime con bajas tasas binarias empleando modelos auditivos. En la implementación de estos modelos se consideran aspectos del sistema auditivo humano como la percepción del nivel o intensidad, la percepción de la frecuencia, así como los efectos de enmascaramiento tanto temporal como en frecuencia. Los métodos PESQ y PEAQ han sido estandarizados (para voz en telefonía y señal de audio) como sistemas de estimación de la calidad percibida, como ya se ha indicado. A partir de éstos, principalmente en la década de los 90, se desarrollaron innumerables modelos de evaluación perceptiva de la calidad cuyo resultado más popular son los códecs de compresión de audio (mp3, aac, wma, atrac, etc...). Estos métodos no se han utilizado para la evaluación de transductores electroacústicos, cuya no linealidad es muy diferente de la de los sistemas de compresión, pero han sido la base de una serie de trabajos encaminados a la evaluación de la percepción de distorsiones no lineales producidas por no linealidades artificiales o por transductores reales. Uno de los principales exponentes de estas metodologías son las métricas DS o  $R_{\text{nonlin}}$ , propuestas por Moore y Tan (Moore, Tan, Zacharov, & Ville-Veikko, 2004).

### **3.3 Sistemas lineales**

Se consideran lineales aquellos sistemas cuyas salidas dependen linealmente de sus entradas. La propiedad de linealidad se basa en dos principios: superposición y homogeneidad. Un sistema obedece al principio de superposición si las salidas resultantes de diferentes entradas pueden sumarse. Por ejemplo, si la salida  $y_1(t)$  corresponde a la entrada  $x_1(t)$  y la salida  $y_2(t)$  corresponde a la entrada  $x_2(t)$ , en el caso de tener como entrada al sistema la señal  $x(t) = x_1(t) + x_2(t)$  se obtendrá una salida  $y(t) = y_1(t) + y_2(t)$ . Un sistema obedece al principio de homogeneidad si la salida resultante al inyectar una señal de entrada  $x(t)$  multiplicada por un escalar resulta también multiplicada por ese mismo factor de escala. Si  $y(t)$  es la salida para una entrada  $x(t)$ , al aplicar al sistema la entrada  $a \cdot x(t)$  (siendo  $a$  una constante) tendremos una salida  $a \cdot y(t)$ .

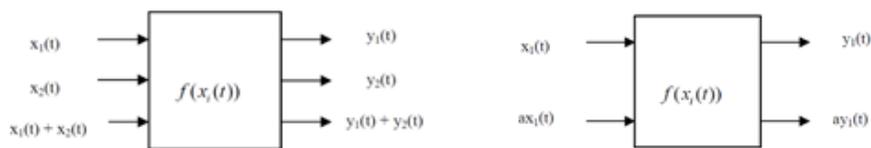
Esto viene a significar que si aplicamos sumas de señales de entrada de tipo

$$x(t) = \sum_{i=1}^P a_i x_i(t) \quad (3.1)$$

Obtendremos salidas como

$$y(t) = \sum_{i=1}^P a_i y_i(t) \quad (3.2)$$

para las que  $y(t) = f(x_i(t))$

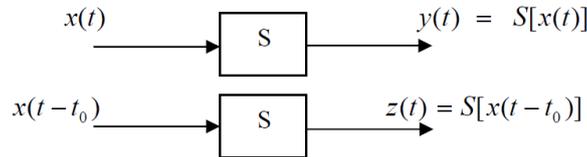


**Figura 3.1. Principio de superposición y homogeneidad**

La salida de un sistema lineal nunca contiene componentes de frecuencia que no estuvieran presentes en la señal de entrada. De esta forma, una entrada sinusoidal tendrá como salida otra senoide de la misma frecuencia, aunque que la fase y la amplitud sí pueden verse modificadas por el sistema, en lo que se denominan distorsiones lineales. Esto no es necesariamente cierto para otros tipos de señales de entrada. Por ejemplo, si en un sistema lineal inyectamos una señal cuadrada la salida no es necesariamente otra señal cuadrada. Este es uno de los factores que explican la popularidad histórica de las sinusoides en la investigación de la audición, entre otros (Zwicker & Fastl, 1990).

Un sistema S:  $S[x(t)]$  es invariante con el tiempo ( también denominado estacionario) si las propiedades o características del sistema no varían con el tiempo. Si después de desplazar la salida  $y(t)$ , el resultado  $y(t-t_0)$  equivale a  $S[x(t-t_0)]=z(t)$  el sistema es invariante con el tiempo.

Los sistemas que además de lineales son invariantes en el tiempo se denominan sistemas LTI. La mayor parte de las propiedades de un sistema LTI se deben a que el sistema puede representarse por ecuaciones en diferencias. Estas propiedades incluyen: respuesta al impulso, convolución, estabilidad, escalado, etc... (Ogunfunmi, 2007)



**Figura 3.2. Propiedad de invariabilidad temporal de un sistema S**

Una de las principales características de un sistema LTI es que puede caracterizarse mediante su respuesta al impulso. Dado que es una herramienta ampliamente utilizada en la acústica de salas para la obtención de parámetros de calidad de esta, en el capítulo correspondiente a la evaluación de salas se hará una breve exposición del concepto de respuesta al impulso.

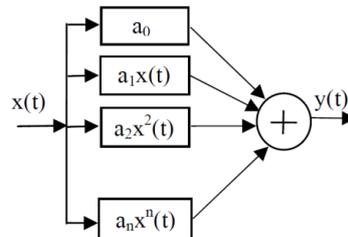
### 3.4 Sistemas no lineales

Cuando nos encontramos con términos como no linealidad, sistema no lineal o distorsión no lineal en literatura concerniente al audio frecuentemente se nos representa una imagen o gráfico en la que el eje X muestra el nivel de entrada de señal mientras que el eje Y corresponde al nivel de salida de señal. La relación entre ambas suele ser polinómica o una función discontinua (característica de efectos como la saturación de señal – ver Figura 3.7, Figura 3.9). Este tipo de gráficas únicamente describen la forma más simple de no linealidad estática y sin memoria, caracterizada por la dependencia instantánea de la señal de salida respecto de la señal de entrada. En otras palabras, el nivel de la señal de salida en un momento determinado es una función no lineal del nivel de la señal de entrada en exactamente el mismo momento. El hecho de que un sistema como este carezca de memoria implica que tanto las características lineales como no lineales del sistema serán independientes de la frecuencia. Para estos casos, la relación entrada-salida del sistema puede expresarse como:

$$f(x) = \sum_{n=0}^N a_n x^n \quad (3.3)$$

En la que el primer elemento  $a_0$  es el término de continua, el segundo término  $a_1x$  representa la ganancia del sistema y pertenece, al igual que  $a_0$ , a la parte lineal de la expansión. El tercer término  $a_2x^2$  es la no linealidad de segundo orden o cuadrática, mientras que el término  $a_3x^3$  es la no linealidad de tercer orden o cúbica. La expansión polinómica constituye el método más sencillo para caracterizar un sistema no lineal aunque, evidentemente, pocos sistemas reales pueden ser modelados con una simple expansión polinómica. Si el nivel de los productos de distorsión en la salida no es alto, es decir, el dispositivo a modelar presenta una no linealidad débil, éste se puede describir con una serie de Volterra (Franz & Scholkopf, 2006). Las series de Volterra pueden considerarse como series de Taylor con memoria o como una extensión del concepto de respuesta al impulso y función de transferencia de un sistema lineal a un espacio

multidimensional (Schetzen, 1989). En un sistema de este tipo, no sólo la componente lineal tiene memoria sino que también la tienen los componentes no lineales de diferentes órdenes, lo que hace que la modelización resulte matemáticamente bastante compleja.



**Figura 3.3. Sistema no lineal estático**

La señal de salida de un sistema como el de la Figura 3.3 es el resultado de elevar a una potencia por cada orden de no linealidad la señal de entrada, multiplicar estos términos por unos coeficientes constantes que ponderan la presencia de productos no lineales de diferentes órdenes en la señal de salida y finalmente, sumar todas estas aportaciones. En cambio, en un sistema con no linealidad débil modelado por una serie de Volterra la elevación a la potencia se sustituye por integrales de convolución de forma que los diferentes órdenes de no linealidad se representan por las correspondientes integrales de convolución multidimensionales. La siguiente ecuación representa la relación entre la entrada y la salida de un sistema no lineal modelado por una serie de Volterra.

$$\begin{aligned}
 y(t) = & \int_{-\infty}^{\infty} h_1(\tau_1) x(t - \tau_1) d\tau_1 \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2) d\tau_1 d\tau_2 \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) x(t - \tau_1) x(t - \tau_2) x(t \\
 & - \tau_3) d\tau_1 d\tau_2 d\tau_3 \dots + \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \tau_2, \dots, \tau_n) x(t \\
 & - \tau_1) x(t - \tau_2) \dots x(t - \tau_n) d\tau_1 d\tau_2 \dots d\tau_n
 \end{aligned} \tag{3.4}$$

En la ecuación 3.4, se cumple que  $h_n(\tau_1 \dots \tau_n) = 0$  para cualquier  $\tau < 0$ . La parte lineal está representada por la respuesta al impulso  $h_1(t)$ . De hecho, el primer término de la serie es la definición de convolución para un sistema LTI. Los términos no lineales están representados por sus correspondientes respuestas al impulso multidimensionales  $h_2(t)$   $h_3(t)$ ...  $h_n(t)$ , que se denominan 'kernels' o núcleos de Volterra. Para transformar estas respuestas al impulso multidimensionales al dominio frecuencial pueden utilizarse transformadas de Fourier múltiples. De esta forma, la respuesta en frecuencia de segundo orden correspondiente a  $h_2(t)$  tendría el aspecto de un gráfico tridimensional con dos ejes horizontales representando ambos frecuencias positivas y negativas (Schetzen, 1989).

Las respuestas en frecuencia de orden superior no tienen representación gráfica dada su multidimensionalidad.

Una notación alternativa a la ecuación 3.4 es la propuesta en la ecuación 3.5

$$y(t) = \mathbf{H}_1[x(t)] + \mathbf{H}_2[x(t)] + \mathbf{H}_3[x(t)] + \dots + \mathbf{H}_n[x(t)] + \dots \quad (3.5)$$

donde cada término

$$\mathbf{H}_n[x(t)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) x(t - \tau_1) \dots x(t - \tau_n) d\tau_1 \dots d\tau_n$$

En esta notación, el término  $\mathbf{H}_n$ , que representa la integral correspondiente a cada orden, se denomina operador de Volterra de orden n-ésimo. La Figura 3.4 es una representación esquemática de la ecuación 3.5 como la suma de las salidas  $\mathbf{H}_n[x(t)]$  producidas por el operador correspondiente a cada orden.

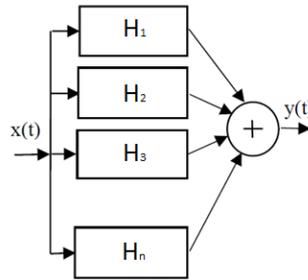


Figura 3.4. Sistema no lineal dinámico

La serie de Volterra expresada en la ecuación 3.4 es una serie funcional, es decir, una operación sobre una función en la que el resultado es un valor numérico. Para una determinada función de entrada  $x(t)$ , la integración para un instante de tiempo particular  $t_0$  de cada uno de los términos de la serie de la ecuación 3.4 tiene como resultado un valor numérico. Por otra parte, desde el punto de vista expresado en la ecuación 3.5 cada operador  $\mathbf{H}_n$  genera una función  $y_n = \mathbf{H}_n[x(t)]$  a partir de la función de entrada  $x(t)$  de forma que se focaliza la atención en la obtención de la función completa de salida.

La reacción de un sistema dinámico débilmente no lineal, que podría modelarse mediante una serie de Volterra, es pues, significativamente más compleja que la de un sistema estático. Además, los sistemas fuertemente no lineales (como podría ser el desplazamiento de la bobina de un *woofer* con niveles de entrada muy altos) presentan un comportamiento no lineal que no puede describirse por una serie de Volterra, dado que ésta diverge en presencia de fuertes no linealidades (Rugh, 1981). Para el caso concreto de transductores acústicos pueden encontrarse caracterizaciones de sistemas de

radiación directa utilizando esta técnica en (Kaizer, 1986) así como también para sistemas de radiación indirecta en (Klippel W., 1996).

### 3.5 Ejemplos de distorsión lineal y no lineal

Los métodos clásicos y más sencillos de medida de la distorsión centran su atención en los resultados obtenidos al inyectar en un sistema una señal sinusoidal o una combinación de éstas. Una de las principales razones del uso de funciones sinusoidales para estas tareas ha sido expuesta más arriba: si en un sistema LTI se inyecta una senoide, la salida de éste será también una senoide de la misma frecuencia a la cual el sistema puede haberle modificado la amplitud, la fase o ambas. En un sistema lineal no aparecerán en la salida componentes de frecuencia que no estuvieran presentes en la señal de entrada. Como también se ha adelantado ya, esto no ocurre con sistemas no lineales, en los cuales a la salida aparecerán componentes de frecuencia no presentes en la entrada. La aparición de estos componentes de frecuencia nuevos sugiere que la medida de éstos podría dar una representación del grado de distorsión introducido por el sistema. De hecho, en este principio se basan las medidas de distorsión armónica total y las de intermodulación, entre otras. Como sus propios nombres indican, se trata de cuantificar la aparición de armónicos en un caso y de productos de intermodulación en el otro.

En este apartado se presentan unos sencillos ejemplos, adaptados de (de Santis & Henin, 2007), que ayudan a clarificar estas ideas. Se muestran tres sistemas, uno lineal y dos con no linealidad estática, es decir, sin memoria. Se representan mediante funciones de transferencia de la amplitud de la señal en el tiempo, es decir, los valores en cada instante de la salida dependen exclusivamente de los valores en la entrada en ese mismo instante. La salida nunca dependerá de un valor anterior. Se somete a estos sistemas a una señal de entrada discreta  $x(n) = \sin(2\pi 1000nT)$ , con una frecuencia de muestreo de 44100Hz. A continuación se representará la señal de salida obtenida para cada sistema tanto en el dominio temporal como en la frecuencia.

#### 3.5.1 Distorsión lineal

Sea un sistema S cuya relación entrada-salida de la amplitud de la señal en el tiempo viene dado por la función de la gráfica mostrada en la Figura 3.5. La simple inspección visual de la función de transferencia ya nos advierte de que se trata de un sistema lineal, cuya relación entrada-salida viene dada por:  $y(t) = 0.5x$

La señal de salida del sistema con distorsión lineal de la Figura 3.6 tiene la misma fase que la señal original, de la cual difiere sólo en amplitud. El cambio en amplitud queda también reflejado en el dominio de la frecuencia como se muestra en la tercera gráfica. Como es de esperar en un sistema lineal, el espectro de frecuencia en la salida contiene el mismo componente que la señal original. La frecuencia original se denomina también fundamental

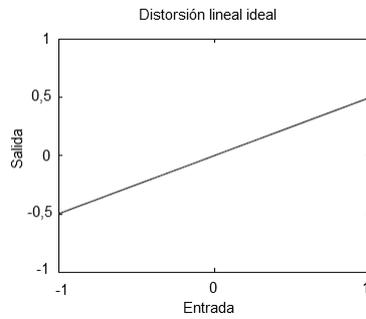


Figura 3.5. Función de transferencia (entrada – salida) de un sistema lineal sin memoria.

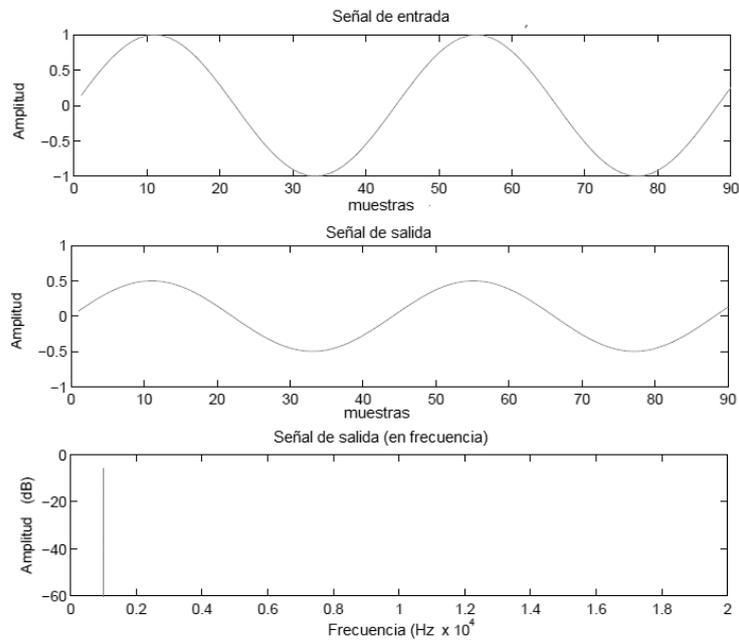
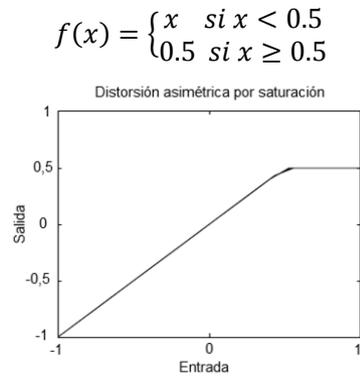


Figura 3.6. Señal de entrada, de salida y espectro de la señal de salida del sistema de la Figura 3.5

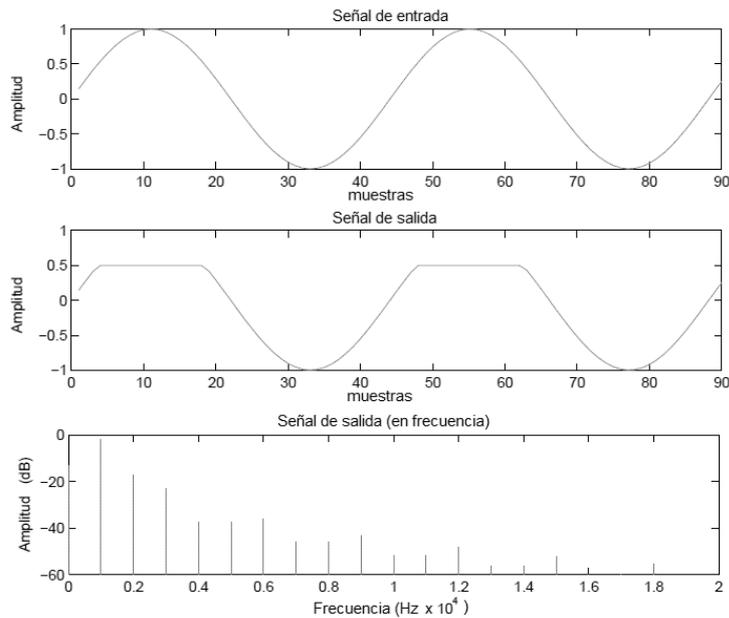
### 3.5.2 Distorsión no lineal asimétrica

En este caso se presenta un ejemplo de un sistema con saturación de señal en la parte positiva del margen dinámico. Si bien la relación entrada-salida en sistemas no lineales sin memoria suele ser polinómica, en los casos de saturación de señal las funciones que vinculan la entrada con la salida pueden ser discontinuas o definidas por tramos, como en este caso. La función que describe el comportamiento del sistema es:



**Figura 3.7. Función de transferencia entrada-salida de un sistema no lineal asimétrico sin memoria**

La señal de salida se muestra en la siguiente figura: tiene el mismo valor que la entrada  $x(t)$  en los ciclos negativos y en los ciclos positivos está saturada. A partir de un valor de amplitud superior a 0,5 en la entrada, la salida se mantiene constante en 0,5.



**Figura 3.8. Señal de entrada, de salida y espectro de la señal de salida del sistema de la Figura 3.7**

Como muestra la última gráfica de la Figura 3.8, este tipo de distorsión inyecta armónicos de orden par e impar a lo largo del espectro que decrecen a medida que va

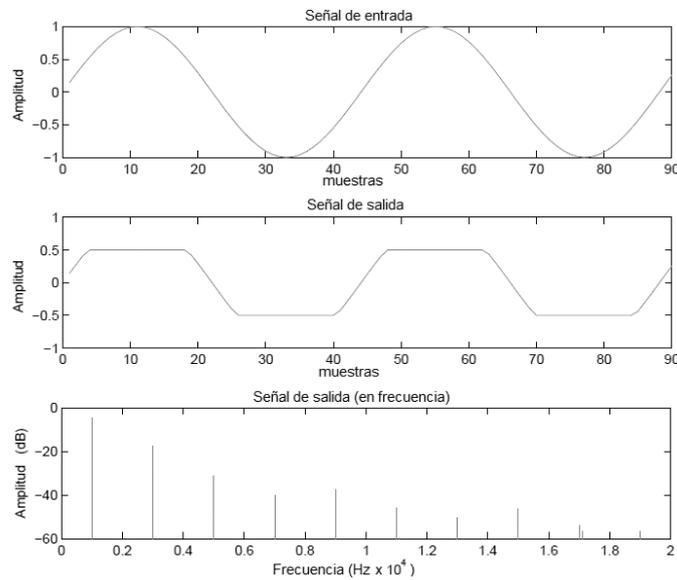
umentando la frecuencia. Los armónicos pares aparecen en los múltiplos pares de la fundamental,  $2f, 4f, 6f, 8f\dots$ , mientras que los armónicos de orden impar aparecen en los múltiplos impares de la frecuencia fundamental  $3f, 5f, 7f, 9f\dots$ . La presencia en la salida de componentes de frecuencia no presentes en la entrada es una característica del comportamiento de un sistema no lineal.

### 3.5.3 Distorsión no lineal simétrica

Veamos ahora otro sistema no lineal sin memoria en el que la saturación se produce tanto en el ciclo positivo como en el negativo.



**Figura 3.9. Función de transferencia entrada-salida de un sistema no lineal simétrico sin memoria**



**Figura 3.10. Señal de entrada, de salida y espectro de la señal de salida del sistema de la Figura 3.9**

La función que proporciona la salida para una entrada  $x(t)$  es:

$$f(x) = \begin{cases} x & \text{si } -0.5 < x < 0.5 \\ 0.5 & \text{si } x \geq 0.5 \\ -0.5 & \text{si } x \leq -0.5 \end{cases}$$

La sobrecarga en dispositivos de estado sólido presenta habitualmente este tipo de distorsión no lineal. Las crestas de la señal de entrada se recortan tanto en los ciclos positivos como negativos a los valores límite del margen dinámico del dispositivo. Como puede observarse en la Figura 3.10, a diferencia de la distorsión asimétrica, los componentes de frecuencia que aparecen en la salida tienen lugar sólo en los múltiplos impares de la frecuencia fundamental.

#### 3.5.4 Productos de intermodulación

En los dos últimos ejemplos, la distorsión no lineal producida se limita a la aparición de armónicos en la salida del sistema. Esto se debe a que la señal inyectada en el sistema es un tono puro (con una sola componente en frecuencia). A medida que vayamos aumentando la complejidad de la señal de entrada al sistema, aumentará también la complejidad para el análisis de los productos de distorsión generados. Cuando en la señal original están presentes varias componentes en frecuencia, aparecerán, además de los armónicos, los denominados productos de intermodulación, que son productos de distorsión que aparecen en la salida en frecuencias suma y diferencia de las fundamentales presentes en la señal original. Por ejemplo, si tenemos un sistema no lineal sin memoria cuya relación entrada-salida viene dada por la expresión polinómica  $f(x)$ .

$$f(x) = a_1x + a_3x^3$$

Y en la entrada del sistema se inyecta la señal  $x(t)$ , que, como muestra la siguiente ecuación, tiene dos componentes frecuenciales ( $\omega_1$  y  $\omega_2$ )

$$x(t) = b \sin(\omega_1 t) + c \sin(\omega_2 t)$$

La salida vendrá dada por la siguiente expresión:

$$f(x) = \begin{cases} a_1(b \sin(\omega_1 t) + c \sin(\omega_2 t)) + \\ a_3 \left( \left( \frac{3b^3}{4} + \frac{3bc^2}{2} \right) \sin(\omega_1 t) + \left( \frac{3c^3}{4} + \frac{3b^2c}{2} \right) \sin(\omega_2 t) \right) - \\ \frac{a^3}{4} (b^3 \sin(3\omega_1 t) + c^3 \sin(3\omega_2 t)) - \\ \frac{3a_3 b^2 c}{4} (\sin((\omega_2 - 2\omega_1)t) + \sin((2\omega_1 + \omega_2)t)) - \\ \frac{3a_3 b c^2}{4} (\sin((\omega_1 - 2\omega_2)t) + \sin((2\omega_2 + \omega_1)t)) \end{cases}$$

La señal de salida obtenida en la ecuación anterior puede describirse de la siguiente manera: la primera línea representa el término lineal, que es simplemente una ganancia determinada por el término  $a_1$ . La segunda línea es el producto de distorsión de primer orden y la tercera línea representa el armónico de tercer orden. Las últimas dos líneas muestran los productos de intermodulación resultantes de la interacción entre las dos frecuencias presentes en la señal de entrada. Como veremos a continuación, los productos de intermodulación tendrán una gran importancia en procedimientos de medida de la distorsión no lineal alternativos a la distorsión armónica total, que únicamente considera los armónicos.

### 3.6 Evolución histórica de las medidas de distorsión

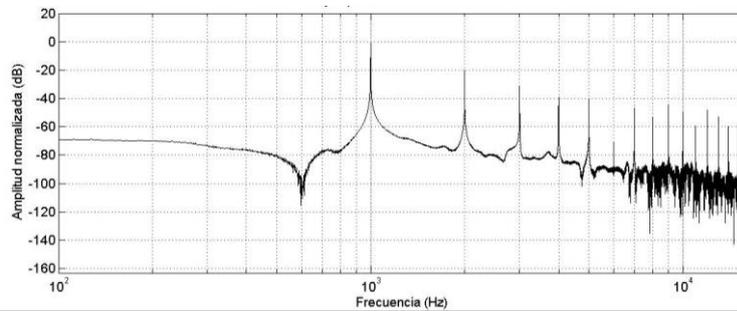
En general, la historia de los métodos de evaluación de las no linealidades en equipos de audio puede dividirse en diferentes etapas. Una primera etapa en los inicios de la industria basada fundamentalmente en las medidas de distorsión armónica total (THD), distorsión armónica total más ruido y medidas de distorsión por intermodulación de dos tonos. La precisión de los test estaba limitada por las modestas posibilidades de los equipamientos de test de la época.

#### 3.6.1 Distorsión armónica total

Para medir la distorsión armónica se hace referencia a la llamada distorsión armónica total, que es precisamente la cantidad de armónicos que el equipo introduce y que no estaban en la señal original. Si en un sistema no lineal introducimos un tono de frecuencia  $f_1$ , en la salida tendremos ese mismo tono (con una amplitud y fase posiblemente diferentes) y sumados a él otros tonos de frecuencia  $2f_1, 3f_1 \dots$  llamados armónicos del tono fundamental  $f_1$ . Para normalizar las medidas se introduce un tono de 1 KHz y se mide la señal de salida. La distorsión armónica total se da en forma de porcentaje. El porcentaje representa la parte del total de la energía a la salida que pertenece a los armónicos, es decir, qué porcentaje es distorsión. Los fabricantes de equipos suelen facilitar este dato. El THD se puede expresar en función de la de la potencia o de la tensión.

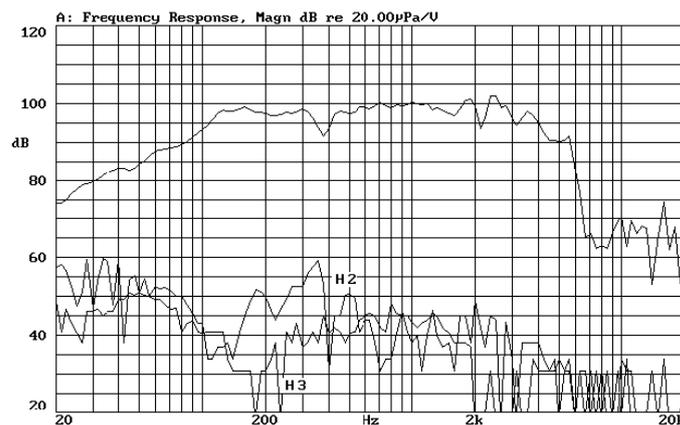
$$\begin{aligned} THD &= 100 \frac{\sum_{i=2}^N P_i}{\sum_{i=1}^N P_i} = 100 \frac{P_2 + P_3 + \dots + P_N}{P_1 + P_2 + \dots + P_N} \\ &= 100 \frac{\sqrt{V_1^2 + V_2^2 + \dots + V_N^2}}{V_T} \end{aligned} \quad (3.6)$$

Donde  $P_i$  es la potencia del armónico  $i$ -ésimo con  $i > 1$  y  $P_1$  es la potencia del tono fundamental de frecuencia 1 KHz. A continuación se muestra una representación del espectro de salida de un aparato con distorsión armónica al introducir un tono de 1 KHz.



**Figura 3.11. Distorsión armónica de un dispositivo en respuesta a una señal de entrada de 1kHz.**

Si intentáramos representar la distorsión armónica para todas las frecuencias de la respuesta de un altavoz de la misma forma que en la figura, tendríamos que hacer muchos gráficos para cada frecuencia de interés o se haría un gráfico muy confuso. Es mucho más cómodo representar el nivel de cada armónico de interés para cada frecuencia como una línea continua por debajo de la curva de presión sonora emitida por el altavoz, tal y como aparece en la siguiente figura. En esta vemos dos curvas, denominadas H2 y H3, correspondientes al nivel de amplitud del segundo y tercer armónico. Estos armónicos son el segundo y tercer múltiplo de la frecuencia fundamental. Se incluyen estos dos armónicos en los catálogos de los fabricantes porque suelen ser los que el altavoz reproduce con mayor amplitud. De esta forma, podemos observar el nivel de ambos armónicos para cada frecuencia.



**Figura 3.12. Representación de la distorsión armónica introducida por los dos primeros armónicos en función de la frecuencia.**

### 3.6.2 Distorsión armónica total más ruido

En ocasiones se proporciona una medida denominada THD+N (distorsión armónica total más ruido) en lugar de la THD. El método para su obtención está basado en el mismo procedimiento que la THD pero este parámetro incluye cualquier otro ruido que esté presente en el sistema. El numerador de la ecuación 3.6 se determina eliminando la fundamental de la salida del sistema mediante un filtro *notch*. De esta forma, la tensión rms de esta señal incluye la contribución de los armónicos y de cualquier otro ruido. El denominador de la ecuación 3.6 será el nivel rms de toda la señal, incluyendo fundamental, armónicos y ruido.

### 3.6.3 Distorsión de intermodulación

La distorsión de intermodulación es debida a que varias frecuencias pertenecientes a una señal de entrada interactúan dentro del dispositivo no lineal generando unas terceras no deseadas. Estas frecuencias adicionales que aparecen en la salida se conocen con el nombre de productos de intermodulación (ver página 52). Considerando que a la entrada del dispositivo no lineal tenemos dos tonos de frecuencia  $f_1$  y  $f_2$ , a su salida se producirán productos de intermodulación a frecuencias suma y diferencia de múltiplos de las frecuencias originales, es decir:

$$f_{out} = mf_1 \pm nf_2$$

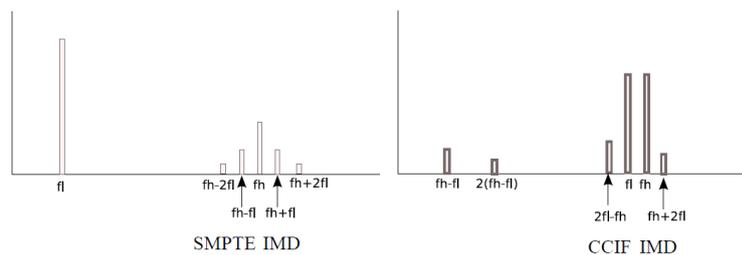
Donde m y n son números enteros mayores o iguales que la unidad. El orden de la distorsión no lineal queda determinado por la suma de m y n. En los casos particulares de  $m=0$  o  $n=0$ , a la salida del dispositivo se obtienen armónicos a frecuencias múltiplos de las frecuencias de entrada:  $f_{out} = mf_1$  ;  $f_{out} = nf_2$ . Este tipo de distorsión se conoce con el nombre de distorsión armónica y ha sido explicada en el apartado anterior. El orden de un determinado producto de intermodulación es la suma de los enteros m y n. La siguiente tabla muestra algunos ejemplos de productos de intermodulación y su orden.

**Tabla 3.1. Orden de productos de intermodulación.**

Producto	Orden
$f_2-f_1$	2
$f_1+f_2$	2
$2f_1-f_2$	3
$f_1-2f_2$	3
$3f_1-f_2$	4
$3f_1+2f_2$	5

Existieron dos métodos ampliamente difundidos para la obtención de medidas de distorsión por intermodulación de dos tonos. El primero de ellos estaba basado en la aplicación de dos tonos muy cercanos en frecuencia y en la medida de diferencias del producto de intermodulación de segundo orden. Este método, conocido como CCIF (*Comité Consultatif International Téléphonique*) tuvo gran popularidad en Europa en primer lugar y posteriormente en Estados Unidos al ser introducido por Scott (Voishvillo A. , 2006). El método CCIF generalmente utilizaba un tono de 3 kHz y otro de 3.05 kHz

de igual amplitud y medía el producto de intermodulación de segundo orden en la diferencia  $f_2-f_1$ . Otro método, concebido inicialmente para medir la distorsión en equipos destinados a salas de proyección cinematográfica fue desarrollado por (Hilliard, 1941) y se conoce como método SMPTE. El método utiliza un tono de baja frecuencia y otro tono de alta frecuencia y menor nivel de amplitud. De esta forma, se produce una modulación del tono de alta frecuencia por el de baja. El método de Hilliard utilizaba generalmente un tono de 60 Hz y otro de 3 kHz, siendo la amplitud del tono de 3 kHz 12 dB inferior respecto de la amplitud del tono de 60 Hz.



Fuente: (de Santis & Henin, 2007)

**Figura 3.13. Medidas de distorsión de intermodulación**

Actualmente, el método más extendido para la medición de la IMD es el SMPTE o el DIN, cuyos estándares RP120-1983 y 45403 son similares. Ambos establecen una señal de test formada por dos tonos: uno de baja frecuencia y mayor amplitud y otro de alta frecuencia cuya amplitud es de  $\frac{1}{4}$  de la amplitud de la señal de baja frecuencia (-12.04 dB). En el caso del estándar americano estas señales son 60 Hz y 7 kHz, mientras que el estándar alemán ofrece más posibilidades de elección en baja y alta frecuencia. Un par muy usado en el estándar DIN es 250 Hz y 8kHz. (Metzler, 1993). Cuando estas señales se inyectan en un sistema no lineal, los productos de intermodulación aparecerán como una serie de bandas laterales alrededor del tono de alta frecuencia. El espacio – en hertzios - entre el tono de alta frecuencia y el primer par de bandas laterales (productos de segundo orden,  $f_2 \pm f_1$ ) es equivalente al tono de baja frecuencia. El segundo par de bandas laterales (productos de tercer orden,  $f_2 \pm 2f_1$ ) estará separado del tono de alta frecuencia dos veces el tono de baja frecuencia. El porcentaje de distorsión de intermodulación se define como el porcentaje de modulación de amplitud que estas bandas laterales representan sobre la portadora de alta frecuencia (el tono  $f_2$ ).

#### 3.6.4 Ponderación de armónicos

Uno de los primeros intentos de mejorar la correlación entre la audibilidad de la distorsión y los datos obtenidos a través de la medida de la distorsión armónica fue introducido en (Shorter, 1950). Shorter puso especial atención al rol desempeñado por los armónicos de orden alto en la percepción subjetiva de la distorsión, llegando a la conclusión de que los sistemas no lineales en cuya salida había presencia de armónicos

de mayor orden sonaban peor que aquellos que generaban armónicos de menor orden teniendo un valor THD similar. Shorter sugirió que los armónicos de mayor orden eran más perceptibles y propuso una ponderación de tal forma que cuanto más alto era el orden del armónico éste recibía un coeficiente de ponderación también más alto.

El experimento de Shorter consistió en la medida de la distorsión armónica y la realización de test de escucha en 6 sistemas que tenían diferentes distribuciones de armónicos y diferentes suma RMS de armónicos. Para el test de escucha estableció cuatro valoraciones posibles: *not perceptible*, *just perceptible*, *perceptible*, *bad*. El test de escucha se realizó sobre una grabación de piano. Como resultado del test de escucha, a cada sistema se le asigna una valoración subjetiva intentando establecer una correlación entre ésta y la suma RMS de armónicos. Los resultados sin realizar ninguna ponderación se muestran en la siguiente tabla.

**Tabla 3.2. Valoración subjetiva vs. Suma RMS de armónicos**

Sistema	Clasificación subjetiva	Suma RMS de armónicos
D	Bad	3.7%
B	Perceptible	3.3%
C	Just perceptible	2.6%
A	Bad	2.3%
E	Just perceptible	0.6%
F	Not perceptible	0.4%

Fuente: (Shorter, 1950)

De la tabla anterior se obtiene una pobre correlación entre la distorsión medida y la evaluación subjetiva. Shorter utilizó dos métodos para ponderar los armónicos de orden alto. Uno de ellos era el recomendado por la *Radio Manufacturers Association* (RMA) y consistía en la multiplicación de cada armónico por el coeficiente  $n/2$ , donde n es el orden del armónico. La aplicación de este coeficiente de ponderación consigue una mejor correlación.

**Tabla 3.3. Valoración subjetiva vs. Suma RMS ponderada  $n/2$**

Sistema	Clasificación subjetiva	Suma RMS ponderada (por $n/2$ )
D	Bad	6.7%
A	Bad	5.1%
B	Perceptible	5.1%
C	Just perceptible	2.8%
E	Just perceptible	1.3%
F	Not perceptible	0.8%

Fuente: (Shorter, 1950)

A pesar de la mejora en la correlación se da el caso de un valor idéntico (5.1%) para dos sistemas clasificados con diferente resultado en el test de escucha, lo cual sugiere al autor la utilización de un coeficiente de ponderación  $n^2/4$ , obteniendo así los resultados de la tabla siguiente, que muestran una escala más precisa.

**Tabla 3.4. Valoración subjetiva vs. Suma RMS ponderada  $n^2/4$**

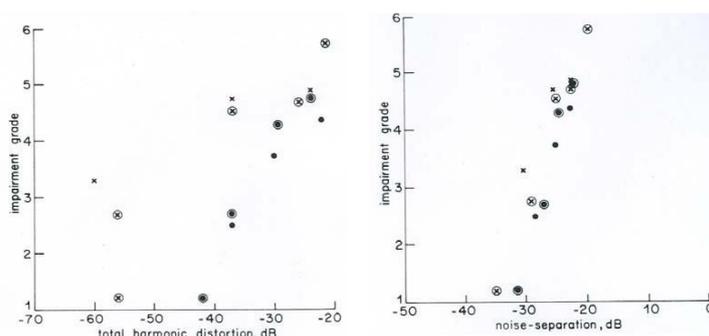
Sistema	Clasificación subjetiva	Suma RMS ponderada (por $n^2/4$ )
A	Bad	19.4%
D	Bad	16.5%
B	Perceptible	8.6%
E	Just perceptible	4.5%
C	Just perceptible	3.3%
F	Not perceptible	2.2%

Fuente: (Shorter, 1950)

Shorter llegó a la razonable conclusión de que la medida de la THD sin detallar su contenido no tenía ninguna utilidad. Los resultados obtenidos por Shorter parecen ser una prueba de la relevancia de los productos de distorsión de orden alto. A pesar de que el autor consideró también los productos de intermodulación llegando a mencionar que su número puede ser significativo en presencia de no linealidades de orden alto, parece subestimar el rol dominante desempeñado por los productos de intermodulación en comparación con los armónicos. Por ejemplo, para el mismo nivel de entrada de una señal multitono el número de productos de intermodulación en un sistema caracterizado por no linealidades de orden alto es considerablemente muy superior que el número de productos de intermodulación en un sistema caracterizado por no linealidades de orden bajo. A pesar de esto, el número de armónicos en ambos sistemas será similar (Czewrwinski & Voishvillo, 2001).

### 3.6.5 Excitación multitono

A partir de la década de los 70 gana interés el uso de señales de test más complejas. Concretamente, autores como Belcher, desde la BBC, abogaron por el uso de señales multitono. En (Belcher, 1978), el autor prueba experimentalmente que las señales multitono producen una mejor correlación entre los datos generados por las medidas y la percepción subjetiva de la distorsión.

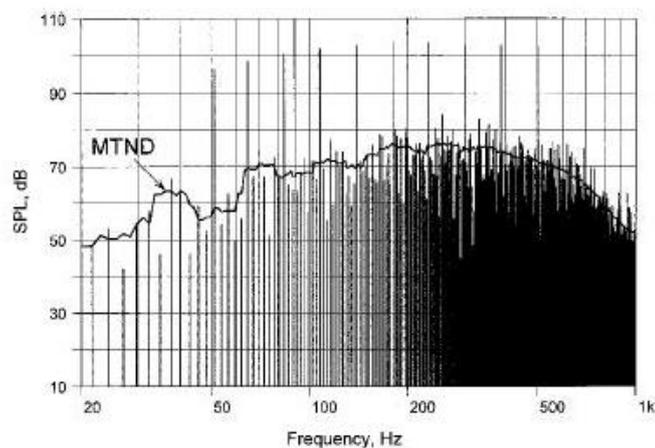


Fuente: (Belcher, 1978)

**Figura 3.14. Evaluaciones objetivo-subjetivas realizadas en (Belcher, 1978)**

Utilizó una medida objetiva de la distorsión con la señal multitono que denominó *noise separation* definida como la relación entre el nivel de pico de todos los productos de distorsión y el nivel de pico de la señal más los productos de distorsión. Además utilizó una escala de degradación subjetiva de seis valores (desde ‘imperceptible’ hasta ‘inservible’). La Figura 3.14 ilustra la comparación de las evaluaciones objetivas-subjetivas correspondientes a la distorsión armónica total y a la métrica propuesta por Belcher.

Como medida objetiva de no linealidad, el uso de una señal de excitación multitono tiene ventajas significativas respecto al uso de tonos puros o *sweeps*. Además de ser más rápida, genera una variedad de armónicos y especialmente de productos de intermodulación que pueden ser clasificados por el orden, el nivel y la fase (Czerwinski & Voishvillo, 2001). Además, con el uso de manipulaciones matemáticas sobre la respuesta al estímulo multitono aplicado con diferentes niveles de entrada, pueden separarse las superposiciones de los componentes espectrales obtenidos. El principal inconveniente del uso de las técnicas con excitación multitono es la generación de un número excesivo de productos de intermodulación que hacen difícil comparar alguna o varias características pertenecientes a dos sistemas diferentes. Con la finalidad de evitar este inconveniente se propone en (Voishvillo, Terekov, & Czerwinski, 2004) una nueva medida, MTND (*multitone total nonlinear distortion coefficient*).



Fuente: (Voishvillo, Terekov, & Czerwinski, 2004)

**Figura 3.15. MTDN obtenida con excitación multitono.**

Esta métrica está basada en un promedio de todos los componentes espectrales armónicos y de intermodulación realizados con la utilización de una ventana que se desplaza en el dominio de la frecuencia, de manera que se obtiene una respuesta continua en función de la frecuencia. El nivel de esta curva de respuesta en una frecuencia concreta dependerá del nivel y la cantidad de componentes espectrales de distorsión

presentes en las cercanías de esa frecuencia. La figura inferior muestra el trazado de una curva MTND a partir de la respuesta en frecuencia a una señal de excitación multitono sobre un altavoz de radiación directa. La curva resultante puede tener un aspecto parecido a las trazadas para representar el THD. La principal diferencia estriba en que la curva MTND tiene en cuenta los productos de intermodulación. En trabajos posteriores no se ha intentado correlar los resultados de esta métrica con apreciaciones subjetivas sobre la calidad de dispositivos.

### 3.6.6 Umbrales de audibilidad. Detectabilidad a partir de modelos

Volviendo a los intentos de correlación entre medidas objetivas y percepción, otra aproximación al problema se ha realizado estableciendo un hipotético umbral de audibilidad de las no linealidades en (Schmitt, 1995). Las aportaciones de Schmitt se diferenciaron principalmente de sus predecesores en que no realizó medidas objetivas al uso para después correlarlas, sino que utilizó modelos no lineales dinámicos con los que simular los efectos físicos no lineales de los transductores mediante DSPs. Con el análisis de los umbrales de audibilidad de la distorsión obtenidos, Schmitt concluye con la existencia de una fuerte interrelación entre la señal de test utilizada y la percepción de la distorsión, si bien establece un umbral mínimo de audibilidad de la contaminación no lineal con señales musicales en dispositivos de radiación directa alrededor del 5% del nivel de pico de la señal y que ésta tiene un efecto negativo en la experiencia auditiva. En lo concerniente a altavoces de radiación indirecta con bocinas acopladas, el porcentaje anterior cae hasta un 2% (Schmitt, 1995).

### 3.6.7 Uso de la función de coherencia

Finalizando la década de los 80 del siglo pasado varios trabajos propusieron el uso de la función de coherencia para la evaluación de la distorsión en sistemas de grabación magnética (Totzek & Preis, 1987) y en sistemas de ayuda a la audición (Kates, 1992). En (Preiss & Gregg, 2000) se introduce la métrica SDR –*signal to distortion ratio*– que pretende ser un indicador de la distorsión presente en sistemas con no linealidades ligeras a moderadas y que operen con señales de banda ancha. Esta métrica está basada en la función de coherencia y el uso de ruido blanco como señal de test.

El módulo de la función de coherencia  $\gamma^2(f)$ , se define, en términos de tres espectros de potencia; la entrada  $P_{xx}(f)$ , la salida  $P_{yy}(f)$ , y el espectro cruzado entre entrada y salida  $P_{xy}(f)$  como:

$$\gamma^2(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)} \quad (3.7)$$

Donde  $f$  es la frecuencia en Hz. La cantidad de potencia coherente a la salida del sistema para una determinada frecuencia  $f$  será igual a  $\gamma^2(f)P_{xx}(f)$ , mientras que la potencia no coherente en la salida o potencia de distorsión para una frecuencia  $f$  vendrá dada por  $(1 - \gamma^2(f))P_{xx}(f)$ . Con lo cual, la relación señal-distorsión en función de la frecuencia será:

$$SDR(f) = \frac{\gamma^2(f)}{1 - \gamma^2(f)} \quad (3.8)$$

Si la coherencia es igual a la unidad, es decir,  $\gamma^2(f) = 1$ , entonces para esa frecuencia toda la potencia de salida es linealmente dependiente de la potencia de entrada y la ecuación 3.8 predecirá un SDR ideal. Cuando se tiene que  $0 \leq \gamma^2(f) < 1$ , el espectro de potencia de la salida contiene una determinada cantidad de potencia no correlada debido a las no linealidades del sistema. Este modelo se compara con la THD obteniendo resultados de entre 2 a 20 dB de empeoramiento de la relación señal-distorsión con el indicador SDR. La conclusión a que llegan los autores es que el modelo proporciona un indicador más ajustado a la realidad que la THD, al considerar todo el ancho de banda de operación del dispositivo. El principal inconveniente de estas técnicas está en que existen otros factores que pueden provocar correlaciones débiles además de las no linealidades. Estos factores pueden ser los ruidos aditivos y los errores de sincronización de las señales. El modelo se basa en el supuesto de que el factor dominante que causa la pérdida de coherencia son las no linealidades del dispositivo bajo estudio. Basadas en estas técnicas surgirán otras como la propuesta en (Temme & Brunet, 2008).

### 3.6.8 La métrica Geddes-Lee

Esta métrica viene a ser una especie de puente entre las medidas de distorsión y los métodos perceptuales. Si bien la métrica no usa explícitamente un modelo del sistema auditivo, ésta se basa en una serie de supuestos psicoacústicos que utiliza de forma indirecta. Se basa en una función de transferencia estática  $T(x)$  – relación entre los niveles de entrada y salida en un dispositivo – que se pondera para obtener una expresión sensible a los bajos niveles de señal – cuya no linealidad es más perjudicial dado que los bajos niveles son pobres enmascarantes – y a las distorsiones de orden alto – dado que producen espectros anchos de productos de distorsión que son débilmente enmascarados -. La métrica fue propuesta en (Geddes & Lee, 2003) y la veremos a continuación con cierto detalle, dado su enfoque a partir de supuestos perceptivos. Una implementación experimental de ésta en transductores reales puede verse en (Cruañes J. , Francés, Ramis, & Alba, 2007).

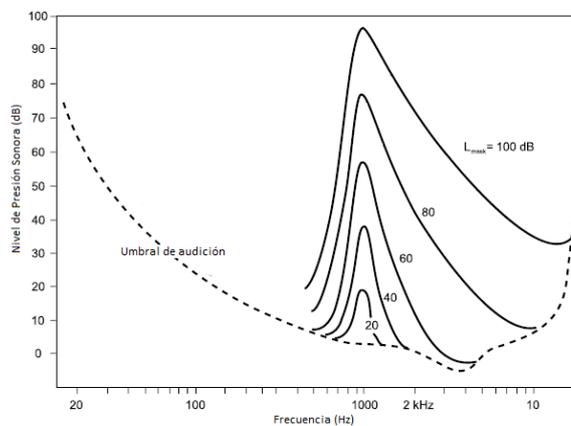
La métrica Geddes-Lee se fundamenta en las propiedades psicoacústicas del enmascaramiento. Este puede definirse, como se verá más adelante, como el proceso por el cual el umbral de audibilidad de un sonido aumenta debido a la presencia de otro sonido enmascarante. O también la cantidad en que se incrementa el umbral de audibilidad de un sonido (en dB) en presencia de otro (enmascarante). En el capítulo dedicado a la percepción auditiva se verá el rol desempeñado por esta característica de la audición a la hora de establecer técnicas experimentales que permiten realizar aproximaciones al filtrado en frecuencia que tiene lugar en la membrana basilar. Pero el enmascaramiento, *per se*, también tiene, evidentemente, sus propias consecuencias perceptivas. La definición de la ASA<sup>4</sup> resulta clarificadora a la hora de interpretar el

---

<sup>4</sup> American Standards Association

enmascaramiento como un mecanismo a partir del cual obtener los anchos de banda críticos, pero los efectos perceptivos del enmascaramiento están presentes en la vida cotidiana de una forma más trivial. El enmascaramiento no es más que el efecto psicoacústico mediante el cual un sonido audible sin presencia de ningún otro sonido se vuelve inaudible en presencia de otro sonido que enmascara al primero. El típico ejemplo cotidiano sería la voz de una conversación que resulta ininteligible o inaudible en presencia de un fuerte ruido de fondo. En general, el enmascaramiento dependerá de la sonoridad y un sonido muy intenso enmascarará a uno débil. Pero el fenómeno estará también en función de la frecuencia y, en general, los sonidos enmascararán de forma mucho más efectiva las frecuencias superiores a la suya propia que las frecuencias por debajo del sonido enmascarante. Además, los sonidos dentro de una misma banda crítica tenderán a enmascararse entre ellos de una forma mucho más pronunciada que aquellos sonidos que caen en bandas críticas diferentes. Si redujésemos el problema a sonidos sinusoidales, un sonido multitono resultaría mucho más difícil de enmascarar que un tono puro. El enmascaramiento resulta un concepto fundamental en la percepción de la distorsión, dado que los productos de distorsión sólo contribuirán a la percepción de ésta si no están enmascarados por el estímulo inicial o por otros productos de distorsión.

A modo de ejemplo, la Figura 3.16, obtenida por Zwicker a partir de datos empíricos (Zwicker & Fastl, 1990) traza el umbral de enmascaramiento de un tono de prueba en presencia de un ruido de banda estrecha centrado en 1kHz.



Fuente: (Zwicker & Fastl, 1990)

**Figura 3.16. Umbral de enmascaramiento de un tono puro en presencia de un ruido de banda estrecha centrado en 1kHz**

Los valores que acompañan a cada curva representan el nivel SPL del ruido de banda estrecha, mientras que las curvas representan el nivel que ha de alcanzar el tono puro en cada frecuencia para resultar audible. El umbral tiene su máximo en la frecuencia central

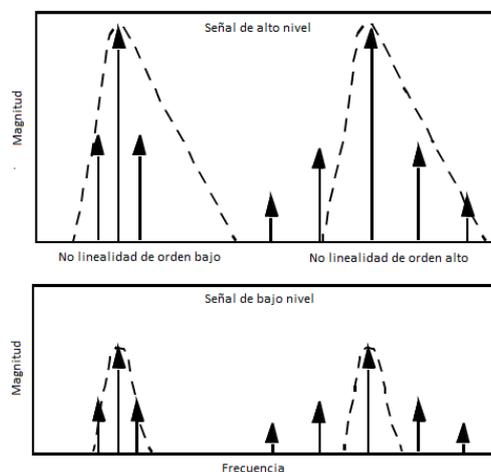
de la banda del ruido enmascarante, 1 kHz. A medida que la frecuencia del test de tono desciende por debajo de 1 kHz también lo hace con rapidez el umbral de audición de éste. Sin embargo, puede verse que este descenso no se produce de la misma forma pronunciada en las frecuencias por encima de la frecuencia central de la banda de ruido. Así, cuando el tono de test está por encima de 1 kHz, el umbral desciende más lentamente, siendo necesario que la señal de test tenga un mayor nivel SPL para resultar audible.

Caben pues destacar dos efectos: el primero es que los efectos del enmascaramiento serán predominantes en alta frecuencia y el segundo que estos efectos incrementan de forma no lineal con el incremento del nivel del sonido enmascarante. Por ejemplo, un incremento de de 60 a 80 dB SPL del nivel del ruido enmascarante no resulta en un incremento de 20 dB en un tono de 3 kHz para resultar audible de nuevo al aumentar el ruido, sino que son necesarios 30 dB. En el caso de la distorsión armónica el umbral de enmascaramiento puede afectar a los diferentes armónicos según sea su distancia a la frecuencia fundamental. Por ejemplo, dado un sistema no lineal que produzca armónicos de segundo y de tercer orden de igual amplitud, los armónicos de segundo orden serían más susceptibles de ser enmascarados que los de tercer orden. Esto puede aplicarse a los armónicos de orden superior, resultando en la idea de que los armónicos de orden más alto pueden resultar más perceptibles.

Geddes y Lee proponen una métrica que tenga en consideración estos dos efectos principales del enmascaramiento: las señales de bajo nivel son difícilmente enmascarables, así las no-linealidades que afectan a estas señales serán peores (es decir, más fácilmente perceptibles) que aquellas que afectan a las señales de nivel alto. Por otra parte, las no-linealidades de órdenes altos producen productos de distorsión más alejados en el espectro y por lo tanto pobremente enmascarados. En cambio los productos de orden bajo tienen más posibilidades de ser enmascarados – y consecuentemente no percibidos -.

Así, los autores sintetizan tres principios de la percepción de la distorsión:

- a) Los productos de distorsión que aparezcan en frecuencias superiores a la de la excitación son menos perceptibles (enmascarados en mayor grado) que aquellos que se producen en las frecuencias inferiores, dada la asimetría de los patrones de enmascaramiento vista anteriormente.
- b) Los productos de distorsión más próximos (en el espectro) a la excitación son menos percibidos (más enmascarados) que aquellos localizados en una zona del espectro más alejada (el enmascaramiento es un efecto localizado, solo ocurre en la vecindad de la máscara).
- c) Los productos de la distorsión de cualquier tipo serán más perceptibles en las señales de bajo nivel que en las señales de nivel alto.



Fuente: (Geddes & Lee, 2003)

**Figura 3.17. Productos de distorsión de diferente orden para señales de alto y bajo nivel y patrones de enmascaramiento.**

La figura anterior resume los principios psicoacústicos tenidos en cuenta por Geddes y Lee a la hora de plantear su métrica. El primer principio se ilustra en la parte derecha del espectro superior de la figura. Dado que los patrones de enmascaramiento decaen más rápidamente en las frecuencias inferiores a la excitación, los productos de distorsión que caigan en esta zona del espectro tendrán más posibilidades de ser audibles, mientras que los que caen en la zona de frecuencias superiores a la excitación serán más fácilmente enmascarables y, consecuentemente, inaudibles. El segundo efecto también se aprecia en la Figura 3.17. Los productos de distorsión cercanos en frecuencia a la excitación – es decir, de orden bajo - caerán por debajo de la curva de enmascaramiento, mientras que a medida que aumenta el orden de no linealidad es más probable que los productos de distorsión ya no sean enmascarados por el patrón de enmascaramiento generado por la excitación. Finalmente, el tercer efecto se ilustra al comparar el espectro superior e inferior de la figura. Si la señal es de nivel alto, el patrón de enmascaramiento tenderá a ocupar una mayor zona del espectro, haciendo inaudibles los productos que aparezcan en esta zona. En cambio, para un nivel de señal bajo, la banda de frecuencias que queda enmascarada es muy estrecha y por lo tanto el enmascaramiento resultante muy pobre.

En resumen, el efecto de enmascaramiento del oído humano hace que las no linealidades de órdenes altos sean más audibles que las de órdenes bajos. Además, los productos de distorsión con nivel alto pueden ser completamente enmascarados si el orden de la no linealidad es bajo. Finalmente, las no-linealidades que tienen lugar para señales de bajo nivel son más audibles que aquellas que ocurren en niveles altos. La métrica debe ajustarse a las premisas anteriores con lo cual deberá ser más sensible a los productos de distorsión de orden alto que a los de orden bajo. Además, deberá ponderar con mayor

peso las no linealidades producidas por señales de nivel bajo. Finalmente, deberá ser inmune a cambios en el nivel de continua o ganancia, ya que éstos son inaudibles.

La ecuación que representa la métrica es la siguiente:

$$G_m = \sqrt{\int_{-1}^1 \left( \cos\left(\frac{x\pi}{2}\right) \right)^2 \left( \frac{d^2}{dx^2} T(x) \right)^2 dx} \quad (3.9)$$

En la expresión que determina  $G_m$ ,  $x$  es la señal de entrada y  $T(x)$  es la función de transferencia no lineal del sistema en cuestión.  $T(x)$  es la función que relaciona el nivel instantáneo de la salida con el nivel instantáneo de la entrada del sistema. Se trata de una función de transferencia no lineal sin memoria dado que es independiente de la frecuencia. También se ha definido como no linealidad estática a la descrita por funciones de transferencia como  $T(x)$ . La relación entre entrada y salida del sistema suele ser polinómica o una función discontinua (característica de efectos como la saturación de señal – ver Figura 3.7, Figura 3.9. En caso de ser una función polinómica se expresará como  $T(x) = \sum_{n=0}^N a_n x^n$ , donde cada término  $n$  representará el orden de no linealidad, excepto  $n=0$  y  $n=1$  que representarán el nivel de continua y la ganancia lineal del sistema.

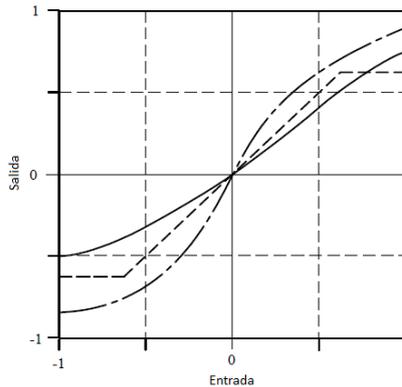


Figura 3.18. Ejemplos de funciones de transferencia no lineales.

Al considerar la segunda derivada de la función de transferencia no lineal, la métrica otorga mayor ponderación a los productos de distorsión de orden más alto. Además, la segunda derivada también permite no considerar la parte continua y la ganancia lineal de  $T(x)$ , dado que al calcularla se eliminarán todos los componentes hasta el segundo armónico. Así, con la segunda derivada se consigue la premisa de otorgar mayor peso a los productos de distorsión causados por órdenes altos. Por otra parte, el término del coseno se aplica para ponderar al alza los niveles de señal más bajos. Para niveles de señal pequeños el término del coseno se aproximará a la unidad, mientras que para niveles altos (próximos o igual a la unidad) el término del coseno se aproximará a cero.

De esta forma, el término proporciona más peso a los niveles de entrada pequeños, para los que el umbral de enmascaramiento es también pequeño, como acabamos de ver.

Finalmente, la ecuación 3.9 se integra en el rango de salida de la señal, entre -1 y 1, dado que en la definición de la métrica el rango dinámico del dispositivo está normalizado. Como se verá en la implementación, este aspecto supone una limitación crucial a la hora de aplicar la métrica en sistemas reales.

### 3.6.9 *Los modelos perceptuales*

Al mismo tiempo que la propuesta anterior, se desarrollarán una serie de trabajos que establecerán el uso de modelos perceptuales, introducidos en la sección 2.2.2, para la evaluación de las no linealidades. Estos trabajos retoman el camino iniciado décadas antes por Shorter y Belcher en el sentido de que hablan abiertamente de percepción de calidad por parte del usuario. No es tan importante la obtención de una medida fenomenológica del sistema como en métodos anteriores. La principal diferencia estriba que si bien hasta ahora se había intentado correlar la medida física sobre el dispositivo con la sensación percibida, con los modelos perceptuales el dispositivo pasa a ser una caja negra generadora de no linealidades sobre la que no se realiza medida alguna. Se actúa directamente sobre las señales de entrada y salida del dispositivo procesadas directamente en el modelo perceptual. En estos casos se trata de conseguir un indicador de la calidad percibida sin entrar en el análisis del dispositivo. De hecho, el más importante de esta serie de trabajos (Tan, Moore, Zacharov, & Ville-Veikko, 2004) desarrollará una métrica para la predicción de la calidad percibida por parte del usuario, la métrica  $R_{\text{nonlin}}$ , que será la que implementaremos experimentalmente en este trabajo. Esta métrica da cuenta de hasta qué punto es significativo este nuevo enfoque, dado que la principal finalidad de la métrica es eliminar los test de audición (de evaluación de calidad). Idealmente, no sería necesario intentar correlar datos objetivos con los resultados de encuestas o test de audición, porque la métrica intenta ser capaz de predecir los propios resultados de los test, con lo cual es, *per se*, un indicador de calidad.

En el primero de estos trabajos (Moore, Tan, & Zacharov, 2003), los autores desarrollaron una métrica denominada DS – *Distortion Score* -. Utilizaron una señal de estímulo multitono como las propuestas en (Czewrinski & Voishvillo, 2001) basada en diez tonos separados logarítmicamente a lo largo del espectro auditivo y no linealidades estáticas generadas artificialmente. Los principales aspectos del procedimiento para la obtención de DS son los siguientes: en primer lugar, tanto la señal de entrada como la de salida del sistema no lineal se alinean en el tiempo y se analizan en *frames* no superpuestas de 30 ms de duración. En segundo lugar, los espectros de entrada y salida se dividen en 40 bandas de frecuencia no superpuestas, con un ancho de 1 ERB cada una de ellas. Para cada frame se calcula la diferencia de nivel en cada ERB y el valor absoluto de la diferencia se suma a lo largo del espectro auditivo. Esto daba una medida perceptualmente relevante de las diferencias de espectro entre las señales de entrada y salida que los autores denominaron *Distortion Score*. En (Tan, Moore, Zacharov, & Ville-Veikko, 2004) se propondrá un modelo perceptual más sofisticado que no utilizará

una señal multitono como única fuente de excitación del sistema. En este caso, se actuará directamente sobre señales musicales o de voz, realizando un procesado similar al anterior sobre las señales de entrada y salida. El hecho de que pueda usarse cualquier señal para la evaluación del sistema, independientemente de su complejidad, es el aspecto verdaderamente innovador de esta métrica y lo que permite presentarla como un método de predicción de la calidad percibida, como se ha mencionado antes. Esta métrica, denominada  $R_{\text{nonlin}}$ , se estudia e implementa en el capítulo 6. Los resultados de su aplicación sobre transductores reales se muestran en el capítulo 7. Las medidas y resultados experimentales de su aplicación en salas se verán en el capítulo 8.

### **3.7 Sistemas de radiación indirecta**

Los altavoces de radiación indirecta se utilizan en aquellas aplicaciones de audio en que son necesarios altos niveles de presión sonora junto con buenas características de directividad. Se caracterizan porque consiguen un mejor rendimiento utilizando una bocina que adapta la alta impedancia del diafragma a la baja impedancia del aire, a diferencia de los de radiación directa, en los que es el diafragma es el elemento que radia directamente al aire. Están formados básicamente por dos partes: la bocina y el motor de compresión.



**Figura 3.19. Altavoz de radiación indirecta con bocina montada sobre el motor de compresión**

El principal inconveniente de estos dispositivos es que presentan niveles de distorsión relativamente elevados, debido principalmente al hecho de trabajar con niveles de presión sonora elevados (Schurer, Berkhoff, Slump, & Herrmann, 1995). Las no linealidades se localizan en el transductor electroacústico así como en la propagación en la bocina (Klippel W. , 1996). A pesar de ser dispositivos concienzudamente estudiados, existe bastante controversia a la hora de valorar las aportaciones de cada una de sus partes (motor de compresión y bocina) al peso final de los productos de distorsión generados. Algunos autores han postulado que la distorsión introducida por el motor de compresión queda enmascarada por la introducida por la bocina, de forma que no tiene sentido mejorar aspectos mecánicos del motor ni implementar sistemas de compensación de la distorsión no lineal para éste (Geddes & Clark, 1995) (Geddes, 2003). Otros autores han

sugerido, que si bien la propagación en la bocina es la fuente de distorsión predominante, la distorsión generada por o en el motor no es despreciable (Schmitt, 1995).

### 3.7.1 El motor de compresión

El motor de compresión es en realidad un altavoz electrodinámico de bobina móvil con ciertas particularidades: una cámara de compresión, un diafragma pequeño y ligero y una estructura para ser anclado a la bocina. Se distinguen las siguientes cuatro partes funcionales: el conjunto móvil, el conjunto magnético, la cámara de compresión y el corrector de fase.

#### 3.7.1.1 El conjunto móvil

Está formado por un diafragma en forma de cúpula, una suspensión en forma de anillo y una bobina. El principio de funcionamiento es el mismo que el del altavoz de radiación directa. La bobina está sumergida en un campo magnético uniforme y permanente producido por el imán. Si una corriente eléctrica atraviesa dicha bobina, se produce una fuerza transversal que la mueve. La cúpula, adherida a la bobina, constituye la superficie radiante del altavoz. Ha de ser lo más ligera y rígida posible para que vibre como una unidad. Esto se consigue a frecuencias bajas pero conforme aumenta la frecuencia las zonas de la cúpula radian independientemente unas de otras lo que produce no linealidades en el funcionamiento del motor. La suspensión controla el desplazamiento de la cúpula, de forma que ésta sólo se pueda desplazar en una única dirección axial. La relación entre el desplazamiento de la cúpula y la fuerza de restauración que ejerce la suspensión solamente es lineal para señales de nivel bajo. La bobina es el elemento motriz del conjunto móvil. El valor de su inductancia varía con su desplazamiento, lo que constituye otra fuente de no linealidad.

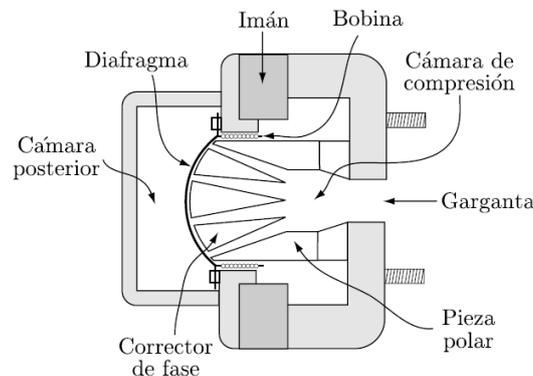


Figura 3.20. Sección transversal de un motor de compresión.

#### 3.7.1.2 El conjunto magnético

La filosofía del conjunto magnético es también idéntica a la del altavoz de radiación directa. El parámetro que caracteriza la capacidad de transformación de la parte eléctrica a la parte mecánica del transductor es el factor de fuerza o  $Bl$ . El conjunto magnético

está formado por un perno o núcleo, una anilla de campo y un imán. Su función es producir un campo magnético  $B$  en el entrehierro en el que está inmersa la bobina

### *3.7.1.3 La cámara de compresión*

La cámara de compresión es el volumen de aire que queda entre la cúpula y el corrector de fase. El corrector de fase es una pieza sólida con una serie de ranuras concéntricas que permiten el paso del aire a través de él. La cúpula está colocada sobre el corrector de fase con una pequeña separación. El aire ha de pasar por las ranuras del corrector de fase, que ocupan una superficie mucho menor a la ocupada por la cúpula, y este cambio de área brusco es lo que produce el efecto de la compresión. Este efecto se traduce en un aumento de la impedancia de radiación del altavoz, con el resultado de que se logra una mejor adaptación de impedancias.

### *3.7.1.4 El corrector de fase*

El corrector de fase es un dispositivo cuya finalidad es conseguir que las contribuciones de cada parte de la cúpula a la radiación se interfieran de forma positiva, es decir, que estén en fase. Se trata de una pieza sólida a la que se le practican una serie de ranuras que guían el paso de las ondas sonoras desde la membrana hasta la salida de la cámara de compresión. Su forma se adapta en la parte superior a la cúpula y en la parte inferior a la cámara de compresión. De este modo, se garantiza por una parte que la distancia entre la membrana y el corrector de fase se mantiene constante y que se rellena por completo la cámara de compresión. El funcionamiento se fundamenta en que en la base del corrector se combinan las diferentes ondas sonoras que vienen de las diferentes zonas de la cúpula, de forma que el corrector asegura que el camino que han recorrido es idéntico, con lo cual se encuentran en fase.

### *3.7.2 La bocina<sup>5</sup>*

La bocina por sí misma es un componente pasivo pero actúa mejorando la eficiencia entre la superficie de radiación y el aire. La bocina puede considerarse como un transformador acústico, que proporciona el acoplamiento de impedancias entre el relativamente denso material que forma el diafragma y el aire (de baja densidad). La parte estrecha que se monta sobre el motor de compresión se denomina garganta y la parte ancha más alejada del motor es la boca. Además de la función de adaptación de impedancias, las bocinas también se han utilizado para extender el rango de baja frecuencia para determinados altavoces. Con la bocina acoplada, un altavoz es capaz de reproducir frecuencias más bajas con mayor potencia. La forma de la bocina y el tamaño de la boca determinan el límite en baja frecuencia. Las bocinas se han usado también

---

<sup>5</sup> Dado que en este trabajo se efectuará la implementación de un método perceptual para la valoración de la distorsión introducida por los motores de compresión únicamente, sin considerar la bocina, nos limitaremos aquí a señalar las funciones principales de ésta en el conjunto del dispositivo sin entrar en consideraciones de diseño, factores de forma, propagación y directividad. Las bocinas, como dispositivos, han generado una ingente cantidad de literatura cuyo estudio dejaremos fuera del ámbito de este trabajo. En éste, se utilizarán únicamente motores de compresión acoplados a un tubo de onda plana con la finalidad de valorar la distorsión introducida por el motor exclusivamente. De hecho, varios autores han propuesto modelos de ambas partes (motor y bocina) por separado (Klippel W., 1996), (Schmitt, 1995)

para modificar las características directivas. La anchura y altura de éstas determinan los ángulos de cobertura horizontal y vertical respectivamente. El comportamiento del dispositivo fuera y dentro del eje de cobertura dependerá de la forma de la bocina. El diseño de la bocina es determinante en el comportamiento del dispositivo dado que las distorsiones del frente de ondas que se producen en la bocina están en gran parte condicionadas por la forma de ésta (Henricksen, 1988).

### 3.7.3 Fuentes de no linealidad en el motor de compresión

El motor de la Figura 3.20 realiza una conversión de las magnitudes eléctricas en el terminal de entrada, vía mecánica, a magnitudes acústicas en la garganta de la bocina. Sin tomar en consideración el efecto de la bocina, el funcionamiento del motor puede ser descrito por su circuito equivalente de la Figura 3.21. El circuito contiene las siguientes variables y parámetros agrupados, que corresponden a los elementos de la Tabla 3.5.

**Tabla 3.5. Variables del circuito equivalente de un motor de compresión**

Variable	Descripción
$U_e$	Tensión eléctrica en los terminales de entrada
$i_e$	Corriente de entrada
$R_e$	Resistencia eléctrica de la bobina
$L_e$	Inductancia de la bobina
$Bl$	Factor de fuerza
$v$	Velocidad del diafragma (dx/dt)
$F$	Fuerza motriz
$m$	Masa del diafragma
$C_D(x)$	Compliancia mecánica de la suspensión
$R_D$	Resistencia mecánica
$S_D$	Área del diafragma
$q$	Velocidad en la cámara de compresión
$p$	Presión sonora
$C_V(p_V)$	Compliancia acústica del aire en la cámara posterior
$C_G(x, p_G)$	Compliancia acústica de la separación diafragma - corrector de fase
$R_C(q_C)$	Resistencia acústica en el corrector de fase
$M_C$	Masa acústica en el corrector de fase
$Z_T$	Impedancia acústica en la garganta de la bocina.

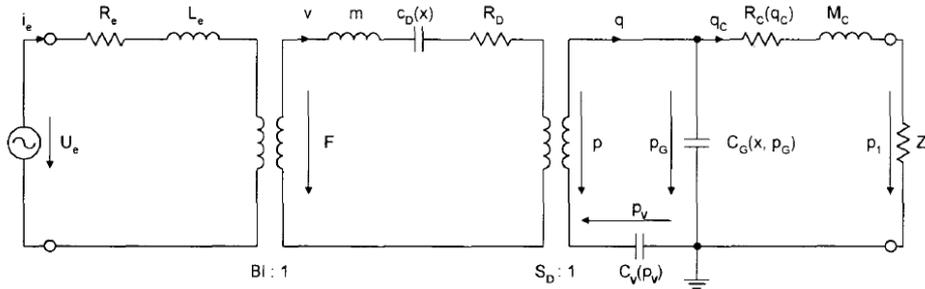


Figura 3.21. Circuito equivalente del motor de compresión

En la tabla anterior puede observarse que se han establecido cuatro parámetros como dependientes de un desplazamiento, una presión o una velocidad. En concreto, la compliancia mecánica de la suspensión estará en función del desplazamiento del diafragma. La compliancia acústica del aire en la cámara posterior dependerá de la presión en ésta. La compliancia acústica del espacio que separa el diafragma y el corrector de fase será función del desplazamiento del diafragma y de la presión en este espacio. Finalmente, la resistencia acústica en el corrector de fase será función de la velocidad de las partículas en éste. En un modelo con no linealidad débil, estos elementos serían considerados como fuentes de no linealidad (Klippel W. , 1996).

En contraste con los resultados de un modelado lineal (Geddes & Clark, 1995), la compliancia de la suspensión  $C_D(x)$  no se asume como constante y será función del desplazamiento del diafragma ( $x$ ).

La compliancia acústica  $C_V(p_v)$  describe el comportamiento del aire en la cámara posterior, detrás del diafragma. Este parámetro puede considerarse constante si los cambios en el volumen ( $V$ ) y en la presión sonora ( $p_v$ ) son pequeños en comparación con el volumen inicial  $V_0$  y la presión estática  $p_0$ . Para grandes compresiones y expansiones del aire contenido en el volumen bajo condiciones adiabáticas ( $\gamma=1,4$ ) la relación no lineal  $(p_0 + p_v)(V_0 + V)^\gamma = p_0 V_0^\gamma$  debe tomarse en consideración.

La compliancia  $C_G(x, p_G)$  del aire entre el diafragma y la entrada del corrector de fase depende de la presión sonora  $p_G$  y del desplazamiento  $x$  dado que el movimiento del diafragma modifica el volumen de aire en ese espacio de acuerdo con la relación  $(p_0 + p_G)(S_D d + S_D x)^\gamma = p_0 (S_D d)^\gamma$ . Donde  $S_D$  es el área del diafragma y  $d$  el ancho del espacio de aire en ausencia de movimiento.

Las propiedades acústicas del corrector de fase estarán modeladas por una masa acústica  $M_C$  y una resistencia acústica  $R_C(q_C)$ . Para velocidades  $q_C$  pequeñas, el flujo es laminar y las pérdidas viscosas se traducen en una resistencia constante. Para una determinada velocidad crítica  $v_{tur}$  el flujo deja de ser laminar y pasa a régimen turbulento. La

velocidad crítica viene dada por  $v_{tur} = \frac{R_E \eta}{\rho r}$ , donde  $R_E$  es el número de Reynolds crítico (1160),  $\eta$  el coeficiente de viscosidad del aire ( $1.7 \cdot 10^{-5}$  Ns/m<sup>2</sup>),  $\rho$  la densidad del aire y  $r$  la dimensión geométrica relacionada. La velocidad crítica para la que el flujo pasa a régimen turbulento puede variar entre  $v_{tur} = 3$  m/s en los canales del corrector de fase (donde  $r=5$ mm) y  $v_{tur} = 0,5$  m/s en la garganta del motor (donde  $r=30$ mm). Cuando ocurre el régimen turbulento, la relación entre la velocidad del volumen  $q_c$  y la presión en éste  $p_c$  deja de ser lineal y la resistencia  $R_c(q_c)$  incrementa con la magnitud de la velocidad  $q_c$ .

En sistemas de radiación directa, principalmente en los de baja frecuencia, las variaciones de la inductancia de la bobina  $L_c$  y del factor de fuerza  $BI$  debidas al desplazamiento de la bobina son una fuente determinante de no linealidades (Klippel W. , 1992), sin embargo en sistemas de radiación indirecta estas fuentes de no linealidad pueden en principio no considerarse, dado que los desplazamientos en estos sistemas suelen ser pequeños. Esto siempre enmarcado en el ámbito de una no linealidad débil y un régimen de señal de entrada no excesivamente alto. En el caso de tener niveles de entrada muy altos, la excursión del conjunto móvil formado por la bobina y el diafragma será mayor, lo cual afectará a las no linealidades ya consideradas (mayor desplazamiento del diafragma y mayores presiones) pero también hará que las no linealidades de la bobina (variación de la inductancia con el desplazamiento) y del producto  $BI$  (producto  $BI$  modulado por la corriente que atraviesa la bobina) tengan que ser consideradas (Voishvillo A. , 2004)

#### 3.7.4 El tubo de onda plana

Los motores de compresión nunca se usan radiando directamente al aire. Su funcionamiento depende estrictamente de la carga acústica proporcionada por la bocina a la que están acoplados. Por otra parte, cada bocina tiene sus propias características y un mismo motor proporcionará resultados significativamente diferentes en función de la bocina que tenga acoplada, lo cual no es útil para tareas de diseño, medición y/o comparación de motores. De esta forma, las técnicas de medida habituales para transductores de radiación directa no pueden utilizarse en altavoces de compresión. Todo esto pone de manifiesto la necesidad de un dispositivo estandarizado para propósitos de medición que sea fácilmente reproducible y además, actúe como carga acústica del motor sobre el que se realice la medida. Ese dispositivo es el tubo de onda plana, cuyas características constructivas y de empleo en mediciones están recogidas en (AES-11D-1991 r 2003 standards and information documents, 1991).

El dispositivo consiste básicamente en un conducto cilíndrico y rígido que se monta sobre el motor de la misma manera que una bocina. El cilindro está parcialmente relleno con material absorbente cuya principal función es la de cancelar las reflexiones en el final del tubo sin modificar la carga en el motor. El cilindro tendrá una perforación lo más cerca posible del sistema de anclaje al motor, que servirá para la colocación de un micrófono de medida, que podrá registrar la onda proveniente del motor sin perturbar

las condiciones del sistema. Este simple dispositivo proporcionará una impedancia acústica independiente de la frecuencia, al menos en el rango de frecuencias de trabajo, que estará determinado por las dimensiones del tubo.

El límite en alta frecuencia del tubo de onda plana es  $f_h = \frac{1.22c}{d}$ , donde  $c$  es la velocidad del sonido en el aire y  $d$  es el diámetro interior del tubo. La respuesta medida en esta frecuencia se caracteriza por una bajada acentuada y estrecha en el espectro (tipo *notch*). A partir de esta frecuencia estas bajadas se repiten armónicamente, por lo que los datos recogidos en esta zona del espectro no son utilizables. En lo que respecta al límite en baja frecuencia, éste viene dado por  $f_l = \frac{c}{4l}$ , donde  $c$  es la velocidad del sonido y  $l$  es la longitud del tubo. La intención del dispositivo es proveer una impedancia acústica constante de valor  $\rho_0 c S_D$  (siendo  $\rho_0 c$  la impedancia del aire y  $S_D$  el diámetro interior del tubo) en la banda de paso determinada por las dos frecuencias  $f_h$  y  $f_l$ . El documento estándar define una aceptabilidad de la banda de paso del dispositivo, que define como SWR (*standing-wave ratio*), que determinará la consistencia de la carga acústica proporcionada por el dispositivo. El SWR en decibelios indica las variaciones del nivel de presión sonora a lo largo del tubo. Esta relación debe medirse en varias frecuencias y puede expresarse como el rango de variación de los niveles de presión sonora a lo largo del tubo. El documento estándar proporciona un procedimiento para la medida de este parámetro. Un método alternativo para la obtención del SWR utilizando señales MLS se propone en (Magalotti, Zuccatti, & Pasini, 1999). Un SWR de 0 dB indica que no se produce ninguna reflexión, de forma que sólo se propagan ondas planas desde el dispositivo hacia el absorbente del tramo final del tubo, donde éstas son completamente absorbidas. Se establece como válido un SWR inferior a 2 dB entre las frecuencias  $f_h$  y  $f_l$  para la obtención de datos fiables.

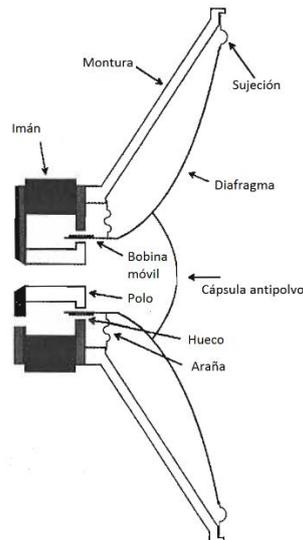
### 3.8 Sistemas de radiación directa

Los sistemas de radiación directa son de uso más habitual que los anteriores en la mayor parte de aplicaciones electroacústicas. El esquema básico de un motor de radiación directa puede verse en la Figura 3.22. En un transductor electrodinámico de radiación directa operan conjuntamente tres subsistemas: el motor, compuesto por el imán, los polos magnéticos, el hueco que habilita el desplazamiento de la bobina – denominado comúnmente entrehierro - y la propia bobina; el diafragma, que es la superficie que produce la radiación en el medio – en este caso aire – que generalmente lleva montada una cápsula antipolvo en la parte central o en ocasiones una cúpula de una sola pieza unida al diafragma para proteger esta parte del dispositivo; finalmente, el tercer subsistema lo conforma la suspensión, formada por la araña, y por la sujeción del diafragma a la montura, generalmente formada por un material similar a la goma. Este tercer conjunto se encarga de dar rigidez al sistema.

#### 3.8.1 El motor

Si en los bornes de la bobina móvil inyectamos una señal sinusoidal, por ejemplo, la corriente que circula por la bobina durante el ciclo positivo producirá el desplazamiento

del diafragma en una dirección. Al cambiar al ciclo negativo de la señal, el desplazamiento del diafragma se producirá en la dirección contraria.



**Figura 3.22. Esquema básico de un motor de radiación directa**

Para poder reproducir de la manera más fidedigna posible el movimiento del diafragma respecto de la señal de entrada, la bobina ha de ser capaz de desplazarse de manera idéntica en ambas direcciones. Para que esto ocurra es determinante que el campo magnético sea lo más simétrico posible, de forma que la fuerza aplicada sea idéntica en una dirección y en otra. De no ser así ya estaríamos hablando de una no linealidad del sistema y de su consecuente distorsión. De forma ideal, el flujo magnético debería estar confinado en el hueco que permite el movimiento de la bobina, de forma que la simetría estaría asegurada. En la práctica, las líneas de fuerza magnética sobrepasan esta área produciéndose los denominados campos de fringe (*fringe fields*) en cada área contigua al entrehierro. Existen diferentes técnicas de diseño para asegurar la simetría del conjunto (Dickason, 2006) (Borwick, 2001).

La fuerza mecánica que desarrollará el motor al introducir una corriente en la bobina vendrá determinada por el factor  $Bl$  – factor de fuerza, definido de igual modo que el caso de los transductores de radiación indirecta -, donde  $B$  es el valor del campo magnético en tesla y  $l$  es la longitud de la bobina en metros. El parámetro que caracteriza pues, la capacidad de transformación de la parte eléctrica a la parte mecánica del transductor es el factor de fuerza o  $Bl$ .

Existen diferentes técnicas de diseño y geometrías concernientes al espacio hueco en el que se ubica la bobina móvil, la longitud de la bobina y la máxima excursión  $X_{max}$  que puede realizar ésta. Diferentes combinaciones entre las dimensiones del espacio y la

longitud de la bobina pueden dar lugar a un mismo valor de  $X_{max}$ , pero darán lugar a comportamientos diferenciados en términos de no linealidad (Dickason, 2006). Otros parámetros que definirán eléctricamente el conjunto del motor, además del factor de fuerza, son la inductancia de la bobina móvil ( $L_e$ ) y la resistencia eléctrica de ésta,  $R_e$ .

### *3.8.2 El diafragma*

Generalmente, la explicación física del comportamiento de un diafragma se inicia utilizando el modelo de pistón plano radiando en una cara y montado sobre pantalla infinita. Este modelo permite determinar la resistencia y la reactancia de radiación, así como el comportamiento según rangos de frecuencia del sistema (Kinsler, Frey, Coppens, & Sanders, 1991). La transferencia de energía desde el pistón al aire vendrá limitada, en términos de frecuencia en la parte baja del espectro, por la frecuencia de resonancia del sistema – por debajo de la cual estará limitada por las restricciones mecánicas del sistema -. En la parte superior del espectro, el límite a partir del cual la transferencia de energía comenzará a decaer vendrá determinada en función de la impedancia de radiación del aire y del radio del diafragma. Así, las superficies radiantes más pequeñas serán capaces de reproducir frecuencias más altas.

Los diafragmas reales no son infinitamente rígidos y presentarán algún grado de flexión en función de su forma y de los materiales con que están contruidos. La flexibilidad del diafragma tendrá un efecto directo sobre la eficiencia del transductor en alta frecuencia y la directividad de éste. A pesar de los diferentes tipos de materiales y diferentes rigideces, de forma general los diafragmas tienden a producir modos propios de vibración estudiados de la misma forma: los modos concéntricos, los modos radiales y la combinación de ambos (Beranek & Mellow, 2012).

### *3.8.3 La suspensión*

Como hemos visto en la Figura 3.22, el sistema de suspensión de un altavoz electrodinámico de radiación directa está compuesto por dos elementos: la araña y la terminación o sujeción del diafragma a la montura. La sujeción exterior ayuda a mantener el diafragma centrado y proporciona parte de la fuerza restauradora que mantiene la bobina en el entrehierro. Además, proporciona un límite amortiguado en la terminación del diafragma, con lo cual se amortiguan los modos propios en la parte exterior de la superficie de radiación. Así pues, sus dimensiones y material pueden alterar notablemente la respuesta del transductor. Suele estar fabricado de diferentes tipos de gomas, espumas o combinaciones de éstas. Un material cada vez más utilizado para la construcción de esta sujeción es un caucho termoplástico vulcanizado denominado Santoprene™.

La araña, por su parte, suele ser de fabricación textil sometida a diferentes tratamientos. Su función es la de mantener la bobina centrada y proporcionar la fuerza restauradora principal que ayuda a mantenerla en su posición. Por lo tanto, es la que en mayor parte determina la rigidez del sistema. La rigidez aportada por el sistema de suspensión se mide en términos de compliancia. Ésta determinara la frecuencia de resonancia mecánica

del dispositivo. Se estima que la aportación de la araña y la sujeción a la compliancia total del sistema está alrededor de un 80% y un 20% respectivamente (Dickason, 2006)

### 3.8.4 Fuentes de no linealidad en un transductor de radiación directa

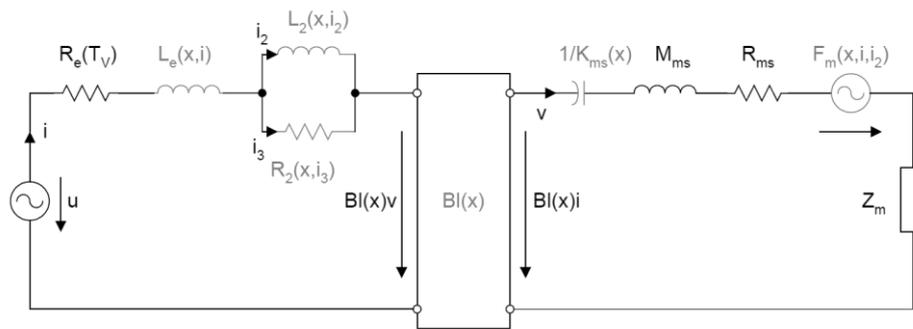
El funcionamiento del transductor de radiación directa puede ser descrito por su circuito equivalente de la Figura 3.23. Al igual que en el caso visto en la sección 3.7.3, se observa que hay variables que a su vez estarán en función de otras, en este caso de un desplazamiento  $x$  o de una corriente  $i_1$ ,  $i_2$  e  $i_3$ . Esta síntesis se basa en (Klippel W. , 1990), (Klippel W. , 1992) y (Klippel W. , 2007) Para un estudio en profundidad del modelado y análisis de no linealidades en altavoces de radiación directa remitimos a estas referencias.

**Tabla 3.6. Variables del circuito equivalente de un motor de radiación directa**

Variable	Descripción
$u$	Tensión eléctrica en los terminales de entrada
$i$	Corriente de entrada
$R_e$	Resistencia eléctrica de la bobina
$L_e(x,i)$	Inductancia de la bobina
$Bl(x)$	Factor de fuerza
$v$	Velocidad de la bobina
$L2(x,i2)$	Modelado del cambio de impedancia en función de $x$
$R2(x,i3)$	Modelado del cambio de impedancia en función de $x$
$1/ Kms(x)$	Compliancia mecánica de la suspensión
$Mms$	Masa mecánica del diafragma
$Rms$	Resistencia mecánica del diafragma
$Fm(x,i,i2)$	Fuerza electromotriz
$Z_m$	Impedancia mecánica de radiación.

Como se observa en la Figura 3.23 y en la Tabla 3.6, el factor de fuerza  $Bl$  dependerá del desplazamiento  $x$  de la bobina, de forma que no será constante. Este aspecto era despreciable en el caso de los motores de compresión debido al bajo orden de magnitud de los desplazamientos, pero en el caso de la radiación directa es una fuente directa de no linealidad, dado que las excursiones realizadas por la bobina son mucho mayores. Así, durante estas excursiones, en la medida en que parte de la bobina sale del entrehierro el factor de fuerza decrece. La forma que tenga la función  $Bl(x)$  dependerá también de la geometría del entrehierro y del campo magnético  $B$  generado por el imán. Obviamente, para desplazamientos pequeños de la bobina el factor de fuerza permanece casi constante, dado que también lo es el número de vueltas de conductor de la bobina que

permanecen en el entrehierro.  $Bl$  tendrá dos efectos no lineales sobre el sistema: como factor de acoplamiento entre la parte eléctrica y mecánica del sistema, cualquier variación de  $Bl(x)$  tendrá su consecuencia en la fuerza electromotriz  $F=Bl(x)i$ , como se observa en el segundo bloque del circuito de la Figura 3.23. El segundo efecto de  $Bl(x)$  es la dependencia con el desplazamiento de la fuerza electromotriz posterior generada por el movimiento de la bobina. En este caso, el factor de fuerza se multiplica por la velocidad causando variaciones de atenuación eléctrica.



**Figura 3.23. Circuito equivalente de un altavoz de radiación directa**

Se ha visto como estos dispositivos disponen de un sistema de suspensión para centrar la bobina en el entrehierro y generar una fuerza de restauración capaz de devolver a la bobina a su posición de reposo. A medida que la superficie de radiación es mayor también lo son los mecanismos de suspensión y sujeción con la finalidad de permitir el movimiento únicamente en el eje centrado en el entrehierro y evitar el balanceo. De nuevo, para desplazamientos cortos el comportamiento de la suspensión puede caracterizarse como un muelle en el que la fuerza de restauración tiene un comportamiento cuasi-lineal. El problema surge con los desplazamientos largos, para los cuales el sistema de suspensión responde con más fuerza de lo que lo haría un muelle. Así, la fuerza restauradora puede definirse como el producto entre el desplazamiento  $x$  y la rigidez no lineal  $K_{ms}(x)$ , con lo cual  $F = K_{ms}(x) \cdot x$ . Dado que la rigidez no es constante sino que está en función del desplazamiento  $x$ , la fuerza de restauración contiene productos del desplazamiento de la bobina que producen no linealidades en la señal. La rigidez varía también con la frecuencia debido al comportamiento viscoelástico del material que conforma la suspensión.

En lo que respecta a las no linealidades relativas a la inductancia, la impedancia eléctrica de entrada dependerá de la posición de la bobina en cada momento. Por encima de la frecuencia de resonancia y para excursiones negativas largas la impedancia eléctrica es significativamente más alta que para una excursión de la misma distancia en sentido contrario (positiva). Esta propiedad puede explicarse por la variación de la inductancia de la bobina en función del desplazamiento. La corriente en la bobina produce un campo

magnético alterno cuyo flujo depende de la propia posición de la bobina y de la magnitud de la corriente. Si la bobina se sitúa en el aire por encima del entrehierro la inductancia será mucho menor que cuando esté en la zona por debajo del entrehierro, donde el material envolvente producirá un decrecimiento de la resistencia magnética. Además de su dependencia del desplazamiento  $x$ , la inductancia dependerá también de la corriente de entrada  $i$ . Cuando no hay corriente de entrada en la bobina, el imán permanente produce una intensidad de campo que determina el punto de trabajo característico. Una alta corriente positiva incrementará la intensidad del campo a un punto en el que el entrehierro estará más saturado y descenderá la permeabilidad  $\mu$ . En cambio, para una alta corriente negativa, descenderá la intensidad del campo aumentando la de la permeabilidad  $\mu$ . El efecto de la variación de la permeabilidad en función de la corriente  $\mu(i)$  se conoce como modulación de flujo y está fuertemente relacionado con el material magnético utilizado en la construcción del motor.

Para modelar los efectos de los cambios de impedancia en función del desplazamiento de la bobina y de la corriente de entrada, (Klippel W. , 2007) propone una simplificación basada en la introducción de una segunda inductancia  $L_2$  en paralelo con una segunda resistencia  $R_2$ , como puede verse en la 0. Esta simplificación propone no considerar las interacciones entre corriente y desplazamiento. Las características no lineales de  $L_e(x)$  y  $L_e(i)$  y los valores  $L_2(0)$  y  $R_2(0)$  en la posición de reposo  $x=0$  son suficientes en la mayoría de las aplicaciones para describir la no linealidad de la inductancia.

$$\begin{aligned} \frac{L_e(x, i = 0)}{L_e(0)} &\approx \frac{L_2(x, i = 0)}{L_2(0)} \approx \frac{R_2(x, i = 0)}{R_2(0)} \\ \frac{L_e(i, x = 0)}{L_e(0)} &\approx \frac{L_2(i, x = 0)}{L_2(0)} \approx \frac{R_2(i, x = 0)}{R_2(0)} \end{aligned} \quad (3.10)$$

# 4 Calidad en salas

## **4.1 La acústica de salas como objeto de estudio**

Hemos visto que, en el trabajo experimental propuesto, el canal de transmisión está conformado por la suma del transductor electrodinámico más el efecto de la sala. En el capítulo anterior se han esbozado las diferentes perspectivas desde las que se ha abordado el problema del estudio de los transductores considerados como canal de transmisión. En este haremos lo propio con las salas. Si el objeto de estudio en el caso de la transducción era la distorsión, tal y como se ha definido en los apartados 3.1, 3.5 y 3.6, en el caso de los recintos, el objeto de estudio será el campo acústico creado en éstos por las fuentes sonoras, y cómo las características del campo acústico de la sala influyen en la percepción de la fuente sonora, de manera que cada recinto imprime al sonido que en él se propaga unas determinadas cualidades que serán las consideradas por el receptor a la hora de evaluar su calidad. El estudio del comportamiento acústico de los recintos constituye por sí sólo, toda una disciplina dentro de la acústica, la denominada acústica de salas. De forma sintética, y estableciendo un paralelismo con el caso de los

transductores visto en la sección 3.2, podría decirse que el problema se ha abordado desde tres ópticas bien diferenciadas:

- a) Desde una perspectiva estrictamente teórica, históricamente se han propuesto tres modelos teóricos para la descripción del campo acústico. Estos modelos son la teoría estadística, la teoría geométrica y la teoría ondulatoria. El objetivo de estos modelos no tiene a priori relación alguna con el establecimiento de una evaluación cualitativa. Se trata de teorías que modelan la formación y el comportamiento del campo acústico a partir de unos determinados supuestos. Las métricas propuestas en el siguiente epígrafe pueden derivarse o explicarse atendiendo a alguna de estas tres teorías. Sin embargo, los modelos teóricos atienden, obviamente, a una representación física de los fenómenos, sin considerar aspectos cualitativos. Veremos que en cambio, las métricas derivadas o justificables mediante estos modelos teóricos sí se han utilizado en ambos sentidos: por una parte, para soportar el modelo teórico y por otra para intentar explicar fenómenos estrictamente cualitativos relacionados con la impresión general que de la sala tiene el receptor.
- b) Metodología y métricas experimentales. Desde los primeros experimentos de Sabine en 1895 y su tiempo de reverberación (RT) se han desarrollado una extensísima cantidad de métricas y procedimientos experimentales orientados a obtener parámetros descriptivos del comportamiento del campo acústico. Como se ha visto en la sección 2.2.4, estos parámetros se pueden clasificar atendiendo a criterios temporales, energéticos y espaciales fundamentalmente, aunque se puede llegar a un mayor grado de concreción en función del contenido de la fuente sonora en el recinto, generalmente música o palabra. Estos parámetros, mensurables mayoritariamente mediante experimentación directa, están orientados a la obtención de una medición que dé cuenta de una determinada fenomenología del campo acústico en un recinto. La casi innumerable lista de parámetros existentes ha puesto en evidencia la necesidad de unificar criterios y establecer un conjunto de parámetros estandarizados para que las comparaciones entre recintos se puedan realizar de forma confiable. Actualmente se ha alcanzado el consenso en una serie de parámetros estandarizados en (UNE EN ISO 3382:1, 2010). Para una revisión histórica de las métricas que han ido surgiendo a lo largo aproximadamente un siglo de estudio puede verse (Lacatis, y otros, 2008). La actual tendencia, como ya se ha mencionado, al diseño de recintos para usos múltiples supone un reto desde el punto de vista del diseño acústico, dado que los indicadores que pueden establecer una determinada idoneidad de la sala para un uso concreto pueden ser contraproducentes al cambiar el uso de la sala. Esto ha dado lugar a establecer una primera división entre valoraciones para la música y valoraciones para la palabra, si bien esta división es en gran medida ampliable en función de la tipología musical o de la palabra. Generalmente se utilizan las mismas métricas objetivas variando aspectos como variables de tiempo, como sucede, por

ejemplo, al evaluar el parámetro de calidad para la música o para la palabra. También existen parámetros dedicados exclusivamente a uno de estos ámbitos, como es el caso del STI, que se describirá en este capítulo, utilizado exclusivamente en el ámbito del habla.

- c) Vinculados con estos últimos y casi siempre a través de técnicas psicoacústicas, se han derivado un conjunto de parámetros denominados subjetivos, encaminados a la ardua tarea de establecer relaciones entre los resultados ofrecidos por las métricas mencionadas en el anterior epígrafe y las descripciones cualitativas aportadas por personas que forman parte del proceso. Así, las evaluaciones cualitativas se han realizado desde el punto de vista del oyente o espectador en el recinto, de los músicos o ejecutantes en el caso de recintos para formaciones musicales, o de los actores u oradores, en el caso de salas de teatro o recintos para conferencias. Los indicadores subjetivos, por definición han de estar condicionados al uso de la sala, dado que no se puede evaluar algo subjetivamente sin la presencia en el contexto real de lo que se está evaluando. Así, una serie de valoraciones subjetivas para la descripción de la calidad de una sala para música sinfónica no tienen prácticamente ninguna aplicación en el ámbito de una sala de teatro, dadas las grandes diferencias existentes en primer lugar entre las características de las fuentes sonoras, y en segundo lugar en las expectativas de los participantes y los oyentes según el evento que tenga lugar. Los estudios subjetivos han generado tres tipos de resultados:

1. Se han establecido parámetros subjetivos que pueden calcularse a partir de los resultados obtenidos en las mediciones de parámetros objetivos. Así, por ejemplo, descriptores subjetivos como la “vivacidad” de una sala, o su “calidez” se calculan a partir de los valores del tiempo de reverberación de una sala en unas determinadas bandas de frecuencia. En el caso de la vivacidad estamos hablando del tiempo medio de reverberación de la sala a frecuencias medias (500 Hz y 1 kHz) y en el caso de la calidez, del refuerzo relativo de la baja frecuencia, estableciendo una relación entre el RT a 125 Hz y 250 Hz y el RT a frecuencias medias.
2. Otro resultado arrojado por los estudios cualitativos es el establecimiento de un intervalo de valores adecuados para cada parámetro objetivo según el tipo de uso que se le dé a un recinto. Así, al establecer comparaciones entre los resultados ofrecidos por los parámetros objetivos y las valoraciones de los usuarios mediante encuestas o experimentos psicoacústicos, se llega a determinar qué intervalos de magnitud de los parámetros objetivos corresponden a una valoración satisfactoria de la sala o los eventos sonoros presentados en ésta por parte de los sujetos bajo experimentación. De esta forma, resulta muy habitual ver, tanto en la literatura como en documentación

referente a procedimientos de medida, los valores recomendados de un determinado parámetro en función del uso de la sala. Por ejemplo, un intervalo de valores recomendado para el parámetro de la Claridad musical media  $C_{80}$  (promediado en 500 Hz, 1 kHz y 2 kHz) está entre -2 y 2 dB. A partir de estas referencias los resultados obtenidos en las métricas objetivas pueden utilizarse para establecer un juicio cualitativo.

3. El tercer aspecto se podría definir como la asignación de descriptores conceptuales a algunos de los parámetros objetivos, de forma que posibilite una traducción al lenguaje ordinario de la abstracción en que se basa una métrica. Términos como “intimidad”, “ancho aparente de la fuente sonora”, “sensación de envolvente del oyente” están asociados a parámetros objetivos como *Initial Time Delay Gap* (ITDG), *Aparent Source Width* (ASW) o *Listener Envelopment* (LEV). En el caso de la “intimidad”, por ejemplo, el término ayuda a dar una descripción cualitativa respecto de una medición estrictamente cuantitativa, como es el intervalo de tiempo entre el sonido directo y la primera reflexión. A priori, una medición estrictamente temporal como es ésta, definida objetivamente, sólo puede hacer referencia al tiempo que separa la llegada del sonido reflejado respecto del directo. A la hora de relacionar esta medición con una valoración cualitativa es cuando surge la necesidad de acuñar términos que expresen de una manera cercana y comprensible por los individuos encuestados aspectos de la experiencia perceptiva a la que se les somete. Siguiendo con los ejemplos propuestos, los parámetros ASW y LEV están definidos objetivamente mediante el uso de la correlación cruzada entre señales binaurales, *Interaural Cross Correlation* (IACC). La definición de los parámetros  $IACC_E$  (*early*) e  $IACC_L$  (*late*) hace referencia a la consideración de las señales binaurales en un período de 0 a 80ms, o entre 80ms y 500ms, respectivamente. Como se ha explicado en el apartado 2.2.4 estas mediciones se han asociado a dos sensaciones perceptivas que describen subjetivamente la experiencia del oyente. En el primer caso, la percepción por parte del espectador de la amplitud –en sentido espacial– de la fuente, y en el segundo la caso, la mayor o menor sensación que pueda tener el oyente el oyente de sentirse inmerso en el campo acústico.

#### **4.2 El problema de la correlación entre parámetros**

El tercer aspecto citado en el apartado anterior es de suma importancia a la hora de intentar establecer correlaciones entre parámetros objetivos y subjetivos. La definición de un término en relación con una determinada sensación perceptiva no siempre está clara, de forma que a la hora de realizar experimentos psicoacústicos para establecer

estas relaciones entre la medida y la impresión perceptual debe establecerse un mínimo consenso respecto del significado de cada término.

#### 4.2.1 Una perspectiva global

Históricamente se han realizado trabajos muy amplios con la finalidad de establecer un conjunto de parámetros que sintetizan la impresión cualitativa de un recinto (Barron, 1993) (Beranek L. , 1996). Este último propone un listado de hasta 18 atributos para salas de audición musical, basándose en consideraciones que atienden tanto a la audiencia como a los ejecutantes. La lista de estos 18 atributos con una breve descripción se enumera en la Tabla 4.1.

**Tabla 4.1. Atributos subjetivos propuestos en (Beranek L. , 1996)**

<b>ATRIBUTO</b>	<b>DESCRIPCIÓN</b>
<b>Intimidad / Presencia</b>	Asociado a la impresión de espacio pequeño, íntimo
<b>Vivacidad</b>	Refuerzo del sonido a frecuencias medias. Asociado al tiempo de reverberación. En contraposición, una sala sin reverberación, en ocasiones, se denomina una sala “muerta”.
<b>Ancho aparente de la fuente</b>	Impresión sobre el ancho espacial de la fuente.
<b>Envolvente</b>	Asociado a la sensación por parte del oyente de sentirse inmerso en el sonido o rodeado por éste.
<b>Claridad</b>	Relacionado con la capacidad para distinguir entre los diferentes sonidos generados por un instrumento
<b>Calidez</b>	Refuerzo del sonido en baja frecuencia. Un exceso de éste puede definir una sala como “oscura”
<b>Sonoridad</b>	Nivel de sonoridad percibido en el punto de audición
<b>Deslumbramiento (<i>Acoustic glare</i>)<sup>6</sup></b>	Generalmente causado por reflexiones laterales en superficies planas, que producen reflexiones cuasi especulares. Un aumento de la difusión en las superficies reduce este efecto.
<b>Brillo</b>	Asociado a un refuerzo del sonido en la alta frecuencia.
<b>Equilibrio</b>	Se entiende por equilibrada una situación en la que todas las fuentes sonoras son audibles para el oyente.
<b>Mezcla</b>	Definido como la mezcla de sonidos en la posición de escucha.
<b>Conjunto (<i>ensemble</i>)<sup>7</sup></b>	Referido a la capacidad de los ejecutantes de escucharse unos a otros con claridad.
<b>Inmediatez de respuesta</b>	Relacionado con la forma en la que los intérpretes perciben la respuesta de la sala. Si la respuesta contiene reflexiones intensas y con cierto retardo puede afectar negativamente en la ejecución.
<b>Textura</b>	Relacionado con el patrón temporal que se obtiene a partir de las reflexiones tempranas en el recinto.

---

<sup>6</sup> Traducción literal, al no encontrar un término equivalente en español en la literatura.

<sup>7</sup> Ídem

<b>Ausencia de eco</b>	En general la presencia de eco, es decir, percibir un mismo evento sonoro como dos eventos diferentes se considera perjudicial.
<b>Rango dinámico y ruido de fondo</b>	El límite inferior del rango dinámico lo constituye el ruido de fondo de la sala. El límite superior vendrá determinado por la sonoridad de las fuentes y la respuesta de la sala.
<b>Efectos no deseados en la calidad tonal</b>	Referente a sonidos producidos por la propia sala, que son no deseables.
<b>Uniformidad</b>	Asociado a la calidad tonal del sonido en todas las posiciones de escucha.

Algunos de estos atributos tienen actualmente un uso habitual y bastante consensuado, como pueden ser el brillo, la calidez o la intimidad. Otros han pasado directamente a formar parte del conjunto de atributos que, relacionados con una magnitud acústica medible experimentalmente, se han estandarizado en la norma (UNE EN ISO 3382:1, 2010). Concretamente, son la sonoridad, la vivacidad (relacionada con el tiempo de reverberación), la claridad, el ancho aparente de la fuente y la percepción de envolvente. Sin embargo, hay otros que no generan un consenso genérico entre los oyentes. Del conjunto de magnitudes acústicas, su ponderación en frecuencia establecida y sus atributos perceptivos asociados en la norma ISO 3382:1 da cuenta la siguiente tabla.

**Tabla 4.2. Parámetros estandarizados en ISO 3382:1**

<b>Atributo perceptivo</b>	<b>Magnitud Acústica</b>	<b>Promediado en frecuencia</b>	<b>Umbral diferencial (JND)</b>
Sonoridad	Fuerza sonora $G$ , en dB	500, 1000	1 dB
Reverberación percibida	EDT, en s	500, 1000	5%
Claridad del sonido percibida	Claridad $C_{80}$ , en dB	500, 1000	1 dB
	Definición $D_{50}$	500, 1000	0,05
	Tiempo central $T_s$ , en ms	500, 1000	10 ms
Ancho aparente de la fuente	Fracción de energía lateral precoz, $J_{LF}$	125 - 1000	0,05
Envolvente del oyente	Nivel sonoro lateral final, $L_J$ , en dB	125 - 1000	Desconocido

En la Tabla 4.2 todos los promediados en frecuencia son aritméticos para las bandas de octava, excepto  $L_J$ , que se promedia energéticamente.

Podría pensarse en un primer momento, que, dada la disponibilidad de un número de métricas notablemente mayor, la norma establece un relativamente pequeño conjunto de parámetros para evaluar un recinto de forma global. El estándar establece que la descripción acústica del recinto puede determinarse con el promediado de estos parámetros obtenidos en diferentes posiciones de medida. Las principales críticas que se

le han hecho al estándar, sin embargo, no se fundamentan en el número de magnitudes acústicas utilizadas, sino en otros aspectos. En primer lugar, por una indefinición a la hora de establecer los algoritmos para el cálculo de los parámetros. En segundo lugar por realizar estimaciones basadas en un ancho de banda demasiado estrecho, en comparación con la audición humana (Bradley, 2011) y finalmente por el uso de una única fuente omnidireccional como fuente de excitación. En (Kirkegaard & Gulsrud, 2011) se considera, además del ya mencionado escaso ancho de banda, que una única fuente de excitación omnidireccional no resulta equiparable a una orquesta como fuente acústica. En (Bradley, 2011) se hace especial hincapié en la necesidad de profundizar en el conocimiento de los datos perceptivos y la relevancia de los umbrales diferenciales (JND) en función de la frecuencia.

Otros ejemplos en la literatura en que las magnitudes acústicas ofrecen una pobre correlación con las valoraciones perceptivas los podemos encontrar en (Barron, 1988), (Farina, 2001), o (Lokki, Vertanen, Kuusinen, Pätynen, & Tervo, 2010)

#### *4.2.2 Nuevas perspectivas de aproximación*

Recientemente han aparecido nuevas perspectivas desde las que abordar el problema de la correlación entre la valoración cualitativa y las magnitudes acústicas.

Una de ellas incorpora directamente un modelo perceptual para la derivación de las magnitudes acústicas a través de éste (van Dorp Schuitman, 2011). Se utilizan respuestas al impulso obtenidas con la recreación de campos acústicos sintetizados. Variando las características acústicas del recinto virtual pueden obtenerse un conjunto diferenciado de respuestas al impulso. Estas respuestas al impulso se someten posteriormente a un modelo perceptual binaural, con lo que el resultado será una representación interna del sistema auditivo modelado de la respuesta al impulso. Las magnitudes acústicas se obtienen posteriormente a partir de estas representaciones. Esta es una de las líneas que resulta más interesante desde el punto de vista de la relación cualitativa-cuantitativa. Si los modelos pudiesen ofrecer resultados tan satisfactorios como ofrecen sus homónimos monoaurales en el ámbito de la codificación, cabría esperar que las magnitudes derivadas a partir de las representaciones internas de la respuesta al impulso guardaran una estrecha relación con los atributos cualitativos. Se abre en este sentido todo un ámbito de investigación cuyos principales retos a solventar son el establecimiento de un modelo auditivo binaural consensuado y los trabajos psicoacústicos de validación.

Otra aproximación desde una perspectiva novedosa es la aportada en (Lokki, 2014). Existen diversas técnicas para la realización de encuestas sobre las cualidades perceptivas referentes a salas para la interpretación musical. Algunas de ellas son la realización de encuestas *in situ* o la escucha de grabaciones binaurales realizadas en las propias salas. Otro procedimiento es la obtención de la respuesta al impulso en varias posiciones de la sala y posteriormente la convolución de una señal anecoica con esa respuesta al impulso, escuchándola con auriculares. En (Lokki, 2014) se lleva esta última técnica bastante más lejos. Se modela una orquesta colocando en el escenario 34

altavoces colocados según las características directivas de los instrumentos que representan. Éstos se utilizan como fuentes acústicas para la obtención de la respuesta al impulso en diferentes posiciones de la sala utilizando un *array* de micrófonos orientados en 6 direcciones: hacia arriba, abajo, hacia derecha e izquierda, hacia el escenario y la parte posterior del recinto. Paralelamente se realiza una grabación anecoica de una orquesta entera instrumento por instrumento. Esta grabación se convoluciona con las respuestas al impulso obtenidas en todas las direcciones en cada punto. La reproducción se realiza en una sala con 24 monitores que reproducen el campo acústico obtenido a partir de la convolución anterior. La principal ventaja de este costoso – en todos los sentidos – procedimiento es que el sujeto bajo test puede cambiar el campo acústico percibido con un solo *click*, es decir, puede comparar diferentes recintos de manera inmediata.

Lokki añade un elemento novedoso más: no se le pide a las personas bajo test que evalúen un determinado aspecto de una sala sino que sean ellos los que aporten los términos para describir la sala, en lo que denomina evaluación sensorial – haciendo una analogía con los test de calidad de alimentos -. Con esto, el autor recopila una serie de términos descriptivos de las impresiones de los encuestados, que posteriormente agrupa bajo diferentes categorías cualitativas. Una de las principales conclusiones del estudio radica en que no siempre se utilizan los mismos términos para describir una misma sensación perceptiva, lo cual hace pensar que en los procedimientos de experimentación psicoacústica puede suceder, que cuando se les pide a diferentes personas que evalúen un determinado aspecto de la audición, estén evaluando cualidades diferentes unos de otros, lo cual nos lleva a retomar las palabras con que iniciábamos esta sección 4.2

### 4.3 La respuesta al impulso

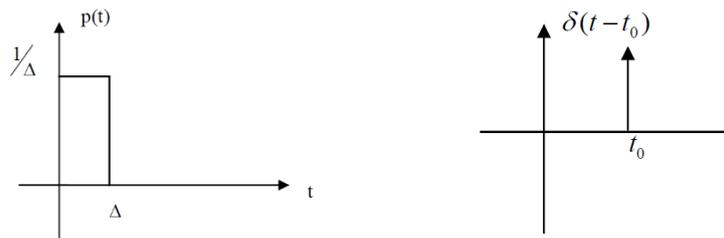
Desde que en (Schroeder, 1965) se definiera el método para la medición del tiempo de reverberación a partir de la respuesta al impulso de la sala, éste método se ha ido convirtiendo progresivamente en el preponderante a la hora de obtener la medición no sólo del tiempo de reverberación sino de otras muchas magnitudes acústicas. El uso de la respuesta al impulso se basa en la aproximación de considerar la sala como un sistema lineal e invariante con el tiempo (LTI), con las características que se definían en la sección 3.3.

La función impulso  $\delta(t)$  es una función ideal que no existe en la realidad. Es ideal porque tiene anchura cero y amplitud infinita. Para su generación en la práctica se han utilizado diferentes métodos orientados a obtener un ruido intenso de muy corta duración y con el mayor ancho de banda posible. Históricamente se utilizaban armas de fuego, y aún hoy el estándar ISO 3382 habla del uso de estos dispositivos. El principal problema que ofrecía este método es su nula reproductibilidad, ya que las variaciones entre diferentes disparos pueden ser muy significativas. Actualmente, las dos técnicas más extendidas para la obtención de la respuesta al impulso son por un lado las basadas en *sweep* sinusoidales, formadas por una señal sinusoidal que aumenta progresivamente de frecuencia de forma que hace un barrido por todo el ancho de banda deseado (Berkhout,

de Vries, & Boone, 1980) (Farina, 2007) y por otra parte las basadas en secuencias pseudo-aleatorias MLS (*Maximum-Length Sequences*) (Rife & Vanderkooy, 1989). Otra técnica de obtención de la respuesta al impulso es mediante impulsos de duración extendida (*Time Stretched Pulses*) y la técnica IRS (*Inverse Repeated Sequence*), que es una variante de la técnica MLS que mejora a priori algunos problemas de distorsión. Las particularidades, ventajas e inconvenientes de estas técnicas pueden compararse en (Stan, Embrechts, & Archambeau, 2002).

La función  $\delta(t)$  puede definirse como muestra la Figura 4.1. Representación para la definición de la función impulso. La función impulso desplazada  $t_0$  s.. Es el límite de la función pulso  $p(t)$  cuando su anchura en el tiempo  $\Delta$  tiende a 0 y su amplitud  $1/\Delta$  tiende a infinito.

$$\delta(t) = \lim_{\Delta \rightarrow 0} p(t) \quad (4.1)$$



**Figura 4.1. Representación para la definición de la función impulso. La función impulso desplazada  $t_0$  s.**

La función impulso cumple las siguientes propiedades

$$\int_{-\infty}^{\infty} \delta(\tau) d\tau = 1$$

$$\int_{-\infty}^{\infty} \delta(t - t_0) dt = 1$$

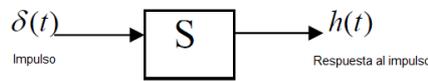
Para una señal  $x(t)$  se cumple que al multiplicar esa señal por la función impulso desplazada  $t_0$  se obtendrá el valor de la función  $x(t)$  en el instante  $t_0$ .

$$\int_{-\infty}^{\infty} x(t) \delta(t - t_0) dt = \int_{-\infty}^{\infty} x(t_0) \delta(t - t_0) dt = x(t_0) \int_{-\infty}^{\infty} \delta(t - t_0) dt = x(t_0)$$

Una consecuencia inmediata de la propiedad de desplazamiento es que cualquier señal de entrada arbitraria  $x(t)$  puede representarse como una suma ponderada de  $x(\tau)$  e impulsos desplazados  $\delta(t-\tau)$ . De esta manera, cualquier señal continua en el tiempo se puede representar de la forma:

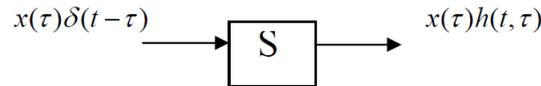
$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau \quad (4.2)$$

Para un sistema S una señal de impulso de entrada proporcionará en la salida la señal respuesta al impulso  $h(t)$ .



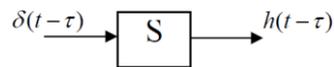
**Figura 4.2. Representación para la definición de la función  $h(t)$  respuesta al impulso.**

Si S es un sistema lineal, la salida para un instante  $\tau$  será proporcional a  $h(t, \tau)$ , tal como muestra la Figura 4.3.



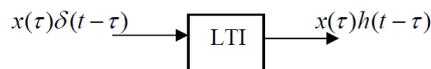
**Figura 4.3. Sistema Lineal.**

Si además, el sistema S es invariante con el tiempo, un desplazamiento  $t_0$  del impulso en la entrada se corresponderá con un desplazamiento  $t_0$  de la función respuesta al impulso en la salida del sistema.



**Figura 4.4. Sistema invariante en el tiempo.**

Es decir, si S es un sistema LTI



**Figura 4.5. Sistema LTI.**

De forma que cualquier entrada arbitraria al sistema LTI generará una salida como la indicada en la Figura 4.6.

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau \rightarrow \text{LTI} \rightarrow y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau$$

**Figura 4.6. Salida de un sistema LTI a una entrada arbitraria  $x(t)$ .**

En definitiva, en un sistema lineal invariante en el tiempo, la salida de éste vendrá dada por la convolución de la señal arbitraria de entrada con la respuesta al impulso del sistema.

$$y(t) = x(t) * h(t)$$
$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau \quad (4.3)$$

Para cualquier instante de tiempo  $t_0$  la salida vendrá dada por

$$y(t_0) = \int_{-\infty}^{\infty} x(\tau)h(t_0 - \tau)d\tau \quad (4.4)$$

La transformada de Laplace de la respuesta al impulso  $h(t)$  particularizada para el caso  $s=j\omega$  (transformada de Fourier) proporciona la respuesta en frecuencia del sistema LTI.

$$H(s) = \mathcal{L}\{h(t)\} = \int_{-\infty}^{\infty} h(t)e^{-st} dt$$
$$H(j\omega) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t)e^{-j\omega t} dt \quad (4.5)$$

En el dominio de la frecuencia, si obtenemos la transformada de Fourier de una señal de entrada arbitraria  $x(t)$  al sistema, la salida se obtendrá mediante

$$X(j\omega) = \mathcal{F}\{x(t)\}$$
$$Y(j\omega) = \mathcal{F}\{y(t)\} \quad (4.6)$$
$$Y(j\omega) = H(j\omega)X(j\omega)$$

Es decir, la convolución en el dominio del tiempo tiene su equivalente en el producto en el dominio de la frecuencia. El espectro de la señal de salida del sistema vendrá dado por el producto entre el espectro de la señal de entrada y la respuesta en frecuencia del sistema. Las propiedades descritas, entre otras, hacen que los sistemas lineales LTI sean fácilmente modelables y analizables. Cuando un fenómeno físico puede modelarse como un sistema lineal, el comportamiento de éste queda totalmente descrito por su modelo. En realidad, la mayoría de procesos y fenómenos físicos reales tienen un comportamiento no lineal. En muchos casos, si el efecto de la no linealidad del proceso no contribuye de forma clara en éste, resulta más cómodo modelar el fenómeno como un sistema lineal teniendo en cuenta las limitaciones de éste (Rugh, 1981).

En acústica de salas, en una respuesta al impulso se distinguen tres partes fundamentales:

- a) Sonido directo. Se considera sonido directo a toda la energía llegada al receptor desde la llegada del frente de onda hasta los primeros 20ms, debido a las características de integración del sistema auditivo humano.
- b) Reflexiones precoces. Esta parte de la respuesta al impulso caracteriza la llegada de reflexiones precoces e influye determinantemente en la claridad. Generalmente se suele considerar el tramo entre 20 y 80-100ms
- c) Cola de reverberación, reverberación tardía o final. A partir de los 100ms la respuesta ya no se describe en términos de reflexiones discretas.

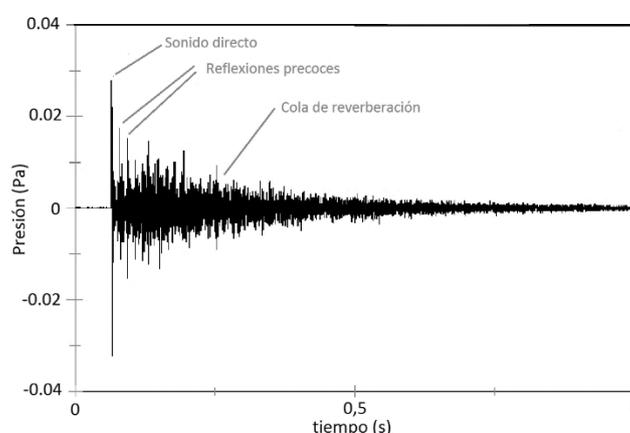


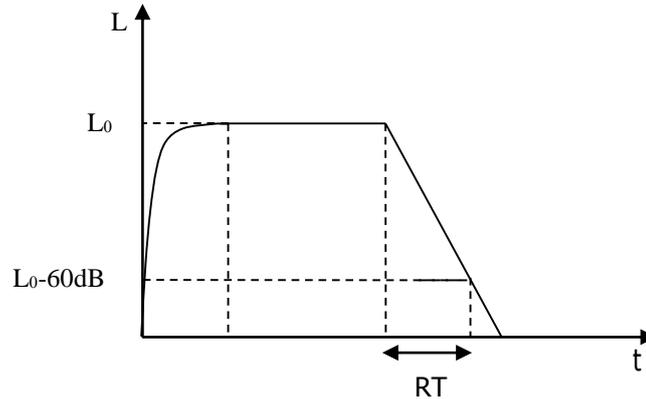
Figura 4.7. Ejemplo de respuesta al impulso

## 4.4 Parámetros temporales

### 4.4.1 Tiempos de reverberación

Si en una sala una fuente acústica está emitiendo un sonido constante y continuo, éste produce un determinado nivel de presión sonora en el interior de la sala,  $L_0$  (dB). Si se produce un cese de la emisión del sonido por parte de la fuente se producirá un descenso gradual de este nivel de presión sonora. La manera en que se producirá este descenso vendrá determinada por las características de la sala, tanto geométricas, como de absorción de los materiales que revisten sus superficies. Se define como tiempo de reverberación al tiempo que transcurre desde el momento en que la fuente cesa la emisión hasta que el nivel de presión sonora en la sala es de  $L_0-60$  (dB), es decir, el tiempo en que el nivel de presión tarda en descender 60 decibelios. Esta es la definición física del parámetro. En la práctica, en ocasiones no es fácil conseguir un margen dinámico de 60 dB, según las condiciones de ruido de fondo y de las fuentes utilizadas para la excitación, con lo cual se considera la caída entre -5 y -35 dB, es decir, considerando un margen dinámico de 30 dB. La pendiente de caída considerada en ese intervalo es la que se extrapola para obtener el tiempo de caída de 60 dB. Esto ha llevado a la definición de

otros tiempos de reverberación, en función del intervalo de caída considerado para extrapolar a una caída de 60 dB.



**Figura 4.8.** Caída de 60 dB respecto del nivel estacionario.

Al representar una de las magnitudes de forma logarítmica, la caída se puede aproximar mediante una recta, definida por una pendiente. Así, se definen el  $T_{30}$ ,  $T_{20}$ ,  $T_{15}$  y el EDT (*Early Decay Time*). El tramo de pendiente de caída considerado en cada caso se resume en la Tabla 4.3.

**Tabla 4.3.** Tramos de la caída energética considerados en los tiempos de reverberación.

Tiempo	Tramo considerado
RT	-5 a -35 dB
$T_{30}$	0 a -30 dB
$T_{20}$	0 a -20 dB
$T_{15}$	-5 a -15 dB
EDT	0 a -10 dB

De todos los anteriores, los más frecuentemente utilizados y estandarizados en ISO 3382:1 son el tiempo de reverberación (RT) y el *Early Decay Time* (EDT). De acuerdo con el propio estándar, el EDT está más relacionado con la percepción de la reverberación, mientras que el tiempo de reverberación lo está con el comportamiento físico de la sala. De acuerdo con (Schroeder, 1965), el tiempo de reverberación puede calcularse a partir de la respuesta al impulso mediante la obtención de la curva de decaimiento energético de la sala (EDC – *Energy Decay Curve*), que se obtiene mediante integración hacia atrás de la respuesta al impulso en términos energéticos. El gradiente

de la caída energética entre los valores correspondientes mostrados en la tabla anterior es el que permite realizar la extrapolación a correspondiente a una caída de 60 dB.

La curva de caída energética viene dada por la expresión

$$EDC(t) = \int_t^{\infty} h^2(\tau) d\tau \quad (4.7)$$

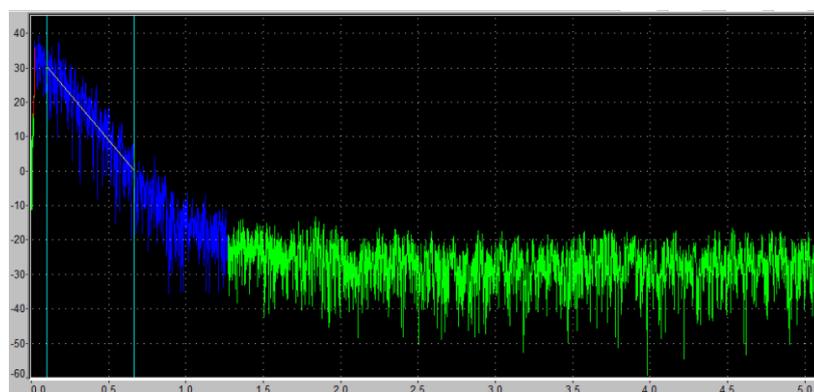


Figura 4.9. Gradiente considerado entre -5 y -35 dB para el cálculo de RT.

La Figura 4.9 muestra una curva EDC(t) y el tramo de caída considerado –con un margen dinámico de 30 dB- para realizar el cálculo del tiempo de reverberación.

#### 4.4.1.1 Atributos cualitativos asociados al tiempo de reverberación

Existen una serie de descriptores cualitativos vinculados al concepto de tiempo de reverberación, directamente calculables a partir de éste. Los más significativos son los siguientes.

- a) Viveza. Una sala anecoica frecuentemente se describe como una sala “muerta”. El adjetivo se debe a que ejerce muy poca o ninguna alteración del campo acústico producido por la fuente sonora. En la medida que disminuya o no exista reverberación, la sala cada vez se parecerá más a las condiciones dadas en campo libre. La manifestación de la sala mediante la presencia de reverberación, generalmente a frecuencias medias, se relaciona con la viveza de ésta. Está asociada a la media aritmética del tiempo de reverberación para las frecuencias de 500 Hz y 1000 Hz.

$$RT_{mid} = \frac{RT_{500\text{ Hz}} + RT_{1000\text{ Hz}}}{2} \quad (4.8)$$

- b) Curva tonal. En ocasiones se oye la expresión de que una sala “colorea” el sonido. Esta sensación está asociada al diferente comportamiento de la

reverberación en la sala en función de la frecuencia. El tiempo de reverberación es diferente en las diversas bandas de octava o tercios de octava en que se mide. Si estas diferencias son muy acusadas en un rango determinado de frecuencias – en el sentido de que los tiempos sean sensiblemente mayores en ese rango –, estas frecuencias serán más audibles, dando lugar a la sensación perceptiva de que la sala “colorea” el sonido. La Figura 4.10 muestra un ejemplo de curva tonal medida en un punto de la Sala de Actos de la Casa de la Cultura de l’Alqueria de la Comessa.

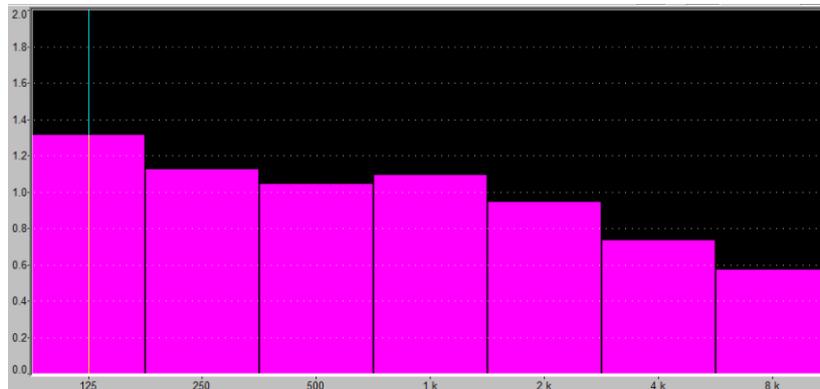


Figura 4.10. Ejemplo de curva tonal - RT (s) vs Banda de octava (Hz) -.

- c) Calidez. En ocasiones también denominada *Bass Ratio* (BR), se trata de una descripción cualitativa asociada al refuerzo de la baja frecuencia, debido a unos tiempos de reverberación mayores para las octavas de 125 y 250 Hz que para las de 500 y 1000 Hz. Cualquier resultado superior a 1 en la 4.9 significará una mayor reverberación en la baja frecuencia. Si ésta es excesiva, la sala puede calificarse como “oscura”. Se define como:

$$BR = \frac{RT_{125\text{ Hz}} + RT_{250\text{ Hz}}}{RT_{500\text{ Hz}} + RT_{1000\text{ Hz}}} \quad (4.9)$$

- d) Brillo. Es un atributo que se define análogamente al anterior, pero en este caso considerando el refuerzo relativo de la alta frecuencia respecto a las medias – de ahí su asociación con el brillo -. Se define como:

$$Br = \frac{RT_{2000 \text{ Hz}} + RT_{4000 \text{ Hz}}}{RT_{500 \text{ Hz}} + RT_{1000 \text{ Hz}}} \quad (4.10)$$

#### 4.4.2 Initial Time Delay Gap (ITDG)

Se define así el tiempo transcurrido desde la llegada del sonido directo al receptor y la llegada a éste de la primera reflexión procedente de alguno de los cerramientos del recinto. Fue introducido por (Davis, 1979) y considerado en ocasiones para salas de audición y en el diseño de salas de monitorización o control de grabación. Sin embargo, el parámetro ha mostrado mayor correlación con la inteligibilidad de la palabra. De la lista de atributos cualitativos mostrados en la Tabla 4.1, el vinculado con esta magnitud temporal es el de intimidad, o sensación de sala pequeña o proximidad de la fuente, que será mayor a medida que disminuya ITDG. En salas de conciertos este valor suele oscilar entre los 15 y los 30ms (Beranek L. , 1996).

### 4.5 Parámetros energéticos

Se trata de un conjunto de métricas que calculan relaciones energéticas delimitadas por una constante de tiempo. La primera de estas magnitudes se introdujo en (Thiele, 1953) y se denominó Definición. Posteriormente se han definido diferentes magnitudes basadas en relaciones energéticas.

#### 4.5.1 Definición

Establece la relación entre la energía llegada durante los primeros 50 ms y el total de la energía recibida en un punto de la sala. Se calcula como

$$D = 10 \log \frac{\int_0^{50ms} P^2(t) dt}{\int_0^{\infty} P^2(t) dt} \quad (dB) \quad (4.11)$$

Inicialmente fue establecido como medida de inteligibilidad de la palabra en salas. En ocasiones no se expresa en dB, sino en %, indicando el porcentaje de energía recibida en los primeros 50ms sobre el total. Su uso está siendo sustituido progresivamente por la magnitud  $C_{50}$ . La definición está asociada a un juicio subjetivo similar al de la claridad, especialmente en el caso de la palabra (distinción).

#### 4.5.2 Claridad $C_{80}$

Se define como la relación entre la energía llegada durante los primeros 80ms y el resto de energía que llega al punto de medida a partir de los 80ms. Viene dada por

$$C_{80} = 10 \log \frac{\int_0^{80ms} P^2(t) dt}{\int_{80}^{\infty} P^2(t) dt} \quad (dB) \quad (4.12)$$

Estableciendo unas condiciones de campo difuso, puede obtenerse un valor teórico estimado de la claridad, mediante

$$C_{80EST} = 10 \log \frac{\left(\frac{r_H}{r_x}\right)^2 + 1 - e^{-\frac{13.8 \cdot 0.08}{RT}}}{e^{-\frac{13.8 \cdot 0.08}{RT}}} \quad (dB) \quad (4.13)$$

Donde  $r_H = 0.057 \sqrt{\frac{V}{RT}}$ , siendo V el volumen de la sala en metros cúbicos y RT el tiempo de reverberación en segundos. La distancia  $r_x$  es la distancia que separa la fuente sonora del receptor, en metros.

Subjetivamente, está asociada a la capacidad para distinguir entre los sucesivos sonidos generados por un instrumento, en el ámbito de salas para conciertos.

#### 4.5.3 Claridad $C_{50}$

Una variante de la claridad asociada a la evaluación de la palabra hablada se define de la misma forma cambiando la constante de tiempo por 50ms. En ocasiones, también se la denomina definición  $C_{50}$

$$C_{50} = 10 \log \frac{\int_0^{50ms} P^2(t) dt}{\int_{50}^{\infty} P^2(t) dt} \quad (dB) \quad (4.14)$$

En condiciones de campo difuso puede utilizarse la misma ecuación 4.13 cambiando la constante de tiempo de 0.08 a 0.05 s para obtener un valor estimado de esta magnitud. Como ya se ha mencionado, perceptivamente está asociada a la inteligibilidad de la palabra, considerando el límite inferior de -2 dB el mínimo aceptable para una buena inteligibilidad.

#### 4.5.4 Tiempo central $T_s$

Viene definido por

$$T_s = \frac{\int_0^{\infty} t \cdot P^2(t) dt}{\int_0^{\infty} P^2(t) dt} \quad (4.15)$$

Lo que corresponde a un cálculo del centro de gravedad de la curva de caída energética. Al igual que las dos magnitudes anteriores, del tiempo central también puede obtenerse un valor teórico en condiciones de campo difuso, dado por

$$T_{Sest} = \frac{RT}{13.8} \quad (ms) \quad (4.16)$$

Donde RT es el tiempo de reverberación en ms. El parámetro fue establecido por (Kürer, 1971), donde señalaba también una correlación con la inteligibilidad silábica  $V_s$  (%) en un recinto que venía dada por

$$V_s = 96 \cdot (1 - T_s^2 \cdot 10^{-5}) \quad (\%) \quad (4.17)$$

Donde  $T_S$  se expresa en ms. De la expresión anterior se deduce que para una inteligibilidad silábica superior al 80%, es decir, considerada de buena a muy buena, se requiere que  $T_S \leq 130$ ms.

#### 4.5.5 Fuerza sonora $G$

La magnitud *Strength* o fuerza sonora (Lehmann, 1976) hace referencia al incremento de sonoridad introducido por la sala al realizar la comparación de los niveles de presión sonora producidos por una misma fuente omnidireccional en el interior de una sala y en campo libre a una distancia de 10m.

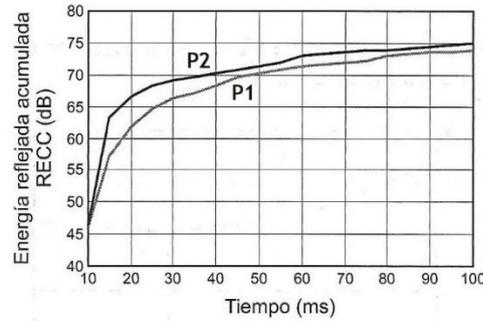
$$G = 10 \log \frac{\int_0^{\infty} P^2(t) dt}{\int_0^{\infty} P_{10}^2(t) dt} \text{ (dB)} \quad (4.18)$$

En la que  $P_{10}$  es la presión sonora de la misma fuente medida en campo libre a 10 metros de distancia. Perceptualmente está asociada a la sonoridad percibida, estando el rango típico de la magnitud entre -2 y 10 dB. Generalmente se recomienda un valor  $G > 0$  dB.

#### 4.5.6 Curvas de energía reflejada acumulada

Las curvas RECC (*Reflection Energy Cumulative Curve excepting direct sound*) grafían la energía reflejada acumulada hasta un instante  $t$ , generalmente 80ms, sin considerar el sonido directo (Toyota, Oguchi, & Nagata, 1988). Las curvas, hasta aproximadamente los 80ms muestran diferentes formas en función de las características del recinto y de los puntos de medición dentro de éste. A partir de los 80ms, la forma de las curvas tiende a converger para todos los recintos. Casi la totalidad de la energía reflejada viene determinada por el parámetro ARE80 (energía reflejada hasta los 80ms).

Son de esperar variaciones significativas en función de la posición de medida dentro de la sala. Así, un punto de medición más cercano a un cerramiento, como es el caso de P2 en la Figura 4.11 acumulará energía de una forma más precoz que un punto más centrado en la sala (P1). Este parámetro mostrará importantes variaciones en función del punto de medida dentro de la sala, independientemente del tiempo de reverberación, que puede ser similar o incluso idéntico para dos puntos donde el parámetro ARE80 varíe significativamente.



Fuente: (Carrión, 1998)

Figura 4.11. Ejemplo de curvas RECC para dos puntos diferentes.

## 4.6 Parámetros espaciales

Todos los parámetros considerados hasta ahora son monoaurales. Establecen un análisis de una señal en un punto a partir de un criterio estrictamente temporal o dividen esa misma señal en función de una constante de tiempo para determinar relaciones energéticas entre esas fracciones de señal. Las métricas espaciales, orientadas a describir la impresión espacial percibida por el oyente se basarán en el análisis de dos señales. Este análisis se abordará desde dos perspectivas fundamentales: por una parte, el establecimiento de relaciones energéticas entre dos señales, o dos respuestas al impulso que llevan implícitas características direccionales (como es el caso de la fracción de energía lateral o la eficiencia lateral). Por otra parte, la obtención de respuestas al impulso binaurales – emulando la llegada del sonido a los dos oídos del receptor – permitirán establecer una comparación directa entre las dos señales mediante técnicas de correlación. La medida de las diferencias entre la señal recibida en uno y otro oído aportará información sobre la percepción de espacialidad, como es el caso de las métricas IACC (*InterAural Cross Correlation*).

### 4.6.1 Fracción de energía lateral

*Lateral Fraction* (fracción de energía lateral), definida en (Barron, 1981) se obtiene a partir de la respuesta al impulso de la sala realizada con un micrófono bidireccional comparándola con la respuesta al impulso medida con un micrófono omnidireccional, de manera que refleja la ratio de la energía lateral en el conjunto de energía recibida en un determinado punto de la sala.

$$J_{LF} = \frac{\int_{5ms}^{80ms} P_8^2(t) dt}{\int_0^{80ms} P^2(t) dt} \quad (4.19)$$

Donde  $P_8$  denota la señal captada por un transductor bidireccional con los lóbulos hacia las direcciones laterales, es decir, con nula respuesta en la dirección frontal a la fuente de sonido. La señal considerada en el denominador de la expresión 4.19 estará captada por un micrófono omnidireccional. De esta manera se establece la proporción de energía procedente desde los laterales sobre la totalidad de energía acumulada en el punto de audición. Según (UNE EN ISO 3382:1, 2010) su valor debe calcularse promediando aritméticamente el resultados en las bandas de octava de 125 Hz a 1000 Hz, fluctuando sus valores típicos entre 0,05 y 0,35.

#### 4.6.2 Eficiencia lateral

En la misma línea (Jordan V. , 1980) propone  $LE$  – (*Lateral Efficiency* – eficiencia lateral -) de forma muy similar a Barron pero están vez considerando la relación entre la energía lateral (respuesta al impulso bidireccional) entre 25 y 80ms y la energía total (respuesta al impulso omnidireccional) considerando para ésta unos límites de integración entre 0 y 80ms.

$$LE = \frac{\int_{25ms}^{80ms} P_8^2(t) dt}{\int_0^{80ms} P^2(t) dt} \quad (4.20)$$

En (Kleiner, 1989) se propuso la consideración del ángulo de incidencia como variación para el cálculo del factor de eficiencia lateral.

En un punto de audición de una sala, las reflexiones procedentes de los cerramientos laterales son de mayor importancia, en comparación con las procedentes de otras direcciones. En el ámbito perceptivo estas magnitudes tienen su influencia en la sensación de ancho aparente de la fuente acústica. Así, cuanto mayor sea la fracción de energía lateral mayor será también el ancho aparente de la fuente (ASW). De acuerdo con (Barron, 1981) la contribución de las reflexiones laterales a las sensaciones perceptivas varía en función de la banda de frecuencias que se evalúe. En el rango hasta 500 Hz influyen sobre la percepción de envolvente. En el rango de 1000 a 4000 Hz definen la impresión de amplitud de la fuente sonora. Por encima de 4000 Hz, si los valores son elevados, pueden influir en la percepción de un desplazamiento de la imagen de la fuente sonora.

Estas dos magnitudes ( $J_{LE}$  y  $LE$ ) focalizan su atención sobre el sonido llegado en los primeros 80 ms. La siguiente magnitud establecerá la misma relación considerando la energía acústica tardía, llegada a partir de los 80ms.

#### 4.6.3 Nivel sonoro lateral final

El nivel sonoro lateral final viene dado por

$$L_J = 10 \log \left( \frac{\int_{80ms}^{\infty} P_8^2(t) dt}{\int_0^{\infty} P_{10}^2(t) dt} \right) \quad (4.21)$$

Considerando igualmente  $P_8$  como la señal o la respuesta al impulso captada bidireccionalmente y  $P_{10}$  la respuesta al impulso de la misma fuente captada omnidireccionalmente en campo libre a una distancia de 10 metros. Es decir, cuanto mayor sea el grado de energía reverberante procedente de los laterales mayor será el valor de  $L_J$ . Generalmente se calcula el promedio energético entre las octavas de 125 a 1000 Hz, considerándose un rango típico entre -14 y +1 dB. Esta magnitud cualitativamente se asocia al establecimiento de la impresión espacial percibida por el espectador, que puede describirse como la sensación de envolvente del sonido, la impresión del tamaño de la sala o de proximidad o no a la fuente.

#### 4.6.4 Correlación cruzada interaural

Como ya se ha anticipado, además de las relaciones energéticas, la otra metodología predominante para la evaluación de la espacialidad se centra en las técnicas de comparación de señales registradas de forma binaural. Para estas grabaciones se utilizan torsos artificiales o *dummies*, cuyas cabezas llevan acopladas unos micrófonos que ejercen la función de “oídos” o bien micrófonos gemelos colocados sobre la cabeza de un oyente real. La técnica de comparación de señales utilizada es la correlación cruzada.

A partir de la respuesta al impulso de la sala obtenida en los oídos izquierdo  $h_L(t)$  y derecho  $h_R(t)$  la correlación interaural se puede obtener utilizando la función de correlación cruzada interaural  $IACF(\tau)$ , que está definida como

$$IACF(\tau) = \frac{\int_{t_1}^{t_2} h_L(t) \cdot h_R(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} h_L^2(t) \cdot \int_{t_1}^{t_2} h_R^2(t)}} \quad (4.22)$$

A partir de la función anterior se obtendrá el coeficiente de correlación cruzada interaural determinado por

$$IACC_{t_1, t_2} = \max |IACF_{t_1, t_2}| \quad -1ms < \tau < 1ms \quad (4.23)$$

Los límites de integración definirán de nuevo los ámbitos de aplicación del parámetro. Así, se definen tres valores diferentes de IACC, cuyos límites de integración y ponderación en frecuencia se resumen en la siguiente tabla.

**Tabla 4.4. Límites temporales y ancho de banda considerado para IACC**

	t1	t2	BW
IACC <sub>A</sub>	0 ms	500 a 2000 ms	125 Hz – 4000 Hz
IACC <sub>E</sub>	0 ms	80 ms	125 Hz – 4000 Hz
IACC <sub>L</sub>	80	500 a 2000 ms	125 Hz – 4000 Hz

Los subíndices A (*all*), E (*early*) y L (*late*) hacen referencia a los tres coeficientes que pueden calcularse. El coeficiente total IACC<sub>A</sub> tiene en cuenta toda la duración de la

respuesta al impulso.  $IACC_E$  compara las señales en el ya conocido intervalo 0-80 ms mientras que  $IACC_L$  realiza la comparación de las colas de reverberación de las respuestas al impulso. Aunque estos procedimientos de medida forman parte de los anexos de (ISO 3382, 1997), el documento no deja claro el límite temporal  $t_2$  para los casos de  $IACC_A$  e  $IACC_L$ , refiriéndose a tiempos mayores que el tiempo de reverberación medio de la sala. (Beranek L. , 1996) determinó que en ambos casos se podía considerar  $t_2=500$ ms.

En lo que respecta a las implicaciones cualitativas de estos coeficientes, en la misma referencia acabada de citar, el autor estableció una correlación entre  $IACC_E$  y el atributo perceptivo de ancho aparente de la fuente ASW (*Aparent Sound Width*), que viene dada por la expresión

$$\rho = 1 - IACC_E \quad (4.24)$$

Según Beranek este parámetro muestra una buena correlación con la amplitud de la fuente sonora percibida por el espectador. Definió también LEV (*Listener Envelopment*) como

$$\varepsilon = 1 - IACC_L \quad (4.25)$$

Parámetro que correlaba con la sensación subjetiva por parte del oyente de sentirse “envuelto” por el sonido.

Como conclusión, en lo a las implicaciones perceptivas de estos coeficientes se refiere, puede decirse que las diferencias binaurales asociadas a la recepción de las primeras reflexiones y la reverberación inicial resultarán determinantes en la formación de una imagen de amplitud de la fuente sonora. Las diferencias respecto de la reverberación tardía se relacionan, en cambio, con la sensación cualitativa de “sentirse envuelto” por el sonido o inmerso en éste.

#### **4.7 El caso de la evaluación de la palabra**

En el apartado 2.2.3 se han enumerado los diferentes métodos y procedimientos tanto cualitativos como cuantitativos que se han utilizado históricamente para la evaluación del habla, tanto en líneas de transmisión como en salas. Se ha visto que frecuentemente se utilizan métodos idénticos en ambos casos, adaptándose a las características del canal. En esta sección se hará una revisión de las magnitudes acústicas que tradicionalmente se han relacionado con la evaluación de la calidad del habla. Algunas de ellas, análogamente al caso de la evaluación de la calidad en recintos para música son particularizaciones de las mismas métricas para el caso de la palabra. Concretamente, las clasificables como magnitudes temporales o relaciones energéticas. Otras, en cambio, son de naturaleza diferente y están encaminadas a obtener una magnitud física que determine la inteligibilidad del habla. Los resultados obtenidos con estas magnitudes son comparadas después con la inteligibilidad medida a través de procedimientos cualitativos como los descritos en la primera parte del apartado 2.2.3

#### 4.7.1 Parámetros temporales

##### 4.7.1.1 Tiempos de reverberación

El tiempo de reverberación medio de la sala  $RT_{mid}$ , tal como se ha definido en la ecuación 4.8 es un primer indicador de la idoneidad de la sala para la palabra. En general se recomiendan tiempos inferiores a 1.2 s. Una aproximación rápida puede calcularse a partir de una expresión denominada tiempo de reverberación óptimo

$$RT_{OP} = 0.32 + 0.17 \log V \quad (4.26)$$

Siendo  $V$  el volumen de la sala en metros cúbicos. Otro aspecto a considerar desde el punto de vista estrictamente temporal sería la evaluación de la curva tonal. Un refuerzo de la reverberación en las frecuencias predominantes en la palabra (1000 – 2000 Hz) podría considerarse perjudicial para la inteligibilidad de ésta.

##### 4.7.2 Parámetros energéticos

Las relaciones energéticas que se consideran para la evaluación de la palabra son análogas a las utilizadas en el caso de la música. Simplemente se adaptan los límites de integración al caso concreto de la palabra. Así, se utiliza la Definición tal y como se ha indicado en el apartado 4.5.1 y también la claridad  $C_{50}$  de idéntica forma a como se ha definido en el apartado 4.5.3. También, de forma análoga a la fuerza sonora  $G$  (vista en la sección 4.5.5 se define el incremento de sonoridad de la palabra producido por la sala como

$$S = 10 \log \frac{\int_0^{\infty} P^2(t) dt}{\int_0^{\infty} P_{10}^2(t) dt} \quad (dB) \quad (4.27)$$

Siendo en este caso  $P_{10}$  la señal de presión generada por el actor u orador en condiciones de campo libre a una distancia de 10 metros.

##### 4.7.3 Medidas de inteligibilidad

Se describirán a continuación los dos procedimientos más habituales para la estimación de la inteligibilidad en salas. El primero de ellos, propuesto en (Peutz, 1971) y reformulado posteriormente en (Davis & Patronis, 2006) es la métrica denominada porcentaje de pérdida en la articulación de consonantes ( $\%AL_{cons}$ ). El segundo es la métrica STI, propuesta por (Hougast & Steeneken, 1984) para su utilización en salas.

###### 4.7.3.1 $\%AL_{cons}$

Esta métrica realiza una aproximación desde dos puntos de vista: en primer lugar se realizó un exhaustivo estudio cualitativo sobre la percepción de las consonantes mediante el uso de logatomos de tipo CVC (ver sección 2.2.3) en diferentes recintos con diferentes tiempos de reverberación. Se realiza un análisis estadístico de estos resultados cuyo indicador es  $\%AL_{cons}$ , el porcentaje de errores en el reconocimiento de los logatomos en los diferentes recintos. En segundo lugar, partiendo de los supuestos de la teoría estadística, se realiza una formulación que a partir de magnitudes físicas de la sala como el volumen, la distancia orador-receptor o el tiempo de reverberación proporcione

una estimación del resultado de  $\%AL_{CONS}$ , correlación establecida a partir de los datos obtenidos en la experimentación llevada a cabo previamente. Con todo esto, se establece que el indicador viene determinado por

$$\%AL_{CONS} = \frac{200 r^2 (RT)^2}{VQ + k} \quad (4.28)$$

Donde  $r$  es la distancia entre el orador y el oyente,  $RT$  es el tiempo de reverberación de la sala,  $V$  el volumen de ésta y  $Q$  es la directividad de la fuente. En el caso de que la fuente sea una persona se considera  $Q=2$ . La constante  $k$  fue introducida en una modificación posterior de la formulación (Davis & Patronis, 2006) y hace referencia a las capacidades auditivas del oyente. Puede tomar valores entre 1.5 y 12.5 en el mejor y peor de los casos respectivamente. Cuando se tiene que

$$r > 0.20 \cdot \sqrt{\frac{V}{RT}}$$

Entonces la ecuación 4.28 se puede expresar como

$$\%AL_{CONS} = 9 \cdot RT + k \quad (4.29)$$

El valor de  $\%AL_{CONS}$  realizará pues una estimación del porcentaje de consonantes que no son percibidas correctamente en una sala. Son aceptables valores de  $\%AL_{CONS}$  relativamente altos – en torno al 20-30% - dado que la inteligibilidad mejora en el contexto de un mensaje concreto. Como se puede ver en la Figura 4.13, los peores resultados de inteligibilidad en experimentos cualitativos siempre son los obtenidos mediante logatomos CVC (por la dificultad articulatoria de la estructura consonante – vocal – consonante y por tratarse de palabras monosilábicas sin significado).

#### 4.7.3.2 STI

El cálculo del índice se basa en la función de transferencia de modulación (MTF). Como material del test se utilizan señales moduladas que imitan el habla humana en diferentes bandas de frecuencia. Dado que la profundidad de modulación de la señal de test es conocida, se compara con la profundidad de modulación de las señales obtenidas una vez han pasado por el canal. En el caso de las salas, en presencia de ruido y de reverberación, se producirá una reducción de la profundidad de modulación. La métrica se obtiene a partir del cálculo de estas reducciones para cada banda de frecuencia.

La función de transferencia de modulación es un buen descriptor de en qué medida la envolvente de modulación de una señal de banda estrecha se ha conservado durante el proceso de transmisión, ya sea en un canal técnico o en un canal puramente acústico. La función de transferencia de la modulación se obtiene habitualmente inyectando al sistema bajo estudio una señal del tipo

$$x(t) = \sqrt{0.5(1 + m_0 \cos(2\pi f_m t))} \cdot s(t) \quad (4.30)$$

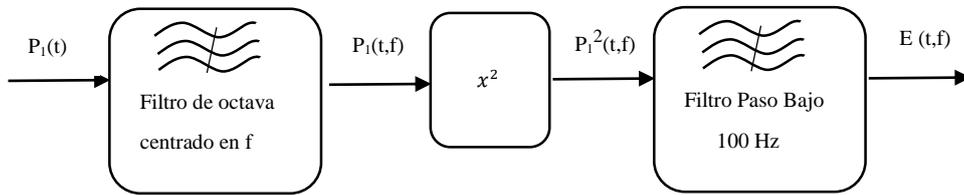
En la que  $s(t)$  es la señal modulada con un estrecho ancho de banda centrado en la frecuencia  $f$  (generalmente el centro de una banda de octava) y  $f_m$  es la frecuencia moduladora. En un canal de transmisión ideal, en el que no hubiese ruido de fondo ni reverberación podríamos obtener una señal de presión en la salida en función de

$$P_o(t) = A_0 \sqrt{0.5(1 + m_0 \cos(2\pi f_m t + \varphi_0))} \cdot s(t) \quad (4.31)$$

Es decir, se introduciría únicamente la ganancia del canal  $A_0$  y el desfase durante la transmisión  $\varphi_0$ . La profundidad de modulación  $m_0$  se conservaría intacta. En cambio, en un canal real, como es el caso de una sala tendremos una salida del tipo

$$P_o(t) = A_1 \sqrt{0.5(1 + m_1 \cos(2\pi f_m t + \varphi_1))} \cdot s(t) \quad (4.32)$$

En la que  $A_1$  representa la ganancia del canal,  $\varphi_1$  es el término que representa el retardo de fase y  $m_1$  representa la nueva profundidad de modulación de la señal, causada por la presencia de reverberación y ruido en el canal. Lo que perseguirá la métrica es establecer la comparación entre la profundidad de modulación  $m_1$  y la profundidad de modulación de la señal original  $m_0$ . Para ello, la señal recibida se somete a un sistema representado por el diagrama de bloques de la siguiente figura



**Figura 4.12. Proceso de extracción de la envolvente de energía de la señal  $P_1(t)$ .**

En primer lugar, la señal transmitida se filtra mediante un filtro de octava centrado en  $f$ , que era la frecuencia central del ancho de banda de la señal  $s(t)$ , en las ecuaciones 4.30, 4.31 y 4.32. La salida de este filtro se eleva al cuadrado posteriormente y se filtra paso-bajo con la finalidad de obtener la envolvente energética  $E(t,f)$ . Como el contenido en frecuencia de  $s(t)$  siempre estará por encima de 100 Hz, sólo su envolvente pasará por el último bloque de la Figura 4.12. Así, la envolvente energética vendrá dada por

$$E(t, f) = E_1((1 + m_1 \cos(2\pi f_m t + \varphi)) \quad (4.33)$$

En la que  $E_1$  es la energía de la señal y  $m_1$  la profundidad de modulación. Según la norma (IEC-60268-16, 2011),  $m_1$  puede calcularse a partir de la siguiente expresión

$$m_1(f, f_m) = 2 \frac{\sqrt{|\int E_1(t, f) \sin(2\pi f_m t) dt|^2 + |\int E_1(t, f) \cos(2\pi f_m t) dt|^2}}{\int E_1(t, f) dt} \quad (4.34)$$

Con lo cual se obtendrá la profundidad de modulación para la frecuencia portadora  $f_m$  de la señal  $E(t,f)$ , normalizada entre 0 y 1. La relación de transferencia de modulación vendrá pues, dada por

$$m(f, f_m) = \frac{m_1(f, f_m)}{m_0(f, f_m)} \quad (4.35)$$

En la medida que  $m(f, f_m)$  se acerca a la unidad mejor habrá sido el proceso de transmisión de la modulación. En cambio, cuanto más cercanos a 0 sean los valores significará que la profundidad de modulación se ha reducido fuertemente debido a la presencia de ruido y reverberación.

Para la medición del índice de transmisión del habla - *Speech Transmission Index* (STI) – se propuso estimar la relación de transferencia de modulación  $m(f, f_m)$  en siete bandas de octava, desde 125 Hz hasta 8000 Hz, que corresponden a las frecuencias  $f$  de las señales  $s(t)$  en las ecuaciones 4.40, 4.31 y 4.32. Cada una de estas señales se modula con 14 diferentes frecuencias moduladoras  $f_m$ , desde 0.63 Hz hasta 12.5 Hz en pasos de tercios de octava, lo cual implica la medición de 98 valores diferentes de  $m(f, f_m)$ . Dado que el índice STI se expresa mediante un único valor entre 0 y 1, se procede de la siguiente manera: para cada uno de los 98 valores de  $m(f, f_m)$  se calcula una relación señal ruido aparente determinada por

$$SNR_{app} = \max \left( -15, \min \left( 15, 10 \log \frac{m}{1-m} \right) \right) \text{ dB} \quad (4.36)$$

Donde  $m$  es cada uno de los 98 valores de  $m(f, f_m)$  obtenidos. Cada uno de los valores de la ecuación anterior estarán dentro del intervalo entre -15 y 15 dB y posteriormente se escalan a un valor entre 0 y 1. En (IEC-60268-16, 2011) define un promediado ponderado de los 98 valores, dando mayor peso a las bandas de octava más relevantes para la inteligibilidad (entre 500 y 2000 Hz), de forma que al final se obtiene un único valor entre 0 y 1, para el que 0 significa ininteligibilidad absoluta y 1 inteligibilidad total.

Conociendo los tiempos de reverberación y las relaciones señal-ruido para cada banda de octava puede estimarse el índice de reducción de modulación para una banda de octava mediante

$$m(f, f_m) = \frac{1}{\sqrt{1 + \left( 2\pi f_m \frac{RT_f}{13.8} \right)^2}} \cdot \frac{1}{1 + 10^{\frac{-SNR_f}{10}}} \quad (4.37)$$

Expresión en la que  $f_m$  es la frecuencia moduladora,  $RT_f$  es el tiempo de reverberación en la octava centrada en  $f$  y  $SNR_f$  es la relación señal ruido en la octava centrada en  $f$ . Obviamente, el primer término de la expresión considera el efecto de la reverberación en la reducción del índice de modulación, mientras que el segundo término aporta la contribución del ruido de fondo.

Con la finalidad de reducir los tiempos de cálculo se desarrollaron dos métodos alternativos al STI: el RASTI -*Room Acoustical STI*- y el STIPA -*STI for Public Address systems*-. RASTI realiza un análisis únicamente en dos octavas: la de 500 Hz y la de 2000 Hz, utilizando unas frecuencias moduladoras de 1, 2, 4 y 8 Hz para la señal de 500 Hz y unas moduladoras de 0.7, 1.4, 2.8, 5.6 y 11.2 Hz para la señal de 2000 Hz, reduciendo el cómputo global a un total de 9 relaciones de transferencia de modulación. En el caso de STIPA se utilizan 14 modulaciones, pero se consideran todas las siete bandas de octava desde 125 Hz hasta 8000 Hz, utilizando tan sólo dos frecuencias moduladoras para cada octava. La elección de las frecuencias moduladoras varía según la octava.

Aunque, de forma general, puede decirse que las simplificaciones RASTI y STIPA obtienen resultados muy similares a STI en su ámbito de aplicación, las técnicas simplificadas están orientadas a situaciones específicas. Así, una aplicación del RASTI fuera del contexto de la acústica de salas diferirá significativamente de la medida STI. STIPA, en cambio, en principio diseñado para sistemas de refuerzo sonoro, ofrece resultados satisfactorios también en la acústica de salas (van Wijngaarden, Verhave, & Steeneken, 2012), aunque en determinadas situaciones –como la presencia de ecos- su uso puede estar desaconsejado. La obtención de las funciones de transferencia de modulación a partir de las respuestas al impulso para la evaluación de sistemas con no linealidades también está desaconsejada, lógicamente, por fundamentarse el método de cálculo a partir de la respuesta al impulso en la consideración de un sistema LTI, como se ha visto en páginas precedentes (Steeneken, van Wijngaarden, & Verhave, 2011).

#### *4.7.3.3 STI e inteligibilidad*

A la hora de evaluar la inteligibilidad de forma cualitativa hay que considerar, evidentemente, que influyen una cantidad significativa de factores perceptivos, como elementos fisiológicos y cognitivos tanto del hablante como del oyente, el mensaje o el tipo de lenguaje utilizado. A pesar de todos estos factores, la métrica propuesta por Hougast y Steeneken ha demostrado tener un comportamiento monótono asociado a la inteligibilidad y se ha venido utilizando durante casi cuatro décadas (Steeneken & Hougast, 2002).

La Figura 4.13 muestra la correlación que se establece entre el índice STI y el porcentaje de inteligibilidad obtenidos en diferentes procedimientos subjetivos de valoración de la inteligibilidad. Como ya se ha anticipado en la sección 2.2.3, la tipología del material de test utilizado para la evaluación subjetiva de la inteligibilidad influye de forma determinante en los resultados de estos experimentos. De esta forma, cuando se utilizan frases con sentido, el contexto semántico de la frase, además de los factores cognitivos del receptor permiten una decodificación del lenguaje de forma más fácil, debilitando así el efecto de la degradación introducida por el canal. Así, puede conseguirse un buen entendimiento del mensaje aunque las condiciones del canal de transmisión sean muy pobres. Puede verse en la gráfica, que con un  $STI = 0.4$  se entienden la práctica totalidad de frases sintácticamente bien formadas, cuando 0.4 representa un valor cualitativamente

malo del índice de transmisión. A medida que nos alejamos del lenguaje formal, la inteligibilidad se vuelve más difícil y las características del canal de transmisión más determinantes.

En el test de rimas se utilizan palabras en las que el cambio de un único fonema altera el significado de la palabra, con lo cual el contexto lingüístico queda delimitado al conocimiento de la propia palabra y las posibles dificultades que pueda introducir el canal para poder reconocer el fonema que cambia, procedimiento más parecido a un test de articulación. Estos factores hacen que para el mismo valor STI que ofrece una inteligibilidad cercana al 100% en el caso de las frases se obtenga ahora una inteligibilidad un 20% menor. El caso que ofrece más dificultad para la recepción correcta del mensaje es el de los test realizados usando logatomos CVC. Al tratarse de palabras monosilábicas sin sentido alguno sería más adecuado incluso hablar de pérdida de articulación de consonantes que de inteligibilidad. La dificultad en la percepción de las consonantes – especialmente en la segunda – crece a medida que empeoran las condiciones del canal. De hecho, en la 0 puede observarse un comportamiento cuasi lineal hasta un valor aproximado de STI=0.6. En la parte superior de la curva se necesitan mayores incrementos de STI para obtener pequeñas mejoras de la pérdida de articulación de las consonantes. Estas correlaciones hacen referencia a experimentos realizados en lengua inglesa. La norma (ISO 9921, 2003) establece un corpus concreto de logatomos fonéticamente equilibrado en inglés correlados con el parámetro STI, de forma que relaciona directamente el valor de articulación obtenido – porcentaje logatomos bien percibidos – con el índice (STI). En el caso del español, adaptado al caso del español hispanoamericano, se han realizado trabajos con este mismo objetivo (Rosas & Sommerhoff, 2008) (Hidalgo, Sommerhoff, & Rosas, 2011).

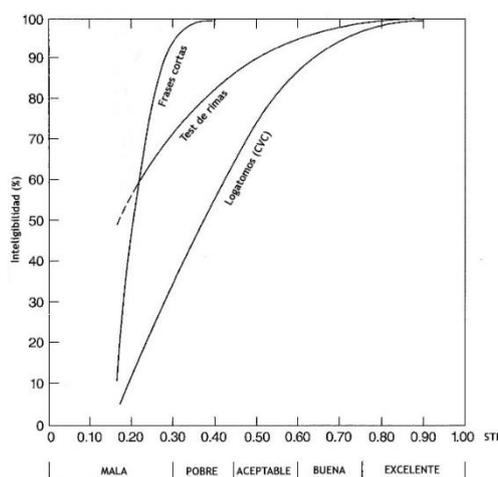


Figura 4.13. STI vs Inteligibilidad percibida en test subjetivos en función del material de test utilizado. Adaptado de (Steeneken & Hougast, 2002)

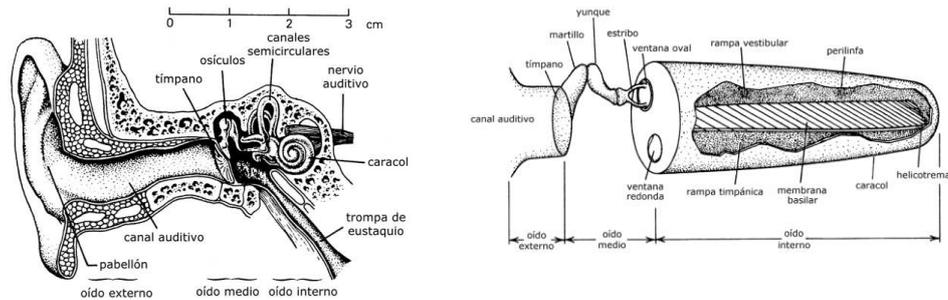
# 5 Percepción auditiva

Se describen a continuación algunos aspectos del funcionamiento del sistema auditivo humano con la finalidad de poder entender los componentes que formarán parte de la simulación que realizan de éste los diversos métodos perceptuales y más concretamente, el método implementado en este trabajo para la predicción de la distorsión percibida en el caso de los transductores y de la degradación acústica en el caso de las salas. Se estudiará la anatomía y la fisiología del aparato auditivo, haciendo énfasis en aquellas partes y estructuras del mismo más importantes para el desarrollo de modelos perceptuales. Además, en la segunda parte de este apartado se estudiarán los principales efectos psicoacústicos involucrados en la percepción auditiva que constituyen la base del modelado de un sistema perceptual.

## **5.1 El sistema auditivo humano**

El sistema auditivo periférico humano se divide en tres partes: el oído externo, medio e interno. Los estímulos sonoros se propagan a través de estas zonas, sufriendo diversas transformaciones hasta su conversión final en impulsos nerviosos. Tanto el procesamiento mecánico de las ondas sonoras como la conversión de éstas en señales

electroquímicas son procesos no lineales, lo cual dificulta la caracterización y modelado de los fenómenos perceptuales.



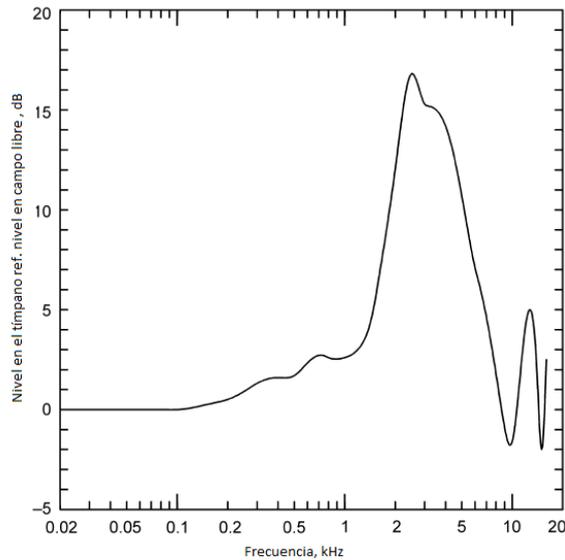
Fuente: (Maggiolo, 2014).

**Figura 5.1. Corte transversal y esquema simplificado del sistema auditivo periférico**

### 5.1.1 Oído externo

El oído externo está compuesto por el pabellón y el conducto auditivo externo. La función del pabellón es la de concentrar las ondas sonoras en el conducto, que termina en el tímpano. La ubicación lateral de los pabellones derecho e izquierdo en el ser humano ha hecho casi innecesaria la capacidad de movimiento de los mismos, a diferencia de lo que sucede en muchos otros animales que tienen una amplia capacidad de movimiento de los pabellones, pudiendo enfocarlos en la dirección de proveniencia del sonido. La modificación de la señal acústica en el sistema auditivo comienza ya en el pabellón, ya que por sus características éste tiene una frecuencia de resonancia entre los 4.500 Hz y los 5.000Hz. El canal auditivo refuerza aún más la presión mediante un mecanismo de resonancia. El conducto auditivo es un "tubo" de unos 2,5 cm de longitud, que influye en la respuesta en frecuencia del sistema auditivo. Dada la velocidad de propagación del sonido en el aire (aprox. 334 m/s), dicha longitud corresponde a 1/4 de la longitud de onda de una señal sonora de unos 3,5 - 4kHz. Este es uno de los motivos por los cuales el aparato auditivo presenta una mayor sensibilidad a las frecuencias cercanas a los 3,5 - 4 kHz. El canal o conducto auditivo está cerrado por una membrana tensada, el tímpano, que separa el canal auditivo del oído medio.

El pabellón de la oreja y el conducto auditivo juntos crean una amplia resonancia que aumenta los niveles de sonido en el tímpano en el rango de frecuencias de 1,5 a 5kHz. La figura 21 muestra el incremento de nivel producido en el tímpano debido a esta resonancia (Shaw, 1974). Las señales sonoras que entran al conducto auditivo externo sufren efectos de difracción debidos a la forma del pabellón auricular y la cabeza, y estos efectos varían según la dirección de incidencia y el contenido espectral de la señal; así, se altera el espectro sonoro debido a la difracción.



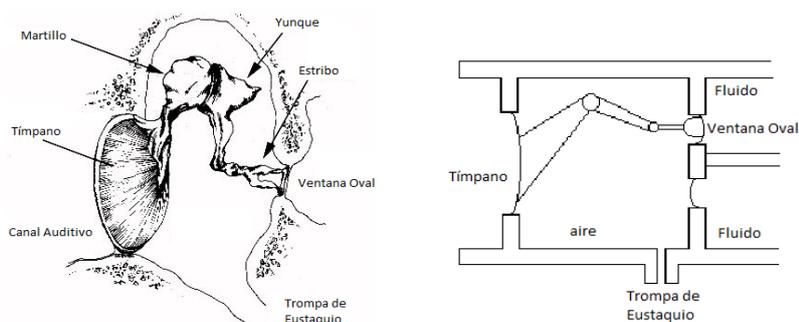
**Figura 5.2. Resonancia del oído externo: pabellón y conducto auditivo**

Estas alteraciones, en forma de "picos" y "valles" en el espectro, son usadas por el sistema auditivo para determinar la procedencia del sonido en el llamado "plano medio" (plano imaginario perpendicular a la recta que une ambos tímpanos) (Moore, Glasberg, & Gaer, 1997)

### 5.1.2 Oído medio

El oído medio está constituido por una cavidad llena de aire, dentro de la cual se encuentran tres huesecillos denominados martillo, yunque y estribo, unidos entre sí en forma articulada. Uno de los extremos del martillo se encuentra adherido al tímpano, mientras que la base del estribo está unida mediante un anillo flexible a las paredes de la ventana oval, orificio que constituye la vía de entrada del sonido al oído interno. El oído medio está conectado a la garganta a través de la trompa de Eustaquio, que es un conducto que llega hasta las vías respiratorias y que permite igualar la presión del aire a ambos lados del tímpano. La principal función del oído medio es asegurar la transferencia eficaz de la energía acústica del aire a los fluidos de la cóclea. Si el sonido incidiera directamente en la ventana oval, la mayor parte de éste simplemente se reflejaría en lugar de entrar en la cóclea. Esto ocurriría porque la resistencia de la ventana oval al movimiento es muy diferente de la del aire, es decir, existe una desadaptación de impedancias. El oído medio actúa como un dispositivo de adaptación de impedancia o un transformador, que mejora la transmisión del sonido y reduce la cantidad de sonido reflejado. Esto se logra principalmente por una relación de 17:1 entre las áreas efectivas

de la membrana del tímpano y de la ventana oval y, en menor medida, por la acción de palanca ejercida por los huesecillos (Zwicker & Fastl, 1990).



Fuente: (Sepúlveda & Cortez, 2013).

**Figura 5.3. Oído medio y modelo mecánico**

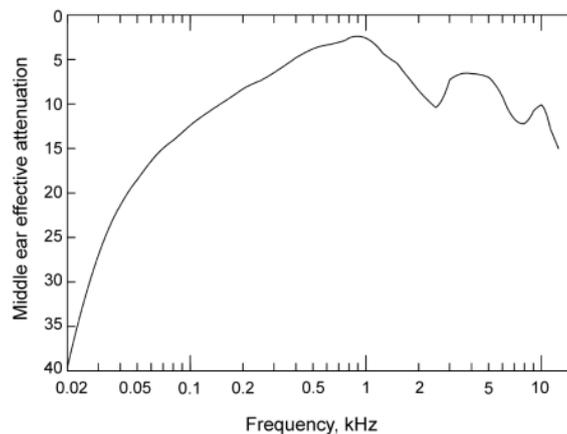
Los cambios de presión en la pared externa de la membrana timpánica asociados a la señal sonora hacen que dicha membrana vibre siguiendo las oscilaciones de dicha señal. Las vibraciones del tímpano se transmiten a lo largo de la cadena de huesecillos, la cual opera como un sistema de palancas, de forma tal que la base del estribo vibra en la ventana oval. Este huesecillo se encuentra en contacto con uno de los fluidos contenidos en el oído interno; por lo tanto el tímpano y la cadena de huesecillos actúan como un mecanismo para transformar las vibraciones del aire en vibraciones del fluido. Ahora bien, para lograr que la transferencia de potencia del aire al fluido sea máxima, debe efectuarse un acoplamiento entre la impedancia mecánica característica del aire y la del fluido, puesto que esta última es mucho mayor que la primera, la cadena de huesecillos actúa como acoplador de impedancias. El máximo acoplamiento se obtiene en el rango de frecuencias medias, en torno a 1kHz. De hecho, éste acoplamiento de impedancia es del orden de 50 a 75% para frecuencias entre los 300 y 3kHz.

En la función de acoplo de impedancias intervienen 2 mecanismos: el sistema de palanca y la relación hidráulica. En primer lugar, debido a la diferencia entre la longitud de la base del martillo y la apófisis larga del yunque, se logra que el sistema de palanca de los huesecillos multiplique la fuerza del movimiento por 1,3 aproximadamente. Mucho más importante que la acción de palanca es la relación entre el área de la membrana timpánica y la diminuta platina o base del estribo. Una relación de 1:17, que permite que la energía de la onda sonora que impacta en la membrana timpánica se aplique a la pequeña base del estribo, causando una presión 17 veces mayor sobre el líquido del caracol que la presión de la onda en el tímpano.

Además de la adaptación de impedancias, el oído medio cumple también una función de protección del oído interno. Ante la llegada de sonidos de gran intensidad al tímpano, los músculos tensores del tímpano y el estribo se contraen de forma refleja, modificando las características de transferencia del oído medio y disminuyendo la cantidad de energía

entregada al oído interno (Moore B. C., 2007). El músculo tensor del tímpano tira del martillo hacia adentro, mientras que el músculo del estribo tira de éste hueso hacia fuera. Estas dos fuerzas se oponen entre sí y, por lo tanto, permiten que todo el sistema de huesecillos logre mucha rigidez, disminuyendo considerablemente la transmisión de las frecuencias bajas al oído interno. El reflejo de atenuación puede disminuir la intensidad de transmisión de sonido hasta 30 o 40 dB. La función de éste mecanismo se basa, en parte, en permitir la adaptación del oído a sonidos de intensidades diferentes, pero sobre todo en proteger a las células receptoras del oído interno frente a sobrecargas. Este efecto se denomina reflejo timpánico. Este reflejo no es instantáneo, tarda de 40 a 160 ms en producirse. El reflejo timpánico es un mecanismo no lineal que introduce un término cuadrático en la relación entrada-salida del oído medio.

El conjunto formado por el oído externo y el oído medio forman un sistema cuya respuesta en frecuencia es de tipo paso bajo. En el intervalo cercano a los 4kHz se observa un pequeño efecto de ganancia, debido a las características del conducto auditivo. Esta respuesta sólo es válida cuando no actúa el reflejo timpánico (Glasberg & Moore, 2006).



**Figura 5.4. Respuesta en frecuencia combinada del oído externo y medio**

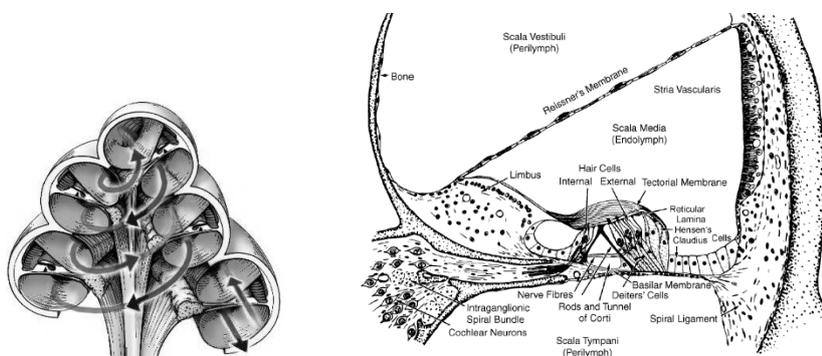
Para niveles de ruido moderado, por debajo de 90 dB SPL, el oído externo y medio se comportan esencialmente como sistemas lineales. Sin embargo, en presencia de altos niveles de presión sonora, tanto la membrana del tímpano como los huesecillos del oído medio pueden comportarse de forma no lineal. Además, el reflejo acústico también introduce no linealidad. Estas no linealidades pueden producir una distorsión armónica y de intermodulación audible.

### 5.1.3 Oído interno

El oído interno también se conoce como la cóclea. Tiene la forma de la concha de un caracol, en espiral. Sin embargo, la forma de espiral no parece tener ningún significado funcional, y la cóclea se describe a menudo como si la espiral estuviera “desenrollada”

(Moore B. C., 2007). La cóclea está llena de fluidos y tiene paredes óseas rígidas. Está dividida longitudinalmente por dos membranas, la membrana de Reissner y la membrana basilar.

Al inicio de la espiral, donde se encuentra la ventana oval, se la denomina base, mientras que el otro extremo, la punta interna, se conoce como el ápex. También es común hablar del extremo basal y el extremo apical. En el ápex hay una pequeña abertura (helicotrema) entre la membrana basilar y las paredes de la cóclea, que conecta las dos cámaras exteriores de la cóclea, la ramba o escala vestibular y la ramba o escala timpánica.

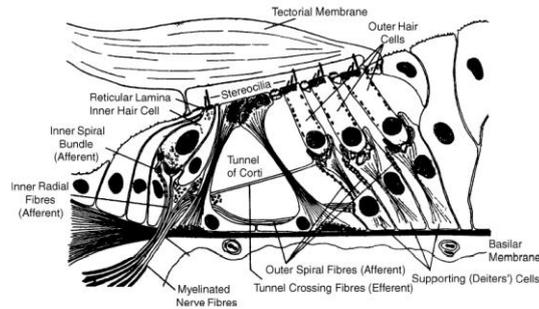


Fuente: (Moore, 2007).

**Figura 5.5. Corte transversal de la cóclea.**

El movimiento hacia el interior de la ventana oval tiene su correspondencia en un movimiento hacia el exterior en una membrana que cubre una segunda abertura en la cóclea - la ventana redonda. Estos movimientos tienen como consecuencia diferencias de presión entre un lado de la membrana basilar el otro (es decir, la presión se aplica en una dirección perpendicular a la membrana basilar), y esto se traduce en el movimiento de la membrana basilar. El helicotrema elimina las diferencias de presión entre la ramba vestibular y la ramba timpánica a frecuencias muy bajas. Esto evita que la membrana basilar se mueva de manera significativa en respuesta a los movimientos de la ventana oval causados por movimientos de la mandíbula o por pequeños cambios en la presión atmosférica (como ocurre cuando se cambia de altitud). El helicotrema también reduce el movimiento de la membrana basilar en respuesta a los sonidos de baja frecuencia.

Por el lado de la cóclea más cercano a la pared exterior, hay una estructura llamada la estría vascular. Ésta juega un papel importante en el metabolismo de la cóclea y en la creación de las tensiones (potencial eléctrico) que son esenciales para el normal funcionamiento de la cóclea. Una tercera membrana, llamada membrana tectorial, se encuentra por encima de la membrana basilar BM, y también discurre a lo largo de la cóclea. Entre la membrana basilar y la membrana tectorial se ubican las células ciliadas, que forman parte de una estructura llamada el órgano de Corti.



Fuente: (Moore, 2007)

**Figura 5.6. Corte transversal del órgano de Corti.**

Las células ciliadas se dividen en dos grupos por un arco conocido como el túnel de Corti. Las situadas en el lado del arco más cercano al exterior de la espiral se conocen como células ciliadas externas, y están dispuestas en hasta cinco filas en los seres humanos, a pesar de que las filas son a menudo un tanto irregulares. Las células ciliadas al otro lado del arco forman una sola fila, y se conocen como células ciliadas internas. En los seres humanos, hay cerca de 12000 células ciliadas externas por oído, cada una con aproximadamente 140 estereocilios que sobresalen de ella, mientras que hay cerca de 3.500 células ciliadas internas, cada uno con aproximadamente 40 estereocilios.

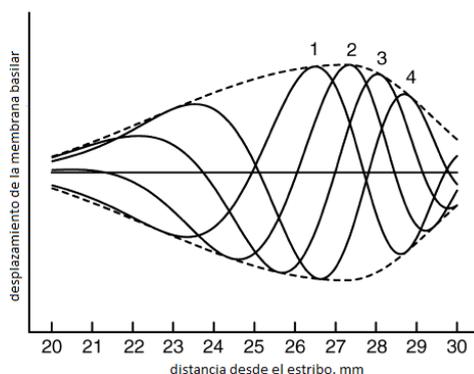
La membrana tectorial, que tiene una estructura gelatinosa, está por encima de las células ciliadas. Parece que los estereocilios de las células ciliadas externas están en contacto con la membrana tectorial, pero esto no ocurre así para las células ciliares internas. La membrana tectorial está ubicada a un lado. Cuando se mueve la membrana basilar arriba y abajo, se crea un movimiento perpendicular de corte y la membrana tectorial se mueve de lado a lado. Como resultado de ello, los estereocilios en la parte superior de las células ciliadas se mueven hacia los lados. El movimiento se produce por contacto directo, en el caso de las células ciliadas externas, pero en el caso de las internas puede ser producido por la fricción viscosa del fluido de transmisión entre la parte superior del órgano de Corti y la membrana tectorial. El movimiento de los estereocilios de las células ciliadas internas produce un flujo de corriente eléctrica a través éstas, que a su vez conduce a la generación de potenciales de acción en las neuronas del nervio auditivo. De esta manera, la función realizada por las células ciliadas internas es la de la transducción entre los movimientos mecánicos y la actividad neuronal.

Las células ciliadas internas y externas tienen funciones muy diferentes. La gran mayoría de los aferentes, las neuronas que llevan la información desde la cóclea hasta los niveles más altos del sistema auditivo, se conectan a células ciliadas internas. Por lo tanto, la mayoría de la información sobre los sonidos se transmite a través de éstas. El papel

principal de las células externas parece ser el de influir activamente en la mecánica de la cóclea. Las células externas pueden cambiar su longitud, forma y rigidez en respuesta a un estímulo eléctrico.

Cuando la ventana oval se pone en movimiento por las oscilaciones del estribo se produce una diferencia de presión entre la superficie superior e inferior de la membrana basilar. La onda de presión viaja casi instantáneamente a través de los fluidos de la cóclea. En consecuencia, la diferencia de presión se aplica esencialmente al mismo tiempo a lo largo de toda la longitud de la membrana basilar. Esto produce un patrón de movimiento en la membrana basilar. El patrón no dependerá de qué extremo de la cóclea es estimulado. Sonidos que llegan a la cóclea a través de los huesos de la cabeza en lugar del través del aire no producen respuestas atípicas.

La respuesta de la membrana basilar a la estimulación con una senoide se traduce en una onda de desplazamiento que se propaga desde la base hacia el ápex. La amplitud de la onda aumenta al principio y luego disminuye de forma abrupta. La Figura 5.7 muestra el desplazamiento de la membrana basilar durante cuatro instantes sucesivos en el tiempo, en respuesta a una senoide de baja frecuencia. Los cuatro picos sucesivos de la onda están etiquetados del 1 al 4.



**Figura 5.7. Desplazamiento de la membrana basilar en respuesta a un tono de baja frecuencia.**

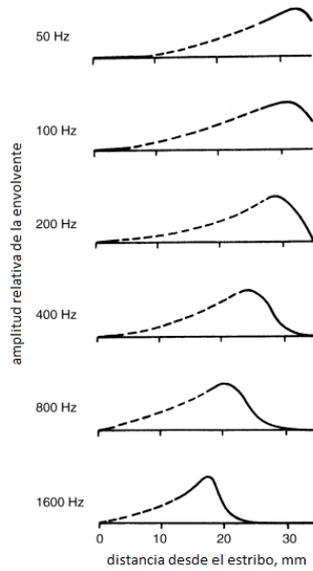
La figura muestra también la unión entre los 4 picos, la envolvente. Esta envolvente muestra un pico en una posición particular de la membrana basilar. La respuesta de la membrana basilar a sonidos de diferentes frecuencias se ve muy afectada por las propiedades mecánicas de la membrana, que varían progresivamente de la base al ápice. En la base, la membrana basilar es relativamente estrecha y rígida. Esto provoca que la base responda mejor a altas frecuencias. A medida que nos acercamos al ápex, la membrana se va haciendo más ancha y flexible, lo que causa que el ápice responda mejor a las bajas frecuencias. Cada punto de la membrana basilar está afinado, es decir,

responde con el mayor desplazamiento a una frecuencia determinada llamada frecuencia característica.

Actualmente se cree que el comportamiento respecto a la frecuencia de la membrana basilar se debe a dos mecanismos. El primero de ellos es conocido como el mecanismo pasivo y depende de las propiedades mecánicas de la membrana basilar y las estructuras circundantes. Éste opera de manera más o menos lineal. El otro es el mecanismo activo, que depende de la intervención de las células ciliares externas y opera de forma no lineal. (Meddis & López Poveda, 2010). El mecanismo activo necesita que la cóclea esté en buenas condiciones fisiológicas y es fácilmente dañable.

La Figura 5.8 muestra las envolventes de los patrones de vibración para diferentes sinusoides de baja frecuencia. Sonidos de diferentes frecuencias producen el máximo desplazamiento en su lugar adecuado de la membrana basilar. En cierta manera, la membrana basilar se comporta como un analizador de frecuencia con una pobre resolución a priori. Entendemos aquí resolución en términos de nitidez de la afinación, es decir, cuanto más estrecho sea un patrón de respuesta mayor resolución de frecuencia se tendrá. Los trabajos pioneros sobre los patrones de vibración de la membrana basilar fueron realizados por (von Békésy, 1960) y muestran patrones bastante anchos (poca resolución de frecuencia). Por ejemplo, el patrón correspondiente a 400 Hz se extiende a lo largo de casi la totalidad de la longitud de la membrana basilar. En la actualidad se cree que estos modelos reflejan únicamente el comportamiento del sistema pasivo. El mecanismo activo no estaba funcionando en el experimento de von Békésy por dos razones: en primer lugar, utilizó niveles muy altos (alrededor de 140 dB SPL) para los cuales el mecanismo activo resulta dañado. En segundo lugar, utilizó oídos de cadáveres, y el mecanismo activo deja de funcionar después de la muerte. Posteriores investigaciones con animales vivos y utilizando técnicas mínimamente invasivas han medido la respuesta de un único punto de la membrana basilar a diferentes sinusoides obteniendo que la resolución en frecuencia de la membrana basilar depende críticamente de las condiciones fisiológicas, siendo más precisa la resolución cuanto mejores son las condiciones fisiológicas (Moore B. C., 2007).

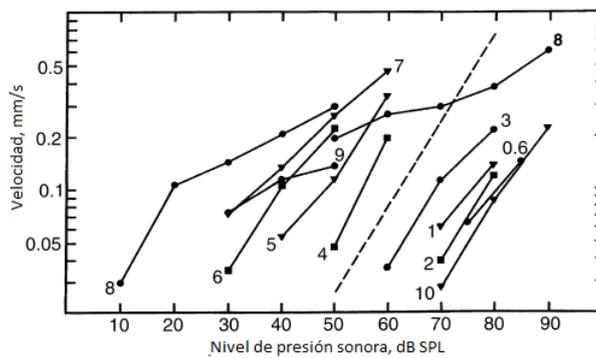
El estado de salud de la cóclea se monitoriza a menudo colocando un electrodo en o cerca del nervio auditivo y midiendo las respuestas combinadas de las neuronas a ráfagas de tonos. Esta respuesta se conoce como potencial de acción compuesto. El nivel más pequeño para el cual se detecta un potencial de acción se denomina umbral del potencial de acción. Normalmente, la membrana basilar está fuertemente sintonizada (i.e. buena resolución en frecuencia) cuando el umbral del potencial de acción es bajo, indicando que la cóclea está en buenas condiciones fisiológicas y que el mecanismo activo está funcionando (Moore B. C., 2007).



Fuente: (von Békésy, 1960)

**Figura 5.8. Desplazamiento de la membrana basilar en respuesta a diferentes tonos.**

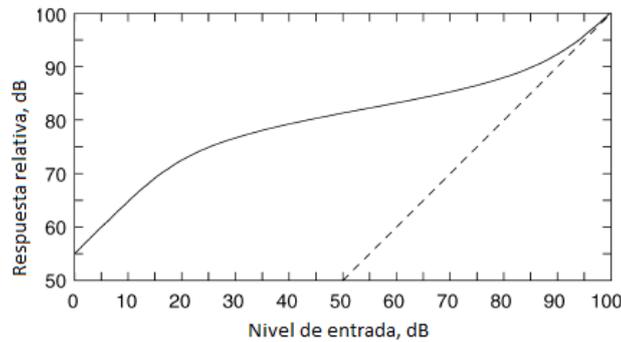
En un oído sano normal, la respuesta de la membrana basilar es no lineal. Cuando se incrementa la magnitud de la entrada, la magnitud de la respuesta no crece proporcionalmente a la entrada (Robles, Ruggero, & Rich, 1986). Esto se ilustra en la Figura 5.9, que muestra las funciones de entrada-salida de la membrana basilar en el punto con frecuencia característica 8 kHz. Cada curva representa una frecuencia de excitación particular, indicada por el número (en kHz) junto a la curva. Si las respuestas fueran lineales, las curvas serían paralelas a la línea de puntos.



Fuente: (Robles, Ruggero, & Rich, 1986)

**Figura 5.9. Velocidad de la membrana basilar en respuesta a un tono en función de la presión sonora de entrada.**

Para la frecuencia característica (8 kHz) se trazan dos curvas, una a niveles altos y otra a niveles bajos. Una función ideal para un tono en el punto de la membrana basilar correspondiente a la frecuencia característica de ese mismo tono es la representada en la siguiente figura, con la velocidad escalada en dB (i.e.  $20 \cdot \log_{10}(\text{velocidad})$ ).



Fuente: (Moore, 2007)

**Figura 5.10. Respuesta ideal de la velocidad de la membrana basilar en respuesta a un tono en el punto de la membrana correspondiente a la frecuencia característica de ese mismo tono.**

Mientras que la función para el tono correspondiente a la frecuencia característica es casi lineal para niveles de entrada bajos (por debajo de 20-30 dB) y se aproxima a la linealidad para niveles de presión altos (por encima de 90 dB), la función presenta una pendiente muy plana en el rango de niveles de presión medios. Esto indica una compresión no lineal: un amplio rango de niveles de entrada se comprimen en un pequeño rango de niveles de salida. La forma de esta función puede explicarse de la siguiente manera: para niveles bajos y medios, el mecanismo activo amplifica la respuesta en la membrana basilar. Esta amplificación puede llegar a ser del orden de 50 dB (Robles & Ruggero, 2001). Para niveles muy bajos, por debajo de 20-30 dB, la amplificación es aproximadamente lineal y a su máximo valor. A medida que se incrementa el nivel de presión sonora la amplificación se reduce progresivamente. De esta manera, la respuesta aumenta mucho más lentamente de lo que lo haría si fuese lineal. Cuando el nivel de presión alcanza valores alrededor de los 90 dB SPL el mecanismo activo deja de actuar y la respuesta se vuelve lineal. Aquí, en los niveles altos, la respuesta del mecanismo pasivo es la dominante.

La no linealidad se produce principalmente cuando la frecuencia de excitación es cercana a la frecuencia característica del punto de la membrana basilar cuya respuesta se está midiendo. Para frecuencias más alejadas de la frecuencia característica las respuestas son más lineales. Así, en la Figura 5.9, las frecuencias de 7 y 9kHz (cercas a la característica) tienen curvas más llanas, mientras que las correspondientes a frecuencias por debajo de 7kHz y por encima de 9kHz muestran pendientes más pronunciadas (lineales). En efecto, la compresión tiene lugar sólo alrededor del pico del patrón de

respuesta de la membrana basilar. Como resultado, el pico en la distribución de la vibración a lo largo de la membrana basilar se aplana para niveles de presión altos, lo cual explicaría en parte la ancha resolución en frecuencia observada en (von Békésy, 1960).

### 5.2 Umbrales de audición absolutos

La sensibilidad de la audición puede variar considerablemente de un sujeto a otro, además de cambiar según las condiciones de propagación del sonido. Por esta razón, resulta conveniente definir un umbral de audibilidad promedio. Éste se representa mediante una curva que indica la presión sonora de un tono puro de larga duración (> 200 ms), el cual se propaga en condiciones de campo libre y en ausencia de cualquier otro sonido, y que puede ser detectado por el 50% de una población de sujetos jóvenes (entre 18 y 25 años) y audiológicamente normales. Los valores medios del umbral de audibilidad han sido objeto de un proceso de estandarización, descrito en el documento *ISO R389-1964*. El umbral de audibilidad no representa pues un límite absoluto, sino una medida estadística asociada con la probabilidad de detección de un tono de determinada frecuencia y amplitud.

La intensidad subjetiva de un sonido determinado queda definida relativamente, comparando la sensación originada por este sonido con la de otro de referencia. Si los dos producen la misma sensación de intensidad, se puede decir que ambos tienen la misma intensidad subjetiva. Para realizar esta comparación se emplea como referencia un tono de frecuencia 1 kHz y de nivel de presión sonora ajustable. Se define la sonoridad de un sonido como el nivel de presión sonora de un tono de 1 kHz en un tubo de onda plana e incidencia frontal que suena igual de fuerte que el sonido a analizar, su unidad es el fonio. Así, un sonido que suena igual de fuerte que un tono de 1 kHz con un nivel de presión de 40 dB tiene un nivel de sonoridad de 40 fonios.

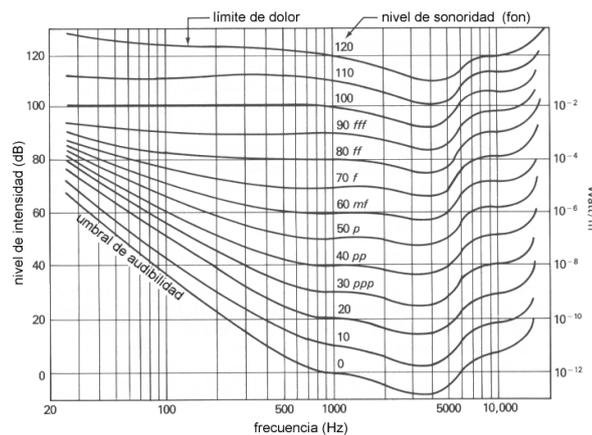


Figura 5.11. Curvas isofónicas descritas por Fletcher y Munson (Fletcher & Munson, 1933)

En (Fletcher & Munson, 1933) se deduce experimentalmente la relación existente entre el nivel de presión sonora, el nivel de sonoridad y la frecuencia sobre un gran número de jóvenes con edades comprendidas entre los dieciocho y los veinticinco años y con audición normal. A las curvas resultantes se las conoce como curvas isofónicas y presentan los niveles sonoros que debe alcanzar un sonido sinusoidal de frecuencia  $f$  para producir la misma sensación auditiva que un sonido sinusoidal de 1.000 Hz de frecuencia y un nivel de presión sonora dado. Estas curvas han sido corregidas y estandarizadas posteriormente. El valor umbral para bajas frecuencias es del orden de 60 dB, pero a medida que la frecuencia aumenta, el oído presenta una mayor sensibilidad, siendo máxima a los 3 kHz, superados los cuales se necesita un nuevo aumento de presión sonora.

### **5.3 Enmascaramiento y banda crítica**

La selectividad en frecuencia hace referencia a la capacidad del sistema auditivo de identificar componentes sinusoidales que se presentan junto a un sonido complejo (en muchas ocasiones ruido). Esta característica está estrechamente relacionada con la resolución en frecuencia de la membrana basilar de la que se hablaba en el apartado anterior. La selectividad en frecuencia juega un papel fundamental en muchos aspectos de la percepción auditiva. Sin embargo, ésta es a menudo demostrada y medida mediante el estudio del enmascaramiento. A partir de la definición de enmascaramiento y sus procedimientos de medida se obtendrán resultados que permitirán desarrollar el concepto de filtro auditivo y de banda crítica, que serán imprescindibles en el desarrollo de los modelos perceptuales.

El enmascaramiento puede definirse como el proceso por el cual el umbral de audibilidad de un sonido se aumenta debido a la presencia de otro sonido (enmascarante). O también la cantidad en que se incrementa el umbral de audibilidad de un sonido (en dB) en presencia de otro (enmascarante)<sup>8</sup>. Durante mucho tiempo se ha observado que una señal es más fácilmente enmascarable por otra señal cuyas componentes en frecuencia sean cercanas o muy similares a la de la señal enmascarada. Esto condujo a la idea de que nuestra habilidad para separar los componentes de un sonido complejo depende, al menos en parte, de la capacidad de resolución en frecuencia de la membrana basilar y condujo también a plantear que el enmascaramiento refleja los límites de la selectividad en frecuencia del sistema auditivo: si la selectividad del oído es insuficiente para separar una señal de otra enmascarante, entonces tiene lugar el enmascaramiento. De esta forma, el enmascaramiento puede utilizarse para cuantificar la selectividad en frecuencia.

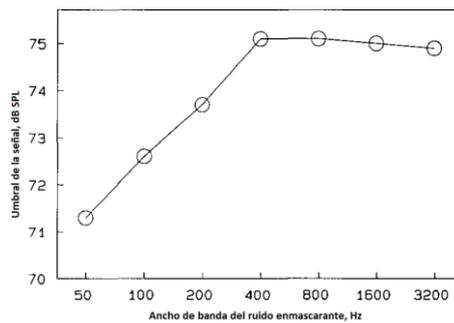
Un importante parámetro físico que afecta al enmascaramiento es el tiempo. Dependiendo de la ubicación temporal de la señal de test (T) con respecto a la señal enmascarante (E), se pueden distinguir tres situaciones posibles:

---

<sup>8</sup> Definiciones de la American Standards Association

- a) Enmascaramiento simultáneo: E y T se presentan simultáneamente en el tiempo (E está presente durante toda la duración de P).
- b) Enmascaramiento previo a la presentación de la señal enmascarante, o pre-enmascaramiento: E se presenta después de T.
- c) Enmascaramiento posterior a la presentación de la señal enmascarante, o post-enmascaramiento: E se presenta antes que T.

En el desarrollo del concepto de banda crítica el enmascaramiento estudiado es el simultáneo. Las teorías actuales sobre el enmascaramiento se derivan de un experimento realizado por (Fletcher, 1940) en el cual se medía el nivel mínimo al que podía detectarse una señal sinusoidal en función del ancho de banda de un ruido enmascarante. La banda de paso del ruido estaba siempre centrada en la frecuencia de la señal sinusoidal y la densidad espectral de potencia del ruido se mantenía siempre constante. De esta forma, la potencia total del ruido se incrementaba a medida que lo hacía el ancho de banda de éste. El experimento se ha repetido en numerosas ocasiones desde entonces confirmando siempre los resultados (Moore B., 1997). Un ejemplo de éstos se muestra en la Figura 5.12.



Fuente: (Moore B., 1995)

**Figura 5.12. Nivel de detección de una señal de 2kHz en función del ancho de banda del ruido enmascarante centrado también en 2kHz**

El nivel de detección de la señal incrementa a medida que lo hace el ancho de banda hasta llegado un punto en que el umbral de detección se vuelve aproximadamente constante de forma que incrementos en el ancho de banda del ruido enmascarante no afectan al nivel de detección de la señal. Como conclusión a sus resultados, Fletcher sugirió que el sistema auditivo periférico humano se comporta como si contuviera un banco de filtros paso-banda, con bandas de paso superpuestas. Estos filtros se denominan en la actualidad filtros auditivos (*auditory filters*). Fletcher pensó que la membrana basilar era la base de los filtros auditivos. Cada localización en la membrana basilar responde a un rango limitado de frecuencias, de forma que cada punto corresponde a un filtro con diferente frecuencia central. Posteriores estudios son consistentes con este postulado respecto a la membrana basilar (Moore B. C., 2007).

El experimento de Fletcher condujo a un modelo de enmascaramiento denominado modelo de espectro de potencia, que se basaba en los siguientes supuestos:

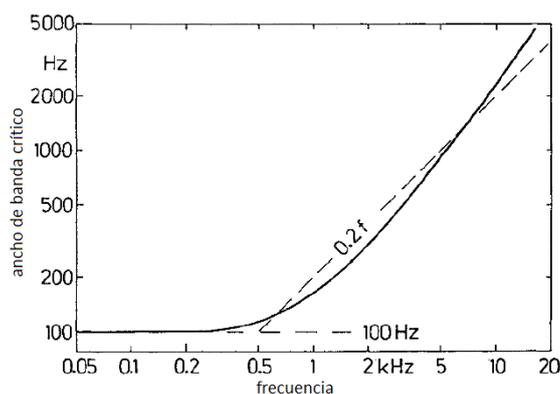
- d) El sistema auditivo periférico contiene una serie de filtros lineales pasa-banda con bandas de paso superpuestas.
- e) Cuando un individuo intenta detectar una señal en presencia de un ruido de fondo se asume que el oyente hace uso únicamente del filtro cuya frecuencia central es más próxima a la de la señal. Este filtro, que deja pasar la señal, elimina gran parte del ruido.
- f) Sólo las componentes del ruido que pasan a través del filtro específico desempeñan un papel en el efecto del enmascaramiento.
- g) El umbral para la detección de la señal se determina a partir de la cantidad de ruido que pasa por el filtro. Se asume que el umbral de detección corresponde a una relación señal-ruido a la salida del filtro.
- h) Los estímulos – tanto la señal como el ruido - se representan en su espectro de potencia, es decir, se ignoran las fases relativas y las fluctuaciones, lo que da nombre al modelo.

Ninguno de estos supuestos es estrictamente correcto. Concretamente, los filtros son en realidad no lineales (Moore & Glasberg, 1987), los oyentes combinan información de más de un filtro para aumentar la detectabilidad de la señal (Spiegel, 1981), el ruido que queda fuera del filtro pasa banda puede afectar a la detección de la señal y las fluctuaciones en la señal enmascarante pueden desempeñar un rol importante (Patterson & Henning, 1977). Sin embargo, el concepto de filtro auditivo está ampliamente aceptado y resulta muy útil. Las desviaciones respecto a la linealidad de los filtros son pequeñas si el rango de niveles de entrada también lo es. Además, aunque los supuestos del modelo fallan en ocasiones, éste funciona correctamente cuando las señales se eligen cuidadosamente (Moore B. , 1995).

En el experimento de ‘ruido inventanado’ de Fletcher, una vez que el ancho de banda del ruido excedía el ancho de banda del filtro auditivo, posteriores incrementos del ancho de banda del ruido no incrementarían el ruido que pasa a través del filtro. Ese ancho de banda a partir del cual el umbral de detección de la señal sinusoidal se mantenía constante fue denominado por Fletcher como ancho de banda crítico - *Critical Bandwidth* - y es muy frecuente referirse a éste como CB. En el análisis de los resultados de su experimento, Fletcher realizó la suposición de que la forma del filtro auditivo podía aproximarse a un simple rectángulo. Para un filtro de este tipo, todos los componentes de la banda pasante son filtrados con la misma magnitud y cualquier componente fuera de la banda pasante es simplemente eliminado. La anchura del rectángulo equivaldrá al CB. De hecho, el término CB se utiliza frecuentemente para hacer referencia a este hipotético filtro rectangular. Desde la introducción del concepto de CB, una gran cantidad de experimentos ha demostrado que las respuestas de los oyentes a sonidos complejos difieren en función del ancho de banda de los sonidos utilizados como estímulos. Aun así, diferentes experimentos han obtenido resultados razonablemente

similares respecto de la estimación del ancho de banda del CB y respecto a la forma con que el CB varía con la frecuencia (Moore B. ,1997). Existen sin embargo algunas diferencias significativas sobretodo en baja frecuencia, como se verá a continuación.

Utilizando cinco métodos<sup>9</sup> para la detección del CB con niveles entre el umbral de audición y 90 dB y promediando los resultados obtenidos en más de 50 individuos, Zwicker estableció la relación entre el ancho de banda crítico y la frecuencia indicada en la Figura 5.13. El CB permanece prácticamente constante (alrededor de 100 Hz) hasta una frecuencia aproximada de 500 Hz. A partir de esa frecuencia, el CB se incrementa casi proporcionalmente a la frecuencia (un poco más lentamente por debajo de 3 kHz y un poco más rápidamente por encima de esa frecuencia). Una aproximación propuesta por el propio autor consiste en considerar anchos de banda de 100 Hz hasta una frecuencia central de 500 Hz y a partir de 500 Hz anchos de banda relativos del 20% del valor de la frecuencia central.



Fuente: (Zwicker & Fastl,1990)

**Figura 5.13. Ancho de banda crítico CB en función de la frecuencia obtenido por Zwicker**

<sup>9</sup>Zwicker utilizó en el primer método tonos uniformemente espaciados como señal a detectar. Un segundo método utilizaba un sonido de banda estrecha como señal de test y dos tonos de la misma potencia como señal enmascarante. El umbral de la señal de test enmascarada por los dos tonos se mide en función de la separación en frecuencia de los dos tonos enmascarantes con la señal de banda estrecha centrada entre ellos. El tercer método estaba basado en la detectabilidad de los cambios de fase. Para ello utilizó señales moduladas en amplitud y en frecuencia. El cuarto procedimiento para la obtención de CB se basaba en medidas de sonoridad (loudness) en función del ancho de banda para un nivel de presión sonora constante (la sonoridad se mantiene constante mientras el ancho de banda del ruido es menor de un determinado valor crítico). Finalmente, el quinto método estaba basado en la percepción binaural, concretamente en la detección de ráfagas de tono puro cortas con un retardo apenas perceptible entre ambas, cada una presentada a un oído y con una diferencia entre la frecuencia de cada una. La sensibilidad decae fuertemente cuando la diferencia entre las frecuencias de ambas ráfagas se hace mayor que el CB. Los detalles de los procedimientos y resultados se muestran en (Zwicker, 1961), (Zwicker, 1965), (Zwicker & Fastl, 1972) y (Zwicker & Fastl,1990).

La curva obtenida por Zwicker puede ser modelada mediante la ecuación 5.1, la cual permite calcular el ancho de banda crítico (en Hz),  $\Delta f_{BC}$ , correspondiente a la frecuencia  $f$  en Hz, con un error inferior al 10%.

$$\Delta f_{BC}(f) = 25 + 75 \cdot \left[ 1 + 1.4 \left( \frac{f}{100} \right)^2 \right]^{0.69} \quad (5.1)$$

Zwicker propuso una escala de tasa de variación de banda crítica en (Zwicker, 1961) cuya unidad era el Bark. La escala está basada en el hecho de que nuestro sistema auditivo analiza un espectro ancho en partes que corresponden a bandas críticas. Añadiendo una banda crítica a la siguiente de manera que el límite superior de una banda crítica se corresponde con el límite inferior de la banda crítica inmediatamente superior se obtiene la escala de tasa de variación de banda crítica.

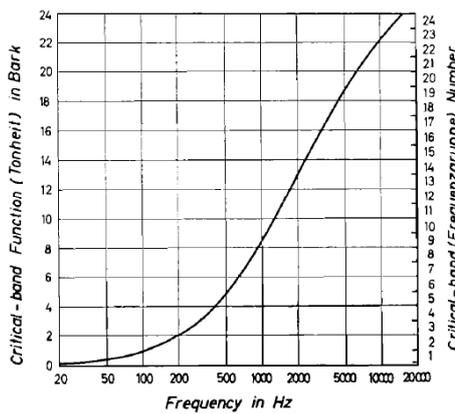


TABLE I.

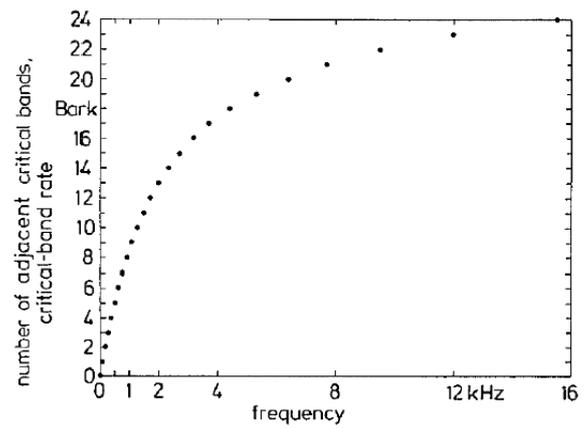
Number	Center frequencies Hz	Cut-off frequencies Hz	Bandwidth Hz
		20	
1	50	100	80
2	150	200	100
3	250	300	100
4	350	400	100
5	450	510	110
6	570	630	120
7	700	770	140
8	840	920	150
9	1000	1080	160
10	1170	1270	190
11	1370	1480	210
12	1600	1730	240
13	1850	2000	280
14	2150	2320	320
15	2500	2700	380
16	2900	3150	450
17	3400	3700	550
18	4000	4400	700
19	4800	5300	900
20	5800	6400	1100
21	7000	7700	1300
22	8500	9500	1800
23	10 500	12 000	2500
24	13 500	15 500	3500

Fuente: (Zwicker E. , 1961)

**Figura 5.14.** Gráfica y tabla adjuntadas por Zwicker en (Zwicker, 1961) para la estandarización de la escala Bark como unidad de medida de CB.

Representando el número ordinal de cada banda crítica en función de la frecuencia se obtiene una serie como la de la Figura 5.15. La mayor parte de las estimaciones de CB se basan en datos de experimentos de enmascaramiento cuyos resultados se utilizan para el cálculo del ancho de banda rectangular equivalente (*Equivalent Rectangular Bandwidth*) - en adelante ERB - de los filtros auditivos. El ERB puede considerarse como una medida del CB. Sin embargo, se ha convertido en una especie de estándar hablar de ERB para diferenciar las estimaciones más recientes de las más antiguas (Zwicker, 1961). El ERB de un filtro específico es igual al ancho de banda de un filtro perfectamente rectangular que tiene una transmisión en la banda de paso igual a la máxima transmisión del filtro específico y transmite la misma potencia de ruido blanco que el filtro específico. En otras palabras, si escalamos el filtro rectangular para obtener

la misma altura máxima y área que el filtro específico, el ancho de banda del filtro rectangular obtenido será el ERB del filtro específico inicial.



Fuente: (Zwicker & Fastl, 1972)

**Figura 5.15. Numeración asociada con la secuencia de bandas críticas adyacentes. Bark vs. frecuencia**

#### 5.4 Estimaciones de la forma de los filtros auditivos

La mayor parte de los métodos para la estimación de la forma de los filtros para una determinada frecuencia central se basan en las suposiciones del modelo de enmascaramiento de espectro de potencia vistas anteriormente. Si el ruido enmascarante está representado por su espectro de potencia  $N(f)$  y la función de ponderación o forma del filtro auditivo es  $W(f)$ , se puede escribir:

$$P_S = K \int_0^{\infty} N(f)W(f)df \quad (5.2)$$

Donde  $P_S$  es la potencia de la señal de prueba en el umbral de detección. Manipulando el espectro de la señal o ruido enmascarante  $N(f)$  y midiendo los cambios correspondientes en  $P_S$ , se puede obtener el perfil o la forma del filtro auditivo  $W(f)$ . La señal enmascarante elegida para realizar este procedimiento debe elegirse de manera que no contravenga los supuestos del modelo de espectro de potencia. Por ejemplo, si la señal enmascarante está formada por varias sinusoides, en función de la distancia entre éstas, se pueden producir batidos que proporcionen una medida incorrecta del umbral de enmascaramiento dado que provocan falsas detecciones. A medida que la distancia entre las sinusoides se incrementa, este efecto será menos perceptible, de forma que modificará de nuevo la medida del umbral. Esto contradice el supuesto del modelo de

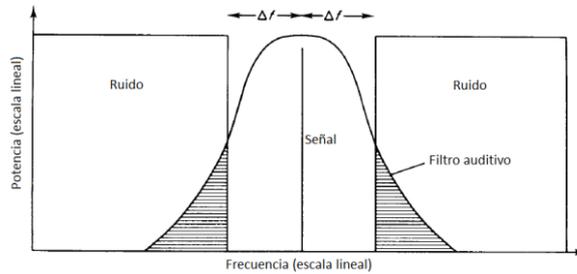
enmascaramiento de espectro de potencia por el cual el umbral corresponde a una relación señal-ruido constante a la salida del filtro. En general, el uso de señales enmascarantes de espectro más ancho –del tipo ruidos– resultan más adecuadas para la realización de experimentos de enmascaramiento encaminados a la obtención de la forma o perfil de filtros auditivos. Aún así, para ruidos de banda estrecha, la interacción entre la señal de prueba y la enmascarante puede llegar a ser audible. Además, fluctuaciones lentas en ruidos de banda estrecha pueden afectar la detectabilidad de la señal de manera que dependa de la diferencia entre la frecuencia central de la banda de paso del ruido y la frecuencia de la señal de prueba (Buus, 1985).

Otra consideración importante en la elección de un ruido para la realización de medidas de la forma del filtro auditivo, es que puede darse el caso de que el filtro que proporcione una relación señal-ruido más elevada no esté centrado necesariamente en la frecuencia de la señal de prueba. Cuando un oyente detecta la señal de prueba con un filtro que no está centrado en la frecuencia de ésta se denomina ‘*off-frequency listening*’ (Patterson & Nimmo-Smith, 1980). Si el espectro enmascarante se concentra principalmente por debajo o por encima de la señal de prueba, puede existir un rango de frecuencias centrales de filtro en el cual la relación señal-ruido es suficientemente alta para proporcionar información útil en la detección de la señal de prueba. Bajo estas condiciones, el escuchante puede estar combinando información procedente de diferentes filtros auditivos para realizar la detección, con lo cual se estaría violando el supuesto de que la escucha se realiza a través de un único filtro auditivo, como sostiene el modelo.

Muchos de estos problemas pueden evitarse si se utiliza un ruido enmascarante con una banda eliminada alrededor de la frecuencia central de la señal de prueba. Para un ruido con estas características, la relación señal-ruido más alta tiene lugar en un filtro centrado en una frecuencia razonablemente cercana a la frecuencia de la señal de prueba. La forma del filtro puede obtenerse midiendo el umbral de detección de la señal en función del ancho de la banda eliminada de ruido alrededor de la frecuencia central. Para niveles intermedios de ruido, el filtro auditivo es casi simétrico en una escala de frecuencia lineal (Patterson & Nimmo-Smith, 1980). Con esta técnica, usada por primera vez en (Patterson R. , 1976), el perfil del filtro auditivo se estimará utilizando un ruido enmascarante con una banda eliminada centrada simétricamente en la frecuencia de la señal de prueba.

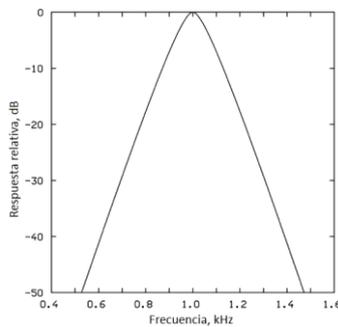
Para un ruido enmascarante con una banda eliminada de ancho  $2\Delta f$  y una frecuencia central  $f_c$ , la ecuación 5.2 puede escribirse como:

$$P_S = KN_0 \int_0^{f_c - \Delta f} W(f)df + KN_0 \int_{f_c + \Delta f}^{\infty} W(f)df \quad (5.3)$$



**Figura 5.16. Ilustración de la técnica utilizada por Patterson para la estimación de la forma del filtro auditivo.**

Donde  $N_0$  es la densidad espectral de potencia del ruido en sus bandas de paso. Las dos integrales de la ecuación 5.3 representan las áreas de la Figura 5.16 en las que las bandas baja y alta de ruido se superponen al filtro. Dado que el filtro y el ruido enmascarante son simétricos respecto de la frecuencia de la señal de test, estas dos áreas son iguales. Así, la función que relaciona  $P_S$  con el ancho de la banda eliminada  $\Delta f$  proporciona una estimación de la integral del filtro auditivo: a medida que el ancho de la banda eliminada se incrementa, cada vez pasa menos ruido a través del filtro auditivo, con lo cual el umbral de detección de la señal de test cae a niveles más bajos. Si se asume que el umbral corresponde a una relación señal-ruido constante a la salida del filtro, la variación del umbral de detección con el ancho de la banda eliminada nos estará proporcionando información de cómo el área debajo del filtro varía con  $\Delta f$ . El área debajo de una función entre determinados límites se obtiene integrando la función entre esos límites. Si derivamos la función que relaciona el umbral de detección con  $\Delta f$  obtendremos la respuesta relativa del filtro para ese valor de  $\Delta f$ . En otras palabras, la respuesta relativa del filtro para una desviación dada  $\Delta f$  desde la frecuencia central, es igual a la pendiente de la función que relaciona umbral con ancho de banda eliminada en ese valor concreto de  $\Delta f$ . Un ejemplo de filtro obtenido mediante esta técnica es el mostrado en la Figura 5.17.



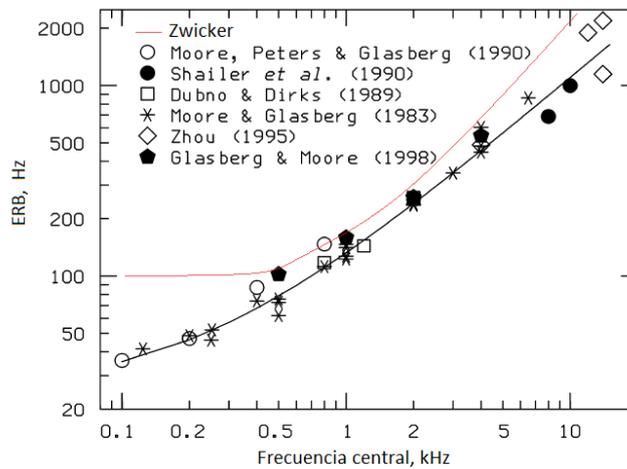
Fuente: (Moore B. C., 2007)

**Figura 5.17. Módulo de filtro auditivo determinado utilizando el método de Patterson.**

A diferencia de un filtro rectangular, un filtro con esta forma no puede quedar completamente especificado con un único número, el CB. Para especificar estos filtros se utiliza la banda de paso de -3dB o el ERB. Los anchos de banda -3dB de los filtros auditivos obtenidos con este método están generalmente entre el 10% y el 15% de la frecuencia central. Los anchos de banda ERB de estos mismos filtros están entre el 11% y el 17% de la frecuencia central. Estos valores están bastante cercanos a los obtenidos utilizando otros métodos anteriores como los de Zwicker. Pero existe una diferencia significativa con éstos: los valores de los anchos de banda de los filtros auditivos en bajas frecuencias tienden a ser menores que los obtenidos con las estimaciones tradicionales de CB. Se ha visto, en la Figura 5.14, que los anchos de banda obtenidos por Zwicker tendían a ser constantes por debajo de 500 Hz. Esto no ocurre con las estimaciones basadas en el método *notched noise* en las que el ancho de banda continúa decreciendo con la frecuencia. En (Glasberg & Moore, 1990) se propone una ecuación que proporciona el valor de ERB en función de la frecuencia, F, en kHz.

$$ERB = 24,7(4,37F + 1) \quad (5.4)$$

La Figura 5.18 representa el ERB del filtro auditivo en función de la frecuencia obtenido en diferentes experimentos utilizando el método de banda de ruido eliminada (*notched noise*), así como la función definida en la ecuación 5.4. Se ha incluido también la función tradicional de CB propuesta por (Zwicker E. , 1961) que difiere significativamente en frecuencias inferiores a 500 Hz.



**Figura 5.18. Estimaciones del ERB en función de la frecuencia obtenidas con el método ‘notched noise’ (excepto la curva roja).**

Fuente: (Glasberg & Moore, 1990)

El trazo negro de la figura anterior es la ecuación 5.4 propuesta en (Glasberg & Moore, 1990). El trazo rojo representa la escala Bark de Zwicker. Obsérvese la diferencia por debajo de 500 Hz.

En ocasiones resulta útil graficar datos psicoacústicos en una escala de frecuencia relacionada con los ERB. El ERB puede utilizarse entonces como una unidad de frecuencia. Por ejemplo, el valor del ERB para una frecuencia central de 1 kHz está alrededor de 130 Hz, de forma que un incremento desde 935 Hz a 1065 Hz correspondería a un paso de 1 ERB. La relación entre el número de ERBs y la frecuencia viene dada por la ecuación 5.5 (Glasberg & Moore, 1990), donde  $F$  es la frecuencia en kHz.

$$\text{Número de ERB} = 21,4 \log_{10}(4,37F + 1) \quad (5.5)$$

Esta escala es conceptualmente similar a la escala Bark propuesta por Zwicker aunque difiere, obviamente, en los valores numéricos.

## 5.5 Modelado de los filtros auditivos

La Figura 5.17 muestra el módulo de un filtro auditivo obtenido a partir de un experimento realizado con el método *notched noise*. El siguiente paso es parametrizar la forma del filtro obtenida en forma de función.

### 5.5.1 Filtros roex

La función *rounded exponential* o exponencial redondeada – en adelante roex- fue propuesta en (Patterson & Nimmo-Smith, 1980) como una representación de los filtros auditivos. Los autores concluyeron que los filtros obtenidos podían ser representados con un par de funciones exponenciales enfrentadas que estuvieran redondeadas de alguna manera en las partes superior e inferior (lo cual explica el nombre de la función). La expresión de una función roex es de la siguiente forma:

$$|H(f)|^2 = (1 + pg)e^{-pg} \quad (5.6)$$

En la que  $g$  es la desviación normalizada de la frecuencia respecto de la frecuencia central del filtro  $f_c$ . Es decir,

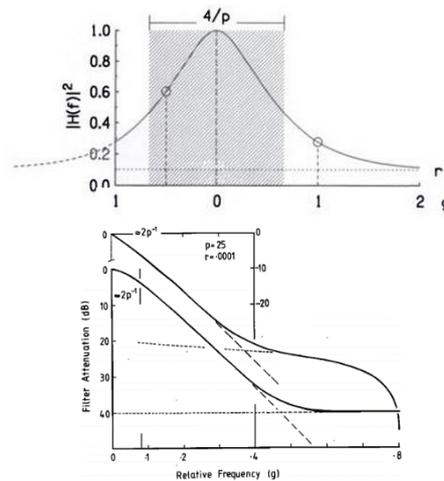
$$g = \left| \frac{f - f_c}{f_c} \right|$$

Este filtro tiene un parámetro adimensional,  $p$ , que describe el ancho de banda crítico determinando la pendiente del filtro en las partes superior e inferior de la frecuencia central. En la representación de la Figura 5.19 se usa una  $p=3$  a efectos de visualización. Los valores de  $p$  en filtros reales suelen ser superiores a 10. Un filtro de esta forma es el filtro roex más simple y se denomina roex( $p$ ). A medida que las señales enmascarantes utilizadas en los experimentos son más intensas o que los escuchantes utilizados en los experimentos son mayores (con la consecuente pérdida de sensibilidad auditiva) se obtienen perfiles de los filtros auditivos con la falda o caída del filtro más anchas. Para modelar estos resultados se propusieron diferentes soluciones como el uso de varios filtros roex( $p$ ) con diferentes pendientes en paralelo (Unoki, Irino, Glasberg, Moore, & Patterson, 2006) hasta que en (Patterson R. , Nimmo-Smith, Weber, & Milroy, 1982) se

propone el filtro  $roex(p,r)$  en la que  $r$  es una constante de pequeño valor que describe el rango dinámico, redefiniendo la ecuación 5.6 como

$$|H(f)|^2 = r + (1 - r)(1 + pg)e^{-pg} \quad (5.7)$$

Los experimentos de enmascaramiento utilizando un filtro *notch* o de banda eliminada simétrico condujeron, obviamente, a una aproximación a los filtros auditivos también simétrica, como es el caso de  $roex(p)$  o de  $roex(p,r)$ . Para evaluar la asimetría de los filtros auditivos es necesaria una modificación del experimento, introduciendo un desplazamiento del centro del filtro *notch* respecto de la frecuencia central de la señal de detección. En este caso, los datos experimentales pueden ajustarse con un filtro del tipo  $roex(p_i, p_u, r)$ , donde el parámetro  $p_i$ , que describe la pendiente del filtro por debajo de la frecuencia central es diferente de  $p_u$ , que determina la pendiente para frecuencias superiores a la frecuencia central del filtro (Glasberg & Moore, 1990). Los resultados muestran entonces que los filtros auditivos son asimétricos, con las pendientes de alta frecuencia más pronunciadas que las de baja frecuencia, de manera que  $p_u > p_i$ . Esta asimetría incrementa a medida que lo hace la intensidad.



**Figura 5.19. Representación del módulo de un filtro  $roex(p,r)$  (Hartmann, 1997).**

En la figura anterior, la pendiente del filtro con línea discontinua indica cómo sería el filtro en ausencia del parámetro  $r$ , es decir,  $roex(p)$  (Patterson R., Nimmo-Smith, Weber, & Milroy, 1982).

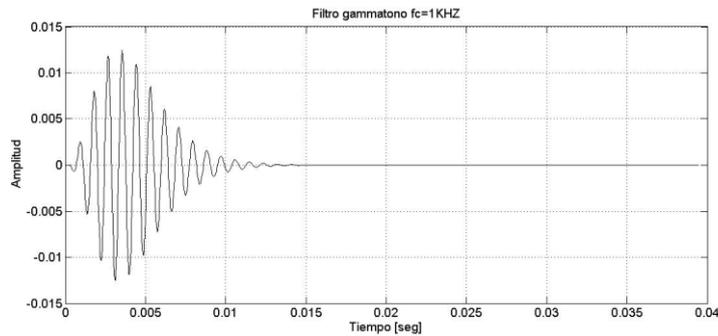
Todos los filtros  $roex$  se definen en el dominio de la frecuencia. El principal inconveniente que presentan estos filtros es la imposibilidad de obtener una respuesta al impulso utilizando la transformada de Fourier inversa de la función en el dominio espectral. Es decir, no se dispone de una versión del filtro en el dominio temporal, lo cual imposibilita el modelado de la percepción de sonidos complejos (Unoki, Irino, Glasberg, Moore, & Patterson, 2006).

### 5.5.2 Filtros gammatono

Esto conducirá a la búsqueda de nuevos filtros que permitan una implementación eficiente. Si bien sugerida en trabajos anteriores (Johannesma, 1972), la función gammatono es propuesta en (de Boer & Jong, 1978) como posible respuesta al impulso de los filtros auditivos. La función gammatono viene dada por la ecuación 5.8:

$$g(t) = \alpha t^{n-1} e^{-2^{nb}t} \cos(2nf_c t + \phi) \quad (5.8)$$

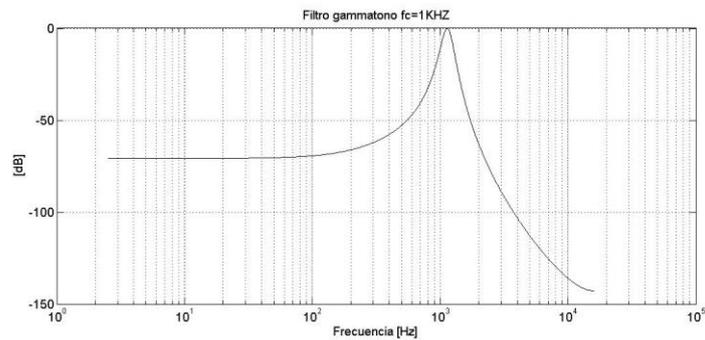
Donde  $\alpha$  es un valor arbitrario utilizado para normalizar el valor de pico de la función de transferencia,  $n$  es el orden del filtro,  $b$  es la duración del impulso,  $f_c$  es la frecuencia central del filtro y  $\phi$  es la fase.



**Figura 5.20. Función gammatono en el dominio temporal para una frecuencia central de 1kHz**

El nombre de la función hace referencia al hecho de que el término en la ecuación 5.8 anterior al coseno es la función gamma utilizada en estadística, que modela la envolvente del tono generado por el coseno, como puede observarse en la Figura 5.20. Las principales virtudes de esta función respecto de las roex podrían resumirse en los siguientes aspectos (Patterson R. , Nimmo-Smith, Holdsworth, & Rice, 1987): en primer lugar, la función representa una respuesta al impulso causal del filtro, de manera que su transformada de Fourier proporcionará una función de transferencia con unas características de amplitud y fase de un filtro implementable, al contrario de lo que sucedía con los filtros roex. Además, la respuesta en amplitud de un filtro gammatono de orden 4 es muy similar a la de un filtro roex(p), teniendo un ancho de banda rectangular equivalente idéntico. La función gammatono proporciona un buen ajuste con la respuesta al impulso de las fibras del nervio auditivo medidas con la técnica *revcor - reverse correlation-* (Johannesma, 1972). Finalmente, y muy importante en el ámbito de este trabajo, la salida de un banco de filtros gammatono se puede calcular con algoritmos eficientes que utilicen un filtrado recursivo (Patterson R. , Nimmo-Smith, Holdsworth, & Rice, 1987), como se verá más adelante, en el apartado dedicado a la implementación de la métrica  $R_{\text{nonlin}}$ , en el que se explicará detalladamente el desarrollo realizado para la

obtención de un banco filtros auditivos basado en la función gammatono. Generalmente, se han utilizado filtros gammatono de 4º orden, los cuales se ajustan correctamente para simular el filtrado coclear de sonido de banda ancha (musical y hablado), especialmente cuando el nivel de sonido se encuentra centrado en el rango audible. Una representación de un filtro de 4º orden se muestra en la siguiente figura.



**Figura 5.21. Respuesta en frecuencia obtenida a partir de la respuesta al impulso representada en la Figura 5.20**



# 6 Un modelo perceptual

## 6.1 La métrica $R_{\text{nonlin}}$

La métrica  $R_{\text{nonlin}}$  fue propuesta en (Moore, Tan, Zacharov, & Ville-Veikko, 2004), y pertenece a la familia de métodos perceptuales denominada de comparación de representaciones internas. Se trata de una evolución de un método anterior de los mismos autores denominado *Distortion Score* (en adelante DS) (Moore, Tan, & Zacharov, 2003). DS se basa fundamentalmente en una comparación espectral entre la señal original y la distorsionada - que incluye los productos de distorsión-. Ambas señales son alineadas en el tiempo y se separan en *frames* de 30 ms. Se calculan los espectros mediante DFT y se normalizan al valor de pico correspondiente para asegurar que los efectos de la distorsión lineal se eliminan de la señal distorsionada. Las muestras de cada espectro son agrupadas en bandas de frecuencia ERB para obtener una representación aproximada del análisis en frecuencia del sistema auditivo. Se calcula la potencia de ambas señales para cada banda y se convierte a decibelios. Finalmente la diferencia de potencias en cada banda se suma para obtener el valor DS. Este valor es promediado finalmente entre todos

los *frames* para obtener el valor DS final. Las señales de test utilizadas eran señales multitono, es decir, señales estacionarias. Como se ha visto también en el capítulo anterior, para este tipo de señales resulta más conveniente el análisis en frecuencia. La diferencia fundamental con la nueva métrica,  $R_{\text{nonlin}}$ , es que ésta utilizará señales complejas, de manera que la comparación entre la señal original y la distorsionada se realizará mediante un procesado en el dominio del tiempo, concretamente mediante el cálculo de la correlación cruzada.

## 6.2 El modelo

La Figura 6.1 muestra el diagrama de bloques del procesado para el cálculo de  $R_{\text{nonlin}}$ . La métrica tiene en cuenta los efectos del oído medio y externo, los cuales modela mediante un filtro FIR. Además, realiza una simulación del análisis en frecuencia de la membrana basilar que tiene lugar en el oído interno usando un banco de filtros gammatono, cuyos anchos de banda varían con la frecuencia. Para cada salida de filtro, se calcula la correlación cruzada entre la señal original y la distorsionada segmentando ambas señales en *frames* de 30 ms de duración. Se determina el valor máximo de la correlación cruzada para cada filtro y *frame*. Los valores máximos obtenidos en cada uno de los filtros se suman, con una ponderación que dependerá de la magnitud de salida de la señal distorsionada en cada filtro. La correlación cruzada ponderada proporcionará una medida de la percepción de la distorsión, denominada  $R_{\text{nonlin}}$ , que puede ser usada para predecir valoraciones subjetivas de calidad de sonido. Cuanto más alta sea la correlación en las salidas de cada filtro entre la señal original y la distorsionada menor será la distorsión –debe entenderse percibida en este caso- y consecuentemente mayor la calidad percibida.

Veamos con algo más de detalle cada uno de los bloques del procesado:

En primer lugar, la señal de test puede ser música o voz. Esta es una de las particularidades de la métrica. No se utiliza una señal estacionaria perfectamente determinada como en la inmensa mayoría de métricas de calidad. Al tratarse de una predicción de una evaluación subjetiva, el procesado puede hacerse sobre cualquier tipo de señal que pudiera ser objeto de un experimento o encuesta psicoacústica. Para la aplicación del modelo es necesario disponer tanto de la forma de onda de la señal original como de la forma de onda de la señal una vez ha pasado por el sistema bajo test, es decir, con los productos de distorsión causados por el sistema no lineal en su caso o la degradación acústica causada por el canal de transmisión. Es necesario que ambas señales estén perfectamente alineadas en el momento de iniciar el procesado.

Ambas señales se inyectan a un filtro FIR de 4097 coeficientes cuyo objetivo es simular los efectos de transmisión en el oído externo y medio. Este filtro, descrito en (Glasberg & Moore, 2002) producirá una atenuación de los componentes de frecuencia por debajo de 500 Hz y por encima de 5000 Hz. Tras la simulación de los efectos de transmisión en el oído externo y medio ambas señales se enrutan separadamente a un banco de 40 filtros gammatono, cada uno de los cuales con un ancho de banda de 1 ERB<sub>N</sub>. Estos filtros

proporcionan una simulación en el dominio temporal computacionalmente eficaz del filtrado en frecuencia realizado en la cóclea. Su respuesta en frecuencia es muy similar a la de los filtros *roex*, frecuentemente usados para modelar la selectividad en frecuencia del sistema auditivo humano como hemos visto en el capítulo anterior. La complejidad del modelado de los filtros gammatono está en función de las características de las señales de test<sup>10</sup>. En este caso, únicamente se implementa la parte principal de cada filtro auditivo, que produce un modelo adecuado del análisis en frecuencia para la mayor parte de las situaciones reales, como las señales musicales o de voz. La modificación de las caídas o colas de la respuesta en frecuencia de estos filtros sólo es relevante cuando se utilizan señales filtradas con bandas de paso muy abruptas (Patterson & Moore, 1986). Las frecuencias centrales de este banco de filtros están espaciadas en intervalos de 1 ERB<sub>N</sub> y cubren el rango de frecuencias desde 50 Hz a 19739 Hz. Se asume así que el filtro auditivo de más baja frecuencia está centrado en 50 Hz.

En la siguiente etapa, para la salida de cada filtro se calcula la correlación cruzada normalizada entre la señal de referencia y la señal distorsionada. La salida de cada filtro se divide en *frames* de 30 ms de duración que no se solapan entre ellos. La correlación cruzada se calcula entre la forma de onda del *i*-ésimo *frame* en la salida del filtro que procesa la señal distorsionada y la forma de onda resultante de concatenar los frames *i-1*, *i* e *i+1* en la salida del filtro que procesa la señal original o de referencia. Para cada salida del banco de filtros y cada *frame* se obtiene el valor  $X_{\max}$ , que es el valor máximo de la correlación cruzada normalizada. De esta manera se tendrá un valor de  $X_{\max}$  por cada *frame* (o 30 ms de señal) y cada filtro auditivo del banco de filtros. Además del valor máximo de la correlación entre la señal distorsionada y la de referencia, para cada *frame* de la señal distorsionada también se calcula y obtiene el nivel máximo en dB. El registro de estos niveles resultará útil a la hora de determinar la ponderación que se le aplicará al valor de  $X_{\max}$  obtenido para el mismo *frame* y filtro, como se verá más adelante con detalle. El promediado ponderado de los valores  $X_{\max}$  obtenidos para cada *frame* considerando el resultado en la salida de cada uno de los 40 filtros y posteriormente el promediado entre todos los frames determinará el valor de  $R_{\text{nonlin}}$ . Hasta aquí hemos visto una síntesis del modelo tal y como se define en (Tan, Moore, Zacharov, & Ville-Veikko, 2004). Veamos a continuación los detalles de nuestra implementación a la hora de aplicarlo en sistemas reales.

---

<sup>10</sup> Se ha visto que en los experimentos de obtención del perfil –respuesta en frecuencia- y ancho de banda de los filtros auditivos humanos basados en el modelo *notched noise*, el perfil de la respuesta en frecuencia estaba en función del tipo de señal utilizada para realizar el experimento y del ancho de banda de ésta. Así, para determinadas señales de test –especialmente señales filtradas con filtros de orden alto que provocan una brusca caída o subida de la respuesta en frecuencia- no es suficiente con modelar un filtro simétrico respecto de la frecuencia central sino que es necesario corregir las caídas a ambos lados de la frecuencia central del filtro solapando respuestas que modifican las colas o caídas de la respuesta en frecuencia haciéndola asimétrica. Para filtros *roex*, únicamente definidos en el dominio de la frecuencia, estas correcciones son fácilmente implementables. En cambio, para filtros gammatono la complejidad es mayor al tener que ser definidos en el dominio temporal.

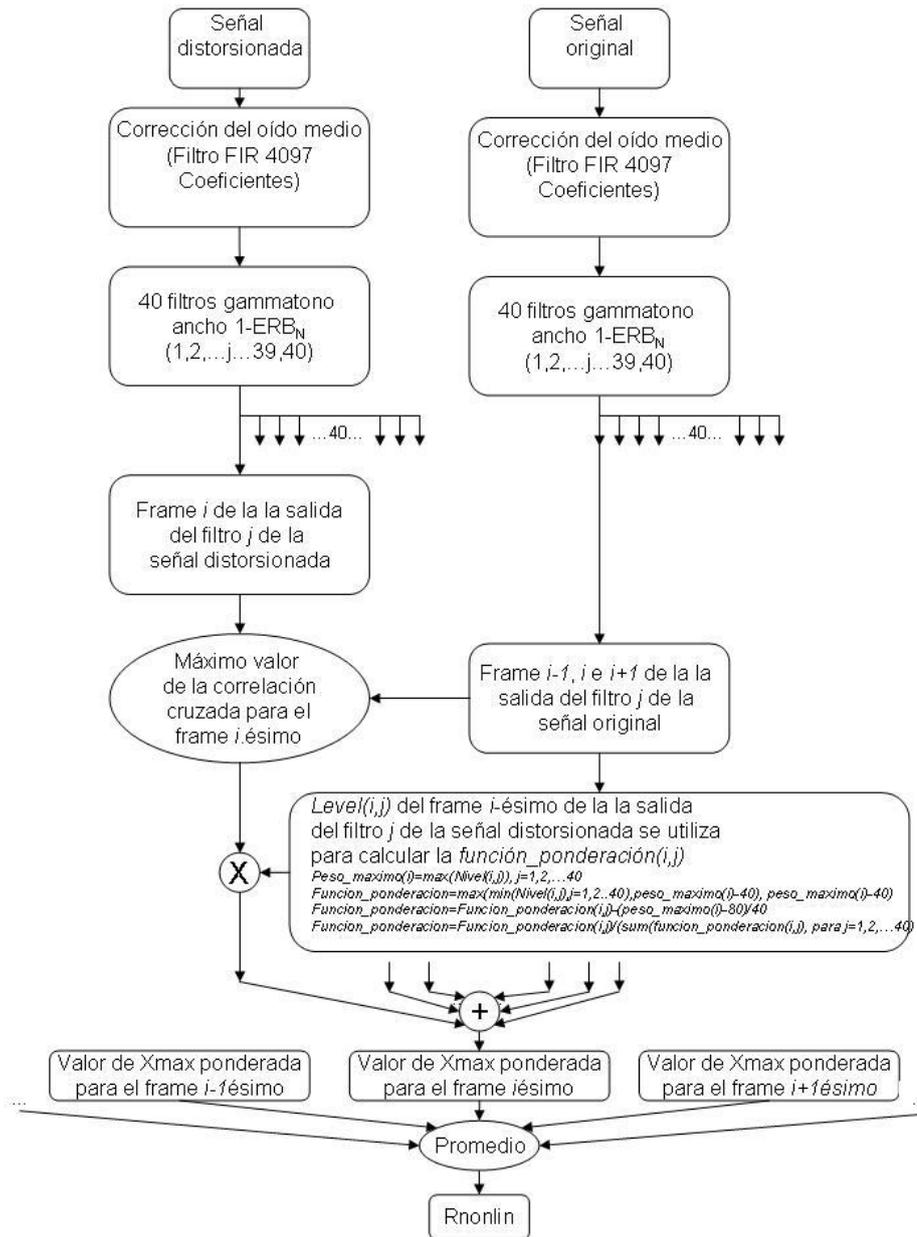


Figura 6.1. Diagrama de bloques del procesado para el cálculo de  $R_{nonlin}$ .

### 6.3 Alineado de señales

El uso de la correlación cruzada como parte fundamental de la métrica en *frames* de tan sólo 30 milisegundos de duración hace necesario que las señales estén perfectamente alineadas. A la hora de realizar cálculos sobre una señal con distorsión generada artificialmente, ésto no supone ningún inconveniente, dado que la distorsión se aplica matemáticamente sobre la señal original. La aplicación en un transductor o un canal de transmisión real exige el registro de la señal emitida por éste, que será la señal distorsionada que correlaremos con la original. Las particularidades de la grabación se considerarán más adelante. Así pues, la señal registrada tendrá que ser alineada con la original antes de realizar el procesado. Esto es estrictamente necesario puesto que las correlaciones se calculan comparando ambas señales en ventanas temporales de 30 ms. Sin un alineado adecuado, los valores máximos de correlación obtenidos en esas ventanas o *frames* no tendrían sentido dado que estaríamos comparando diferentes fragmentos de señal. En nuestro caso, para la implementación del alineado se ha hecho uso de la correlación cruzada. Ésta no es más que una medida de la similitud entre dos formas de onda o series temporales expresada en función de un retardo que se aplica a una de las señales. Veremos esta definición con detalle más adelante, dado que la correlación cruzada constituye una parte fundamental del procesado de la métrica, como ya se ha adelantado. En el caso del alineado de señales, puede resumirse de la siguiente forma. Si nuestra señal de referencia es  $x[n]$  y la señal que queremos comparar con  $x[n]$  es  $y[n]$ , a ésta última la someteremos a una serie de desplazamientos temporales  $\eta$  para cada uno de los cuales se calculará la siguiente expresión

$$R_{xy}(\eta) = \sum_{n=-\infty}^{\infty} x[n]y[n - \eta], \quad \eta = 0, \pm 1, \pm 2 \dots \quad (6.1)$$

De manera que la correlación cruzada se expresará en función del retardo  $\eta$ . En el caso de tener dos señales idénticas cuya única diferencia es un retardo temporal  $\tau$ , como en el ejemplo de la Figura 6.2, el máximo valor de  $R_{xy}$  ocurrirá cuando se tenga  $\eta = \tau$ , de forma que inmediatamente se determinará el desfase entre las dos señales. Cuando  $x[n]$  e  $y[n]$  son dos series discretas con un número de elementos  $N$  y  $M$  respectivamente, entonces  $R_{xy}$  tendrá una longitud (número de elementos) de  $N+M-2$  y los límites del sumatorio se establecerán entre  $-(N-1)$  y  $M-1$ .

En nuestro caso, la señal de referencia se inyectará en bornes del transductor en un caso o se radiará a la sala en el otro y la señal a comparar será la registrada con un micrófono en la garganta del motor de compresión o en un punto de medida de la sala. Al realizar las grabaciones de forma manual, inevitablemente la señal registrada tiene que ser de mayor duración que la de referencia para que el registro incluya toda la señal de referencia reproducida por el transductor, con lo cual existirá un retraso entre ambas señales. Antes de empezar el procesado las dos señales se alinearán utilizando la correlación cruzada. La Figura 6.3 muestra un ejemplo real de las señales original y

registrada sin alinear, el resultado del cálculo de  $R_{xy}$  y el alineado de las formas de onda a partir del resultado obtenido en la correlación cruzada.

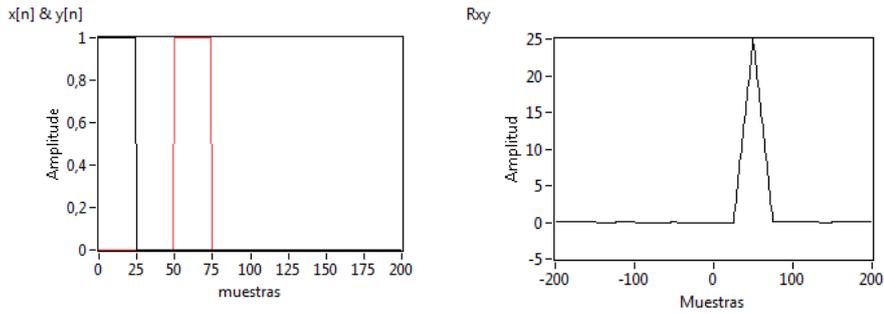


Figura 6.2. Ejemplo de correlación cruzada

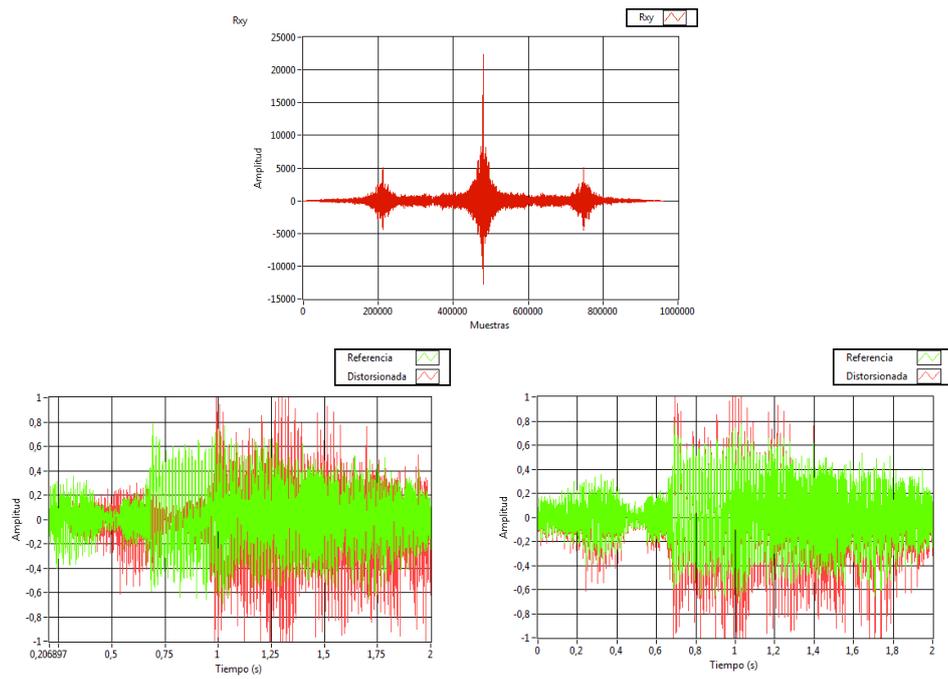


Figura 6.3. Alineado de las señales de referencia y distorsionada. Arriba, correlación cruzada. Inferior izquierda, Señales sin alinear. Inferior derecha, Señales alineadas.

## 6.4 Emulación del oído externo y medio

En (Tan, Moore, Zacharov, & Ville-Veikko, 2004) los efectos del oído medio y externo se modelan utilizando un filtro FIR de 4097 coeficientes. Este filtro es el resultado de la combinación de las funciones de transferencia del oído externo y del oído medio definidas y descritas en (Moore, Glasberg, & Gaer, 1997). El filtro resultante fue publicado en (Glasberg & Moore, 2002) como una aproximación a los datos obtenidos de forma empírica en audiencias y tratados estadísticamente. La finalidad principal de este filtro es atenuar las frecuencias por debajo de 500Hz y por encima de 5000 Hz. El relativo alto orden del filtro se asume para asegurar suficiente atenuación en baja frecuencia aunque los autores señalan que no es estrictamente necesario un orden de 4097. Pueden implementarse filtros IIR con órdenes mucho menores y resultados similares, lo cual agiliza el tiempo de cálculo. Como se observa en la Figura 6.4, entre la utilización de un FIR de 4097 coeficientes y la de un IIR de 10 más otro de 4 en cascada aparece tan sólo una zona entre 100 y 500 Hz en la que cabría evaluar las diferencias. Nuestra implementación posibilita la elección del tipo de filtro a utilizar en esta etapa, pudiéndose comprobar que su efecto en el resultado final es insignificante.

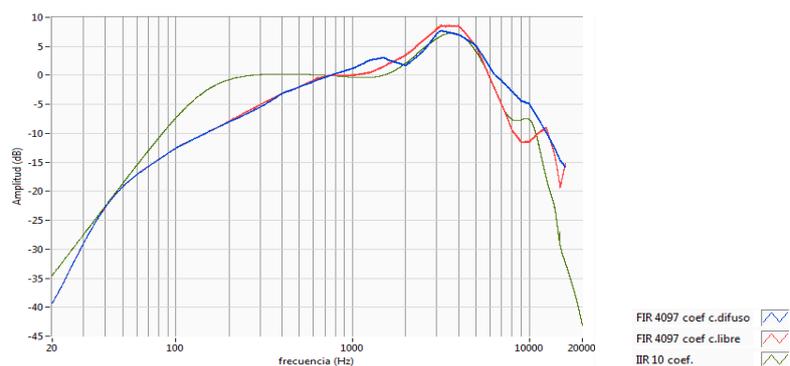


Figura 6.4. Respuesta en frecuencia de los tres filtros emuladores del oído externo y medio implementados.

## 6.5 Simulación de los filtros auditivos. Frecuencias centrales, ERB y factor de solapamiento

Para la implementación del banco de filtros gammatono se ha partido del modelo propuesto en (Patterson R.,1992). El modelo coclear de Patterson está basado en un banco de filtros paso banda independientes organizados tonotópicamente<sup>11</sup> desde la alta frecuencia en la base de la cóclea hasta la baja frecuencia en el ápex. En este modelo, el ancho de banda de cada filtro coclear se define mediante su ERB, o ancho de banda

<sup>11</sup> La organización tonotópica hace referencia al fenómeno mediante el cual las propiedades mecánicas de la membrana basilar hacen que cada zona de ésta sea sensible a un rango de frecuencias en función de la distancia de esta zona a la base, tal como se ha explicado en el capítulo 5, creando lo que se denomina gradiente tonotópico.

rectangular equivalente. Además, el espaciado entre los canales cocleares tendrá que obtenerse de manera que cada filtro solape a sus vecinos en la misma proporción.

Se ha visto que el ERB es una medida psicoacústica del ancho del filtro auditivo en cada punto a lo largo de la coclea. La siguiente es una expresión genérica para determinar el ERB dada una frecuencia central:

$$ERB = \left[ \left( \frac{f}{EarQ} \right)^n + (minBW)^n \right]^{\frac{1}{n}} \quad (6.2)$$

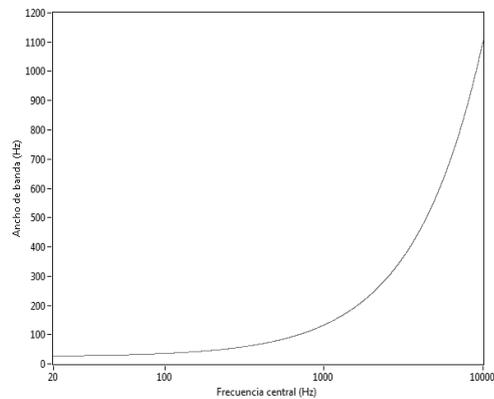
Donde  $f$  es la frecuencia central (en Hercios),  $EarQ$  es el factor de calidad del filtro en alta frecuencia (adimensional),  $minBW$  es el ancho de banda mínimo en baja frecuencia (en Hercios) y  $n$  es el orden (también adimensional). Para los canales de baja frecuencia resultará preponderante el valor que se le otorgue al término  $minBW$ . A medida que aumenta la frecuencia, el término  $minBW$  deja de ser relevante siendo el factor de calidad el que determina en mayor parte el ancho de banda equivalente correspondiente a una determinada frecuencia central. Varios autores han recomendado diferentes valores para los parámetros de la ecuación 6.2. No olvidemos que en definitiva se trata de obtener una ecuación que se ajuste a los valores experimentales que se han ido obteniendo a lo largo de la historia (ver Figura 5.18), de manera que los valores sugeridos para los parámetros de la ecuación 6.2 se han ido adaptando a medida que se obtenían nuevos resultados. Hemos visto que en (Glasberg & Moore, 1990) se recomienda una expresión para el cálculo de ERB<sup>12</sup>. En realidad, esta expresión se obtiene de aplicar a la ecuación 6.2 el valor de los parámetros sugerido por los autores. Así, Glasberg y Moore proponen  $EarQ=9,26449$ ;  $minBW=24,7$  y  $n=1$ . Con lo cual, el ancho de banda de un filtro coclear según su frecuencia central vendrá dado por

$$ERB = \frac{f}{9.26449} + 24.7 \quad (6.3)$$

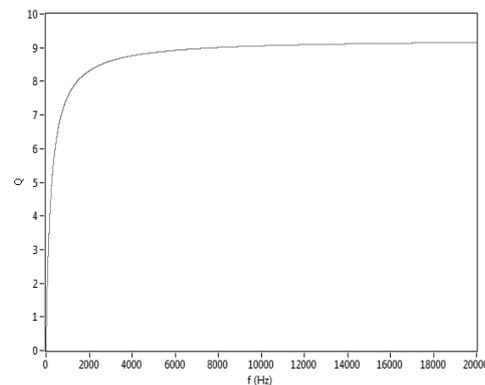
Expresión que resulta equivalente a la ecuación 5.4. La Figura 6.5 muestra el ancho de banda del filtro auditivo correspondiente a cada frecuencia central y la Figura 6.6 el factor de calidad del filtro auditivo en función de su frecuencia central. Se observa, como se ha comentado anteriormente, que a medida que la frecuencia central disminuye por debajo de 500 Hz el ancho de banda tiende asintóticamente a 24.7 Hz. A estas frecuencias el factor de calidad del filtro es menor (ver Figura 6.6). Al incrementarse la frecuencia central del filtro, especialmente a partir de 4 kHz, el factor de calidad del filtro tiende asintóticamente a 9.2645.

---

<sup>12</sup> Ver ecuación 5.4



**Figura 6.5. ERB en función de la frecuencia central**



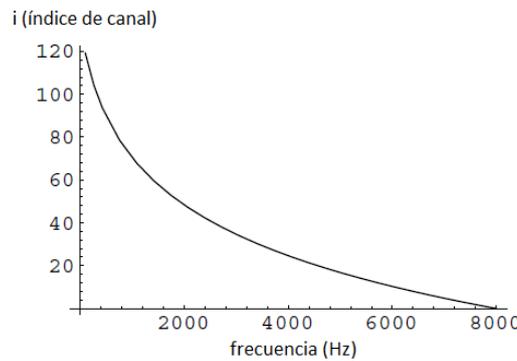
**Figura 6.6. EarQ en función de la frecuencia central.**

En el modelo de sistema auditivo que estamos implementando cada filtro posee un ancho de banda de 1 ERB, pero aún no se ha determinado el espaciado entre los canales. Evidentemente, en una simulación, por grande que sea ésta, no se podrá discretizar la totalidad de puntos de la membrana basilar que son sensibles a cada frecuencia. El sistema real – la cóclea – es una distribución continua de células ciliadas que tiene lugar a lo largo de la membrana basilar conformando un gradiente tonotópico. La simulación hace necesario discretizar este sistema de análisis en frecuencia, con lo cual habrá que acotar el número de filtros que posee el sistema y establecer la separación entre éstos. En (Slaney, 1993) se introduce un concepto que resulta útil en el diseño de bancos de filtros auditivos: el factor de solapamiento o *step factor*. Un factor de solapamiento cercano a cero indica que los filtros se superponen casi completamente. Un factor de solapamiento igual a 0,5 indicará que cada frecuencia en la señal de entrada estará en la banda de paso de dos canales cocleares. Con un *step factor* de valor 1 tendremos un banco de filtros en el que prácticamente no habrá solapamiento entra filtros. Dado un

determinado banco de filtros, Slaney define una función que relaciona la frecuencia central de un filtro ( $f_c$ ) con su correspondiente índice de posición en la cóclea. Integra desde la frecuencia de Nyquist ( $f_N$ ) hasta una frecuencia arbitraria ( $f_c$ ) para modelar la propagación de energía a lo largo de la membrana basilar desde la base (alta frecuencia) hasta el ápex (baja frecuencia). El resultado es el índice del canal que está centrado en la frecuencia arbitraria  $f_c$ .

$$i = \frac{Q \ln(f_N + Q \cdot BW_{\min})}{\text{stepfactor}} - \frac{Q \ln(f_c + Q \cdot BW_{\min})}{\text{stepfactor}} \quad (6.4)$$

La Figura 6.7 muestra la relación entre frecuencia central y número de índice para un banco de filtros con una frecuencia de muestreo de 16 kHz (la frecuencia de Nyquist es de 8 kHz) y un factor de solapamiento de 0,25.



**Figura 6.7. Índice del filtro en función de su frecuencia central.  $F_N=8$  kHz y  $\text{step factor}=0,25$ .**

Si se invierte la ecuación 6.4 haciendo que la frecuencia central del filtro sea una función del índice  $i$ , podremos calcular la frecuencia central exacta para cada uno de los filtros de nuestro banco.

$$f_c = -Q \cdot BW_{\min} + \frac{f_N + Q \cdot BW_{\min}}{e^{\frac{i \cdot \text{stepfactor}}{Q}}} \quad (6.5)$$

Así, la ecuación 6.5 permite calcular las frecuencias centrales ( $f_c$ ) correspondientes a un banco de  $N$  filtros ( $n=1, 2, \dots, N$ ), con un factor de solapamiento ( $\text{stepfactor}$ ) y una frecuencia de Nyquist ( $f_N$ ) determinadas. En el modelo propuesto por (Moore, Tan, Zacharov, & Ville-Veikko, 2004) se utiliza un banco de 40 filtros y una frecuencia central mínima de 50 Hz. Todo ello con una frecuencia de Nyquist de 22050 Hz. Dado que conocemos la frecuencia de Nyquist, la frecuencia central mínima ( $f_c=50$  Hz) y su índice ( $i=40$ ) podemos obtener el factor de solapamiento que necesitaremos utilizar para cubrir todas las frecuencias entre 50 y 22050 Hz con nuestro banco de 40 filtros.

$$\text{stepfactor} = \frac{Q \ln(f_N + Q \cdot BW_{\min})}{i} - \frac{Q \ln(f_C + Q \cdot BW_{\min})}{i} = 1,015$$

De lo que se deduce que el banco propuesto para el cálculo de  $R_{\text{nonlin}}$  está formado por filtros que no llegan a solaparse o lo hacen en las inmediaciones de las frecuencias de corte  $-3$  dB de cada filtro, dada la cercanía a la unidad del factor de solapamiento. Con esto ya podemos calcular las frecuencias centrales para cada uno de los 40 canales utilizando la expresión 6.5 y sus anchos de banda correspondientes mediante la ecuación 6.3. La siguiente tabla muestra la frecuencia central y el ERB para cada uno de los 40 filtros.

**Tabla 6.1. Fc y ERB para cada filtro del banco.**

i	fc	ERB	i	fc	ERB	i	fc	ERB	i	fc	ERB
1	19739	2155	11	6450	721	21	2005	241	31	518	81
2	17668	1932	12	5757	646	22	1773	216	32	441	72
3	15811	1731	13	5136	579	23	1566	194	33	371	65
4	14147	1552	14	4580	519	24	1379	174	34	309	58
5	12656	1391	15	4081	465	25	1213	156	35	253	52
6	11319	1246	16	3634	417	26	1063	139	36	203	47
7	10121	1117	17	3233	374	27	929	125	37	158	42
8	9048	1001	18	2874	335	28	809	112	38	118	37
9	8085	897	19	2552	300	29	701	100	39	82	34
10	7223	804	20	2264	269	30	605	90	40	50	30

## 6.6 Simulación de los filtros auditivos. Implementación de los filtros gammatono

En nuestro caso se ha realizado una implementación siguiendo las directrices propuestas en (Slaney, 1993), si bien utilizando herramientas diferentes, concretamente el módulo *Digital Filter Desing* del lenguaje gráfico de programación LabVIEW™. Recordemos la ecuación 5.8 vista en el capítulo anterior. Es la definición de la respuesta al impulso de un filtro gammatono.

$$g(t) = \alpha t^{n-1} e^{-2nbt} \cos(2nf_c t + \varphi)$$

$\alpha$  es un valor arbitrario utilizado para normalizar el valor de pico de la función de transferencia a 0 dB,  $n$  es el orden del filtro,  $b$  es el ancho de banda,  $f_c$  es la frecuencia central del filtro y  $\varphi$  es la fase. El orden de los filtros gammatono con el cual se considera que éstos se ajustan correctamente a la simulación del filtrado coclear es de  $n=4$

(Patterson R. , Nimmo-Smith, Holdsworth, & Rice, 1987). Con lo cual nuestra respuesta al impulso de cada filtro coclear será:

$$g(t) = \alpha t^3 e^{-8\pi b t} \cos(2\pi f_c t + \phi) \quad (6.6)$$

En la que sólo habría que sustituir el valor de la frecuencia central y del ancho de banda para cada frecuencia central obtenidos en el apartado anterior. En este sentido, Patterson recomienda utilizar un valor de  $b=1,019 \cdot \text{ERB}$  (Patterson R., 1992).

Considerando las siguientes transformadas de Laplace

$$\begin{aligned} \mathcal{L}\{f(t)\} &= F(s) \\ \mathcal{L}\{t \cdot f(t)\} &= -\frac{d}{ds} F(s) \\ \mathcal{L}\{t^n \cdot f(t)\} &= (-1)^n \frac{d^n}{ds^n} F(s) \\ \mathcal{L}\left\{\frac{\cos(\omega t)}{e^{Bt}}\right\} &= \frac{s+B}{(s+B)^2 + \omega^2} \end{aligned}$$

Teniendo en cuenta que posteriormente se reemplazará el término B por  $2\pi b$  o  $2\pi \cdot 1,019 \cdot \text{ERB}$  en su caso, podemos expresar la transformada de Laplace de nuestra respuesta al impulso utilizando los pares transformados anteriores.

$$\mathcal{L}\left\{\frac{t^n \cdot \cos(\omega t)}{e^{Bt}}\right\} = (-1)^n \frac{d^n}{ds^n} \left[ \frac{s+B}{(s+B)^2 + \omega^2} \right] \quad (6.7)$$

Puesto que el orden del filtro es  $n=4$ , al aplicar este valor en la ecuación 5.8 tendremos la ecuación 6.6, con lo cual la ecuación 6.7 quedará como

$$\mathcal{L}\left\{\frac{t^3 \cdot \cos(\omega t)}{e^{Bt}}\right\} = (-1)^3 \frac{d^3}{ds^3} \left[ \frac{s+B}{(s+B)^2 + \omega^2} \right] \quad (6.8)$$

Lo que significa que tendremos que obtener la tercera derivada respecto de la variable compleja s de la función  $\frac{s+B}{(s+B)^2 + \omega^2}$ . Podemos obviar el término  $(-1)^3$  dado que al final se ajustará la ganancia para que el pico de la función de transferencia se ajuste a 0 dB. Derivando se tendrá:

$$\begin{aligned} \frac{d^3}{ds^3} \left[ \frac{s+B}{(s+B)^2 + \omega^2} \right] &= F(s) = \\ &= \frac{-48(s+B)^4}{((s+B)^2 + \omega^2)^4} + \frac{48(s+B)^2}{((s+B)^2 + \omega^2)^3} - \frac{6}{((s+B)^2 + \omega^2)^2} = \\ &= \frac{6 \cdot (-B^4 - 4B^3s - 6B^2s^2 - 4Bs^3 - s^4 + 6B^2\omega^2 + 12Bs\omega^2 + 6s^2\omega^2 - \omega^4)}{(B^2 + 2Bs + s^2 + \omega^2)^4} \end{aligned}$$

El orden más alto que aparece en la expresión obtenida al derivar es ocho. Así pues, puede utilizarse un filtro analógico de orden ocho para implementar un filtro gammatono de cuarto orden. La expresión en el denominador tiene dos únicas soluciones al igualarla a cero, pero estas mismas raíces se repiten cuatro veces. De esta forma tendremos un par de polos conjugados que se repite en cuatro ocasiones siempre en la misma posición, concretamente el par  $(-B \pm j\omega)$ . Para encontrar los valores de la frecuencia compleja que hacen que el numerador sea igual al cero utilizamos una herramienta como la función *solve* en *MathScript*<sup>13</sup>, obteniendo las siguientes soluciones.

$$\begin{aligned} s &= -B + \left(2\sqrt{3 + 2^{3/2}}\right) \omega ; s = -B - \left(2\sqrt{3 + 2^{3/2}}\right) \omega ; \\ s &= -B + \left(2\sqrt{3 - 2^{3/2}}\right) \omega ; s = -B - \left(2\sqrt{3 - 2^{3/2}}\right) \omega \end{aligned}$$

Este sería el análisis de polos y ceros de nuestro filtro en el dominio de la frecuencia compleja. Lamentablemente, este análisis sólo es válido para filtros analógicos, en los que el tiempo no está discretizado, de manera que no es implementable como filtro digital. Necesitaremos convertir los filtros obtenidos en tiempo continuo en su equivalente en tiempo discreto. Para ello existen diversas técnicas, si bien se ha seguido lo propuesto en (Cooke, 1993) al respecto de los filtros gammatono: utilizar la técnica de la invarianza al impulso. Dado que la respuesta en frecuencia de los filtros está definida en función de su respuesta al impulso, la técnica de la invarianza al impulso parece la opción más adecuada. La técnica se fundamenta en que la respuesta al impulso del sistema discreto es una versión muestreada de la respuesta al impulso del sistema continuo. Esta técnica requiere reescribir la expresión del filtro obtenida al derivar la ecuación 6.8 en su expansión en fracciones parciales y posteriormente reemplazar cada polo por su equivalente discreto.

Dado que tenemos múltiples polos con el mismo valor no se puede realizar una simple expansión en fracciones parciales, en (Slaney, 1993) se propone como solución, agrupar los polos complejos obtenidos en pares conjugados y asignar cada uno de esos pares conjugados a un filtro diferente, con lo cual se tendrán cuatro filtros. Considerando que hay cuatro ceros, Slaney asigna un cero a cada uno de los filtros. Así un filtro continuo con raíz compleja se transforma en un filtro digital con coeficientes complejos. Afortunadamente, combinando los resultados de transformar dos polos complejos

<sup>13</sup> Mathscript es un lenguaje de scripting integrado dentro de LabVIEW™, con funciones similares a Matlab™ o Mathematica™.

conjugados, el filtro digital resultante tendrá coeficientes reales (Cooke, 1993). El filtro digital estará formado al final por un conjunto de cuatro filtros bicuadráticos en cascada por cada canal coclear. Cada una de las cuatro etapas del filtro se expresará de la forma:

$$\frac{r_1}{s - p_1} + \frac{r_2}{s - p_2} \quad (6.9)$$

Donde  $r_1$  y  $r_2$  son los residuos obtenidos al evaluar la función original en  $p_1$  y  $p_2$  respectivamente. La técnica de la invarianza al impulso se basa en que la transformada de Laplace de un impulso muestreado equivale a la transformada Z de la secuencia correspondiente cuando se tiene  $s = \frac{\ln(z)}{T}$ , donde T es el intervalo o periodo de muestreo. De forma que si a las expresiones obtenidas al expandir el filtro original en cuatro filtros de segundo orden, que tienen la forma general de la ecuación 6.9, se aplica la transformación

$$H(s) = \frac{r}{s - p} \rightarrow H(z) = \frac{Tr}{1 - \frac{e^{Tp}}{z}} = \frac{Trz}{z - e^{Tp}} \quad (6.10)$$

Obtendremos la transformada Z de cada una de las etapas del filtro.

$$\begin{aligned} H_1(z) &= \frac{-2Tz^2 + z \cdot \left[ \frac{2T\cos(2\pi f_c T)}{e^{BT}} - \frac{2\sqrt{3 + 2^{3/2}T\sin(2\pi f_c T)}}{e^{BT}} \right]}{\frac{-2}{e^{2BT}} - 2z^2 + \frac{4z\cos(2\pi f_c T)}{e^{BT}}} \\ H_2(z) &= \frac{-2Tz^2 + z \cdot \left[ \frac{2T\cos(2\pi f_c T)}{e^{BT}} - \frac{2\sqrt{3 + 2^{3/2}T\sin(2\pi f_c T)}}{e^{BT}} \right]}{\frac{-2}{e^{2BT}} - 2z^2 + \frac{4z\cos(2\pi f_c T)}{e^{BT}}} \\ H_3(z) &= \frac{-2Tz^2 + z \cdot \left[ \frac{2T\cos(2\pi f_c T)}{e^{BT}} + \frac{2\sqrt{3 - 2^{3/2}T\sin(2\pi f_c T)}}{e^{BT}} \right]}{\frac{-2}{e^{2BT}} - 2z^2 + \frac{4z\cos(2\pi f_c T)}{e^{BT}}} \\ H_4(z) &= \frac{-2Tz^2 + z \cdot \left[ \frac{2T\cos(2\pi f_c T)}{e^{BT}} - \frac{2\sqrt{3 - 2^{3/2}T\sin(2\pi f_c T)}}{e^{BT}} \right]}{\frac{-2}{e^{2BT}} - 2z^2 + \frac{4z\cos(2\pi f_c T)}{e^{BT}}} \end{aligned} \quad (6.11)$$

Siendo  $f_c$  la frecuencia central del filtro,  $B$  el ancho de banda y  $T$  el intervalo de muestreo. Las ecuaciones anteriores se pueden expresar como ecuación en diferencias, de manera que puedan ajustarse a una implementación como la mostrada en el diagrama de bloques de la Figura 6.8, donde cada  $z^{-1}$  equivale al retardo de una muestra,  $A_i$  son los coeficientes aplicados a la entrada  $x[nT]$  y  $B_i$  los coeficientes aplicados a la salida  $y[nT]$  del filtro.

$$H(z) = \frac{A_0 + A_1z^{-1} + A_2z^{-2}}{1 + B_1z^{-1} + B_2z^{-2}} = \frac{A_0z^2 + A_1z + A_2}{z^2 + B_1z + B_2} \quad (6.12)$$

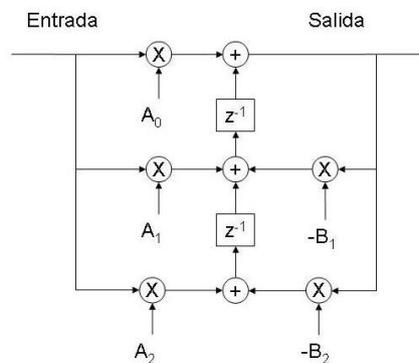


Figura 6.8. Diagrama de bloques que implementa cada una de las etapas.

Comparando cada una de las expresiones de las ecuaciones 6.11 con la forma canónica de la ecuación 6.12 se obtendrán los coeficientes necesarios para implementar el diagrama de bloques anterior.

Para  $H_1(z)$

$$A_0 = T \quad ; \quad A_1 = -\frac{1}{2} \cdot \left( \frac{2T \cos(2\pi f_c T)}{e^{BT}} + \frac{2\sqrt{3 + 2^{3/2} T \sin(2\pi f_c T)}}{e^{BT}} \right) \quad ; \quad A_2 = 0$$

$$B_1 = \frac{-2 \cos(2\pi f_c T)}{e^{BT}} \quad ; \quad B_2 = e^{-2BT}$$

Para  $H_2(z)$

$$A_0 = T \quad ; \quad A_1 = -\frac{1}{2} \cdot \left( \frac{2T \cos(2\pi f_c T)}{e^{BT}} - \frac{2\sqrt{3 + 2^{3/2} T \sin(2\pi f_c T)}}{e^{BT}} \right) \quad ; \quad A_2 = 0$$

$$B_1 = \frac{-2\cos(2\pi f_c T)}{e^{BT}} ; B_2 = e^{-2BT}$$

Para  $H_3(z)$

$$A_0 = T ; A_1 = -\frac{1}{2} \cdot \left( \frac{2T\cos(2\pi f_c T)}{e^{BT}} + \frac{2\sqrt{3 - 2^{3/2}}T\sin(2\pi f_c T)}{e^{BT}} \right) ; A_2 = 0$$

$$B_1 = \frac{-2\cos(2\pi f_c T)}{e^{BT}} ; B_2 = e^{-2BT}$$

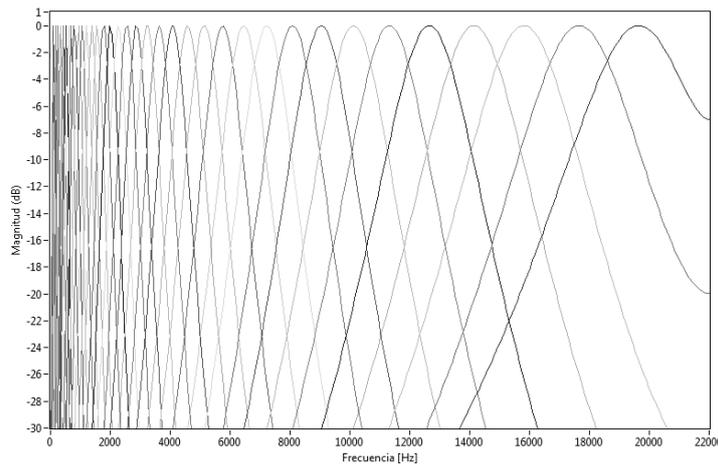
Para  $H_4(z)$

$$A_0 = T ; A_1 = -\frac{1}{2} \cdot \left( \frac{2T\cos(2\pi f_c T)}{e^{BT}} - \frac{2\sqrt{3 - 2^{3/2}}T\sin(2\pi f_c T)}{e^{BT}} \right) ; A_2 = 0$$

$$B_1 = \frac{-2\cos(2\pi f_c T)}{e^{BT}} ; B_2 = e^{-2BT}$$

La siguiente imagen representa las funciones de transferencia obtenidas para cada frecuencia central al sustituir los valores de los coeficientes calculados según las expresiones anteriores en la ecuación 6.12 para cada uno de los cuatro filtros y colocar éstos en cascada, es decir:  $H(z) = H_1(z) \cdot H_2(z) \cdot H_3(z) \cdot H_4(z)$

Como se observa en la Figura 6.9, las funciones de transferencia vecinas se cortan en las proximidades de -3dB lo cual significa que el valor de *step factor* es prácticamente 1, es decir, cada ancho de banda de 1 ERB está cubierto por un único filtro a lo largo del rango audible.



**Figura 6.9.** Funciones de transferencia del banco de filtros gammatono generado.

## 6.7 Correlación cruzada normalizada

La salida de cada filtro auditivo  $j$  se divide en *frames* de 30 ms no superpuestos, con índice  $i$ . Se calcula la correlación cruzada entre la forma de onda de cada *frame*  $i$  de la señal distorsionada y la forma de onda resultante de la concatenación de *frames*  $i-1$ ,  $i$  e  $i+1$  de la señal de referencia. La correlación cruzada normalizada para el filtro  $j$ -ésimo y *frame*  $i$ -ésimo se expresará en función del índice del retardo  $\eta$  según la ecuación 6.13.

$$r_{xy}(i, j, \eta) = \frac{\sum_{n=(i-1)L+\eta}^{iL+\eta} x(n; j)y(n - \eta; j)}{\sqrt{\left(\sum_{n=(i-1)L+\eta}^{iL+\eta} x(n; j)x(n; j)\right)\left(\sum_{n=(i-1)L+\eta}^{iL+\eta} y(n - \eta; j)y(n - \eta; j)\right)}} \quad (6.13)$$

Para una frecuencia de muestreo  $f_s=44.1$  kHz, la longitud  $L$  de cada *frame* será de  $L=1323$  muestras (30 ms). El intervalo de retardos  $\eta$  a que se somete la señal  $y[n]$  para el cálculo de la relación es el equivalente a un desplazamiento temporal entre -10 ms y 10 ms (Tan, Moore, Zacharov, & Ville-Veikko, 2004), con lo cual, para  $f_s=44.1$  kHz se tendrá que  $\eta$  variará entre -441 y 441 muestras. Los valores de  $R_{xy}$  obtenidos variarán entre 1 y -1. Para cada *frame* y filtro se registra el máximo valor obtenido en la correlación cruzada normalizada entre todos los valores de  $\eta$  – de -10 ms a 10 ms -. Esto permite tener en cuenta pequeños retardos temporales dependientes de la frecuencia en el sistema no lineal (Yoo, 2009). La métrica se fundamenta en el supuesto de que el valor máximo de la correlación cruzada normalizada,  $X_{max}$ , proporciona una medida del grado en que la señal distorsionada fue afectada por componentes que no estaban presentes en la señal original. Así, cuanto mayor sea la influencia de los productos de distorsión o la degradación de la señal menor será el valor  $X_{max}$  obtenido. Una  $X_{max}=1$  indicará la inexistencia de distorsión o degradación, dado que ambas señales se superponen perfectamente para un determinado valor de  $\eta$ .

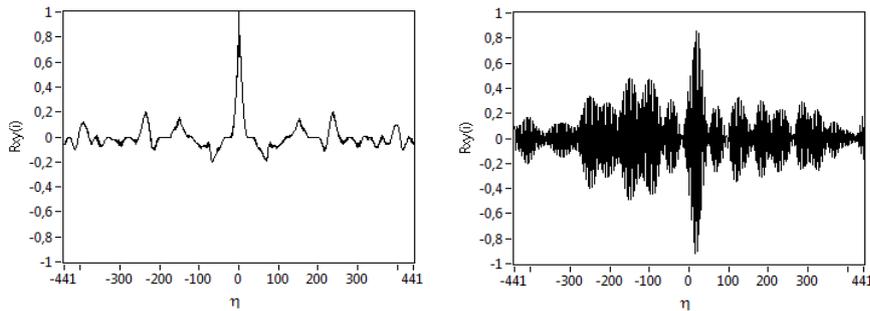


Figura 6.10. Ejemplos de correlación cruzada normalizada.

La Figura 6.10 muestra dos ejemplos de correlación cruzada normalizada entre dos señales en función del retardo  $\eta$ , con  $X_{max} = 1$  en  $\eta=0$  para la imagen de la izquierda y  $X_{max} = 0,85$  en  $\eta=25$  en la imagen de la derecha.

### 6.8 Matriz de ponderación

Una vez procesados todos los *frames*  $i$  de las señales de referencia y distorsionada – o registrada - para cada uno de los filtros auditivos  $j$ , que en nuestro caso serán 40, obtendremos una matriz con los valores máximos de cada correlación cruzada normalizada  $X_{max}$  tal como la mostrada en la Figura 6.11.

$$\begin{array}{c}
 \begin{array}{cccccc}
 & \text{Frame 0} & \text{Frame 1} & \text{Frame 2} & \dots i \dots & \text{Frame n-1} & \text{Frame n} \\
 \text{Canal 0} & \left( \begin{array}{cccccc}
 X_{max\ 0,0} & X_{max\ 1,0} & X_{max\ 2,0} & \dots & X_{max\ n-1,0} & X_{max\ n,0} \\
 X_{max\ 0,1} & X_{max\ 1,1} & X_{max\ 2,1} & \dots & X_{max\ n-1,1} & X_{max\ n,1} \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 X_{max\ 0,38} & X_{max\ 1,38} & X_{max\ 2,38} & \dots & X_{max\ n-1,38} & X_{max\ n,38} \\
 X_{max\ 0,39} & X_{max\ 1,39} & X_{max\ 2,39} & \dots & X_{max\ n-1,39} & X_{max\ n,39}
 \end{array} \right)
 \end{array}
 \end{array}$$

Figura 6.11. Matriz con los valores de  $X_{max}$  para cada frame  $i$  y filtro  $j$ .

La métrica considera también el supuesto de que la percepción de la degradación de la señal en la salida de un determinado filtro estará relacionada con la magnitud relativa de la señal en la salida de ese filtro. Así, para los filtros en los que la salida tenga un nivel muy bajo se asumirá que el efecto perceptual de la distorsión será escaso o nulo. Un nivel de salida bajo generalmente tendrá lugar en las frecuencias muy bajas o muy altas, dado el efecto del filtro simulador del oído externo y medio que produce fuertes atenuaciones de señal en esas partes del espectro. Los niveles de señal muy bajos también pueden deberse a las limitaciones en frecuencia del dispositivo que queramos evaluar o la respuesta en frecuencia del canal de transmisión, si es el caso. En nuestro caso concreto, para los altavoces de radiación indirecta, esta limitación es especialmente remarcable en la baja frecuencia como se verá, siendo menos importante en los dispositivos de radiación directa. En el caso de las salas, no existen a priori limitaciones en la parte baja del espectro. Para implementar el promediado de la matriz de la Figura 6.11 de manera que se tenga en consideración el nivel de señal de cada *frame* a la salida de cada filtro evaluado, se calcula la potencia en cada salida y se convierte a decibelios según la expresión:

$$Nivel(i, j) = 10 \log \left( \frac{1}{L} \sum_{(i-1)L+1}^{iL} y(n; j)y(n; j) \right) \tag{6.14}$$

El nivel calculado con la expresión 6.14 se utilizará para determinar una ponderación que se aplicará al valor de  $X_{max}$  obtenido para el mismo *frame* y filtro. A cada *frame* se le asignará un valor de ponderación obtenido mediante una función ya propuesta en un trabajo anterior de los autores de la métrica (Moore, Tan, & Zacharov, 2003). El valor del coeficiente de ponderación de cada *frame* y filtro,  $W_{ij}$ , se determinará de la siguiente manera: para un determinado *frame*  $i$  de la señal distorsionada, se busca el valor más elevado de nivel de entre todos los filtros  $j$ . Denominaremos a este valor  $Nivel_{max}(i)$ . A la salida del filtro en el que se da este valor máximo se aplicará la ponderación máxima. Para todos los filtros de ese mismo *frame* cuyo  $Nivel(i,j)$  cumpla que esté dentro de un rango de 40 dB por debajo de  $Nivel_{max}(i)$  se aplicará la misma ponderación máxima a  $X_{max}$ . Cuando  $Nivel(i,j)$  esté por debajo de 80 dB de  $Nivel_{max}(i)$  se aplicará un peso nulo a  $X_{max}$ . Para valores entre 40 y 80 dB por debajo de  $Nivel_{max}(i)$  el peso será linealmente decreciente con el nivel. Finalmente, los pesos se normalizarán de manera que la suma de todos ellos es la unidad, como muestra la Figura 6.12.

$$\begin{array}{c}
 \begin{array}{cccccc}
 & \text{Frame 0} & \text{Frame 1} & \text{Frame 2} & \dots & \text{Frame n-1} & \text{Frame n} \\
 \text{Canal 0} & W_{0,0} & W_{1,0} & W_{2,0} & \dots & W_{n-1,0} & W_{n,0} \\
 \text{Canal 1} & W_{0,1} & W_{1,1} & W_{2,1} & \dots & W_{n-1,1} & W_{n,1} \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \text{Canal 38} & W_{0,38} & W_{1,38} & W_{2,38} & \dots & W_{n-1,38} & W_{n,38} \\
 \text{Canal 39} & W_{0,39} & W_{1,39} & W_{2,39} & \dots & W_{n-1,39} & W_{n,39}
 \end{array} \\
 \left( \begin{array}{cccccc}
 \sum W_{0,i} = 1 & \sum W_{1,i} = 1 & \sum W_{2,i} = 1 & \dots & \sum W_{n-1,i} = 1 & \sum W_{n,i} = 1
 \end{array} \right)
 \end{array}$$

Figura 6.12. Matriz de coeficientes de ponderación  $W_{ij}$ .

La ponderación se realizará multiplicando cada  $X_{maxij}$  con su correspondiente  $W_{ij}$ , es decir, multiplicando cada elemento de la matriz mostrada en la Figura 6.11 con su homólogo en la matriz de la Figura 6.12. A continuación se realizará la suma de los valores no nulos de cada columna (*frame*) con lo que obtendremos el promedio global de  $X_{max}$  en cada *frame*, que denominaremos  $Xw_n$ , siendo n el número de *frames*.

$$\begin{array}{cccccc}
 \text{Frame 0} & \text{Frame 1} & \text{Frame 2} & \dots & \text{Frame n-1} & \text{Frame n} \\
 \left[ \begin{array}{cccccc}
 Xw_0 & Xw_1 & Xw_2 & \dots & Xw_{n-1} & Xw_n
 \end{array} \right]
 \end{array}$$

Figura 6.13. Vector de valores máximos de correlación ponderados entre todos los filtros para cada *frame*.

Finalmente se calculará el promedio de todos los  $Xw_n$ , cuyo resultado se denomina  $R_{nonlin}$  y constituye el indicador de la métrica.

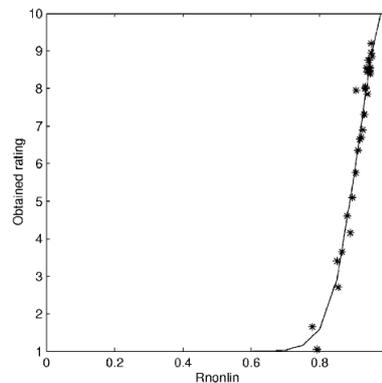
$$R_{nonlin} = \frac{\sum_0^N Xw_n}{N} \quad (6.15)$$

### 6.9 Predicción

Con el fin de predecir evaluaciones subjetivas de distorsión realizadas en experimentos psicoacústicos como, por ejemplo, test de escucha, los autores de la métrica proponen utilizar una expresión que relacione los valores de  $R_{nonlin}$  con una escala subjetiva de percepción de distorsión, graduada de 1 a 10, en la que 10 indica ausencia total de distorsión y 1 señal muy distorsionada. Esto se consigue ajustando una curva a los datos obtenidos en test subjetivos de audición. La curva tiene la forma propuesta por la ecuación 6.16.

$$Predicted\ Rating = \frac{a + b \cdot (R_{nonlin})^c}{a + (R_{nonlin})^c} \quad (6.16)$$

Siendo  $a$ ,  $b$  y  $c$  parámetros de la función que se utilizan para un ajuste fino de la curva a los datos. Se muestra un ejemplo de curva ajustada de valoraciones subjetivas respecto valores de  $R_{nonlin}$  en la siguiente imagen.



Fuente: (Tan, Moore, Zacharov, & Ville-Veikko, 2004)

**Figura 6.14. Curva  $R_{nonlin}$  ajustada a valoraciones subjetivas.**

La Figura 6.14 muestra un ajuste de resultados de la métrica comparados con valoraciones subjetivas realizadas mediante experimento psicoacústico sobre señales con distorsión generada artificialmente en (Tan, Moore, Zacharov, & Ville-Veikko, 2004).

## **6.10 Desarrollo del software**

Sin ánimo de extendernos en los detalles de implementación del procesado expuesto hasta aquí, sí creemos necesario explicar brevemente los aspectos más relevantes de la aplicación desarrollada para realizar el cálculo de  $R_{\text{nonlin}}$ . El desarrollo de la aplicación ha exigido un importante esfuerzo de investigación y testeo de los procesados implementados, dado que el trabajo que fundamenta la métrica (Tan, Moore, Zacharov, & Ville-Veikko, 2004) expone los rasgos generales de ésta sin entrar en detalles de ejecución o implementación, lo que hace necesario recurrir a una considerable y diversa cantidad de fuentes para su desarrollo, especialmente en lo concerniente al diseño e implementación de los filtros auditivos.

Se ha utilizado el entorno de programación LabVIEW™ y se ha generado el código de manera modular, adaptándose a la estructura del procesado de la métrica. Así pues, no se ha desarrollado una gran aplicación sino una serie de componentes básicos que conforman una librería de instrumentos con la que poder realizar las diferentes operaciones que conforman la métrica. Estos instrumentos pueden utilizarse de forma individual, para realizar un cálculo o un procesado específico o bien pueden integrarse en una secuencia mayor que incluya todos los pasos conducentes a la obtención de  $R_{\text{nonlin}}$ . Esta forma de proceder tiene la ventaja de poder compilar en futuras ocasiones una aplicación adaptada a unas necesidades específicas en poco tiempo, dado que ya dispondremos de la mayor parte – y la más compleja – del código fuente. Además, tiene la ventaja de poder integrar nuevos módulos dentro de la estructura sin tener que reescribir código ya compilado.

Los módulos de procesado implementados trabajan con archivos de sonido *.wav* por ser un formato sin compresión. Todo el procesado implicado en el cálculo de la métrica se realiza sobre archivos de audio. La mayoría de módulos básicos de filtrado tienen como entrada un archivo de audio y generan en la salida otro archivo de audio con la forma de onda procesada. Así ocurre con el módulo para alinear la señal de referencia con la señal a analizar, el filtrado del oído medio y externo o cada uno de los filtrados que representan los canales auditivos. En el caso del alineado, los parámetros de entrada al programa son las rutas de la señal de referencia y de la señal distorsionada – que en nuestro caso, recordemos, es una señal registrada con un micrófono-.

Dado que ambas señales son monofónicas y que una vez alineadas se procesarán en paralelo, se optó por generar un único archivo estéreo que contiene las dos señales alineadas (el canal izquierdo contendrá la señal de referencia y el canal derecho la señal a evaluar). La salida del módulo será pues la ruta del fichero estéreo generado. Esta ruta puede entregarse al módulo que realizará el procesado emulador del oído medio y externo, que una vez realizado, generará un nuevo fichero con ambas señales procesadas. Este nuevo fichero se entregará al programa encargado de realizar los filtrados cocleares. Cada filtrado coclear se almacenará en un fichero estéreo siguiendo la misma estructura (canal izquierdo referencia y canal derecho señal a evaluar). El programa encargado de realizar la correlación cruzada de ambas señales irá llamando uno a uno a cada uno de

estos ficheros correspondientes a cada canal auditivo, y realizando los cálculos descritos en la ecuación 6.13 siendo la señal  $x[n]$  la del canal izquierdo del fichero –referencia- e  $y[n]$  la del canal derecho.

Una vez procesados todos los filtros cocleares se generará una matriz como la de la Figura 6.11 y otra con los valores obtenidos de aplicar la ecuación 6.14 a cada *frame* y filtro. Éstas se almacenarán en dos archivos, esta vez de datos. Posteriormente se utilizará otro módulo, que a partir de los ficheros con las matrices de máximos de correlación cruzada y niveles generarán la matriz de ponderación de Figura 6.12 y el resto de cálculos conducentes a la obtención de  $R_{\text{nonlin}}$ .

Además de los módulos de procesado se han implementado otros módulos para el cálculo y diseño del banco de filtros cocleares, así como diversas utilidades. Si bien, en este caso concreto se ha utilizado un banco de 40 filtros con un factor de solapamiento igual a 1, con las herramientas implementadas sería posible obtener el banco de filtros correspondiente a cualquier configuración que deseáramos implementar, variando frecuencias de muestreo, factor de solapamiento, número de canales, etc.

En resumen, se ha desarrollado una librería para poder implementar cualquier rutina para el cálculo de  $R_{\text{nonlin}}$  de forma que pueda adaptarse a las necesidades específicas de un tipo concreto de señal, estudio o medida. Los vi's<sup>14</sup> implementados se agrupan en 5 grupos, como muestra la Figura 6.15. El primer grupo, *gammatone*, contiene los módulos desarrollados para el diseño e implementación del banco de filtros cocleares. El segundo grupo, *archivos*, agrupa módulos encargados de organizar la estructura de archivos que se crea al realizar un procesado, explicada dos párrafos más arriba. El tercer grupo recoge los programas que implementan los filtros emuladores del oído medio y externo para diferentes frecuencias de muestreo y con las respuestas en frecuencia mostradas en la Figura 5.4. La cuarta carpeta, *procesado*, incluye todos los módulos encargados de secuenciar las tareas de procesado y está formada por programas que contienen, a su vez, módulos de las otras cuatro carpetas. Finalmente, el último grupo contiene módulos auxiliares, como por ejemplo, el vi encargado de generar el nombre del fichero a guardar en función del índice del filtro coclear o una calculadora del número de filtros necesarios en función de la frecuencia de muestreo, el factor de solapamiento y/o la frecuencia central mínima.



**Figura 6.15. Paleta con los grupos de instrumentos desarrollados para el cálculo de  $R_{\text{nonlin}}$ .**

---

<sup>14</sup> Virtual Instrument, siguiendo la terminología de LabVIEW™, equivale a un módulo o programa que puede ser llamado por o incluido en otro de mayor envergadura o complejidad.

En total se han desarrollado 32 módulos, agrupados en la jerarquía que muestra la Figura 6.16<sup>15</sup>. Como ya se ha mencionado, esta forma de implementación deja abierta la posibilidad de incluir nuevos módulos que permitan añadir nuevas funcionalidades al conjunto de herramientas. Para procesado de las grabaciones realizadas se ha implementado una secuencia a partir de los módulos desarrollados que calcula  $R_{\text{nonlin}}$  para una serie de señales cuyas rutas (de la señal de referencia y de la señal de test) se indican en un vector. De esta forma, se pueden cargar tantas señales como deseemos analizar y los cálculos se realizarán secuencialmente. Dado que los cálculos conllevan un tiempo considerable, la rutina puede dejarse en ejecución en segundo plano o programarse para ser ejecutada cuando desee el usuario. La Figura 6.17 A muestra la ventana en que se introducen las rutas de los archivos a analizar. Para cada medida, habrá siempre un fichero con la señal de referencia y un fichero de ‘test’ – la señal supuestamente distorsionada o degradada-. El control numérico a la izquierda de las rutas de las señales es el índice del vector que las contiene, de manera que para cada índice indicaremos una señal de referencia y una de test, obteniendo al final tantos valores de  $R_{\text{nonlin}}$  como índices no vacíos tengamos en el *array* con las rutas de los ficheros. Para cada par de señales referencia-test se realizan los 5 pasos mostrados en el panel y explicados en el presente capítulo.

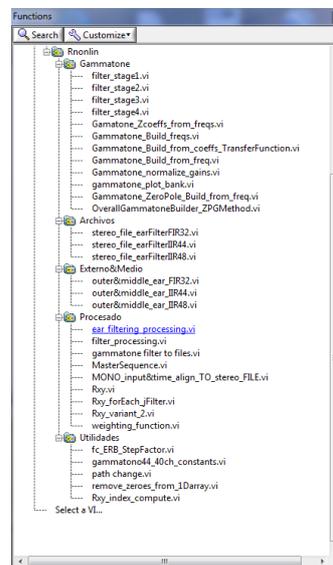
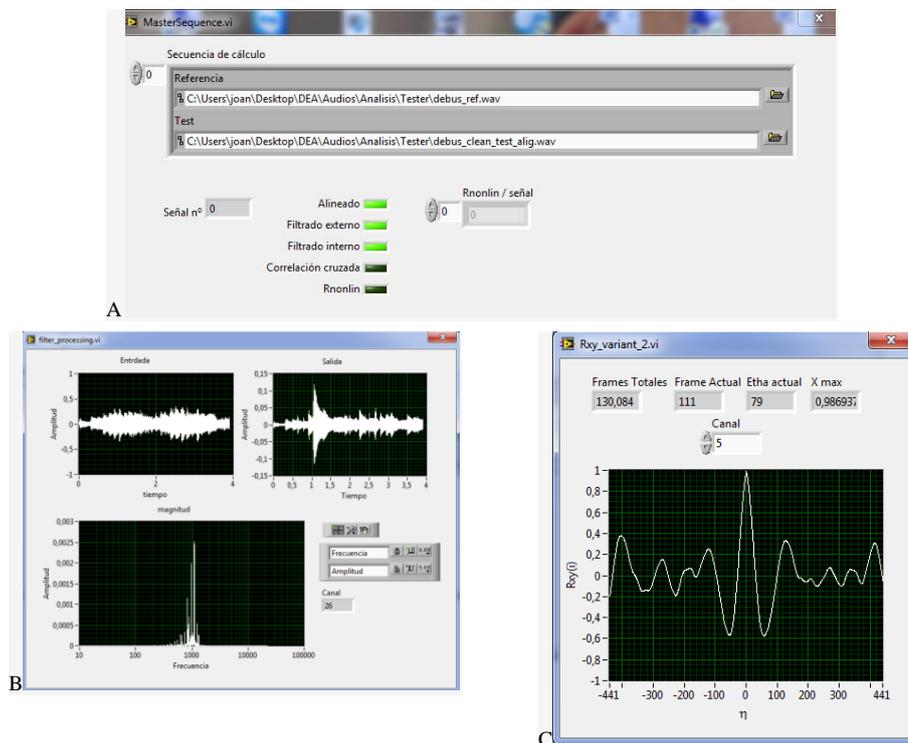


Figura 6.16. Módulos (vi's) desarrollados para el cálculo de  $R_{\text{nonlin}}$ .

<sup>15</sup> Los instrumentos virtuales desarrollados se han nombrado y documentado en inglés, con la finalidad de utilizar la misma nomenclatura que las fuentes bibliográficas consultadas.

La Figura 6.17 B y C muestran los otros dos paneles de la aplicación que representan información a medida que se desarrolla el procesado. La ventana B es un ejemplo del filtrado realizado por uno de los filtros del banco de filtros coclear y la C muestra el resultado de la correlación cruzada en función de  $\eta$  para un determinado *frame* y filtro – en el caso de la imagen *frame* 110 y canal 5 (12656 Hz).



**Figura 6.17.** Ventanas de la aplicación. A. Panel principal con vector de rutas a los archivos, indicadores de evolución del cálculo y vector de resultados. B. Ejemplo de filtrado de las señales en un canal del banco de filtros. C. Ejemplo de cálculo de correlación cruzada.

# 7 Evaluación en transductores

Se describe en este capítulo el procedimiento y resultados experimentales llevados a cabo con la implementación de la métrica en transductores reales, tanto de radiación indirecta como de radiación directa.

## **7.1 Evaluación en transductores de radiación indirecta**

Se han utilizado tres transductores de radiación indirecta de gama alta cuyas características principales se resumen en la Tabla 7.1. Son transductores comerciales y modelos actualmente vigentes en el mercado. Los denominaremos A, B y C.

### *7.1.1 Procedimiento*

Para la realización de las medidas, los motores de compresión se han acoplado a un tubo de onda plana, con la finalidad de proporcionar una carga acústica a los dispositivos. El tubo de onda plana consiste básicamente en un conducto cilíndrico y rígido que se monta sobre el motor de la misma manera que una bocina. El cilindro está parcialmente relleno con material absorbente cuya principal función es la de cancelar las reflexiones en el final del tubo. El cilindro tiene una perforación cerca del sistema de anclaje al motor que servirá para la colocación del micrófono de medida, que podrá registrar la onda proveniente del motor sin perturbar las condiciones del sistema.

**Tabla 7.1. Características de los motores de compresión utilizados**

	<b>A</b>	<b>B</b>	<b>C</b>
			
<b>Diámetro de garganta</b>	49 mm	36 mm	50 mm
<b>Impedancia</b>	8Ω	8Ω	8Ω
<b>Potencia</b>	90 W	80 W	100 W
<b>Potencia de programa</b>	180 W	160 W	200 W
<b>Rango de frecuencia</b>	600-20000 Hz	500-18000 Hz	500-20000 Hz
<b>Sensibilidad en eje 1W/1m</b>	110 dB	109.5 dB	110 dB
<b>Material del diafragma</b>	Titanio	Titanio	Titanio
<b>Tipo de imán</b>	Ferrita	Ferrita	Neodimio



**Figura 7.1. Dispositivo utilizado para el registro de señal.**

La imagen derecha de la Figura 7.1 muestra el detalle de la perforación en la que se inserta el micrófono para la medida o registro de la señal. Una vez montados los dispositivos, a éstos se les ha inyectado una serie de señales con diferentes niveles en la entrada. Para la generación de la señal se ha usado una tarjeta de sonido Tascam US122 que proporcionaba la señal de test a un amplificador BGW 7500, marca comercial y modelo cuyo mercado son los sistemas de refuerzo sonoro. Se trata de un amplificador comúnmente usado para el tipo de transductores que vamos a utilizar. Una vez inyectada

la señal en el motor de compresión, ésta es reproducida por el dispositivo, reproducción que registramos utilizando un micrófono de alta presión Brüel&Kjaer 4041 conectado a una unidad de adquisición 01dB Symphonie. La imagen de la izquierda de la Figura 7.1 muestra el montaje previo de estos dispositivos para nuestra configuración.

Respecto de las señales que se han utilizado en el experimento, han sido un total de cuatro. Todas estas señales son monofónicas, se han normalizado en amplitud y se han filtrado con un filtro paso alto Butterworth de tercer orden y frecuencia de corte 800 Hz con la finalidad de adaptar las señales al rango de frecuencia de trabajo de los transductores de radiación indirecta. Las señales de prueba se han inyectado a diferentes niveles de tensión en bornes del transductor. En primer lugar, se ha utilizado un simple tono puro de 1 kHz. La utilidad de esta señal radica en poder establecer el nivel de los ajustes de amplificación como referencia para los tres transductores y el resto de señales. Necesitamos una señal periódica cuyo cálculo de potencia sea inmediato. De esta forma, los ajustes de amplificación se realizarán tomando como referencia la señal de 1 kHz. Una vez establecidos éstos se inyectará el resto de señales sin modificarlos. La señal de 1 kHz también se ha utilizado para el cálculo de la distorsión armónica total para cada nivel de entrada y transductor, cuyos resultados pueden verse en la Tabla 7.3.

Para analizar cómo responde la métrica según el tipo de señal de test, se ha utilizado, además del tono puro de 1kHz, otra señal estacionaria consistente en otro tono de 1 kHz modulado en amplitud por una señal moduladora sinusoidal de 70 Hz, con una profundidad de modulación igual a la unidad. Este tipo de señal está relacionada con un descriptor psicoacústico denominado *roughness*, o rugosidad, y se utilizó en (Cruañes J. , Francés, Ramis, & Alba, 2008). Pero en el caso que nos ocupa la utilizaremos simplemente como señal estacionaria potencialmente capaz de introducir más productos de distorsión que un solo tono puro, dadas las características espectrales de una señal modulada en amplitud.<sup>16</sup> Las otras dos señales utilizadas han sido dos pasajes musicales. Se han seleccionado dos señales con características bien diferenciadas: el primero de ellos es una grabación de guitarra clásica, de 25 s. de duración. Al ser la guitarra el único instrumento presente en el pasaje, el contenido espectral de la señal estará limitado al propio de este instrumento. Teniendo en cuenta, además, que se han eliminado las componentes por debajo de 800 Hz, resultará que la mayor parte de la información se concentrará entre los 1000 y 5000 Hz, es decir, las frecuencias que quedan inalteradas o ligeramente reforzadas por el filtrado realizado para la emulación del oído externo y medio. En lo sucesivo a esta señal la denominaremos señal G.

El otro pasaje musical utilizado consiste en un fragmento de 59 s. de una pieza jazzística en la que están presentes una gran cantidad de instrumentos percusivos, además de bajo eléctrico, guitarra, saxo y voces. Esta variedad instrumental tiene como resultado una distribución mucho más uniforme de los componentes frecuenciales de la señal como puede apreciarse en la Figura 7.2. En adelante, a esta señal la denominaremos señal J.

---

<sup>16</sup> El espectro de la señal moduladora se desplaza a la frecuencia de la portadora.

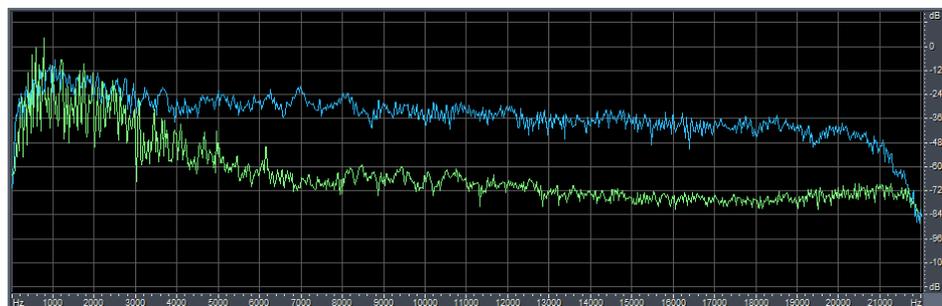


Figura 7.2. Análisis FFT de los dos pasajes musicales empleados. Verde señal G. Azul señal J.

La Figura 7.3 muestra las formas de onda de las señales G y J. Puede apreciarse por simple inspección visual que ambas señales tendrán un comportamiento dinámico también diferenciado.

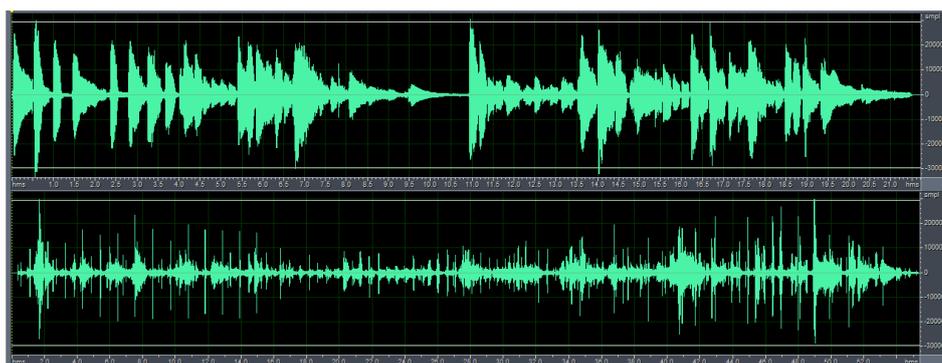


Figura 7.3. Formas de onda de las señales G (superior) y J (inferior).

En el caso de la señal G, la forma de onda está constituida por una sucesión de acordes de guitarra con un fuerte ataque y una caída lenta. Dadas las características de la pieza y del instrumento cada uno de los fragmentos de señal formado por un rápido ataque y una caída lenta tienen un fuerte contenido armónico, como puede deducirse de la densidad de picos en la FFT entre 1000 y 5000 Hz (a pesar de que la FFT de la Figura 7.2 es un promedio de toda la señal). En cambio, la señal J se caracteriza por su abundancia de transitorios (variaciones dinámicas muy rápidas) lo cual explica en parte su contenido en alta frecuencia. Por otro lado, ya se ha comentado que el experimento se ha realizado con diferentes niveles de tensión de entrada. El ajuste de amplificación de referencia se ha realizado con la señal de 1 kHz. Los valores mostrados en la Tabla 7.2 son válidos para esa señal. Cuando representemos en adelante resultados en función de la tensión de entrada, ésta estará referida a los datos mostrados en la Tabla 7.2 para 1 kHz. En la misma tabla también puede apreciarse que todo el experimento se ha realizado en un rango cómodo de potencias para los dispositivos empleados. Es evidente que la utilización de cualquier dispositivo en los límites de su rango de trabajo se traducirá en un rendimiento

deficiente y en el caso de los motores de compresión cabe esperar la aparición de una fuerte distorsión.

En el caso que nos ocupa se trata de aplicar la métrica dentro de los márgenes de trabajo de los dispositivos. Trataremos de analizar los resultados de la métrica según el tipo de señales utilizadas, según los transductores – comparando los resultados entre éstos – y finalmente, los resultados en función del nivel de la señal de entrada a los dispositivos. En este último sentido, posiblemente sí hubiera sido deseable ensayar con algún nivel de entrada superior a los 21,21 V rms. La limitación en este aspecto venía dada por el rango dinámico del micrófono con el que se realizaron los registros, cuyo nivel de presión máximo es de 160 dB. Evidentemente, si hacemos trabajar al micrófono en el límite o ligeramente por encima de su rango obtendremos una distorsión que invalidará todo el proceso con lo cual se opta por establecer como niveles de tensión máximos aquellos que generen niveles de presión sonora alrededor de 6 dB por debajo del límite del margen dinámico del micrófono. Esto considerando la señal estacionaria de 1 kHz. Una vez establecido el nivel deseado, podemos comprobar que el nivel de presión sonora generado por cada transductor es muy similar para un mismo nivel de tensión de entrada, tal como era de esperar a la vista de los datos de sensibilidad de la Tabla 7.1. En la Tabla 7.2 puede apreciarse que las variaciones en el nivel de presión sonora generado no oscilan más allá de 0,5 dB.

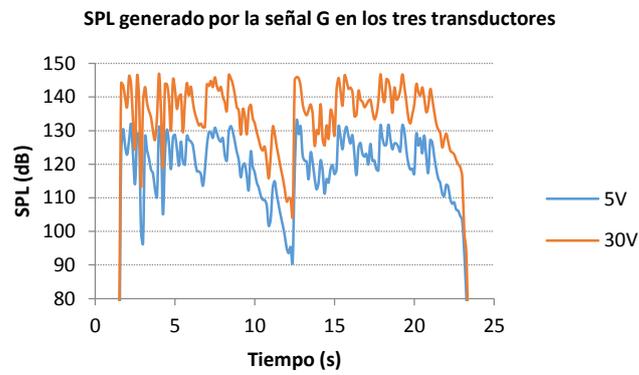
**Tabla 7.2. . Niveles de tensión en bornes y SPL para cada transductor con la señal de 1 kHz.**

V pico (V)	V RMS (V)	P (W)	Transductor A	Transductor B	Transductor C
			L SPL (dB A)	L SPL (dB A)	L SPL (dB A)
5	3,54	1,56	141,1	140,7	141,2
10	7,07	6,25	147,3	147	146,8
15	10,61	14,06	150,5	150,1	149,7
20	14,14	25,00	152,6	152,3	152,1
25	17,68	39,06	154,3	154	153,7
30	21,21	56,25	155,1	155,3	154,9

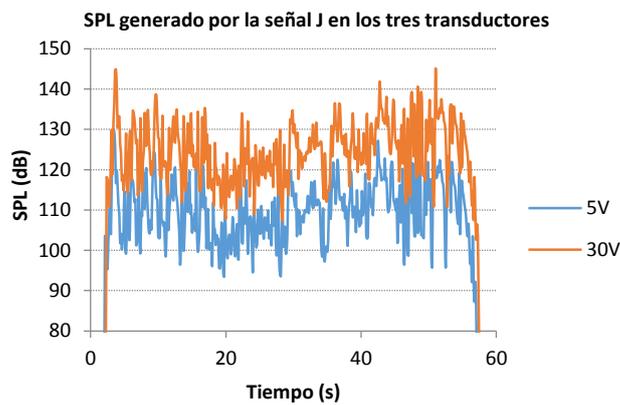
**Tabla 7.3. THD (%) para cada nivel de entrada y transductor.**

V RMS (V)	Transductor A	Transductor B	Transductor C
	THD (%)	THD (%)	THD (%)
3,54	1,54	1,33	1,10
7,07	3,92	2,85	2,15
10,61	6,61	4,26	3,21
14,14	11,94	5,93	4,53
17,68	15,13	8,41	6,37
21,21	15,75	10,85	10,26

Cuando utilizemos señales complejas, como es el caso de las señales musicales, sólo se alcanzará la tensión de pico tomada como referencia cuando la señal ocupe todo el margen dinámico. Viendo la Figura 7.3, esto sucede tan sólo en tres ocasiones para la señal G y en dos para la señal J (cuando para la señal sinusoidal sucede en cada ciclo), con lo cual la potencia generada por estas señales será menor y consecuentemente también lo será en nivel de presión sonora generado por los motores. La Gráfica 7.1 y la Gráfica 7.2 muestran los niveles mínimos y máximos de presión sonora generados con cada señal musical en el rango de niveles de entrada utilizados.



**Gráfica 7.1. Promedio SPL generado por la señal G en los tres transductores**



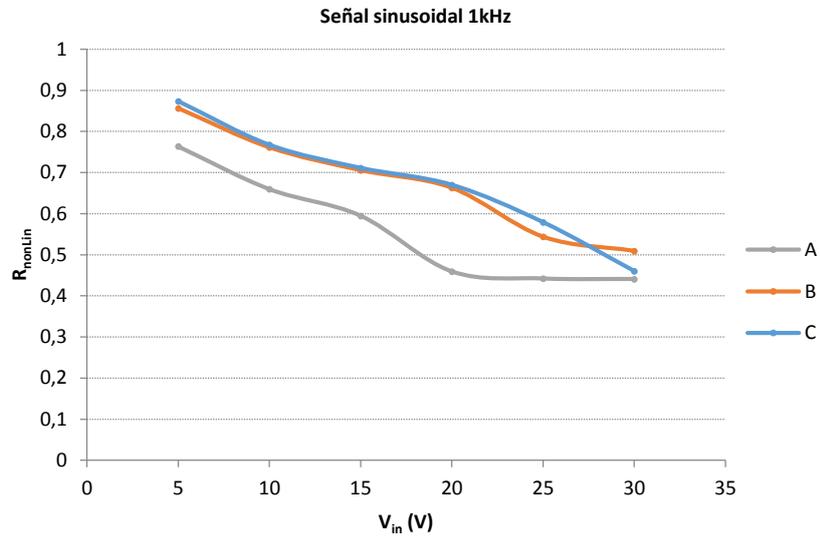
**Gráfica 7.2. Promedio SPL generado por la señal J en los tres transductores**

### 7.1.2 Resultados

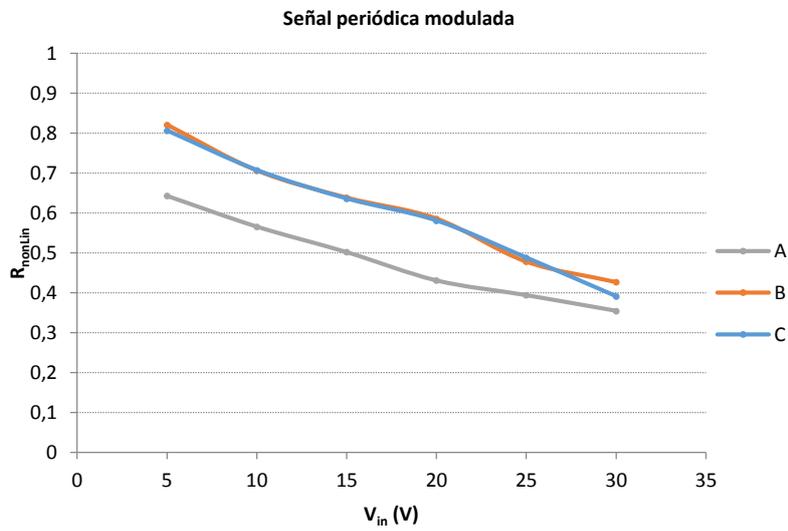
Desde la Gráfica 7.3 hasta la Gráfica 7.8 se muestran los resultados obtenidos al aplicar el algoritmo de cálculo de  $R_{\text{nonlin}}$  desarrollado a las señales registradas en los transductores bajo test. En primer lugar, se grafían los resultados obtenidos en cada transductor para las señales estacionarias (Gráfica 7.3, Gráfica 7.4). A continuación se hace lo mismo con las señales complejas (Gráfica 7.5 y Gráfica 7.6). Finalmente, la Gráfica 7.7 muestra todos los resultados obtenidos y la Gráfica 7.8 el promedio del valor de  $R_{\text{nonlin}}$  obtenido por cada señal de test en los tres transductores. En la siguiente tabla se resumen todos los resultados

**Tabla 7.4.  $R_{\text{nonlin}}$  para todos los transductores, señales y niveles**

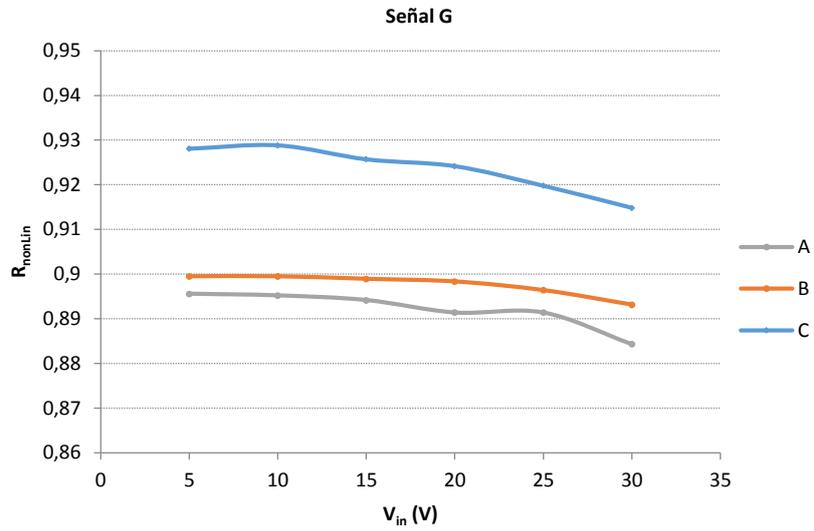
<b>Ik</b>	<b>5V</b>	<b>10V</b>	<b>15V</b>	<b>20V</b>	<b>25V</b>	<b>30V</b>
A	0,763518	0,659671	0,594830	0,459295	0,442166	0,440882
B	0,856442	0,761254	0,705740	0,662866	0,543833	0,509298
C	0,873616	0,767557	0,711372	0,670133	0,579315	0,460431
<b>Mod.</b>	<b>5V</b>	<b>10V</b>	<b>15V</b>	<b>20V</b>	<b>25V</b>	<b>30V</b>
A	0,643285	0,566145	0,502151	0,431140	0,394199	0,354873
B	0,821123	0,706642	0,638558	0,585918	0,478272	0,427100
C	0,806847	0,707769	0,636419	0,581112	0,488177	0,391415
<b>Señal G</b>	<b>5V</b>	<b>10V</b>	<b>15V</b>	<b>20V</b>	<b>25V</b>	<b>30V</b>
A	0,894616	0,894248	0,894101	0,891226	0,89132	0,884067
B	0,899525	0,899523	0,898950	0,898364	0,896429	0,893190
C	0,928100	0,928851	0,925719	0,924179	0,919795	0,914850
<b>Señal J</b>	<b>5V</b>	<b>10V</b>	<b>15V</b>	<b>20V</b>	<b>25V</b>	<b>30V</b>
A	0,913823	0,913993	0,914550	0,914753	0,914438	0,914086
B	0,920649	0,920554	0,920299	0,920740	0,919383	0,920281
C	0,926606	0,927572	0,926986	0,926904	0,927046	0,927745



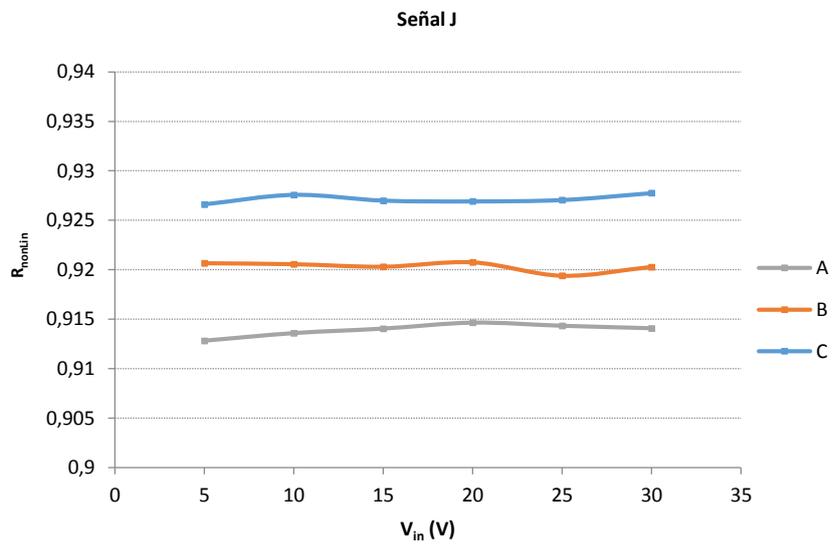
Gráfica 7.3.  $R_{nonlin}$  en función del nivel de entrada para la señal sinusoidal.



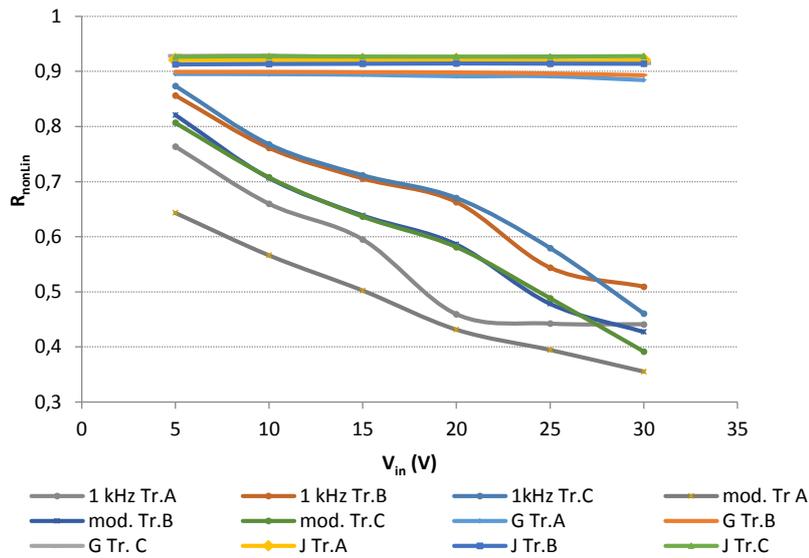
Gráfica 7.4.  $R_{nonlin}$  en función del nivel de entrada para la señal periódica modulada.



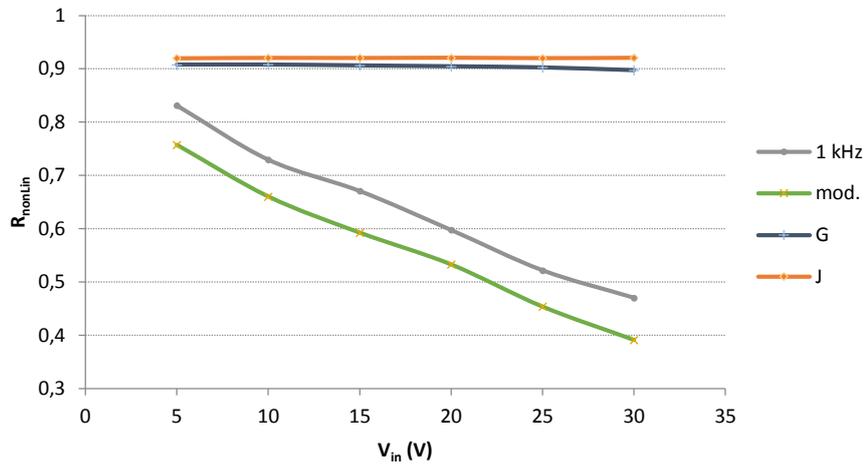
Gráfica 7.5.  $R_{nonlin}$  en función del nivel de entrada para la señal G.



Gráfica 7.6.  $R_{nonlin}$  en función del nivel de entrada para la señal J.



Gráfica 7.7.  $R_{nonlin}$  en función del nivel de entrada para las cuatro señales y tres transductores.



Gráfica 7.8.  $R_{nonlin}$  promediado para las cuatro señales de test.

### 7.1.3 Análisis

A la vista de las gráficas anteriores es evidente que las características de la señal elegida pueden resultar determinantes en los resultados obtenidos. Es muy significativa la diferencia de comportamiento de la métrica si se utilizan señales de test estacionarias o complejas. En la Gráfica 7.7 y Gráfica 7.8 se comparan todas las señales utilizadas. La Gráfica 7.7 muestra el valor de  $R_{\text{nonlin}}$  para todas las señales y transductores. Puede observarse que en los seis test realizados con señales complejas (las denominadas G y J) se han obtenido resultados con un valor de  $R_{\text{nonlin}}$  alrededor de 0,9 con una horquilla de variación entre los diferentes transductores entre 0,89 y 0,93 y que apenas experimentan variación al incrementar el nivel de entrada de la señal.

En cambio, en el caso de las señales estacionarias, los test muestran que los valores de  $R_{\text{nonlin}}$  por un lado son menores que los obtenidos con señales complejas y por otro descienden rápidamente, casi de forma lineal, a medida que se incrementa el nivel de entrada. Si tenemos en cuenta que en última instancia la finalidad de la métrica es proporcionar una predicción de una valoración subjetiva por parte de una audiencia, y que esta predicción se ajusta a una curva (dada por la ecuación 6.16, que es de tipo exponencial), se puede concluir que en el caso de las señales estacionarias prácticamente todos los valores obtenidos proporcionarían una valoración muy deficiente de los dispositivos, independientemente del ajuste que se llevara a cabo con los parámetros a, b y c de la ecuación 6.16<sup>17</sup>.

Este aspecto, junto con la diferencia observada respecto al comportamiento de la métrica con señales complejas, nos lleva a la conclusión de que las señales estacionarias son inadecuadas para su uso en esta métrica. La explicación a este comportamiento puede radicar en la poca o casi nula información que contienen estas señales en un gran número de filtros del banco de filtros auditivos. Cuando, debido a la introducción de los productos de distorsión causados por el dispositivo, en la señal contaminada aparecen armónicos o productos de intermodulación, éstos son la única información presente – y consecuentemente no enmascarada- en determinados canales auditivos, de manera que al compararla con la señal original se obtienen correlaciones muy bajas. A medida que aumenta el nivel de la señal del test, es más probable también que los niveles alcanzados por los productos de distorsión entren en el umbral de -80 a -40 dB o incluso que éstos sean inferiores a -40 dB respecto al filtro auditivo con el nivel de salida máximo, con lo cual, los valores de correlación cruzada obtenidos en estos casos tendrán un peso mayor en el promedio total del cálculo de  $R_{\text{nonlin}}$ .

---

<sup>17</sup> Por ejemplo, tomando los valores de a, b y c propuestos en (Tan, Moore, Zacharov, & Ville-Veikko, 2004) para ajustar un experimento psicoacústico realizado en el mismo trabajo se tiene que las predicciones de las valoraciones subjetivas varían entre una valoración de 1 y 1,4 – en una escala de 1 a 10 - para valores de  $R_{\text{nonlin}}$ , de 0,1 y 0,8 respectivamente. Las valoraciones entre 2 y 10 se producen para variaciones de  $R_{\text{nonlin}}$  entre 0,85 y 0,99. Con este ajuste, las variaciones más significativas en la valoración se producen para  $R_{\text{nonlin}}$  entre 0,96 y 0,99.

En definitiva, correlaciones muy bajas pueden tener una ponderación considerable o incluso igual a la del filtro auditivo con mayor nivel<sup>18</sup>. Esto se reflejará en el promediado final para la obtención del parámetro, obteniendo valores de  $R_{\text{nonlin}}$  muy bajos. La Gráfica 7.8 muestra el promedio de los  $R_{\text{nonlin}}$  obtenidos para cada señal y nivel de entrada con los tres transductores utilizados. Observando las curvas relativas a las señales estacionarias, veremos que los resultados obtenidos para la señal con modulación de amplitud son sensiblemente menores que los obtenidos por el tono puro. Teniendo en cuenta las propiedades espectrales de una señal modulada en amplitud, en la que el espectro de la señal moduladora se desplaza a la frecuencia de la portadora, es de esperar que al aparecer los productos de distorsión causados por el dispositivo, la dispersión de éstos a lo largo del espectro de la señal distorsionada sea mayor que en el caso de un tono puro. Con esto, el número de filtros auditivos en que se producen correlaciones bajas con niveles significativos será también mayor, lo cual generará necesariamente valores de  $R_{\text{nonlin}}$  inferiores a los obtenidos con la señal sinusoidal pura.

En lo que respecta a las señales complejas, los resultados obtenidos con éstas muestran a priori un comportamiento coherente con lo que se espera de la métrica<sup>19</sup>, con valores de  $R_{\text{nonlin}}$  ajustados a una dispersión mucho más limitada como se observa en la Gráfica 7.8, en comparación con las señales periódicas. De hecho, para un mismo transductor y señal las diferencias entre los seis test realizados (según el nivel de entrada) se producen a partir del tercer decimal de  $R_{\text{nonlin}}$  en la mayor parte de los casos (ver Tabla 7.4). Ahora bien, suponiendo constatada la idoneidad de las señales complejas para su empleo en la métrica, sí cabría matizar las diferencias entre las dos señales que hemos denominado ‘complejas’ en este experimento y entre los resultados obtenidos al aplicarles la métrica.

Dado que la métrica evalúa diferencias entre dos señales, cabe esperar posibles variaciones en los resultados en función de la estructura espectral y/o temporal de la señal de test utilizada. Es por esta razón que se han utilizado dos señales complejas con estructuras muy diferenciadas, como se ha visto anteriormente. Si clasificáramos las señales de test según su estructura temporal en una escala que variara desde una estructura estacionaria hasta una estructura transitoria, la señal G estaría cerca del origen de esa escala debido a su composición a base de acordes descrita unas páginas atrás, mientras que la señal J estaría alejada del origen, dado su alto contenido en transitorios. En general, se obtienen mejores resultados de  $R_{\text{nonlin}}$  para la señal J que para la señal G – excepto para los dos primeros valores de tensión de entrada en el transductor C- . La Gráfica 7.8 muestra los valores de  $R_{\text{nonlin}}$  promediados entre los tres transductores para

---

<sup>18</sup> Ocurrirá cuando en los filtros auditivos en que recaigan productos de distorsión, el nivel máximo de la señal de salida esté entre 0 y -40 dB de diferencia respecto del nivel máximo de la señal de salida de todos los filtros, que obviamente corresponde al filtro auditivo en cuya banda de paso cae la frecuencia fundamental de la señal de test, en este caso 1kHz.

<sup>19</sup> Correlaciones más cercanas a la unidad y menores diferencias entre los valores obtenidos, dada la relación entre predicción y  $R_{\text{nonlin}}$  proporcionada por la ecuación 6.16, en la que una variación en la escala de predicción de una unidad puede corresponderse en una variación de 0,01 en el valor de  $R_{\text{nonlin}}$  – según el ajuste que se realice -.

ambas señales. Vemos que la señal J obtiene valores entre 0,2 y 0,3 por encima de la señal G, lo cual puede corresponderse con diferencias de percepción significativas. Esto sugiere que los umbrales de percepción de la distorsión pueden ser más elevados en caso de utilizar señales con estructuras temporales en las que predominan los transitorios, lo cual coincidiría con los resultados de (Schmitt, 1995) obtenidos a partir de simulaciones de distorsiones<sup>20</sup> modelizando motores de compresión<sup>21</sup>. Sin embargo, matizando más este resultado, se observa que la diferencia entre los valores de  $R_{\text{nonlin}}$  de las señales G y J es significativa en los transductores A y B, no así en el transductor C, en el que se obtienen resultados más similares (ver Tabla 7.4).

En lo que respecta a la posibilidad de uso de la métrica para la comparación entre dispositivos, a la vista de los resultados, éstos parecen confirmar este extremo. De hecho, en todas las pruebas – incluidas las realizadas con señales estacionarias - se repite prácticamente el mismo orden en cuanto a valores de  $R_{\text{nonlin}}$ , siendo el mayor valor siempre para el transductor C, seguido del B y el A. La repetición de este patrón en todas las pruebas lleva a pensar en la viabilidad de la métrica en la comparación de dispositivos<sup>22</sup>.

Por otra parte, resulta llamativo que el transductor que obtiene mejor puntuación en todos los casos sea el que también presenta menores diferencias entre los valores obtenidos para la señal G y la señal J. Vista la evidente vinculación de la métrica con la señal de test, cabría pensar en principio en el uso de una amplia tipología de señales para la realización de comparaciones entre dispositivos. Pero el detalle de la escasa diferencia entre las señales G y J en el caso del transductor C puede llevarnos a pensar en la respuesta de la métrica a señales a priori más sensibles a ser percibidas como distorsionadas como un indicador de calidad del dispositivo bajo estudio. Es decir, si el resultado con una señal con una estructura temporal con muchos componentes estacionarios es bueno, cabe esperar un resultado mejor con otra señal más compleja.

Otro posible indicador podría ser la distancia entre las valoraciones obtenidas por una señal pseudo-estacionaria y una más compleja para un mismo transductor. De hecho, promediando las distancias obtenidas en este sentido en nuestro experimento obtenemos que la menor distancia tiene lugar para el transductor C (0,004485), seguido del B (0,022654) y el A (0,022677), curiosamente el mismo orden que se obtiene en todos los cálculos realizados de  $R_{\text{nonlin}}$ . Todos estos aspectos relacionados con el tipo de señal a utilizar y las diferencias entre las respuestas a diferentes señales para un mismo

---

<sup>20</sup> Inyectando la señal original a un DSP que procesa una función de transferencia que simula el comportamiento del motor de compresión.

<sup>21</sup> Es decir, utilizando técnicas pertenecientes a los denominados métodos de identificación explicados en el primer capítulo, sin basarse en ninguna consideración psicoacústica, más allá de la encuesta de percepción para determinar los umbrales.

<sup>22</sup> Si bien este aspecto debería ratificarse correlando con una encuesta real de evaluación las grabaciones realizadas con el método seguido en este trabajo.

transductor abren más interrogantes a corroborar y/o investigar mediante procedimientos más específicos y acotados que los llevados a cabo en este trabajo.

Finalmente, realizando un breve análisis de los resultados en función del nivel de la señal de entrada a los dispositivos, tenemos de nuevo un comportamiento muy diferenciado entre las señales periódicas y las complejas. Vemos en el caso de las señales periódicas que  $R_{\text{nonlin}}$  desciende casi de forma lineal a medida que aumenta la tensión de entrada. Esto se corresponde en principio con la idea ampliamente extendida de que a mayor nivel de la señal de entrada mayor distorsión generada. Esto es cierto en términos de THD (ver Tabla 7.3), pero no tiene una relación directamente proporcional con la percepción de la distorsión. Ya se han analizado las causas de la inverosimilitud de estos valores de  $R_{\text{nonlin}}$  dos páginas más atrás<sup>23</sup>.

Los resultados obtenidos con las señales complejas vienen a corroborar la idea, de nuevo, de que en términos perceptivos no es tan significativa la importancia de la amplitud de la señal de entrada<sup>24</sup> como cabría esperar a la vista de los resultados calculados mediante THD. Contra este supuesto podría argumentarse que el nivel generado por las señales periódicas es mayor que el de las señales complejas (a la vista de la Gráfica 7.1 y Gráfica 7.2), pero en términos de tendencia de las curvas obtenidas en función de la señal de entrada, la Gráfica 7.8 es bastante clarificadora. Sin embargo, se observa claramente una diferencia de comportamiento en función, de nuevo, de las características de la señal compleja utilizada. Así, en el caso de la señal G se aprecia un descenso de  $R_{\text{nonlin}}$  en función del nivel de entrada que podría traducirse en unas valoraciones subjetivas a la baja razonables – a diferencia del caso de las señales periódicas – a medida que aumenta la señal. Esta tendencia no ocurre para la señal J, al menos en el rango de trabajo estudiado, poniendo de manifiesto la importancia de la señal con la que se evalúa un sistema. Parecen ser, otra vez, las características estructurales de las señales las causas de este comportamiento.

Partiendo del supuesto de que la distorsión resulta más difícil de percibir con señales con alto contenido en transitorios resulta lógico pensar que la percepción de la degradación causada por la distorsión en relación con el nivel de entrada ocurra antes con una señal con más componentes estacionarios y que sean necesarios niveles bastante más elevados para producir el mismo efecto con señales con mucho contenido en transitorios. Esta hipótesis explicaría los resultados obtenidos.

---

<sup>23</sup> Los resultados obtenidos con el tono puro y el tono modulado.

<sup>24</sup> Siempre que ésta esté dentro del rango de trabajo del dispositivo, lógicamente.

## **7.2 Evaluación en transductores de radiación directa**

Para la aplicación del algoritmo desarrollado en transductores de radiación directa, se han utilizado tres monitores autoamplificados de gama alta cuyas características principales se resumen en la Tabla 7.5. Se trata de modelos actualmente vigentes en el mercado. Al igual que en el caso de los transductores de radiación indirecta, los denominaremos A, B y C.

### *7.2.1 Procedimiento*

A diferencia del caso anterior, para los transductores de radiación directa no se utilizará, obviamente, ningún dispositivo acoplador. La peculiaridad de los motores de compresión radica en que la carga acústica la proporciona la bocina que se les acopla en una situación real. En un motor de radiación directa, la carga acústica la proporciona el propio medio de radiación. Con esto, las medidas se realizarán colocando un micrófono en el eje de radiación del dispositivo y registrando la señal reproducida por éste. Para asegurar que la señal registrada es lo más fiel posible a la reproducida por el dispositivo, la grabación debe efectuarse en un entorno que minimice los efectos de cualquier ruido de fondo presente o de reverberación, con lo cual, las grabaciones se han efectuado en cámara anecoica. Para ello, se han utilizado micrófonos omnidireccionales Brüel&Kjaer Type 4190, junto con el acondicionador de señal de la misma firma Nexus 2690. Este proporciona la polarización necesaria para el funcionamiento de los micrófonos además de establecer una relación de salida de la señal en mV/Pa determinada por el usuario, con la finalidad de poder realizar posteriormente mediante procesado los cálculos de nivel de señal, entre otros, en caso de ser necesario. La señal se ha adquirido utilizando una interfaz de audio AVID Mbox 3G. La Figura 7.4. Montaje para el registro de señal en transductores de radiación directa muestra el montaje de los dispositivos empleados en este procedimiento.



**Figura 7.4. Montaje para el registro de señal en transductores de radiación directa**

Tabla 7.5. Características de los transductores de radiación directa utilizados.

	A	B	C
			
<b>Wofer</b>	6"	5"	6.5 "
<b>Tweter</b>	1"	1"	1 "
<b>Potencia HF</b>	25 W	30 W	35 W
<b>Material HF</b>	Seda	Seda	-
<b>Potencia LF</b>	48 W	40 W	60 W
<b>Material LF</b>	Fibra de aramidia	Kevlar	-
<b>Rango de frecuencia</b>	38-35000 Hz	56-22000 Hz	52-24000 Hz
<b>Frec. crossover</b>	2500 Hz	3000 Hz	2000 Hz

Entre las particularidades de los dispositivos bajo test cabe destacar el amplio rango de frecuencias en que trabajan. Esto se consigue mediante el uso de dos transductores: uno para la baja frecuencia y otro para la alta frecuencia. El rango de frecuencias en que trabaja cada uno de los transductores viene determinado por la frecuencia de corte (*crossover frequency*) (ver Tabla 7.5). Por debajo de esta frecuencia la señal será reproducida por el transductor de mayor tamaño, o *woofer*, mientras que por encima de la frecuencia de corte, la radiación la realizará el transductor de 1" de diámetro, también denominado *tweeter*. Cabe destacar también que se trata de dispositivos autoamplificados, los circuitos de entrada, acondicionamiento y amplificación de la señal para cada uno de los dos transductores se encuentran integrados en el propio sistema, con lo cual, a la hora de establecer una valoración utilizando alguna métrica, dicha valoración hará referencia a todo el conjunto: los amplificadores y los transductores de cada dispositivo.

En el caso anterior utilizábamos el mismo amplificador para todos los transductores, con lo cual, las diferencias relativas de la métrica podían ser asociadas directamente al motor de compresión, que es el único elemento que varía en toda la cadena. En este caso es diferente, ya que los amplificadores forman parte del dispositivo, las diferencias en la métrica se establecerán respecto del conjunto. Al tratarse de dispositivos autoamplificados, la comparación de los resultados de la métrica no se ha hecho midiendo la tensión en bornes del transductor, dado que esto no es posible. En su lugar,

se ha actuado sobre la ganancia de señal introducida por el amplificador incorporado, midiéndose todas las señales en un rango entre -6 dB y +6 dB en intervalos de 3 dB.

Se ha elegido este tipo de dispositivos por dos razones fundamentales: en primer lugar por sus grandes diferencias respecto de los motores de compresión, con la finalidad de poder constatar una diferenciación en los resultados obtenidos con el algoritmo implementado. En segundo lugar, porque se trata de transductores de monitorización de mezcla, con una respuesta fundamentalmente plana en todo el rango audible de frecuencias (lo cual evita coloraciones de la señal), que se caracterizan además por una reproducción fidedigna de la señal con muy bajos niveles de distorsión. Este segundo argumento es el que nos ha llevado a pensar en la utilización de estos dispositivos para la evaluación del algoritmo en salas.

En esta sección estudiaremos el comportamiento de estos transductores de manera análoga a como se ha hecho con los de radiación indirecta. En el siguiente capítulo realizaremos un análisis de la respuesta de estos dispositivos con las señales de test que se utilizarán para la evaluación en salas, escogiendo el que presente mejores resultados. El transductor elegido se utilizará como fuente de señal en el procedimiento experimental para la aplicación del algoritmo en salas, como se verá más adelante. En lo que respecta a las señales utilizadas en este experimento, se han utilizado las mismas cuatro señales que en el caso de los transductores de radiación indirecta, si bien esta vez las señales no han sido filtradas paso-alto como en el caso anterior, puesto que estos transductores responden perfectamente por debajo de los 800 Hz. Como ya se ha comentado, en este caso no se ha tomado como referencia el nivel de la señal de 1000 Hz en bornes del transductor sino la ganancia introducida por el amplificador del dispositivo. Además del tono puro de 1000 Hz se ha utilizado la señal modulada de 1 kHz en amplitud por una moduladora sinusoidal de 70 Hz, y también las señales G y J tal y como se han descrito en el apartado 7.1.1, con la salvedad de que en este caso no se ha eliminado el contenido por debajo de 800 Hz, tal como se ha comentado.

Además de estas señales, en este mismo procedimiento se han registrado diferentes audios que se utilizarán en el siguiente capítulo: un pequeño corpus de palabra hablada, compuesto por frases psicoacústicas y textos fonéticamente equilibrados y cuatro señales musicales, que se analizarán en el capítulo 8.

### *7.2.2 Resultados*

De forma análoga a la sección 7.1.2 se muestran los resultados obtenidos en el procesado de la métrica. La Gráfica 7.9 y Gráfica 7.10 muestran los resultados de las señales estacionarias. La Gráfica 7.11 y Gráfica 7.12 de las complejas. La Gráfica 7.13 muestra todos los resultados y, finalmente, la Gráfica 7.14 el promediado de cada señal en los tres dispositivos.

**Tabla 7.6.  $R_{nonlin}$  obtenidos en los transductores de radiación directa**

<b>1kHz</b>	<b>-6 dB</b>	<b>-3 dB</b>	<b>0 dB</b>	<b>+3 dB</b>	<b>+6 dB</b>
A	0,848454	0,82345	0,756432	0,734523	0,653432
B	0,912343	0,912442	0,900231	0,843214	0,742414
C	0,908765	0,897698	0,885643	0,853452	0,751232

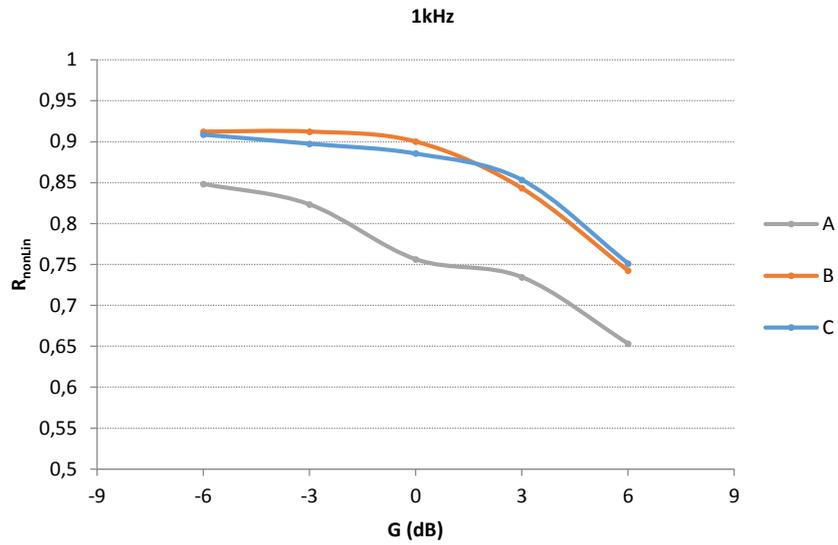
<b>Mod.</b>	<b>-6 dB</b>	<b>-3 dB</b>	<b>0 dB</b>	<b>+3 dB</b>	<b>+6 dB</b>
A	0,771243	0,723452	0,643459	0,597865	0,53423
B	0,876754	0,834553	0,789784	0,75342	0,661232
C	0,853546	0,824569	0,789754	0,731245	0,64321

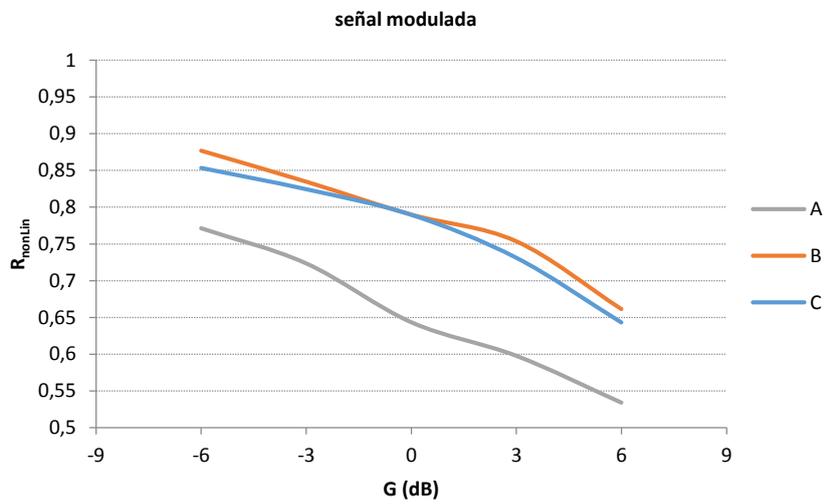
<b>Señal G</b>	<b>-6 dB</b>	<b>-3 dB</b>	<b>0 dB</b>	<b>+3 dB</b>	<b>+6 dB</b>
A	0,928765	0,929002	0,927865	0,926068	0,9187
B	0,938997	0,939654	0,937643	0,934231	0,933998
C	0,936523	0,936324	0,9338	0,934001	0,933989

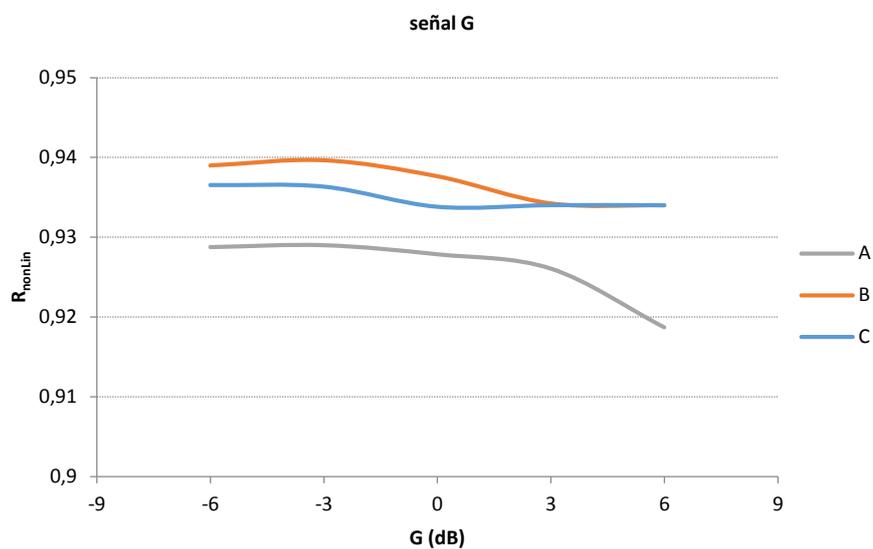
<b>Señal J</b>	<b>-6 dB</b>	<b>-3 dB</b>	<b>0 dB</b>	<b>+3 dB</b>	<b>+6 dB</b>
A	0,934283	0,931965	0,92978	0,929212	0,927991
B	0,940123	0,940012	0,939781	0,939945	0,939032
C	0,941214	0,940005	0,938217	0,938964	0,93901



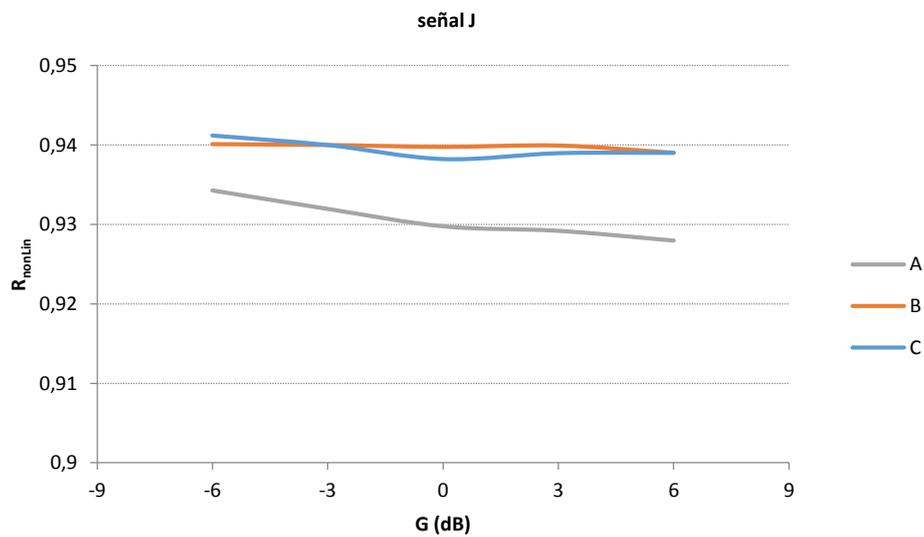
Gráfica 7.9.  $R_{nonlin}$  en función de la ganancia de entrada para la señal sinusoidal.



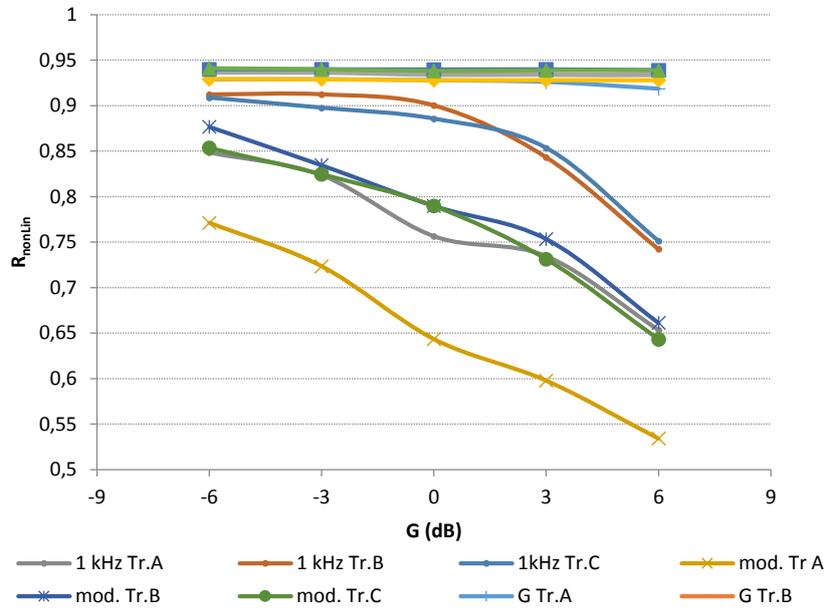
Gráfica 7.10.  $R_{nonlin}$  en función de la ganancia de entrada para la señal modulada.



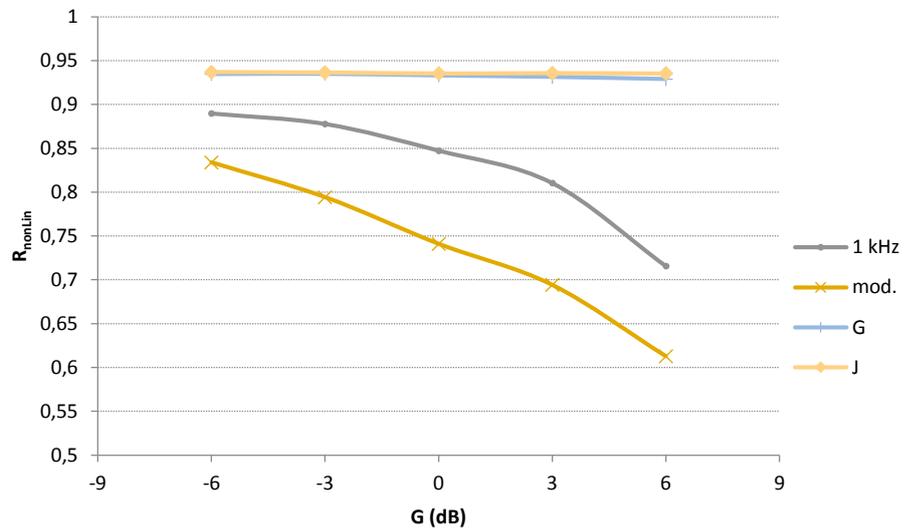
**Gráfica 7.11.**  $R_{nonlin}$  en función de la ganancia de entrada para la señal G.



**Gráfica 7.12.**  $R_{nonlin}$  en función de la ganancia de entrada para la señal J.



Gráfica 7.13.  $R_{nonlin}$  en función de la ganancia de entrada para las cuatro señales y los tres transductores.



Gráfica 7.14.  $R_{nonlin}$  obtenido en función de la ganancia de entrada promediado para cada señal

### 7.2.3 Análisis

A modo de referencia, antes de entrar en el análisis de resultados, se muestra la siguiente tabla, que presenta los niveles de presión sonora medidos a 1m de distancia en el eje de radiación de cada transductor según la ganancia de amplificación ajustada. Estos niveles hacen referencia a una señal sinusoidal de 1 kHz, en la que se alcanza el valor de pico en todos los ciclos. A la hora de utilizar señales complejas, de igual manera que se ha visto en la sección 7.1.1 y en la Gráfica 7.1 y Gráfica 7.2 los valores de pico se alcanzarán sólo en unas pocas ocasiones, según la distribución dinámica de las señales, con lo que los niveles de presión sonora equivalente para las señales complejas serán sensiblemente menores que los mostrados en la Tabla 7.7.

**Tabla 7.7. SPL para la señal de referencia de 1 kHz**

<b>G (dB)</b>	<b>Transductor A</b>	<b>Transductor B</b>	<b>Transductor C</b>
	L SPL (dB A)	L SPL (dB A)	L SPL (dB A)
<b>-6</b>	75,3	76,2	76,9
<b>-3</b>	79,0	80,3	81,6
<b>0</b>	83,8	84,6	85,3
<b>3</b>	87,9	88,0	89,2
<b>6</b>	92,07	92,3	92,8

Lo primero que se observa a la vista de los resultados de la métrica aplicada a los transductores de radiación directa es un comportamiento a priori análogo al caso de los de radiación indirecta en lo que se refiere a la tipología de la señal utilizada. Las curvas de la Gráfica 7.13 representan el valor de  $R_{\text{nonlin}}$  para todas las señales y transductores en función de la ganancia de amplificación utilizada, mientras que las curvas de la Gráfica 7.14 el promedio de los resultados obtenidos en los tres transductores para cada señal, también en función de la ganancia. De la lectura de estas gráficas, se desprende que las señales complejas (las que hemos denominado G y J) obtienen una correlación  $R_{\text{nonlin}}$  que siempre oscila entre 0,92 y 0,94, con muy pocas variaciones con respecto al incremento de la ganancia de entrada. En este sentido, cabe destacar que las variaciones son mayores en el caso de la señal G, especialmente en el caso del transductor A, cuyos valores sufren una bajada más pronunciada con el incremento de nivel de presión sonora, si bien siempre se mantienen dentro de la horquilla mencionada anteriormente. En lo referente a la señal J, el valor es prácticamente constante en función del nivel.

Por otra parte, en lo que respecta a las señales estacionarias, se obtienen, de igual manera que ocurría en la sección 7.1.2, valores menores que los obtenidos con las señales complejas, que descienden también a medida que aumenta el nivel de presión. En este punto cabe destacar una diferencia con respecto a los resultados con dispositivos de radiación indirecta. El descenso de  $R_{\text{nonlin}}$  en aquel caso se producía de manera cuasi lineal, mientras que en el caso de los dispositivos de radiación directa el descenso es progresivo pero menos pronunciado y siempre obteniendo valores significativamente

superiores al caso visto en el epígrafe anterior. Aun siendo valores superiores a los obtenidos para los motores de compresión ocurre igualmente que una valoración de los transductores realizada a partir de señales estacionarias basada en este modelo perceptual proporcionaría un resultado igualmente deficiente de los dispositivos independientemente del ajuste que se llevara a cabo con los parámetros  $a$ ,  $b$  y  $c$  de la ecuación 6.16, como se ha explicado en la sección 7.1.3. En este caso, el de los dispositivos de radiación directa, se observa una sensible diferencia de comportamiento entre el tono puro de 1000 Hz y la señal modulada. El valor de  $R_{\text{nonlin}}$  en el caso del tono puro mantiene valores más similares entre sí para los niveles de presión sonora más bajos y empieza a caer de forma más pronunciada en los dos niveles más altos. No se trata pues de una caída con una pendiente casi constante como ocurría con los motores de compresión, sino que la correlación baja ligeramente en los niveles más bajos manteniéndose en un rango de variación pequeño para decaer después más fuertemente en los dos últimos niveles medidos. Esto no ocurre con la señal modulada, que presenta, además de unos valores significativamente menores que los del tono puro, una caída progresiva y constante de los resultados de  $R_{\text{nonlin}}$  en función del nivel de presión radiado.

En esencia, para el caso de las señales estacionarias, la explicación a este resultado es, como ya se ha comentado en la sección 7.1.3 la aparición de productos de distorsión en bandas de filtros auditivos con una energía mínima o casi nula en la señal original. Esto producirá una correlación muy baja de la señal registrada con el producto de distorsión y la señal original. Al incrementar el nivel de emisión se hace más probable que los niveles alcanzados por estos productos de distorsión empiecen a ser considerados por la matriz de ponderación explicada en el apartado 6.8, de manera que estas distorsiones tienen un peso progresivo en el resultado final a medida que aumenta el nivel de presión sonora generado por los dispositivos. Este fenómeno también podría explicar el peor resultado obtenido para la señal modulada, dada la presencia de un mayor número de productos de distorsión debido a las propiedades espectrales de la señal.

Las principales diferencias que se observan en los resultados obtenidos con las señales estacionarias entre los transductores de radiación directa y los de radiación indirecta son:

- a) Los valores de  $R_{\text{nonlin}}$  son significativamente más altos en el caso de la radiación directa, lo cual nos lleva a pensar en una menor presencia de productos de distorsión. Para la radiación directa estos valores fluctúan entre 0,65 y 0,91 para la señal de 1000 Hz, mientras que para la indirecta lo hacían entre 0,44 y 0,87. En el caso de la señal modulada las horquillas son (0,53-0,87) para radiación directa y (0,35-0,82) para radiación indirecta. Puede observarse que la dispersión de valores con el incremento de nivel es mucho mayor para los motores de compresión. Las diferencias obtenidas entre un tipo de motores y otros están, en el rango de menor distorsión, entre 0,04 y 0,05 por encima para los motores de radiación directa mientras que en el rango de mayor distorsión las variaciones están entre 0,11 y 0,18 de diferencia entre ambos tipos de motores. Estas diferencias corresponden a variaciones perceptivas considerables

atendiendo a lo visto en el apartado 6.9. Los motores de radiación directa son dispositivos a priori con mucha menor distorsión que los motores de compresión, por las características de funcionamiento explicadas en los apartados 3.7 y 3.8. y las diferencias propias de la tipología del motor. Los transductores de radiación directa también trabajan a niveles de presión mucho menores que los existentes en la garganta de una cámara de compresión de un motor de radiación indirecta, como se puede ver comparando la Tabla 7.2 y la Tabla 7.7. Los altísimos niveles de presión sonora generados en ese punto de medición serían otra explicación a la presencia de un mayor número de no linealidades que quedarán reflejadas en los resultados de la métrica.

- b) La degradación de la señal del tono puro de 1000 Hz reflejada por la métrica es mucho menos pronunciada en los motores de radiación directa. Así, las correlaciones obtenidas para los niveles de presión más bajos presentan un descenso muy suave hasta la mitad del rango de niveles medido. El gradiente de degradación de la señal mostrado por la métrica se incrementa a medida que aumenta el nivel generado por el motor. La menor presencia de productos de distorsión o el bajo nivel relativo de éstos para los menores niveles de presión podría explicar estos resultados. A partir de que los dispositivos introducen una ganancia de amplificación superior a 0 dB los productos de distorsión empiezan a ser computados en la matriz de ponderación, como ya se ha explicado.
- c) La degradación de la señal modulada sigue un patrón similar en ambos tipos de motores, si bien los resultados son significativamente mejores en el caso de la radiación directa, como hemos visto en el apartado a). El descenso más acusado a partir de los niveles más bajos de presión puede deberse a la mayor presencia de productos de distorsión que genera esta señal.

En lo que respecta a las señales complejas, se observa también un comportamiento similar en ambos tipos de motores, obteniéndose, de nuevo, resultados sensiblemente mejores en el caso de la radiación directa. Al igual, que en el caso visto en la sección 7.1.3, en la mayoría de los casos las variaciones para un mismo transductor en función del nivel se producen a partir del tercer decimal del resultado de la métrica. En el caso de la radiación indirecta se han observado diferencias respecto de los resultados obtenidos con las señales G y J. La señal con mayor ancho de banda y mayores variaciones temporales (J) obtenía en casi todos los casos mejor resultado que la señal G (formada por un único instrumento). Esto sucede también en el caso de los motores de radiación directa, aunque las diferencias son mucho menores. Así, si en el caso de los motores de compresión, la señal J obtenía valores entre 0,2 y 0,3 por encima de la señal G, lo cual puede corresponderse con diferencias de percepción significativas, en el caso de la radiación directa estas diferencias se reducen a entre 0,05 y 0,1. Esto puede hacernos suponer una menor dependencia de la tipología de señal para el caso de estos dispositivos, en parte posiblemente explicable por los menores niveles de presión sonora generados.

Veamos ahora, tal como hemos hecho en la sección 7.1.3, la posible viabilidad de la métrica para su uso en la comparación entre transductores. De forma similar a la vista para los motores de compresión, se produce una repetición de patrones en cuanto a los resultados obtenidos por los diferentes transductores, de manera que el mayor valor casi siempre suele corresponder al transductor B, seguido del C y el A, si bien las diferencias entre el B y el C resultan mínimas, obteniendo este último mejor resultado en una única medición (ver Gráfica 7.12). La repetición de este orden en todas las medidas lleva a pensar, al igual que en los motores de compresión, en la viabilidad de la métrica en la comparación de dispositivos.

En la sección dedicada a los transductores de radiación indirecta habíamos observado que para el transductor mejor valorado el promedio de la diferencia de los resultados obtenidos para las señales G y J también era el menor de los tres dispositivos. Este resultado se repite curiosamente en los transductores de radiación directa, si bien las diferencias son mucho menores, como por otra parte cabría esperar a raíz de los mayores valores obtenidos. Así, promediando las diferencias de los resultados obtenidos para cada transductor entre la señal J y G, se obtiene que la diferencia media es de 0,002874; 0,0045546 y 0,0045662 para los transductores B, C y A respectivamente, lo cual corresponde al mismo orden de la valoración obtenida por los dispositivos al aplicarles la métrica.



## 8 Evaluación en salas

Si bien en principio la métrica implementada está pensada para la evaluación de productos de distorsión lineal y no lineal a partir de distorsiones generadas artificialmente simulando un canal de transmisión, los resultados obtenidos en transductores reales como los presentados en el capítulo 7, nos llevan a plantearnos la posibilidad de utilización de la métrica para la evaluación de canales de transmisión que presenten otras formas de degradación de la señal. Así, las correlaciones relativamente altas obtenidas en el caso de los transductores de radiación directa hace pensar en su uso para la evaluación de calidad en salas utilizando la misma métrica. El problema se puede enfocar desde dos puntos de vista. Utilizar el transductor como fuente de excitación utilizando señales complejas de forma que la métrica establezca una evaluación de la degradación de la señal en el canal de transmisión, en este caso, la sala, o bien, otra perspectiva sería utilizar la métrica para la evaluación de la calidad en salas con refuerzo sonoro. En estas salas, normalmente se dispone siempre de la señal que se envía al sistema de refuerzo sonoro, de forma que esta señal podría utilizarse como referencia de la métrica. La evaluación proporcionada por el algoritmo contendría tanto la degradación introducida por los transductores, en forma de productos de distorsión lineal y no lineal,

como la introducida por la sala, en forma de ruido de fondo y reverberación. En este capítulo presentaremos los resultados obtenidos realizando la primera aproximación.

## 8.1 Procedimiento

Las señales de excitación utilizadas se presentarán en la sección 8.2. De los transductores de radiación directa utilizados en el capítulo anterior se elegirá el que mejores resultados proporcione para el conjunto de materiales de test utilizados – sección 8.3 –, y éste se utilizará como fuente de excitación de las salas. Se han elegido un total de 7 salas con características acústicas bien diferenciadas para poder evaluar la respuesta de la métrica en un rango amplio de valores de parámetros objetivos. Para cada sala, se han obtenido los tiempos de reverberación, RT y EDT, la definición D, la claridad  $C_{80}$ , el índice de transmisión de la palabra STI y la versión del STI reducida para acústica de salas, RASTI. Puede verse que se ha optado por los parámetros objetivos de la norma ISO 3382:1 que son de carácter monofónico, dado que la métrica implementada también lo es. De esta manera se tratará de establecer una relación entre los resultados obtenidos con las métricas objetivas y los resultados de aplicar el algoritmo desarrollado a las señales de excitación reproducidas en las salas.

Las medidas de los parámetros objetivos comentados en el párrafo anterior se han realizado según la norma UNE EN ISO 3382-1:2010, en lo que respecta a equipos de medida utilizados, distancias y ubicaciones de las posiciones de medida. En las salas más grandes, se han realizado mediciones en seis puntos de medida, tal como establece la norma. Dado que interesaba tener un rango amplio de valores de los parámetros objetivos, se han realizado medidas en salas más pequeñas, en las que se han empleado cuatro puntos de medida. La medición de los parámetros objetivos se ha realizado mediante obtención de la respuesta al impulso de la sala en el punto de medida, utilizando como señal de excitación una secuencia MLS, promediada 16 veces en cada punto de medida. Para ello se ha utilizado el analizador Symphonie, del fabricante 01 dB (s/n 282), de tipo I y homologado para este tipo de mediciones. Este equipo se ha utilizado también para la generación de la señal de excitación, que era enviada a una fuente de señal dodecaédrica, Type 4292 de la marca Brüel&Kjaer. El amplificador utilizado para el suministro de potencia a la fuente dodecaédrica es el modelo 2716 también de Brüel&Kjaer. Los micrófonos utilizados para la obtención de las medidas objetivas son los omnidireccionales Type 4190 de la marca Brüel&Kjaer.

Una vez realizadas las mediciones según el procedimiento descrito en el párrafo anterior se procedía a colocar en la misma posición que había ocupado la fuente dodecaédrica el transductor elegido para la reproducción de señal musical o hablada en la sala. Las señales que se detallan en la siguiente sección eran reproducidas en la sala y registradas en los mismos puntos que se habían utilizado para efectuar las mediciones. Para el registro de estas señales se han utilizado dos micrófonos omnidireccionales Type 4190 de Brüel&Kjaer –los mismos utilizados en el procedimiento anterior– esta vez conectados a un acondicionador de señal Nexus 2690 de Brüel&Kjaer. El acondicionador microfónico proporciona la polarización necesaria para el

funcionamiento de los micrófonos además de establecer una relación de salida de la señal en mV/Pa facilitando el cálculo de niveles posteriormente mediante procesado en caso de ser necesario. La salida del acondicionador de señal era enviada a un interfaz de audio AVID MBox 3G conectado al ordenador desde el que se enviaba la señal de excitación y registraban las señales captadas por los dos micrófonos. De esta forma, se realizaba el registro de señal en dos puntos de medición simultáneamente.

Posteriormente, las señales registradas, una vez transmitidas por el canal, es decir, la sala, se han sometido al procesado del algoritmo implementado. Los resultados obtenidos en este procesado para cada tipo de señal, en cada punto de cada una de las salas utilizadas son los que intentaremos relacionar en las páginas sucesivas con los resultados de las métricas objetivas efectuadas.

## **8.2 Señales utilizadas**

Se describen a continuación el conjunto de señales que se han utilizado en el procedimiento descrito en el epígrafe anterior. Se han utilizados cuatro señales musicales con grandes diferencias en cuanto a contenido espectral y características temporales, como se verá a continuación, así como unas grabaciones de palabra de un corpus formado por 20 frases psicoacústicas, procedentes de la adaptación realizada al español por (Valero, 1991) de las frases psicoacústicas de Harvard, entre otros materiales.

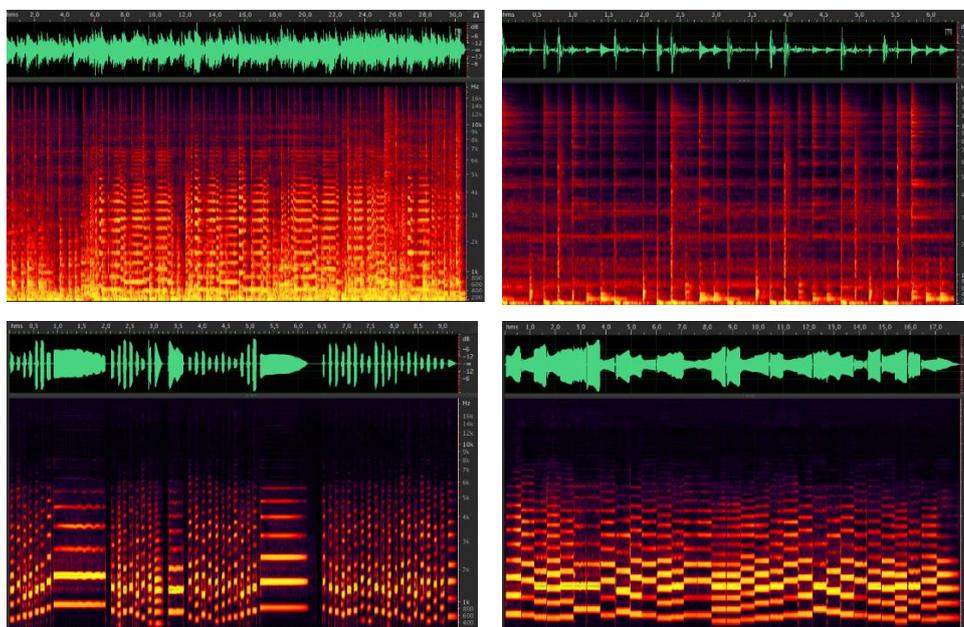
### *8.2.1 Señales musicales*

La Figura 8.1 muestra las señales utilizadas tanto en tiempo como en contenido espectral. En estas representaciones, el eje x representa el tiempo, mientras que el eje y la amplitud de la señal – en el caso de la representación temporal – o la frecuencia – en el caso de los espectrogramas de la parte inferior. El contenido en frecuencia se representa mediante una escala con un gradiente de color entre el negro (ausencia de contenido) y el amarillo (máximo nivel de contenido). La simple inspección visual de los espectrogramas constata las diferencias de contenido y estructura temporal entre las señales.

La primera señal elegida es un fragmento jazzístico de 30 segundos de duración en la que intervienen instrumentos de percusión, bajo acústico, piano y saxo. En la Figura 8.1, es la representada en la parte superior izquierda. La presencia de todo el conjunto instrumental durante casi toda la extensión de la pieza hace posible que la señal contenga información en todo el espectro de 20 Hz a 20 kHz prácticamente durante toda su duración. El ensamble instrumental hace que la señal no tenga transitorios excesivamente marcados, como puede observarse en la representación temporal, más allá de algunos acentos percusivos que quedan integrados en el conjunto. En las representaciones gráficas sucesivas nos referiremos a esta señal como  $S_1$ .

La segunda señal que se utilizará se representa en la parte superior derecha de la Figura 8.1 y corresponde a un fragmento de 16 segundos de un pasaje formado estrictamente por instrumentos de percusión: un bombo bastante tenue – puede observarse el escaso contenido en baja frecuencia, a diferencia de la señal anterior -, una caja golpeada sobre aro que aporta los transitorios más remarcados que se observan en la representación

temporal de la señal, un plato a modo de *riders* que aporta el contenido de más alta frecuencia y finalmente unas congas que introducen contenido espectral en las frecuencias medias. De todas las señales utilizadas es la que más fluctúa en el margen dinámico, ya que el nivel de señal es normalmente bajo produciéndose transitorios en los golpes de caja que recorren todo el margen dinámico de la señal para volver rápidamente al nivel medio. Del análisis del espectrograma se deduce el escaso contenido en frecuencias medias, una importante presencia de silencios a lo largo de la señal, en la que prácticamente no hay contenido espectral (franja negra que se repiten periódicamente en la representación en frecuencia) y como el contenido en frecuencia de los transitorios ocupa todo en ancho de banda. En las representaciones gráficas sucesivas nos referiremos a esta señal como  $S_2$ .



**Figura 8.1. Señales musicales representadas en tiempo y frecuencia**

Las dos siguientes señales musicales empleadas corresponden a dos grabaciones de oboe, con el ancho de banda limitado por el del propio instrumento (puede verse en las imágenes inferiores de la Figura 8.1 que éste es de alrededor de 6 - 7 kHz) pero diferenciadas en su estructura temporal.

La tercera señal es un pasaje de 9 segundos de duración del segundo movimiento de la tercera sinfonía de Mendelssohn, “La escocesa”. Está interpretado a un tiempo rápido pero que permite utilizar el picado simple, en *staccato*: nota que está separada de las demás en su ejecución por un silencio de articulación dado con cierto énfasis. La presencia de notas picadas se puede observar en la imagen inferior izquierda de la Figura

8.1. Estas escalas de notas picadas introducen componentes transitorios en la señal temporal, si bien, delimitados por el propio ancho de banda del instrumento. Cada nota ejecutada por el instrumento se corresponde en el espectrograma con la componente en frecuencia de mayor intensidad, correspondiendo las demás componentes simultáneas a los armónicos. En las presentaciones de resultados nos referiremos a esta señal como  $S_3$ .

Sinfonía n° 3 "Escocesa"  
(2° mov)

F. Mendelssohn  
op.56

**Vivace**



Fuente: Juan Pedro Romero

**Figura 8.2. Pasaje de la señal  $S_3$**

Por último, la cuarta señal es un fragmento de 18 segundos de duración correspondiente al segundo movimiento de la Sinfonía n° 4 de Tchaikovsky, consistente en un fraseo melódico de notas largas lo cual, dinámicamente se asemeja a una forma de onda pseudo-estacionaria, en la medida que es una sucesión lenta de ondas sinusoidales, formadas por la frecuencia fundamental más los armónicos del instrumento (como puede verse en la parte inferior derecha de la Figura 8.1). Al hacer referencia a esta señal en gráficos nos referiremos a ella como  $S_4$ .

Sinfonía n° 4  
(2° mov.)

P.I.Tchaikovsky  
op.36

**Andantino**



Fuente: Juan Pedro Romero

**Figura 8.3. Pasaje de la señal  $S_4$**

### 8.2.2 Señales para palabra

Se ha realizado una grabación en cámara anecoica de un corpus formado por veinte frases psicoacústicas, adaptadas a la lengua española por (Valero, 1991). Las grabaciones han sido realizadas con voz masculina y femenina para su posterior reproducción en la sala con un transductor de radiación directa. El conjunto de estas frases está fonéticamente equilibrado, es decir, el contenido fonético de éstas se ajusta a la distribución estadística del contenido fonético de la lengua española. Las frases se muestran en la siguiente tabla.

**Tabla 8.1. Corpus de frases psicoacústicas adaptadas al español**

Frases Psicoacústicas	
La oscuridad me da mucho miedo	He comido carne con patatas fritas
Podéis venir con nosotros al cine	El balazo se desvió de la diana
Se ha roto la tapa del bote de mermelada	Se pegaron en medio de la curva
Ganará lo que ganaba en su anterior trabajo	Esta sustancia nos prolonga la memoria
Carlos es capaz de no saludarnos	Recoge su llavero y pónselo dentro del cesto
Se tiene por forzuda, pero no resiste nada	Las oscuras pestañas enmarcan su dulce mirar
Las llaves están en mi bolsa	Una de las redes ya está en mi barco pesquero
Pili quiere que tú y yo cojamos zarzamoras	Te vencí y di la copa al orfanato
Le apetece un poco de vino blanco	Beber anís puede quemarte el gaznate
Las notas de diseño que te faltan no las sabrás por mí	Pon mis discos y cállate

Cualquier estudio psicoacústico que utilice palabra hablada debe utilizar material fonéticamente equilibrado para que sea representativo del idioma en que se realiza. Aunque el experimento llevado a cabo en este trabajo no es un procedimiento psicoacústico, la evaluación de la métrica con materiales de corta duración podría no ser representativa del idioma en que se realiza si éstos no están equilibrados, como se ha explicado en la sección 2.2.3. Si se usara un fragmento muy largo de palabra hablada, la distribución fonética tendería a ser la propia del idioma. Por cuestiones de eficiencia, tanto en el caso de encuestas psicoacústicas como en nuestro caso (procesado de señal) se utilizan fragmentos cortos, con lo cual, éstos deben ser representativos fonéticamente del idioma utilizado en el experimento. Con esto, los resultados de aplicar el algoritmo implementado a las frases psicoacústicas se representarán de forma estadística, -ya que lo que es representativo del habla es el conjunto de las frases, no una o varias de ellas- mostrando la dispersión obtenida y los valores medios para todo el conjunto, como veremos en páginas sucesivas.

Una de las alternativas al uso de frases cortas para la aplicación de la métrica a la señal hablada consiste en utilizar párrafos de texto que también estén fonéticamente

equilibrados. La diferencia con respecto al uso de frases radica en que la estructura temporal de la señal puede ser más representativa del habla espontánea en el sentido en que hay presentes pausas, silencios, así como cambios de entonación. La presencia de silencios puede alterar a priori los resultados ofrecidos por el algoritmo implementado. Para estudiar estos posibles efectos en el caso de la palabra hablada se ha registrado también un párrafo fonéticamente equilibrado en español, obtenido de (Bruyninckx, Harmegnies, Llisteri, & Poch, 1994), que se presenta en la siguiente tabla. La duración de la señal registrada es de 30 segundos para el caso de la voz masculina y de 29 segundos para la voz femenina. Otra fuente de textos fonéticamente equilibrados en español es la base de datos “Ahumada” (Ortega, Gonzalez, & Marrero, 2000).

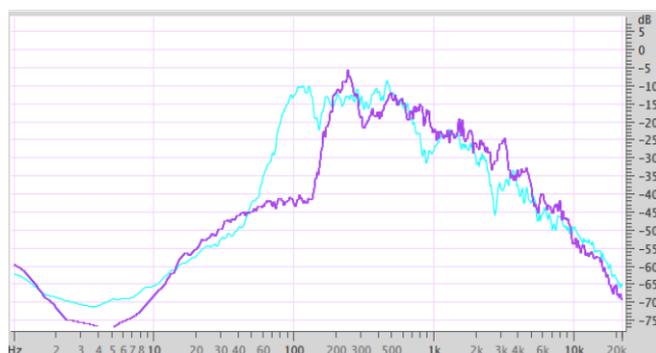


Figura 8.4. Comparación de espectros de una señal de voz masculina (azul) y femenina (violeta)

Tabla 8.2. Párrafo fonéticamente equilibrado usado como señal de palabra

Texto fonéticamente equilibrado
El joyero Federico Vanero ha sido condenado por la audiencia de Santander a ocho meses de arresto mayor y cincuenta mil pesetas de multa por un delito de compra de objetos robados. La vista oral se celebró el miércoles pasado y, durante ella, uno de los fiscales, Carlos Valcárcel, pidió para el joyero tres años de prisión menor y una multa de cincuenta mil pesetas. Gracias a las revelaciones de Vanero de hace dos años y medio se llegó a descubrir la existencia de una sospechosa mafia policial en España, parte de la cual se vio envuelta en el llamado “caso el Nani”.

La Figura 8.4 muestra el análisis en frecuencia del párrafo de texto registrado para una voz masculina y una femenina. Puede observarse que la principal diferencia entre ambas estriba en las formantes de baja frecuencia. En el caso de la voz masculina, éstas se encuentran alrededor de los 115 Hz, parte del espectro en el que la voz femenina no presenta energía. Las formantes de baja frecuencia para la voz femenina se centran alrededor de los 250 Hz, siendo también significativas las diferencias de contenido espectral entre los 800-900 Hz y entre los 3 y 4 kHz.

### 8.3 Elección del transductor

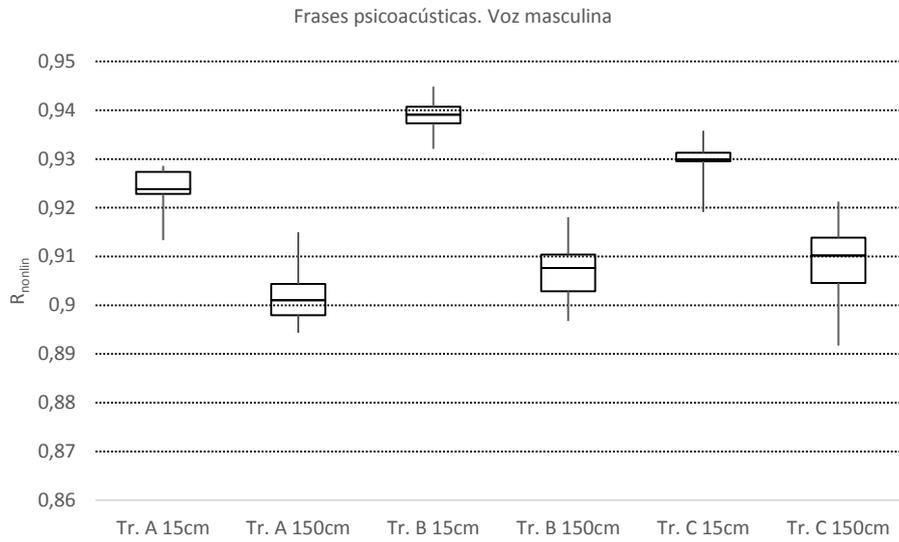
De los tres transductores de radiación directa utilizados en la sección 7.2 se escogerá uno para la realización del procedimiento explicado en la sección 8.1. Así, los tres transductores utilizados han sido excitados con las señales de palabra y música indicadas en la sección anterior. La grabación de la reproducción de estas señales por parte de los dispositivos se ha realizado en cámara anecoica siguiendo el mismo procedimiento explicado en la sección 7.2.1. Se han realizado registros de señal a 15 cm del transductor en el eje central de radiación y a 150 cm de distancia en el mismo eje. Para el caso de las señales de voz, se ha utilizado el corpus de frases psicoacústicas, mientras que en el caso de las señales musicales se han reproducido las cuatro señales que se han visto en el apartado 8.2.1.

Las grabaciones se han sometido a la métrica implementada y los resultados obtenidos pueden verse en las gráficas representadas desde la Gráfica 8.1 a la Gráfica 8.4. La Gráfica 8.1 representa la dispersión de valores de la métrica obtenidos para el corpus de frases psicoacústicas con voz masculina. La barra superior de cada caja indica el valor máximo obtenido en el conjunto de frases, la barra inferior indica el valor mínimo. Cada caja engloba los valores entre el percentil 25 y el 75 de la serie de resultados obtenidos, siendo la línea que divide cada caja el valor de la mediana de la serie. En lo que a esta gráfica se refiere, puede observarse que las correlaciones más altas se obtienen con el transductor B, en los registros realizados en la proximidad de la fuente, siendo también éste el que menor dispersión de resultados presenta. A 150 cm de distancia, los resultados del transductor C son ligeramente superiores a los del B pero presentan una mayor dispersión (la mayor de los tres transductores comparados).

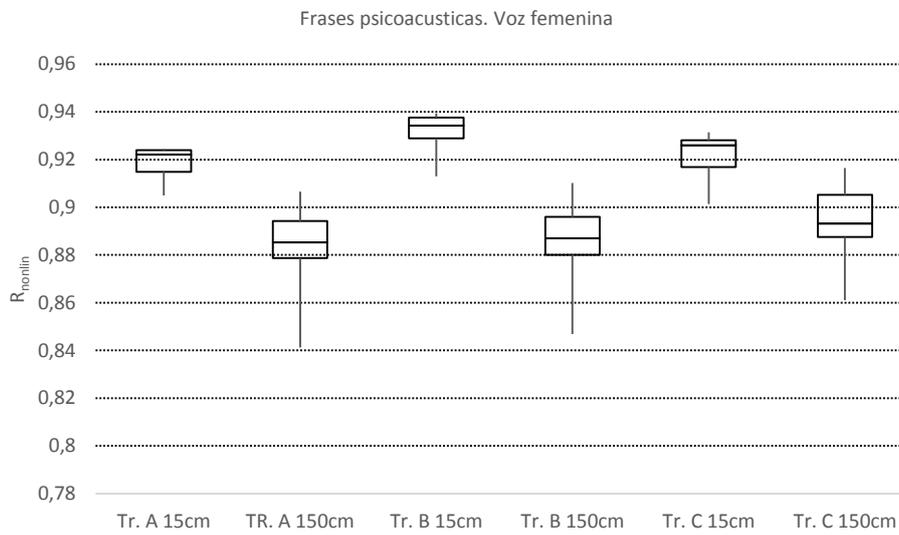
En el caso de la voz femenina (Gráfica 8.2) lo primero que se observa es que las correlaciones obtenidas son ligeramente menores que para la voz masculina. Para la medida en la proximidad de la fuente sigue siendo el transductor B el que presenta valores más altos. La menor dispersión en este caso se obtiene en el transductor A pero en torno a valores de  $R_{\text{nonlin}}$  sensiblemente menores que para los otros dos transductores. En las medidas a 150 cm el transductor C vuelve a obtener valores ligeramente más altos que el B, con dispersiones muy similares entre los tres dispositivos.

La Gráfica 8.3 representa los resultados de la métrica para el conjunto de todas las frases psicoacústicas incluyendo las señales tanto de voz masculina como femenina, con lo cual, las dispersiones obtenidas son ligeramente mayores, especialmente en el caso de mayor distancia al dispositivo.

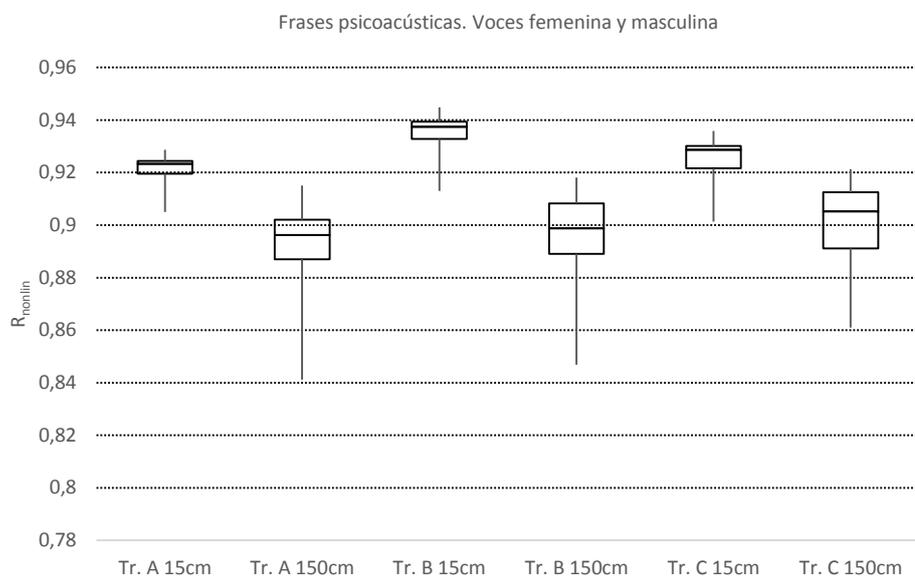
Se repite aproximadamente el mismo patrón que en los casos anteriores, una mejor correlación y menor dispersión para el transductor B en la proximidad del transductor y unos valores ligeramente superiores del transductor C a 150 cm del dispositivo.



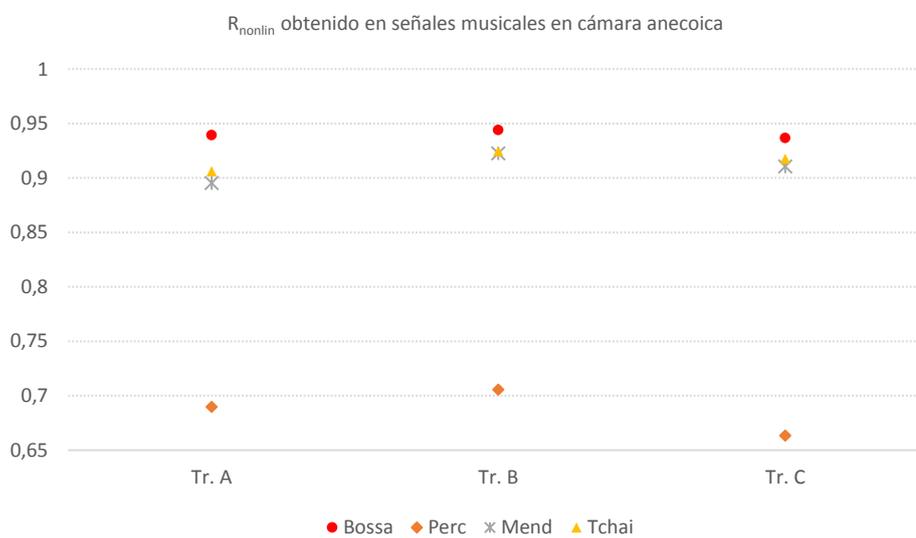
Gráfica 8.1.  $R_{nonlin}$  obtenido en frases psicoacústicas, voz masculina, 3 transductores a 2 distancias



Gráfica 8.2.  $R_{nonlin}$  obtenido en frases psicoacústicas, voz femenina, 3 transductores a 2 distancias



**Gráfica 8.3.  $R_{nonlin}$  en frases psicoacústicas, voz masculina y femenina, 3 transductores a 2 distancias**



**Gráfica 8.4.  $R_{nonlin}$  obtenido en las cuatro señales musicales, 3 transductores a 15 cm de éstos**

A la vista de las tres gráficas dedicadas a la evaluación de la señal de voz pueden concluirse las siguientes observaciones:

- a) La señal de voz masculina presenta resultados más altos y menor dispersión de éstos.
- b) La dispersión de valores aumenta al incrementar la distancia a la fuente, luego, son de esperar dispersiones más elevadas al emplear la métrica en salas con condiciones acústicas muy diferentes de las de una cámara anecoica, teniendo en cuenta además, que las distancias serán mucho mayores.
- c) En todos los casos el transductor B obtiene los mejores resultados en las medidas realizadas en la proximidad de la fuente.

La Gráfica 8.4 muestra los resultados de correlar con la métrica implementada las señales musicales originales y las radiadas por los dispositivos registradas en cámara anecoica a 15 cm de éstos en el eje central de radiación. Lo primero que llama la atención es la gran diferencia existente para las dos señales con más ancho de banda. Así, si la señal musical  $S_1$  obtiene en todos los casos el valor más alto, la señal musical  $S_2$  obtiene valores considerablemente inferiores a los obtenidos por las otras tres señales. Esto puede explicarse en parte por la gran cantidad de pasajes de la señal con muy bajo nivel (silencios) y el poco contenido espectral en frecuencias bajas y medias, como se ha explicado en el apartado 8.2.1. Dado que la métrica funciona dividiendo la señal temporal en *frames* de 30 ms (ver sección 6.7), en pasajes con silencios prolongados en los que todo el *frame* está conformado por silencio, la correlación se establecerá entre el silencio de la señal de referencia y el ruido de fondo de la sala donde se registre la señal, por lo que el resultado tenderá a cero necesariamente. Este aspecto también lo consideraremos en el caso del habla, comparando los resultados entre las frases psicoacústicas y el párrafo, para determinar en qué medida la presencia de silencios puede influir en el resultado de la métrica.

Volviendo a las señales musicales, vemos que la señal  $S_1$ , la que posee un mayor ancho de banda es la que mejores resultados obtiene para los tres transductores, obteniéndose en el transductor B el resultado más alto de nuevo. Las señales de oboe ofrecen resultados muy similares en los tres casos (casi idénticos en el transductor B), siendo ligeramente superiores los valores en la señal  $S_4$ . Si volvemos de nuevo a la Figura 8.1 (la imagen inferior derecha corresponde a la señal  $S_4$ ) observaremos que en esta señal el fraseo es continuo, sin ninguna interrupción, es decir no hay ningún silencio, a diferencia de la señal  $S_3$ , compuesta por notas picadas que interrumpen su ejecución en el momento de ataque de la siguiente nota. Este matiz podría explicar la sutil diferencia de resultados obtenidos al procesar las señales.

En todos los casos se obtienen las correlaciones más altas con el transductor B, siendo las distancias entre éstas – a pesar de los bajos valores de la señal  $S_2$  – también menores para el caso del mismo transductor. En vista de estos resultados y los obtenidos en el caso de las señales de palabra, se utilizará el transductor B como fuente de excitación de

las salas en la parte experimental relativa a la aplicación de la métrica implementada para su evaluación en salas.

## 8.4 Salas

El objetivo de esta parte experimental no es realizar una evaluación de unas determinadas salas, sino más bien, obtener una serie de datos a partir de salas bien diferenciadas que puedan permitirnos establecer una posible relación entre los parámetros tradicionalmente denominados objetivos de carácter monoaural contemplados en la norma ISO 3382-1:2009 y los resultados obtenidos con la aplicación del algoritmo  $R_{\text{nonlin}}$  en los mismos puntos de medición donde se han obtenido las medidas de parámetros objetivos, introduciendo además la nueva variable de la tipología de señal que se evalúa: palabra, música (con señales con características muy diferenciadas, como se ha visto en la sección 8.2 ). Por ello, nos entraremos en una descripción detallada de las salas ni en un análisis particularizado de los resultados obtenidos en cada una de ellas. El planteamiento consistirá en utilizar los datos obtenidos en el conjunto de salas diferenciadas para tener un rango lo más extenso posible de tiempos de reverberación, valores de claridad, valores de definición e índices de transmisión de la palabra obtenidos a partir de situaciones reales con la finalidad de poder confrontarlos con los resultados del algoritmo implementado.

En este sentido, se han utilizado salas con diferencias tan significativas como una cámara reverberante y una sala de locución para doblaje y grabaciones de voces en *off*. Los resultados completos para cada sala pueden consultarse en el Anexo I. En los subapartados sucesivos se realizará una breve descripción de las características geométricas, de materiales y principales usos de las salas, así como de los valores ponderados en frecuencia y en puntos de medición – de acuerdo con ISO 3382-1:2009 - de los parámetros que estudiaremos posteriormente.

### 8.4.1 Sala 1. Cámara Reverberante Campus de Gandia UPV

El caso de la cámara reverberante es el más particular de todas las salas que se verán en este capítulo. Obviamente, una sala de estas características no está pensada para funciones acústicas comunicativas, como pueden ser la palabra o la música. En nuestro caso, utilizaremos los resultados obtenidos en este recinto para disponer de datos en uno de los extremos del rango de valores. Si en el caso del estudio específico de los transductores se han elegido también como dispositivos de estudio los transductores de radiación indirecta es porque éstos son, a priori, los que van a proporcionar una mayor degradación de la señal – en forma de distorsión - debido a sus características de funcionamiento. Podemos establecer cierto paralelismo con la cámara reverberante. Se realizarán mediciones en un entorno de estas características porque en principio será el que introducirá una mayor degradación de la señal, desde el punto de vista comunicativo. Es de esperar que los mayores tiempos de reverberación, los menores valores de claridad y definición, así como los menores índices de transmisión se midan en esta sala, luego de estas medidas obtendremos el extremo inferior – en cuanto a calidad de la comunicación se refiere – del rango de datos total que compararemos.

La principal utilidad de una sala de este tipo es la realización de ensayos para la medida de la potencia acústica de fuentes de ruido, así como para la medida de coeficientes de absorción de materiales. Acústicamente, la sala se caracteriza por tiempos de reverberación muy largos y una muy baja absorción de sus cerramientos. Esto, junto con las características geométricas del recinto permite la creación de un campo acústico difuso y homogéneo. La adición de elementos difusores en el interior mejora la homogeneidad del campo, reduciendo las diferencias de niveles de presión entre diferentes puntos de la sala (Hernández, 2012).

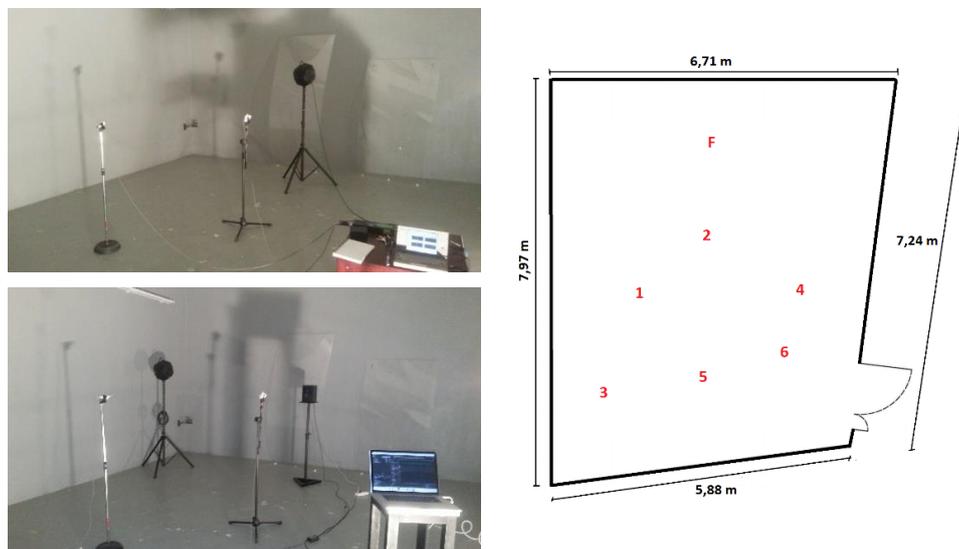


Figura 8.5. Planta, dimensiones, puntos de medida e imágenes de la sala 1

La figura anterior muestra la planta del recinto y sus dimensiones. Se trata de un recinto en forma de poliedro irregular de seis caras, en el que cada una de las superficies tiene diferentes dimensiones e inclinaciones. Para minimizar la influencia de los modos propios de la sala se evita que las superficies confrontadas de ésta sean paralelas, incluyendo también el techo respecto del suelo, con lo que entre las superficies se establece una angulación de  $8^\circ$  aproximadamente. Los cerramientos están acabados con enlucido liso de cemento y pintados con pintura Epoxi para minimizar la absorción acústica. El volumen de la sala es de  $238 \text{ m}^3$ . Se han realizado medidas y grabaciones de señal en los 6 puntos indicados en la Figura 8.5, siendo la posición de las fuentes, la marcada con una F.

La Tabla 8.3 indica los resultados de las medidas de carácter monofónico realizadas en este recinto, con un promediado en frecuencia considerando las bandas de 500 Hz y 1000 Hz y realizando el promedio aritmético de los seis puntos de medición en la sala. La sala se caracteriza por unos tiempos de reverberación extremadamente largos, unas relaciones energéticas en las que el peso de la energía temprana que llega al receptor es muy débil

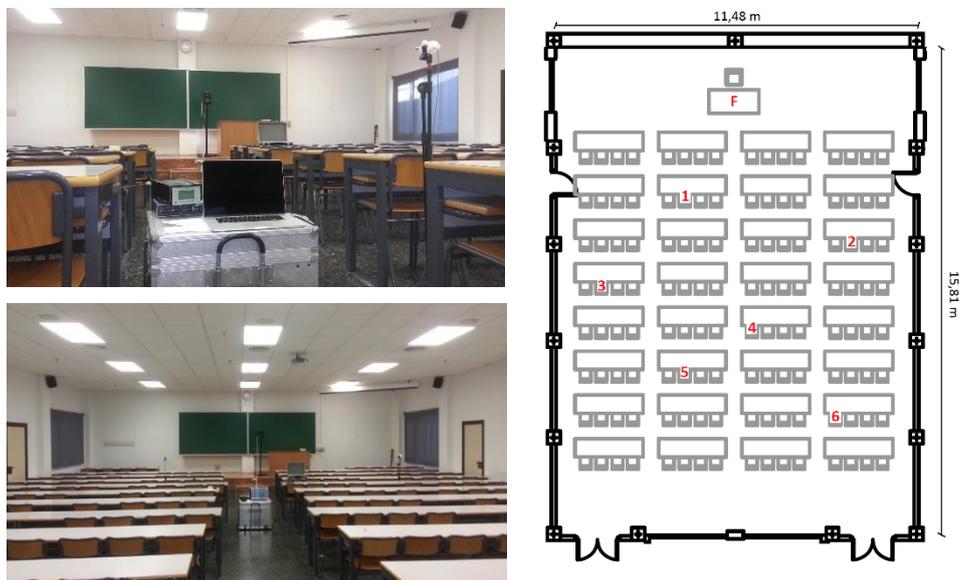
(como se observa de los valores D y  $C_{80}$  promediados) y un índice de inteligibilidad de la palabra deficiente.

**Tabla 8.3. Medidas ponderadas en la Sala 1**

Sala 1. Cámara Reverberante	
Ruido de fondo	23,3 dB(A)
RT	7,16 s
EDT	7,46 s
D	0,08
$C_{80}$	-8,1 dB
STI	0,3
RASTI	0,29

#### 8.4.2 Sala 2. Aula Docente Campus de Gandia

En este caso ya se trata de una sala concebida en principio con finalidades comunicativas, concretamente, se trata de un aula docente del Campus de Gandía de la UPV. Las dimensiones y puntos de medición y colocación de la fuente de excitación de la sala se muestran en la Figura 8.6. El volumen aproximado de la sala es de 690 m<sup>3</sup> y tiene una capacidad máxima de 175 personas.



**Figura 8.6. Planta, dimensiones, puntos de medida e imágenes de la sala 2**

A pesar de las relativamente grandes dimensiones del recinto, los cerramientos de éste no tienen ningún tratamiento acústico. Las paredes laterales están recubiertas con enlucido liso de yeso proyectado. El suelo es de terrazo y el cerramiento superior lo forma un techo suspendido formado por placas de escayola totalmente opacas. Otras superficies particulares del recinto son la tarima de madera en uno de los extremos del aula y las superficies de la gran cantidad de mesas que hay en el interior de la sala. De forma genérica, puede decirse que las superficies que conforman el recinto se caracterizan por una baja absorción acústica. El uso casi exclusivo de esta sala es el de la palabra hablada, generalmente sin refuerzo sonoro.

**Tabla 8.4. Medidas ponderadas en la Sala 2**

Sala 2. Aula docente	
<b>Ruido de fondo</b>	29,3 dB(A)
<b>RT</b>	1,98 s
<b>EDT</b>	1,99 s
<b>D</b>	0,31
<b>C<sub>80</sub></b>	-1,2 dB
<b>STI</b>	0,49
<b>RASTI</b>	0,48

#### 8.4.3 Sala 3. Sala de actos Casa de la Cultura de L'Alqueria de la Comtessa

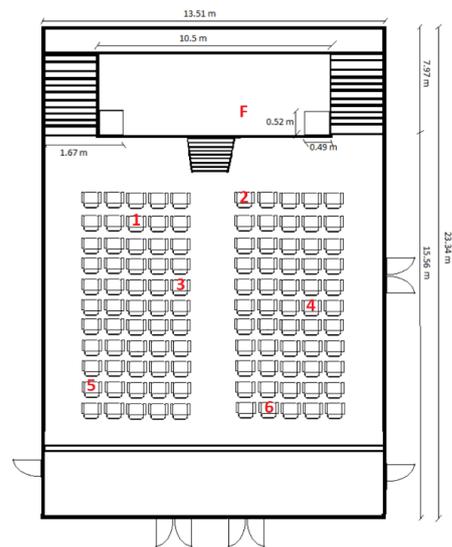
Se trata de una sala de uso polivalente en la localidad de l'Alqueria de la Comtessa. Las medidas de la planta, así como los puntos de medición y de la fuente se muestran en la Figura 8.7. En uno de los extremos, la sala tiene un escenario de unos 8 metros de profundidad que ocupa gran parte del ancho del recinto, cuyas superficies más destacadas son el suelo de madera y los telones que circundan el escenario. Ya en la zona destinada a la audiencia, las paredes están revertidas hasta una altura de 1,10 m por un zócalo de madera lisa, sin perforaciones. El resto de pared, desde el zócalo hasta el techo está recubierto por una moqueta. La altura del techo varía de forma escalonada a lo largo de la sala, siendo en la parte más alta de unos 6,5 m hasta los 3,12 m de altura de la parte posterior de la sala. El techo está formado por una superficie de enlucido liso y el suelo es de terrazo. Hay una disposición de 200 butacas colocadas en la platea de la sala, como puede verse en las imágenes de la Figura 8.7.

Como ya se ha indicado, se trata de una sala multiusos, que alberga actuaciones musicales con diferentes tipos de formaciones, así como teatro y otros eventos culturales. Se estima un volumen aproximado de unos 1480 m<sup>3</sup>, incluyendo la cavidad del escenario. De los elementos que conforman el recinto, la mayor absorción acústica la proporcionan los telones del escenario, así como la moqueta que recubre todas las paredes laterales y

la posterior. También puede resultar significativo el efecto de las butacas en el resto de la sala, dado que sus materiales son absorbentes y ocupan una gran parte de la superficie de la sala. La Tabla 8.5 muestra los resultados promediados para este recinto.

**Tabla 8.5. Medidas ponderadas en la Sala 3**

Sala 3. Sala Actos	
<b>Ruido de fondo</b>	25,4 dB(A)
<b>RT</b>	1,04 s
<b>EDT</b>	1,01 s
<b>D</b>	0,43
<b>C<sub>80</sub></b>	2,7 dB
<b>STI</b>	0,62
<b>RASTI</b>	0,63



**Figura 8.7. Planta, dimensiones, puntos de medida e imágenes de la sala 3**

#### 8.4.4 Sala 4. Aula Magna Campus de Gandia UPV

La cuarta sala que se ha utilizado en el procedimiento experimental es el Aula Magna del Campus de Gandia de la Universitat Politècnica de València. La sala está formada

por una platea ligeramente inclinada con una disposición de butacas para unas 260 personas y un escenario de unos 3,75 metros de profundidad ampliable mediante la retirada del panel trasero de la concha del escenario. El revestimiento de toda la concha escénica es de madera, al igual que lo son los revestimientos de la mayor parte de sus superficies laterales y el techo, situado a una altura de 9 metros. Una de las paredes laterales tiene una gran superficie revestida de moqueta acolchada, como se puede observar en las imágenes de la Figura 8.8. El suelo está tapizado con moqueta y el material de las butacas es también absorbente. El volumen aproximado de la sala es de unos 2200 m<sup>3</sup>. El uso más frecuente de la sala es el de la palabra – conferencias / teatro, aunque en ocasiones también aloja actuaciones musicales.

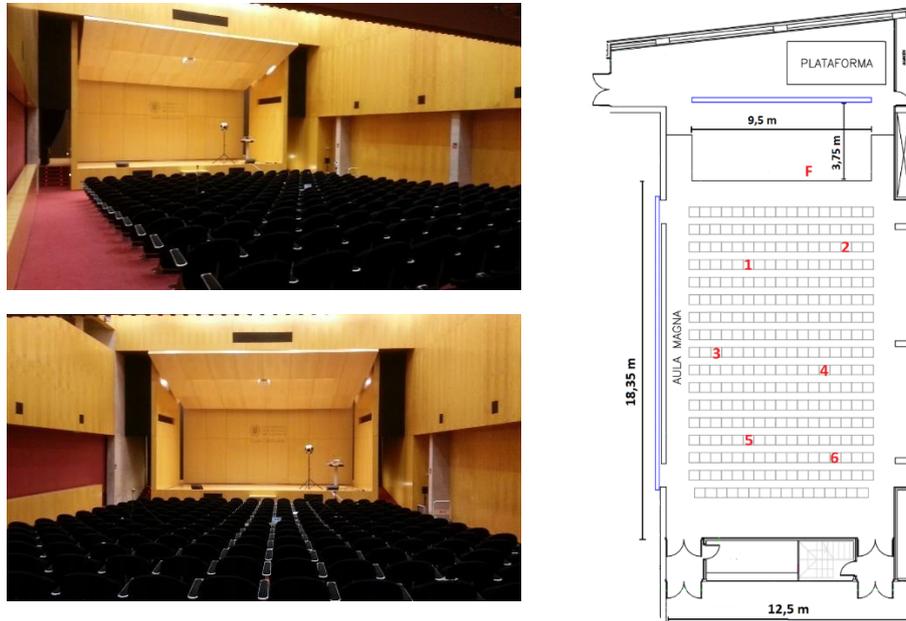


Figura 8.8. Planta, dimensiones, puntos de medida e imágenes de la sala 4

Tabla 8.6. Medidas ponderadas en la Sala 4

Sala 4. Aula Magna	
Ruido de fondo	28,1 dB(A)
RT	1,25 s
EDT	1,27 s
D	0,45
C <sub>80</sub>	2,4 dB

<b>STI</b>	0,58
<b>RASTI</b>	0,60

#### 8.4.5 Sala 5. Sala de Conferencias Campus de Gandia

Con esta sala iniciaremos la batería de medidas en recintos con volúmenes más reducidos. Cabe esperar pues, menores tiempos de reverberación y relaciones energéticas de la energía temprana más elevadas. Esta sala presenta una planta casi cuadrada y tiene una capacidad para unas 22 personas aproximadamente. El suelo está revestido de moqueta, siendo ésta la única superficie que conforma los cerramientos con algún coeficiente de absorción acústica considerable. El resto de cerramientos lo conforman paneles de madera desmontables en dos de las superficies – indicados en azul en el plano de la Figura 8.9, una tercera pared de ladrillo revestida también de madera, y la cuarta es completamente de cristal, con lo cual los cerramientos laterales son muy reflectantes. El techo lo conforma una moldura perimetral de 1,10 m de ancho acabado en enlucido liso y un hueco central a 70 cm más de altura rematado por escayola. El volumen aproximado de la sala es de 120 m<sup>3</sup>. El uso de la sala es para palabra.



Figura 8.9. Planta, dimensiones, puntos de medida e imágenes de la sala 5

Tabla 8.7. Medidas ponderadas en la Sala 5

Sala 5. Sala conferencias	
<b>Ruido de fondo</b>	25,0 dB(A)
<b>RT</b>	0,85 s
<b>EDT</b>	1,08 s
<b>D</b>	0,49
<b>C<sub>80</sub></b>	2,7 dB
<b>STI</b>	0,63
<b>RASTI</b>	0,63

#### 8.4.6 Sala 6. Locutorio Estudio de radio EPSG

El locutorio del estudio de radio del Campus de Gandía es una sala con un volumen aproximado de 82 m<sup>3</sup>, cuyo resto de dimensiones se pueden ver en la Figura 8.10. De sus cuatro cerramientos laterales, tres están revestidos por un zócalo de madera perforada con velo acústico de 1,80 metros de altura, que aloja una cámara rellena de guatas sintéticas de diferentes densidades. Esto le aporta a la sala una considerable absorción en las frecuencias medias. El resto de estas superficies hasta el techo está terminado en enlucido de yeso con pintura negra hasta el techo. El cuarto cerramiento es un panel de yeso que delimita dos cabinas adyacentes. El suelo es de resina y el cerramiento superior lo conforma un techo suspendido con placas de escayola.

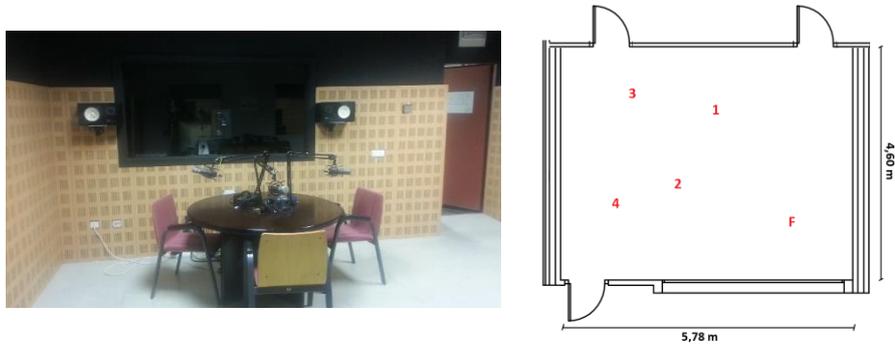


Figura 8.10. Planta, dimensiones, puntos de medida e imagen de la sala 6

Tabla 8.8. Medidas ponderadas en la Sala 6

Sala 6. Estudio de radio	
Ruido de fondo	24,2 dB(A)
RT	0,38 s
EDT	0,44 s
D	0,82
C <sub>80</sub>	11,9 dB
STI	0,81
RASTI	0,81

#### 8.4.7 Sala 7. Locutorio de doblaje EPSG

Se trata de un espacio de reducidas dimensiones – alrededor de 30 m<sup>3</sup> - dedicado principalmente a la grabación de voz en *offo* para doblaje. Dos de los cerramientos están formados por un zócalo de 1,80 metros de altura de madera perforada con cámara posterior rellena de guata sintética de diferente densidad. El resto del cerramiento hasta el techo está revestido por una fibra absorbente rematada con terciopelo negro. La parte inferior proporciona absorción a frecuencias medias, mientras que la superior es más absorbente en la alta frecuencia. El tercer cerramiento lo forma un panel separador con una ventana con cristal con visión directa a la cabina adyacente. La parte superior de este cerramiento tiene instalado un panel formando un ángulo con el techo de unos 30°, el panel está conformado por una estructura de madera forrada del mismo material absorbente que las paredes mencionadas anteriormente. El cuarto cerramiento es simplemente otro panel separador. El suelo es de resina y el techo es suspendido con placas de escayola opacas.



Figura 8.11. Planta, dimensiones, puntos de medida e imagen de la sala 7

Tabla 8.9. Medidas ponderadas en la Sala 7

Sala 7. Locutorio voz	
Ruido de fondo	23,3 dB(A)
RT	0,16 s
EDT	0,15 s
D	0,98
C <sub>80</sub>	29 dB
STI	0,91
RASTI	0,90

## **8.5 Resultados**

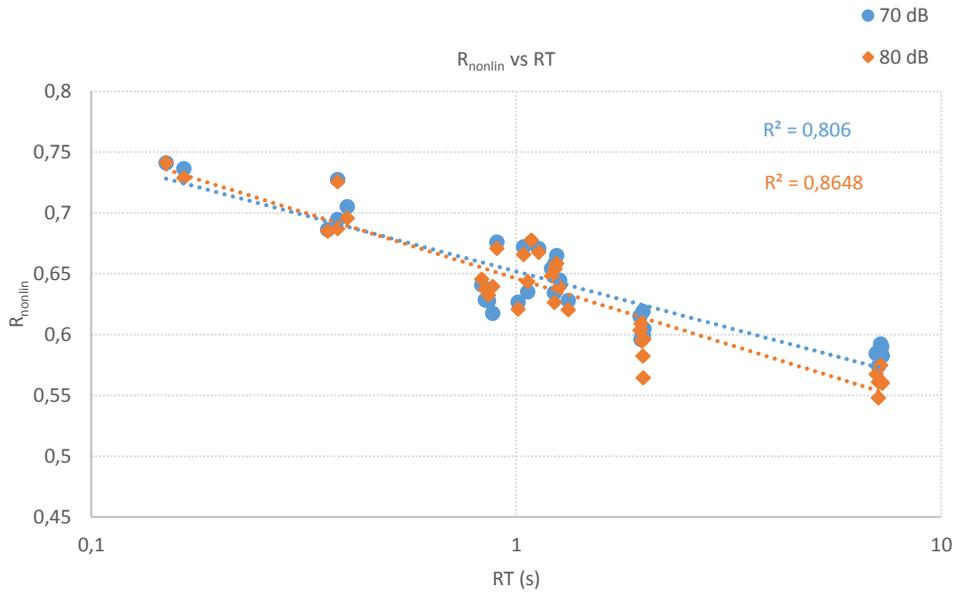
Se presentan en esta sección los resultados obtenidos en las mediciones *in situ* realizadas en las salas detalladas en la sección anterior. Como se ha comentado anteriormente, se han obtenido mediciones de parámetros objetivos monoaurales en diferentes puntos de cada sala. Además de éstos, las grabaciones de las señales de test empleadas se han realizado en los mismos puntos de medida y en la misma posición de la fuente de excitación de la sala. Estas grabaciones se han procesado posteriormente utilizando la herramienta implementada en el capítulo 6. Para el caso de las señales musicales, en el que se han utilizado cuatro señales de prueba se han procesado un total de 256 archivos, incluyendo los 34 puntos de medición utilizados en las diferentes salas con dos niveles de excitación de la sala en cada punto. En el caso de las señales de palabra, la evaluación se ha realizado utilizando dos tipos de materiales diferenciados: las frases psicoacústicas y los párrafos fonéticamente equilibrados, lo que constituye un *corpus* de 22 señales de prueba y un total de 1408 archivos procesados – teniendo en cuenta que también se han utilizado dos niveles de excitación de las salas en cada punto.

Los resultados obtenidos con el procesado de esos 1664 archivos se muestran en este apartado organizados de la siguiente manera: en primer lugar, se representarán los valores de la métrica obtenidos en el caso de las señales musicales comparándolos en gráficas de dispersión de pares con los resultados de las métricas objetivas obtenidas para cada punto de medición. En el caso de las señales musicales se representarán los valores de  $R_{\text{nonlin}}$  obtenidos con respecto a los tiempos de reverberación RT y EDT y a los parámetros energéticos de claridad  $C_{80}$  y definición D, tal como se han definido en las secciones 4.4 y 4.5 respectivamente.

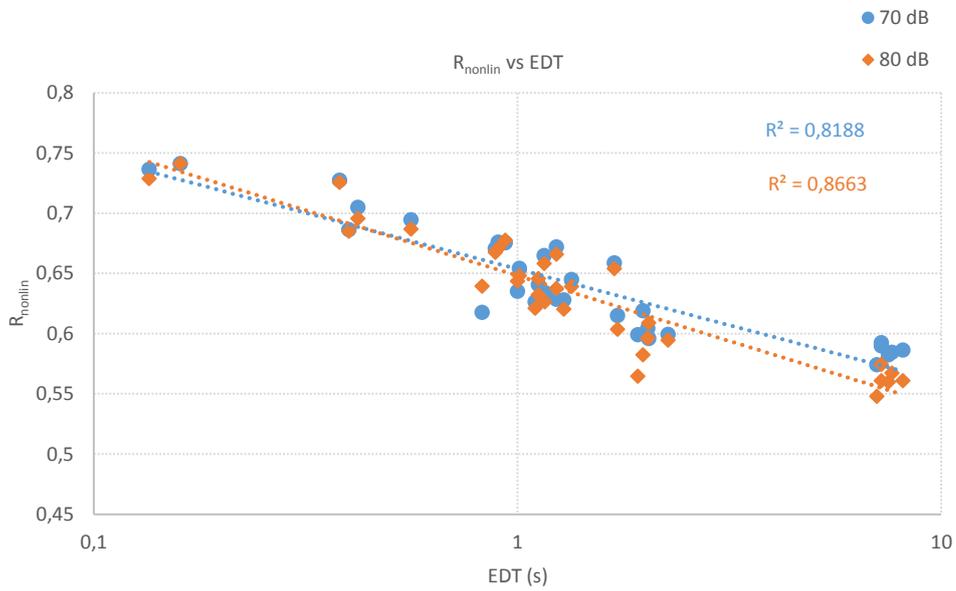
Respecto de las señales de test de palabra, las comparaciones se establecen entre los valores obtenidos con la métrica implementada y los parámetros de tiempo de reverberación EDT, la definición D y los índices de transmisión de la palabra STI y RASTI en el caso de la señal conformada por un párrafo de texto de 30 segundos de duración y en lo concerniente a las frases psicoacústicas, se representan en esta sección la dispersión entre los resultados de la métrica perceptual y los parámetros STI y RASTI. Resultados adicionales con estas frases pueden consultarse en el Anexo I.

### *8.5.1 Señales musicales*

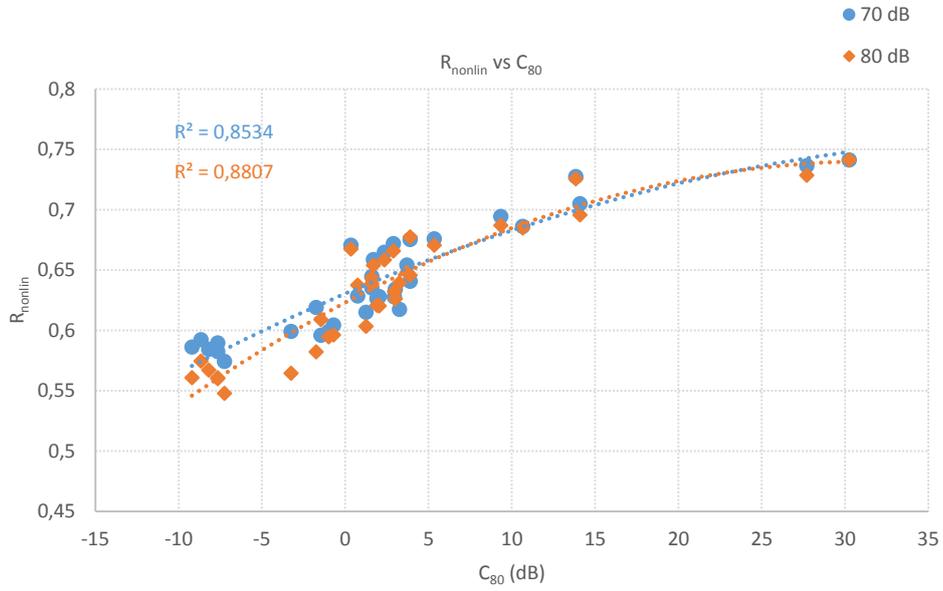
Para las señales musicales se presentan los pares de dispersión obtenidos en los 34 puntos de medida empleados en las diferentes salas. Cada uno de los registros se ha realizado a dos niveles de presión sonora diferentes en la sala. Es por ello que se representan dos nubes de pares de puntos, indicados con diferente color y marcador. El nivel de presión sonora indicado en cada gráfica corresponde al nivel equivalente en dB obtenido al excitar el transductor con la señal correspondiente durante toda la duración de ésta, medido en el punto más cercano al transductor utilizado en cada sala. La ganancia se ha ajustado para que el nivel generado por cada señal en las diferentes salas sea mismo, de forma que éste sólo variará para cada señal, en función de la densidad espectral de potencia de cada una de ellas.



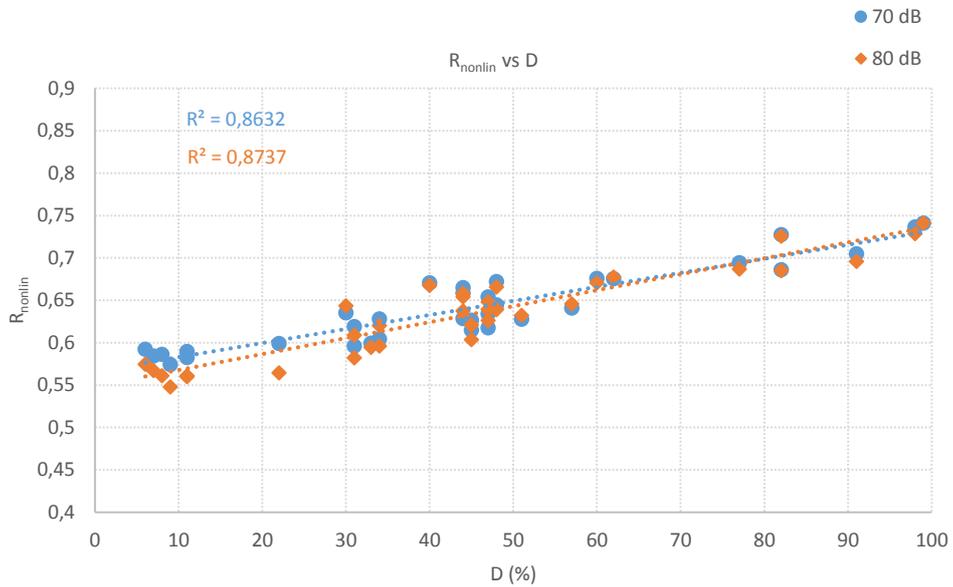
Gráfica 8.5. Pares RT - R<sub>nonlin</sub> con la señal de test S<sub>1</sub>



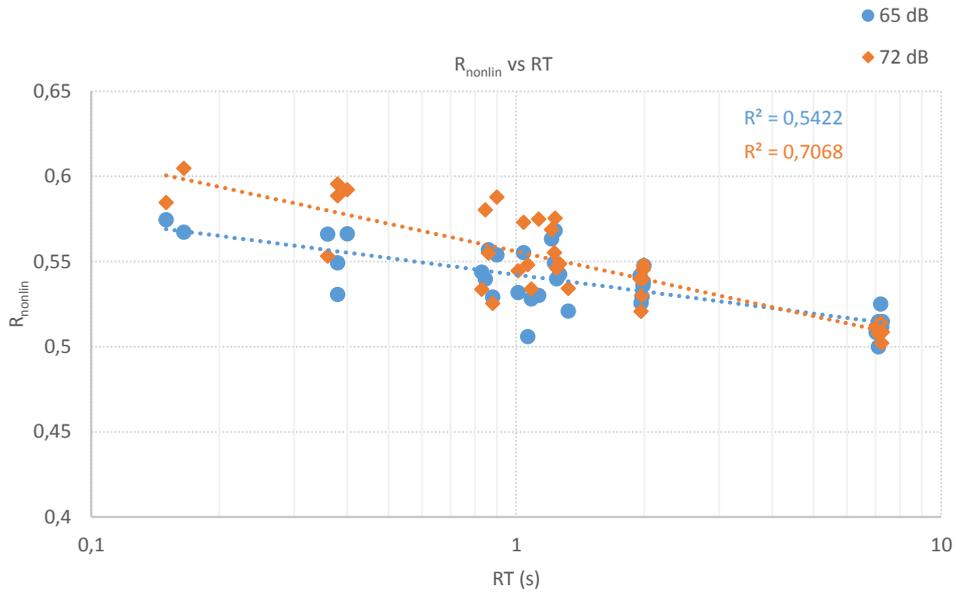
Gráfica 8.6. Pares EDT - R<sub>nonlin</sub> con la señal de test S<sub>1</sub>



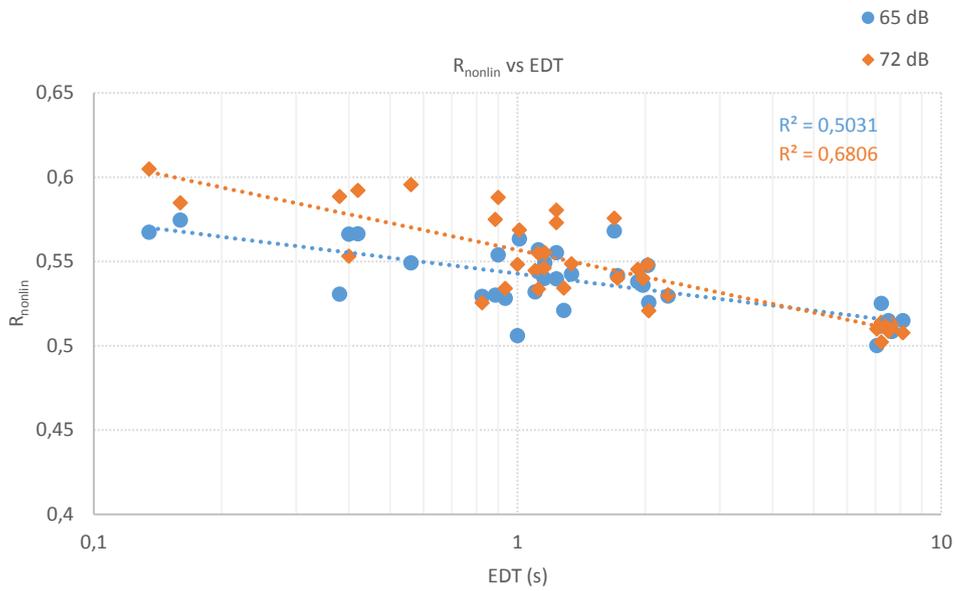
Gráfica 8.7. Pares C<sub>80</sub> – R<sub>nonlin</sub> con la señal de test S<sub>I</sub>



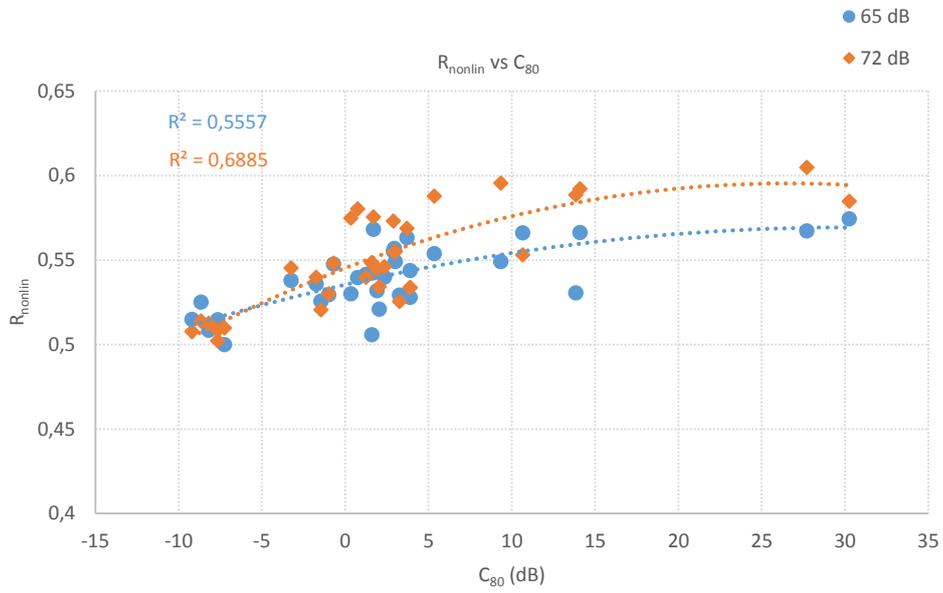
Gráfica 8.8. Pares D – R<sub>nonlin</sub> con la señal de test S<sub>I</sub>



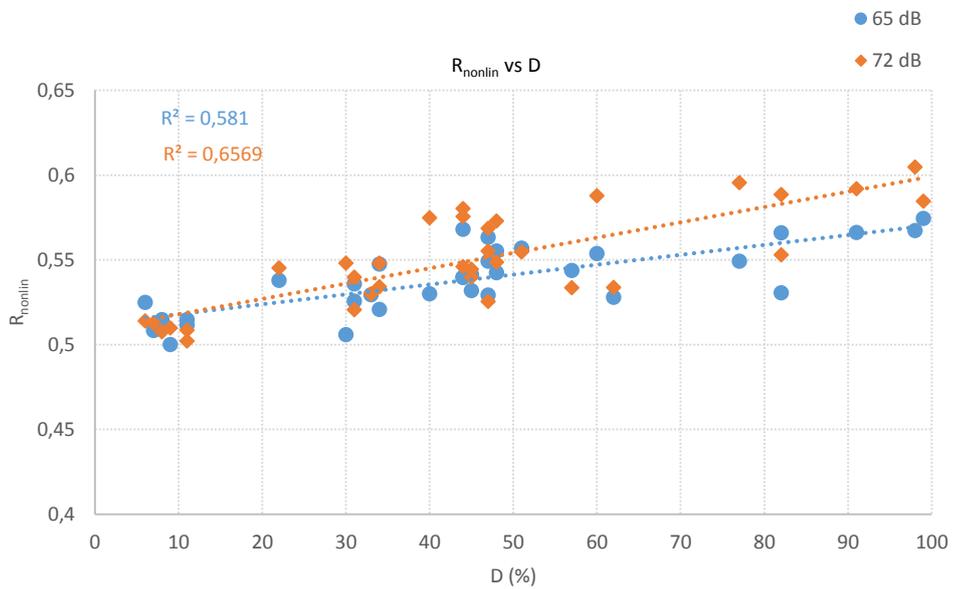
Gráfica 8.9. Pares RT - R<sub>nonlin</sub> con la señal de test S<sub>2</sub>



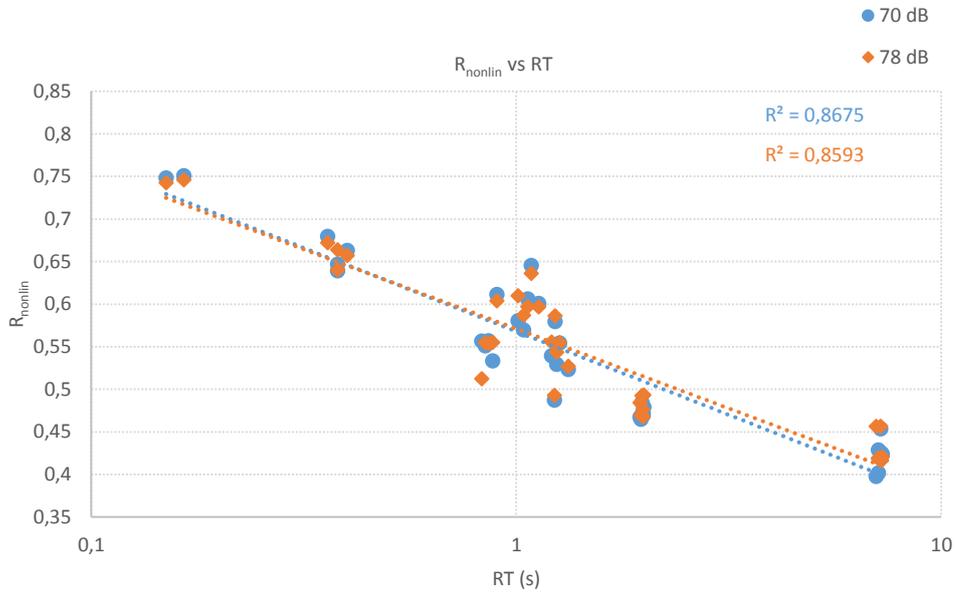
Gráfica 8.10. Pares EDT - R<sub>nonlin</sub> con la señal de test S<sub>2</sub>



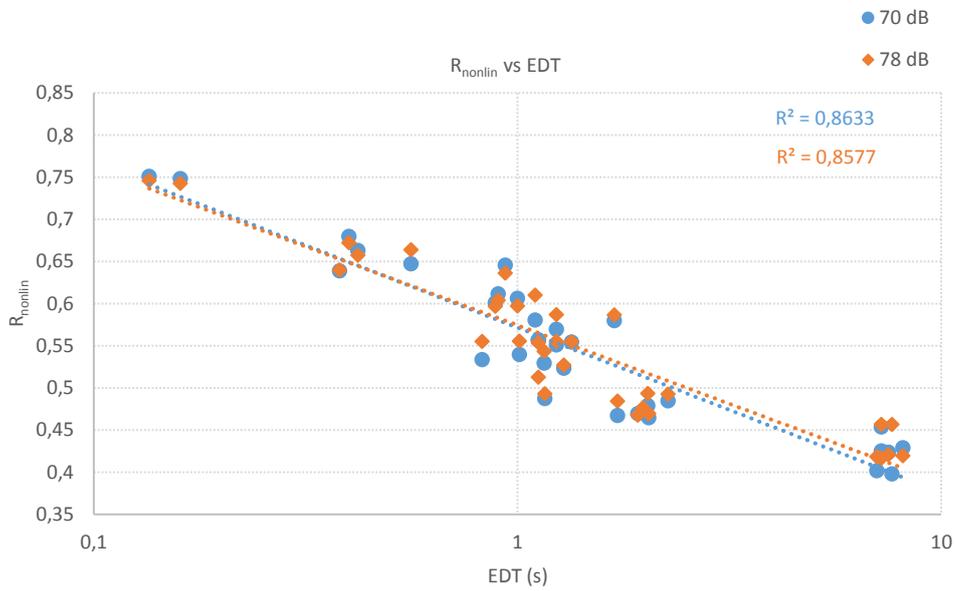
Gráfica 8.11. Pares C<sub>80</sub> - R<sub>nonlin</sub> con la señal de test S<sub>2</sub>



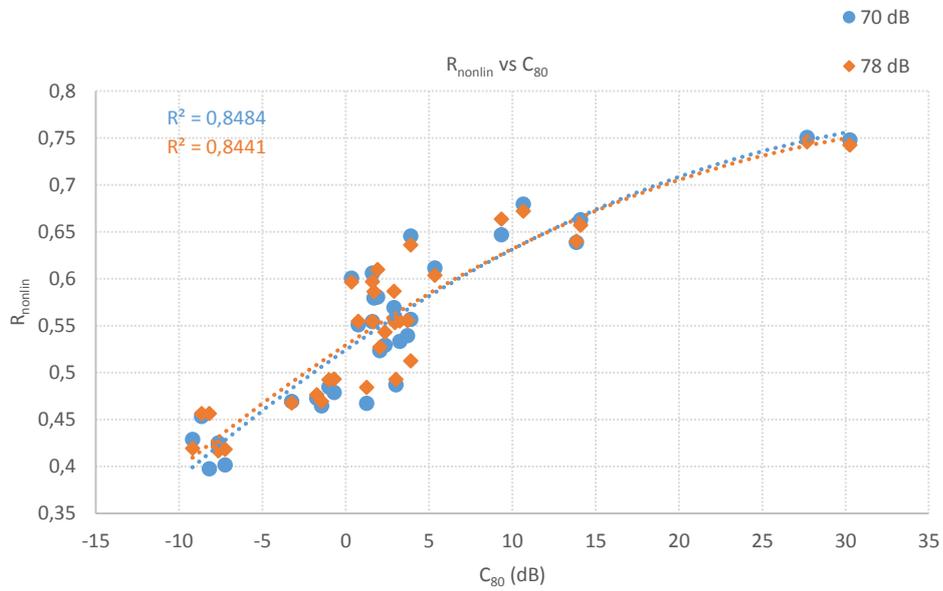
Gráfica 8.12. Pares C<sub>80</sub> - R<sub>nonlin</sub> con la señal de test S<sub>2</sub>



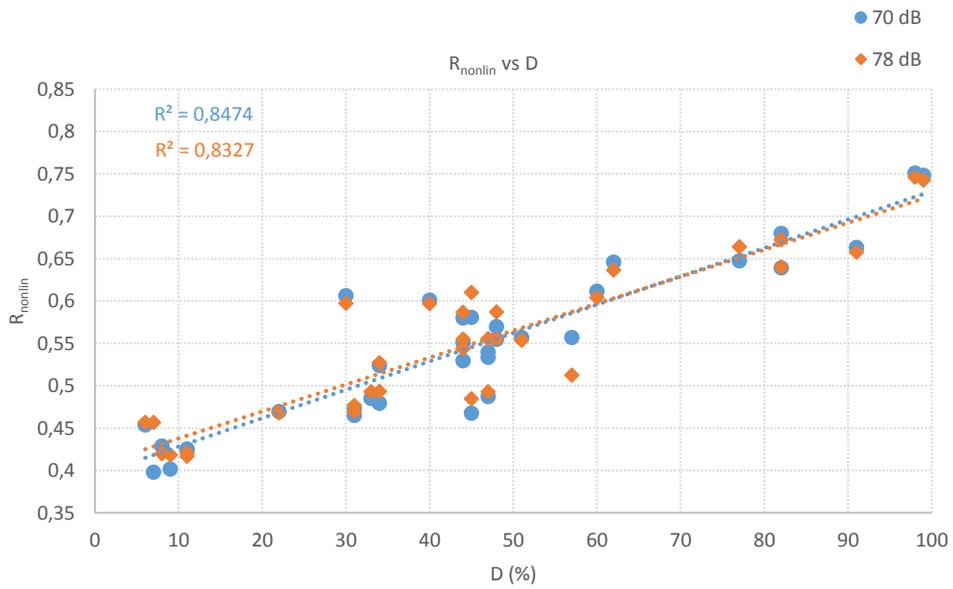
Gráfica 8.13. Pares RT - R<sub>nonlin</sub> con la señal de test S<sub>3</sub>



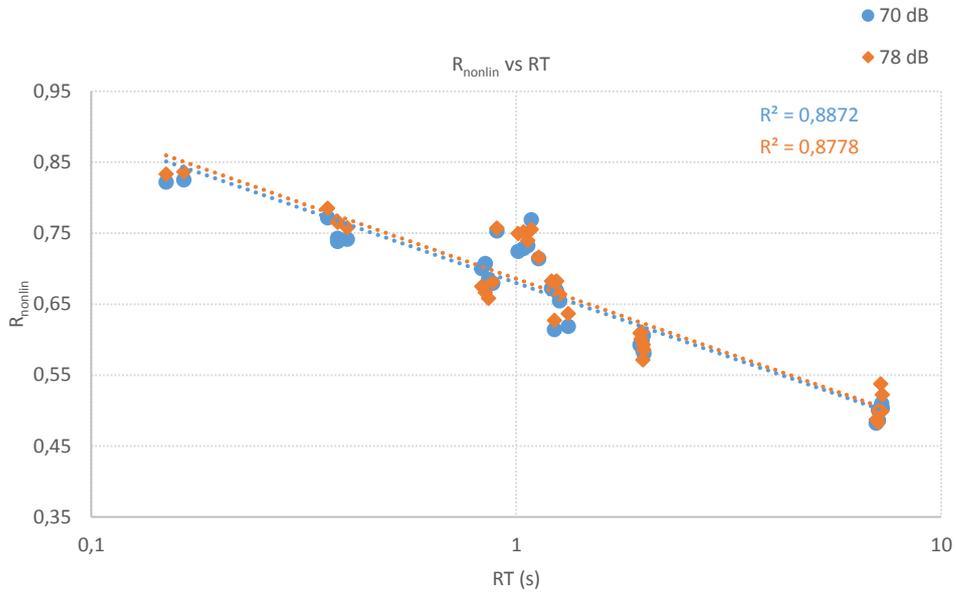
Gráfica 8.14. Pares EDT - R<sub>nonlin</sub> con la señal de test S<sub>3</sub>



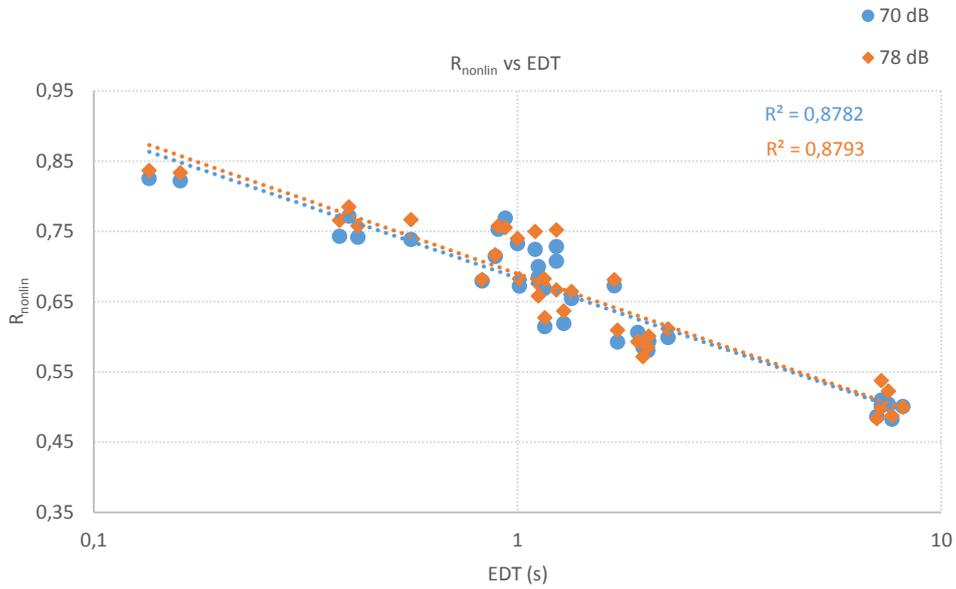
Gráfica 8.15. Pares C<sub>80</sub> - R<sub>nonlin</sub> con la señal de test S<sub>3</sub>



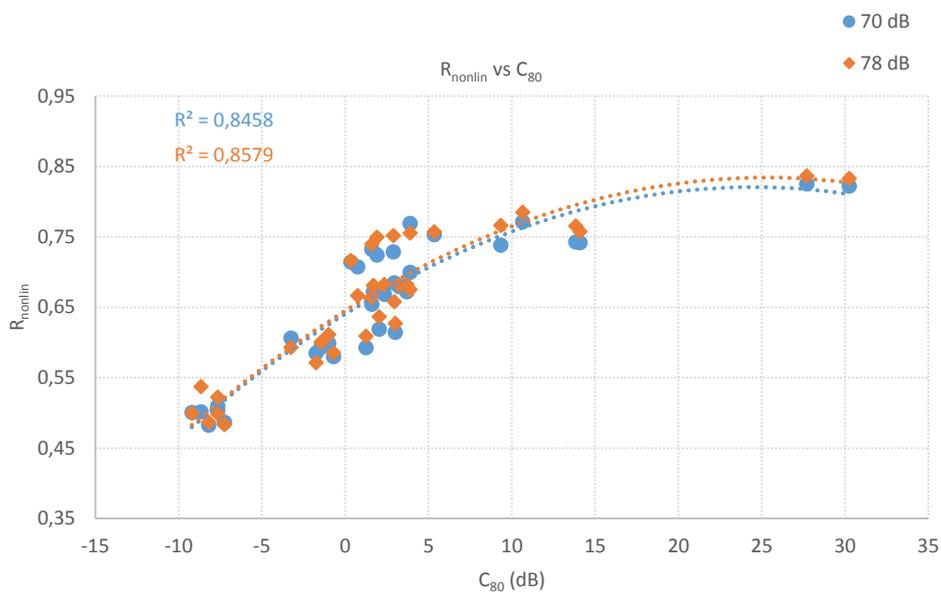
Gráfica 8.16. Pares D - R<sub>nonlin</sub> con la señal de test S<sub>3</sub>



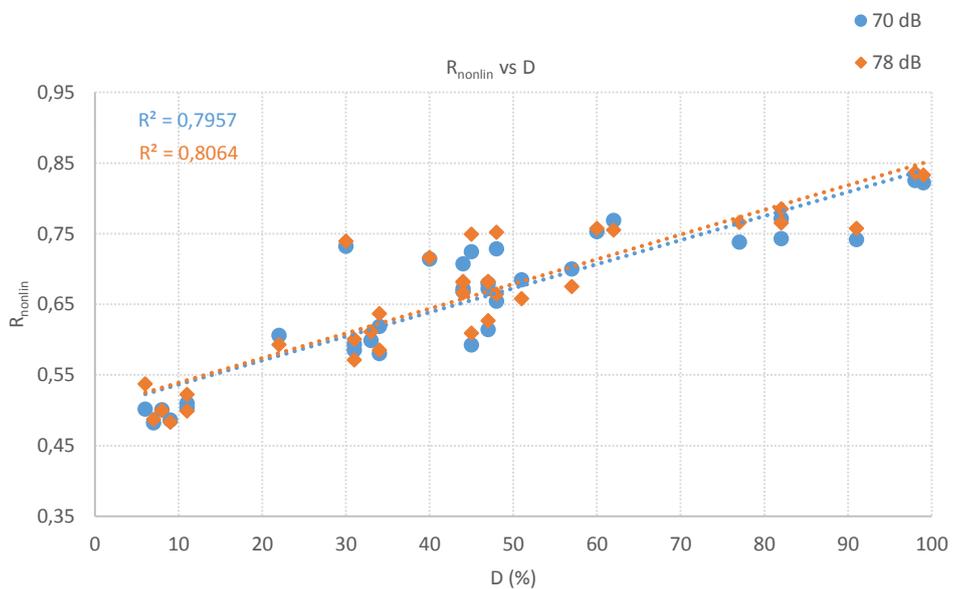
Gráfica 8.17. Pares RT - R<sub>nonlin</sub> con la señal de test S<sub>4</sub>



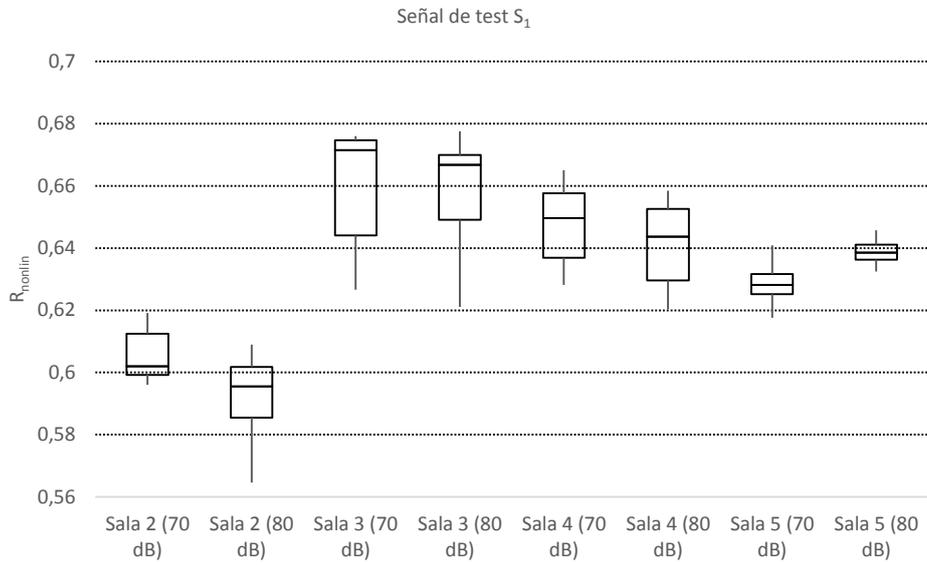
Gráfica 8.18. Pares EDT - R<sub>nonlin</sub> con la señal de test S<sub>4</sub>



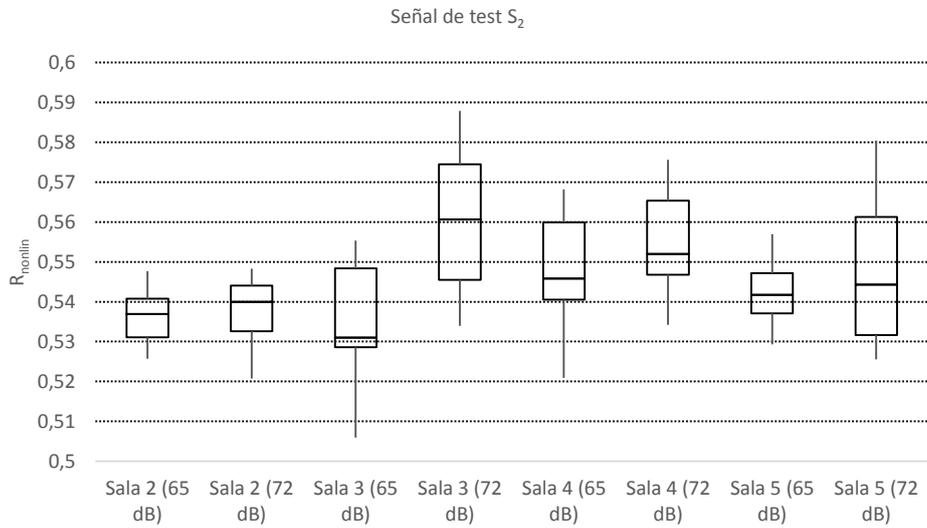
Gráfica 8.19. Pares C<sub>80</sub> - R<sub>nonlin</sub> con la señal de test S<sub>4</sub>



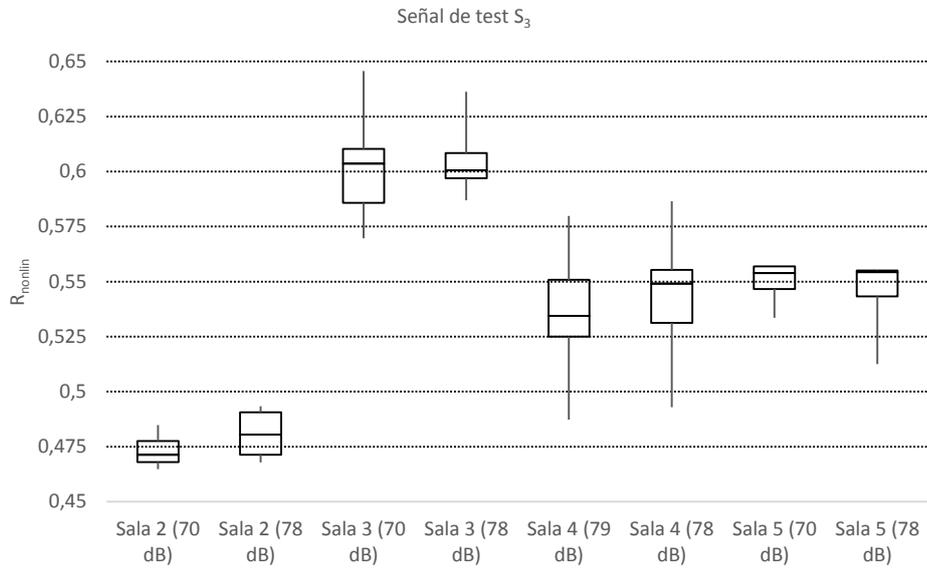
Gráfica 8.20. Pares D - R<sub>nonlin</sub> con la señal de test S<sub>4</sub>



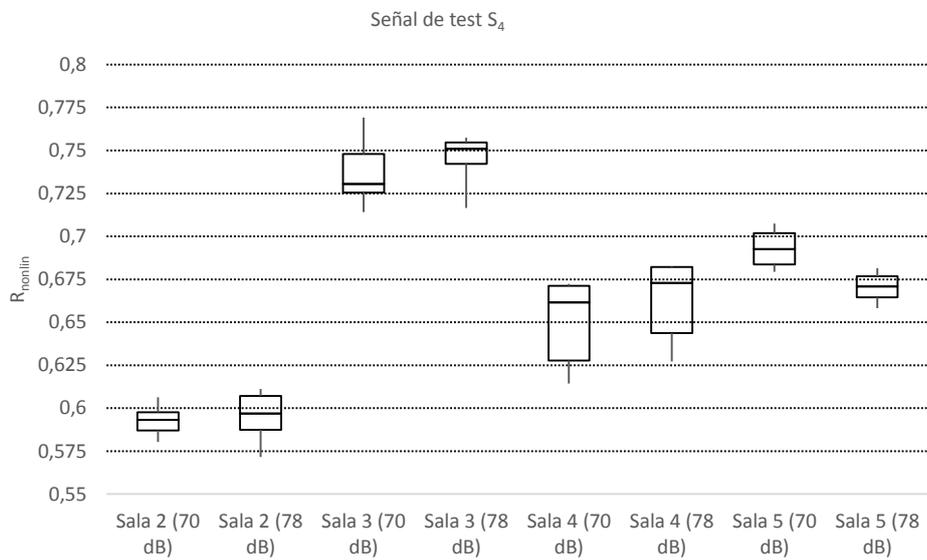
**Gráfica 8.21. Resultados de la señal de test  $S_1$  para las salas 2-5 con los dos niveles**



**Gráfica 8.22. Resultados de la señal de test  $S_2$  para las salas 2-5 con los dos niveles**



**Gráfica 8.23. Resultados de la señal de test  $S_3$  para las salas 2-5 con los dos niveles**



**Gráfica 8.24. Resultados de la señal de test  $S_4$  para las salas 2-5 con los dos niveles**

#### 8.5.1.1 *Discusión de resultados*

Las gráficas representadas en las páginas anteriores corresponden a las comparaciones realizadas entre los parámetros objetivos monoaurales establecidos en la norma ISO 3382-1:2009 y los resultados de la métrica implementada una vez procesadas las señales registradas en los mismos puntos de medición de los parámetros objetivos. La fuente de excitación dodecaédrica utilizada en la medición de los tiempos de reverberación, los parámetros energéticos y los índices de transmisión de la palabra, fue sustituida en el mismo emplazamiento por el transductor de radiación directa elegido tal como se ha descrito en la sección 8.3 para proceder a la excitación de la sala con las señales musicales  $S_1 - S_4$  descritas en el apartado 8.2.1.

En el capítulo 4, dedicado a establecer una visión general de la evaluación de la calidad en salas y recintos se ha visto que dos de las principales aproximaciones al problema consisten en el análisis temporal de la señal acústica propagada en el recinto y en la comparación de relaciones energéticas estableciendo diferentes constantes de tiempo, generalmente 50 u 80 ms, que vienen a relacionar la capacidad de integración del sistema auditivo humano y su resolución de distinción entre sonido directo y reverberado. Se ha visto que los parámetros denominados energéticos están formulados en este sentido: la relación entre sonido precoz –incluyendo el directo y las reflexiones tempranas- y el sonido reverberante. La mayor o menor presencia de sonido precoz respecto de reverberante otorgará un determinado grado de definición y de claridad. Estas técnicas de análisis exigen la comparación entre partes de señal de diferente longitud temporal – como es el caso de la definición, en la que se compara la energía de los primeros 50 primeros ms con el total de la respuesta al impulso- o bien entre diferentes tramos temporales de la señal obtenida – como ocurre con la claridad, en la que se compara la energía del primer tramo de la respuesta al impulso hasta los 80 ms con el resto de la energía contenida en la señal desde los 80 ms hasta el final de la duración de la respuesta al impulso de la sala.

Por otra parte, se ha visto en el capítulo 6 que la métrica implementada se fundamenta en una correlación cruzada de señales perfectamente alineadas en el tiempo, con lo cual no se establece ninguna relación entre diferentes partes o tramos de señal. El uso de técnicas de correlación cruzada de señales en el ámbito de la acústica de salas se ha realizado predominantemente en el ámbito de los parámetros de análisis espacial, como se remarca en el apartado 4.6. Así, mediante la técnica de correlación se establece la diferencia entre las señales llegadas al oído izquierdo y derecho, diferencia analizada principalmente desde un punto de vista temporal.

En nuestro caso, no se realiza un análisis temporal de las señales registradas sino una comparación del contenido de éstas con la finalidad de establecer en qué grado la métrica es capaz de reflejar la degradación de la señal provocada en el canal de transmisión, ya sea un transductor, una sala o el conjunto transductor-sala. Así, un requisito fundamental para proceder al procesado de la señal es que las muestras, la de referencia y la transmitida por el canal estén alineadas. En nuestra aplicación práctica esto se ha

realizado transmitiendo un *burst* corto de apenas 20 ciclos de una sinusoidal de 440 Hz. El espaciado entre este *burst* y el inicio de la señal está predeterminado, de forma que en la señal registrada podemos acceder al mismo punto de inicio que la señal de referencia.

Con esta aproximación podríamos pensar que estamos perdiendo información en la señal analizada, como los tiempos de retardo entre diferentes puntos de la sala, pero recordemos que el modelo se fundamenta en la comparación temporal de *frames* de 30 ms de duración. Se espera que la métrica refleje la degradación de la señal original en función de las modificaciones temporales sufridas por la señal transmitida en un mismo punto que la señal de referencia. Por ejemplo, un instante  $t$  de la señal original se comparará con el mismo instante de la señal transmitida, en la que para ese instante se habrá acumulado la energía reverberada anterior a este, con lo cual se establecerá una diferencia entre las señales en función del grado de degradación de la señal que introduzca la sala.

Como hemos visto también en el capítulo 6, este análisis se realiza en un banco de 40 filtros cuyo ancho de banda equivale a 1 ERB. Este planteamiento de análisis de la degradación tiene cierto parecido con el de la métrica STI, si bien en esta no se utiliza la técnica de la correlación. En cierta manera, el parámetro STI realiza una comparación de una señal original con una señal degradada en diferentes bandas de frecuencia ponderadas según su relevancia en la inteligibilidad de la palabra. La medida de la magnitud de la degradación se establece en función de la variación de la profundidad de modulación introducida por el canal o la sala, dado que la señal original es una señal modulada con una profundidad de modulación conocida. En nuestro caso, se realiza igualmente una comparación entre una señal original y una degradada en los mismos instantes de tiempo, sin considerar las variaciones temporales de propagación introducidas por la sala. La principal diferencia estriba, además de en el procesamiento perceptual relativo a los filtros del oído externo, medio e interno en que la comparación no se realiza con una señal determinista, sino que puede realizarse con cualquier tipo de señal aleatoria. Es por esto que la técnica de comparación no se limita a un detector de envolvente sino que se utiliza la técnica de correlación cruzada de señales.

Se han utilizado cuatro señales musicales con diferencias significativas en cuanto a contenido espectral, características temporales y densidades espectrales de potencia con la finalidad de poder establecer cómo responde la métrica a estas diferentes características. En primer lugar, para cada una de las cuatro señales se han establecido comparaciones entre los resultados de la métrica perceptual y los parámetros RT, EDT,  $C_{80}$  y D medidos en cada punto. Los resultados de las mediciones de los parámetros objetivos figuran en el Anexo III. El Anexo II recopila las funciones de correlación entre estos parámetros y la métrica  $R_{\text{nonlin}}$ , así como los índices de correlación obtenidos en cada caso.

Las gráficas desde la 8.5 hasta la 8.8 muestran los resultados para la señal  $S_1$ . En ellas puede observarse que los resultados ofrecidos por la métrica presentan un intervalo de

valores delimitado entre 0,54 en el peor de los casos y 0,75 en la mejor comparación obtenida para la señal  $S_1$ . Al representar la distribución de estos valores en función del tiempo de reverberación medido en cada punto obtenemos la Gráfica 8.5, en la que se establece una traslación entre los tiempos de reverberación medios medidos en los 34 puntos cuyo intervalo se establece entre un mínimo de 0,15 s y 7,24 s y unos valores de  $R_{nonlin}$  en el intervalo mencionado anteriormente. Se establece una correlación logarítmica entre ambos parámetros, que viene dada por una ecuación de la forma

$$RT = -a \cdot \ln(R_{nonlin}) + b \quad (8.1)$$

En la que  $a$  y  $b$  son dos constantes de ajuste de la función de correlación. En las gráficas relativas a los tiempos de reverberación, tanto  $RT$  como  $EDT$ , el eje temporal se ha representado en todos los casos en escala logarítmica, de forma que el ajuste se representa visualmente como una recta. A la vista de la nube de puntos pueden diferenciarse claramente los obtenidos en las salas con mayores tiempos de reverberación (Sala 1 y Sala 2), que son los dos grupos de puntos representados en la parte derecha de la gráfica y los medidos en las salas más anecoicas (Salas 6 y 7) representados en la parte izquierda de la gráfica. La nube de puntos central, que oscila en torno a tiempos de reverberación cercanos a 1s incluye puntos de las salas 3, 4, 5. Como se discutirá más adelante, la sala 5, a pesar de sus reducidas dimensiones, presenta unos resultados en cuanto a parámetros objetivos un tanto deficientes, en cuanto a tiempos de reverberación más altos de lo que cabría esperar en principio dadas sus dimensiones e índices de transmisión de la palabra un tanto pobres a pesar del poco volumen de la sala. Esto puede deberse a sus cerramientos extremadamente reflectantes – incluido uno entero de cristal –, con lo cual los datos obtenidos en esta sala no se diferencian claramente en estas gráficas de los valores obtenidos con las salas de mucho mayor volumen como son la 3 y la 4.

Se ha establecido también una comparación entre los resultados obtenidos con dos niveles de excitación de la sala. Los niveles presentados en la leyenda corresponden al nivel equivalente de la señal entera en el punto de medición más cercano al transductor en cada sala. Para la señal  $S_1$ , la que posee mayor contenido espectral de todas las utilizadas, la diferencia de nivel entre ambos ajustes de excitación es de 10 dB. De los resultados representados tanto en la Gráfica 8.5 como en la Gráfica 8.6 puede deducirse, en lo que respecta a los niveles de excitación de la sala, que en este caso la diferencia de nivel de excitación únicamente es relevante en las salas con mayores tiempos de reverberación, siendo claramente visible la diferente distribución de la nube de puntos para las salas 1 y 2 en función del nivel. De esta forma, se obtienen resultados de la métrica sensiblemente inferiores con un mayor nivel de excitación, lo cual es a priori consecuente con el fundamento de la métrica, tal como se ha planteado más arriba. En una sala muy reverberante, un aumento de nivel de excitación producirá un incremento del campo reverberante, que contribuirá a una mayor degradación de la señal original.

En ambos casos, tanto en lo que respecta al RT como al EDT, las funciones de ajuste para ambos niveles tienden a converger en los tiempos de reverberación por debajo de 1 s, produciéndose las mayores diferencias en los tiempos de reverberación altos. En cada una de las gráficas representadas, así como en el Anexo II se presentan los índices de correlación de las funciones de ajuste obtenidas, también para ambos niveles de excitación de la sala. Puede observarse que estos índices son significativamente mayores en el caso del nivel de excitación más alto, en lo que respecta a la señal  $S_1$ .

En lo que respecta a la Gráfica 8.6, que relaciona el EDT con  $R_{nonlin}$  para la señal  $S_1$  se observa un comportamiento casi idéntico al analizado respecto del tiempo de reverberación RT. El ajuste también viene dado, obviamente por una función logarítmica.

$$EDT = -a \cdot \ln(R_{nonlin}) + b \quad (8.2)$$

Continuando con los resultados obtenidos para la señal  $S_1$  pasamos a evaluar ahora los parámetros energéticos de la claridad  $C_{80}$  y la definición D. Los resultados se muestran en la Gráfica 8.7 y en la Gráfica 8.8, respectivamente.

En el caso del parámetro  $C_{80}$  la traslación se produce entre un amplio intervalo de valores -10 y 29 dB en el caso del parámetro objetivo, al que corresponde la variación de la métrica perceptual indicada anteriormente entre 0.54 y 0.75. Al igual que en el caso de los tiempos de reverberación pueden distinguirse claramente las nubes de puntos asociadas a las salas 1 y 2 – en la parte baja – y 6, 7 – en la parte de mayor claridad. A la vista de la distribución de puntos el que en principio parece que vaya a ser un ajuste lineal se convierte en polinómico, dado que los valores de claridad tan altos obtenidos en la sala 7 no se corresponden con valores de la métrica perceptual proporcionalmente – de forma lineal - más altos. Parece que la métrica tiende a saturar llegado un determinado nivel. Por otra parte un valor de claridad de 29 dB es en realidad un caso extremo, cercano a un campo anecoico. Con esto, se tiene un ajuste polinómico de segundo orden de la forma

$$C_{80} = -a \cdot (R_{nonlin})^2 + b \cdot R_{nonlin} + c \quad (8.3)$$

Siendo igualmente a, b y c las constantes de ajuste de la función. Respecto de los resultados en función del nivel de excitación, ya hemos visto anteriormente que los valores de la métrica perceptual difieren de forma significativa en el caso de las dos salas más reverberantes. Trasladado a la comparación con el parámetro de claridad, el ajuste entre ambos niveles de excitación tiende a converger a partir de valores de  $C_{80}$  cercanos a 0 dB. Es decir, para aquellos casos en que la energía precoz es igual o superior a la reverberante la métrica no responde significativamente a variaciones del nivel de excitación. En cambio, si la relación entre la energía directa o precoz y la reverberante es favorable a esta última un incremento de nivel se traducirá en una mayor degradación de la señal, que quedará reflejada por la métrica. Al igual que en el caso de los tiempos

de reverberación se obtiene un mejor índice de correlación del ajuste en el caso del nivel de excitación de la sala más alto.

En lo referente a la definición D, el ajuste se ha realizado para un intervalo de valores de definición entre el 6% y el 99% y la horquilla obtenida con la señal S<sub>1</sub> mencionada anteriormente. El comportamiento es análogo a los casos anteriores en lo que respecta a la diferenciación clara de los pares obtenidos en las salas 1, 2, 6 y 7 en los extremos inferior y superior de la definición D, así como en los peores resultados obtenidos con el nivel de excitación más bajo en las salas 1 y 2. El ajuste entre los valores de definición y los resultados de R<sub>nonlin</sub> presentan una relación lineal, de forma que ésta viene dada por

$$D = a \cdot R_{nonlin} + b \quad (8.4)$$

Obteniéndose unos índices de correlación de 0,87 y 0,86 para los niveles alto y bajo respectivamente. Las funciones de ajuste tienden a converger para valores de D cercanos y superiores al 50%, pudiéndose establecer la misma conclusión respecto al comportamiento de la métrica al compararla con parámetros energéticos que la explicada en el párrafo anterior concerniente a C<sub>80</sub>.

Para todas las señales musicales utilizadas en este experimento el mejor ajuste respecto de los parámetros objetivos se obtiene utilizando expresiones como las de las ecuaciones 8.1, 8.2, 8.3 y 8.4, es decir, la relación entre los tiempos de reverberación y R<sub>nonlin</sub> corresponderá a un ajuste logarítmico, el mejor ajuste al relacionarlo con C<sub>80</sub> será de tipo polinómico<sup>25</sup>, mientras que se establecerá una relación lineal entre R<sub>nonlin</sub> y la definición D. A pesar de que este rasgo es común a todas las señales, se verá a continuación que la respuesta de la métrica puede variar significativamente en función de las características de la señal de excitación. Veremos primero los casos de las señales S<sub>3</sub> y S<sub>4</sub>, que tienen un comportamiento similar a lo visto hasta ahora en el análisis de resultados realizados sobre la señal S<sub>1</sub>. Finalmente estudiaremos el caso de la señal S<sub>2</sub>, que presenta un comportamiento diferente obteniendo resultados bastante pobres en cuanto a índices de correlación con los parámetros objetivos.

En lo referente a la señal S<sub>3</sub>, a la vista de los resultados desde la Gráfica 8.13 hasta la Gráfica 8.16, se observa un comportamiento muy similar al obtenido con la señal S<sub>1</sub>, si bien en este caso el intervalo de valores de R<sub>nonlin</sub> obtenido oscila entre 0,4 y 0,75. En este sentido, la métrica ofrece un mayor recorrido de magnitud del parámetro en la parte baja, al obtenerse resultados alrededor de 0,14 puntos por debajo de los mínimos obtenidos con la señal S<sub>1</sub>. En la parte alta, los valores del parámetro son muy similares a los obtenidos con la primera señal analizada.

En lo que respecta a los ajustes con los parámetros objetivos se obtienen también resultados muy similares a los de la primera señal si bien en este caso los índices de correlación de los ajustes realizados son ligeramente inferiores a los obtenidos con la

---

<sup>25</sup> Se ha realizado siempre un ajuste polinómico de segundo orden dada la escasa variación del índice de correlación obtenido al incrementar el orden.

señal  $S_1$ . También se aprecia, en lo relativo a los dos niveles de excitación de la sala, que estos influyen en el resultado ofrecido por la métrica únicamente en el caso de las salas con mayor reverberación, tal como ocurría también con  $S_1$ . Sin embargo, a la vista de la dispersión de puntos mostrada en las gráficas relativas a la señal  $S_3$ , estas variaciones causadas por las diferencias de nivel no son tan pronunciadas como en el caso de la señal  $S_1$ . La diferenciación entre la nube de puntos de los diferentes niveles resulta sólo visible para el caso de la sala 1 – la que posee un mayor campo reverberante –. En los puntos asociados a la sala 2, las diferencias no son tan evidentes como lo eran con la señal  $S_1$ . Esto produce que las distancias entre las funciones de ajuste a los diferentes niveles sean menores, de forma que éstas divergen menos entre ellas, haciéndolo únicamente al acercarse a los valores extremos de reverberación obtenidos en el rango de medidas realizadas, así como a los valores más bajos de  $C_{80}$  y  $D$ .

Los datos representados entre la Gráfica 8.17 y la Gráfica 8.20 corresponden a los resultados obtenidos con la señal  $S_4$ . En este caso, la horquilla de valores ofrecida por la métrica está entre un mínimo de 0,48 y un máximo de 0,83, de forma que para esta señal la principal diferencia con respecto a  $S_1$  y  $S_3$  estriba en un incremento del rango de resultados de la métrica en la parte superior – en torno a 0,08 puntos por encima de los valores máximos obtenidos con las dos señales vistas anteriormente –. Otra particularidad de los resultados obtenidos con esta señal está relacionada con el comportamiento respecto de los dos niveles de excitación. Hemos visto, que en el caso de las señales  $S_1$  y  $S_3$ , especialmente en la primera, un incremento del nivel de excitación conllevaba un peor resultado de la métrica en los casos relacionados con los tiempos de reverberación más largos y las relaciones energéticas más favorables a la energía reverberante. Esto producía que las funciones de ajuste calculadas sobre la nube de puntos divergieran a medida que los tiempos de reverberación se acercan a valores muy altos o que los parámetros energéticos lo hacen a relaciones muy bajas. Este fenómeno no se produce con la señal  $S_4$ .

Dicho de otra manera, si bien los valores de  $R_{\text{nonlin}}$  obtenidos en la sala 1 son, al igual que con las otras señales, los más bajos de toda la serie, estos resultados no disminuyen al aumentar el nivel de presión sonora en una sala muy reverberante, como sí ocurría con las señales  $S_1$  y  $S_3$ . De hecho, al observar las funciones de ajuste graficadas entre la Gráfica 8.17 y la Gráfica 8.20 vemos que estas son prácticamente idénticas en casi todo el rango de valores. Otro rasgo diferencial de la señal  $S_4$  es la presencia de una ligera divergencia de los valores de la métrica obtenidos en función del nivel de excitación de la sala en el rango opuesto al que producía este fenómeno para las señales  $S_1$  y  $S_3$ . Así, la variación de la métrica con la diferencia de nivel, aunque mucho menor que la observada para las dos señales anteriores, se produce en las salas 6 y 7, es decir, las que tienen menor tiempo de reverberación y una relación energética muy favorable a la energía precoz. De esta forma, los resultados de la métrica obtenidos con un nivel de excitación alrededor de 8 dB superior al inicial son ligeramente inferiores en el caso de las salas más apagadas. Creemos que este comportamiento tiene su interpretación en la

estructura temporal de la señal  $S_4$ , como concluiremos en este mismo apartado un poco más adelante.

Antes de llegar a esto se analizarán los resultados obtenidos con la señal  $S_2$ , que son los más diferenciados de la batería de señales musicales utilizadas y los que peor correlación establecen con los parámetros medidos. Veamos lo que ocurre con la señal  $S_2$ . Los resultados obtenidos con esta señal se muestran entre la Gráfica 8.9 y la Gráfica 8.12. Lo primero que llama la atención es la reducción del rango de valores de  $R_{\text{nonlin}}$  obtenidos en este caso, comparado con el resto de señales utilizadas. Así, para  $S_2$  los resultados oscilan entre un valor mínimo de 0,50 obtenido en la sala 1 y un valor máximo de 0,60 obtenido en la sala 7. De esta forma, el ancho del rango de valores obtenido se reduce a una variación máxima de 0,1 puntos entre el peor y el mejor de los casos. Por otra parte, existe una dispersión de valores mucho mayor que las obtenidas en los casos anteriores de forma que los ajustes realizados presentan los peores índices de correlación obtenidos con todas las señales de test incluidas las de voz. Además, la diferencia de comportamiento de la métrica en relación con el nivel de excitación de la sala es notable en este caso. Así, como puede observarse en las gráficas mencionadas, el caso de esta señal es el que presenta unas curvas de ajuste más diferenciadas y distanciadas en función del nivel de excitación.

Normalmente, en los casos anteriores se produce una ligera diferencia de comportamiento en función del nivel de excitación en uno de los extremos de las curvas y el comportamiento converge posteriormente a una respuesta similar. Sin embargo, la simple inspección visual de las gráficas relativas a la señal  $S_2$  ya muestra dos nubes de puntos bien diferenciadas en función del nivel de excitación de la sala, de forma que las funciones de ajuste divergen, intersectándose en un punto alrededor de 4 segundos, en lo que respecta a los tiempos de reverberación RT y EDT, en un valor de claridad alrededor de -5 dB y en un valor de definición cercano al 15%. Por debajo de estos valores de tiempos de reverberación y por encima de estas relaciones energéticas los resultados varían significativamente en función del nivel de excitación. Como última característica de los resultados obtenidos con esta señal, cabe mencionar que el ajuste a una función en relación con los parámetros objetivos mejora significativamente al aumentar el nivel de excitación de la sala, si bien, incluso a este mayor nivel, los índices de correlación obtenidos son los más bajos de todas las señales utilizadas en el experimento.

La interpretación de los resultados obtenidos con las diferentes tipologías de señales la haremos en función de un análisis del contenido espectral y de la estructura temporal de las señales de test utilizadas, de la misma manera que se ha hecho en el caso de la evaluación de la métrica en transductores y como ya se ha adelantado en la sección 8.2 de este capítulo. La Figura 8.12 compara los espectros en frecuencia de las señales  $S_1$  y  $S_2$  mientras que la Figura 8.13 hace lo propio con las señales  $S_3$  y  $S_4$ .

Respecto de la señal  $S_2$ , podemos observar en la Figura 8.12 su espectro en comparación con la señal  $S_1$ . Se ha elegido esta comparación, ya que ambas señales son las que a priori presentaban un mayor ancho de banda. Sin embargo, la simple comparación de espectros realizada en la mencionada figura sobre las señales de referencia muestra que la energía de la señal  $S_2$  es mucho menor que la contenida por la señal  $S_1$  en todo el ancho de banda, excepto en la banda entre 20 y 80 Hz. De hecho, puede observarse que la mayor parte del contenido espectral de la señal  $S_2$  se concentra en esa banda. Este análisis está realizado sobre las señales utilizadas como referencia. En la Figura 8.14, se representa la densidad espectral de potencia de las cuatro señales de test utilizadas pero en este caso en una de las grabaciones reales efectuadas para la parte experimental de este trabajo, concretamente en la sala 7.

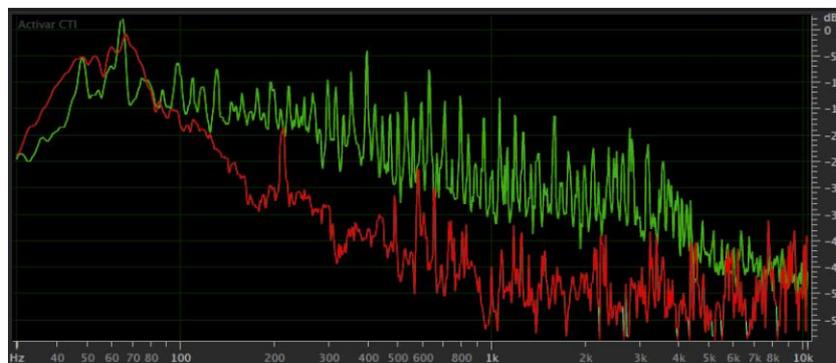


Figura 8.12. Espectros promediados de las señales  $S_1$  (verde) y  $S_2$  (rojo)

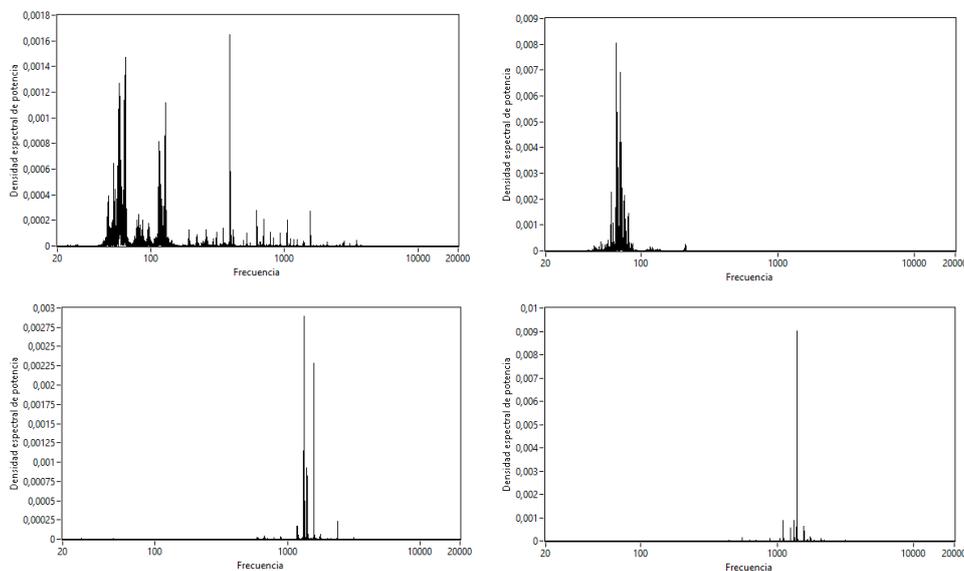


Figura 8.13. Espectros promediados de las señales  $S_3$  (rojo) y  $S_4$  (verde)

La densidad espectral de potencia es una función que da cuenta de cómo está distribuida la potencia de la señal en el ancho de banda de ésta. Así, las funciones representadas en la Figura 8.14 están calculadas sobre las señales de presión medidas en la sala citada y son proporcionales al cuadrado de la presión. La gráfica superior derecha de la Figura 8.14 es la que da cuenta de la densidad espectral de potencia de la señal  $S_2$  emitida por

el transductor utilizado en este experimento en una sala real. Puede observarse que la práctica totalidad de la energía se acumula por debajo de los 100 Hz, lo cual es consecuente con el espectro de la señal de referencia representado en la Figura 8.12 y puede ayudarnos a entender los pobres resultados obtenidos con la métrica estableciendo comparaciones entre la señal de referencia y la transmitida por el canal.

En primer lugar hay un problema de nivel de la señal. Ya se ha visto que el nivel de presión sonora equivalente generado por esta señal en el punto más cercano al transductor en cada sala era de 65 dB. Sin embargo, si medimos el nivel sonoro equivalente de la misma señal registrada aplicando una red de ponderación A, obtendremos valores de nivel de presión sonora alrededor de 52 dB(A). La red de ponderación A es un ajuste del valor del nivel de presión sonora adaptado aproximadamente a las características de las curvas isofónicas vistas en el capítulo 5 (Figura 5.11) para niveles de presión bajos y medios. Si medimos los niveles de presión sonora equivalente con la red de ponderación A para las demás señales utilizadas en el test apenas obtendremos variación respecto a los niveles de presión obtenidos sin ponderar. Esta diferencia se explica porque la mayor parte de la potencia de la señal  $S_2$  está contenida en frecuencias ponderadas a la baja en relación con la respuesta del sistema auditivo humano.



**Figura 8.14. Densidades espectrales de potencia de las señales  $S_1$ ,  $S_2$ ,  $S_3$  y  $S_4$**

Como se ha visto en el capítulo 6, la propia métrica implementa un filtrado para la simulación del oído medio y externo cuya respuesta en frecuencia es muy similar al de las curvas isofónicas, atenuando fuertemente todo el contenido espectral por debajo de 300 Hz (ver Figura 6.4). Creemos que este déficit de potencia es uno de los factores que

explica los pobres resultados obtenidos con esta señal. Bajo este precepto también se podría justificar que un aumento del nivel de excitación contribuyese a una mejor correlación de los resultados obtenidos, como de hecho ocurre a la vista de las curvas representadas entre la Gráfica 8.9 y la Gráfica 8.12. Dadas las características espectrales de la señal, el nivel de excitación hubiera debido ser considerablemente mayor para obtener datos comparables al del resto de señales. Existe además, otro importante aspecto a considerar en relación con el resultado arrojado por  $S_2$ , también relacionado con la composición espectral. Como se observa en la Figura 8.12 el pico de contenido espectral de la señal se centra en la banda entre 40 y 80 Hz. La respuesta en frecuencia del transductor utilizado, según datos del fabricante, es de 56 Hz – 2kHz en un rango de  $\pm 3$  dB, con lo cual en el canal de transmisión constituido por el conjunto transductor-sala ya estaríamos suprimiendo la casi mitad de esa banda de frecuencia. Dado que en la señal de referencia este contenido no se suprime – sólo se atenúa en función de la emulación del sistema auditivo – es de esperar que las comparaciones entre la señal de referencia y la registrada sean deficientes.

En lo que respecta a las señales  $S_3$  y  $S_4$ , dado que tienen como origen el mismo instrumento musical, es de esperar que tengan un espectro muy similar como se corrobora en la Figura 8.13. De esta forma, las diferencias en cuanto a respuesta de la métrica pueden abordarse desde el punto de vista de la estructura temporal de la señal. Se obtienen valores ligeramente superiores de  $R_{nonlin}$  con la señal  $S_4$  en comparación con la señal  $S_3$ . Si retomamos la representación temporal de las señales vista en la Figura 8.1 –las dos imágenes inferiores izquierda y derecha corresponden a  $S_3$  y  $S_4$  respectivamente- veremos que la primera está formada por una sucesión de notas picadas y silencios y la segunda por una melodía de notas ligadas. A nivel de estructura temporal de las señales esto se traduce en una presencia de transitorios y silencios en el caso de  $S_3$  y en una forma de onda cuasi estacionaria en el caso de  $S_4$ .

Hemos visto, al analizar la respuesta de la métrica con los transductores de forma aislada, que las señales estacionarias presentaban una peor respuesta de la métrica, dado que posibilitaban la comparación con los armónicos emitidos por los transductores en registros realizados muy cerca de la superficie de radiación de éstos y en total ausencia de campo reverberante, tanto en el caso de los transductores de radiación indirecta – medidos con el tubo de onda plana – como los de radiación directa – medidos en cámara anecoica -. Al añadir el efecto de la sala, es de esperar que la presencia de un campo reverberante en el canal de transmisión altere este comportamiento, ya que la degradación sufrida por los transitorios debidos a la reverberación puede resultar mucho más significativa que la experimentada por una señal con características parecidas a una estacionaria. Además, los registros de señal se realizan a distancias muy superiores, con lo cual, con los niveles de presión sonora utilizados en el experimento, la influencia de los armónicos emitidos por el transductor es seguramente despreciable en comparación con la degradación de la señal introducida por la sala.

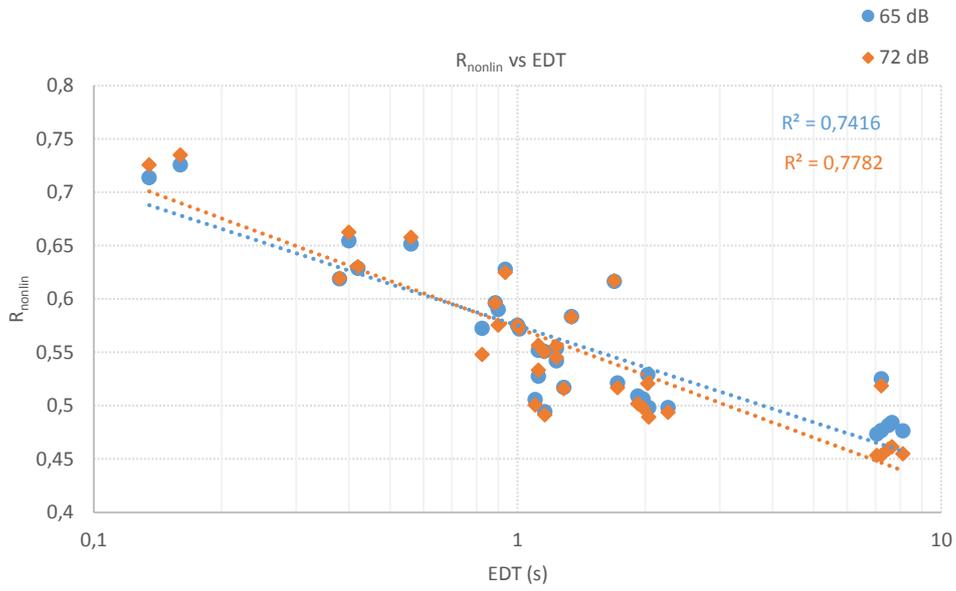
Finalmente, en lo que respecta a los resultados obtenidos con las señales musicales comentaremos brevemente los valores obtenidos en relación a la posible utilidad de la métrica como recurso de comparación entre recintos o entre canales de transmisión. Para ello, remitimos al lector a los datos representados entre la Gráfica 8.21 y la Gráfica 8.24. En ellas se representan únicamente las dispersiones de valores de  $R_{\text{nonlin}}$  obtenidas con cada señal en las diferentes salas sin referirlos a ningún otro parámetro. No se representan en estas gráficas los resultados de la sala 1, la sala 6 ni la sala 7 cuyas diferencias son evidentes y constatables tanto con los parámetros clásicos medidos como con la propia métrica implementada.

Hemos querido centrarnos en las salas con características de volumen o de tiempos de reverberación más similares – aquellas que no se diferencian claramente en las nubes de puntos de las curvas de ajuste representadas en las gráficas anteriores. Así, en estas gráficas se establece la comparación de los resultados obtenidos entre las salas 2, 3, 4 y 5 con las cuatro señales de test y los dos niveles de excitación utilizados. Las salas 2, 3 y 4, con sus diferencias, son comparables en cuanto a dimensiones y parámetros objetivos obtenidos mientras que la sala 5 a pesar de sus reducidas dimensiones presenta unos resultados en cuanto a parámetros objetivos un tanto deficientes seguramente por las características de sus cerramientos, tal como se ha comentado anteriormente. Los resultados se presentan en forma de diagramas de cajas y bigotes con la finalidad de representar gráficamente la distribución de los valores obtenidos en los diferentes puntos de medición de cada sala, los valores intermedios y los extremos. A la vista de la Gráfica 8.22 se vuelve a poner de manifiesto lo inadecuado de la señal  $S_2$  para su uso con la técnica implementada por todos los motivos argumentados anteriormente, por lo que centraremos la discusión sobre los resultados obtenidos con las señales  $S_1$ ,  $S_3$  y  $S_4$ , representados respectivamente en la Gráfica 8.21, Gráfica 8.23 y Gráfica 8.24. En lo referente a las salas de mayor volumen (2, 3 y 4) puede observarse que la métrica proporciona la misma valoración relativa entre las salas, obteniendo siempre mejor resultado la sala 3, posteriormente la 4 y siempre la 2 como peor valorada. La sala 5 obtiene en todos los casos una valoración similar a la 4, pero su posición relativa respecto a ésta depende de la señal utilizada y del nivel de excitación. El rango de diferencia de valores entre las salas es menor con la señal  $S_1$  y más remarcado con las señales  $S_3$  y  $S_4$ .

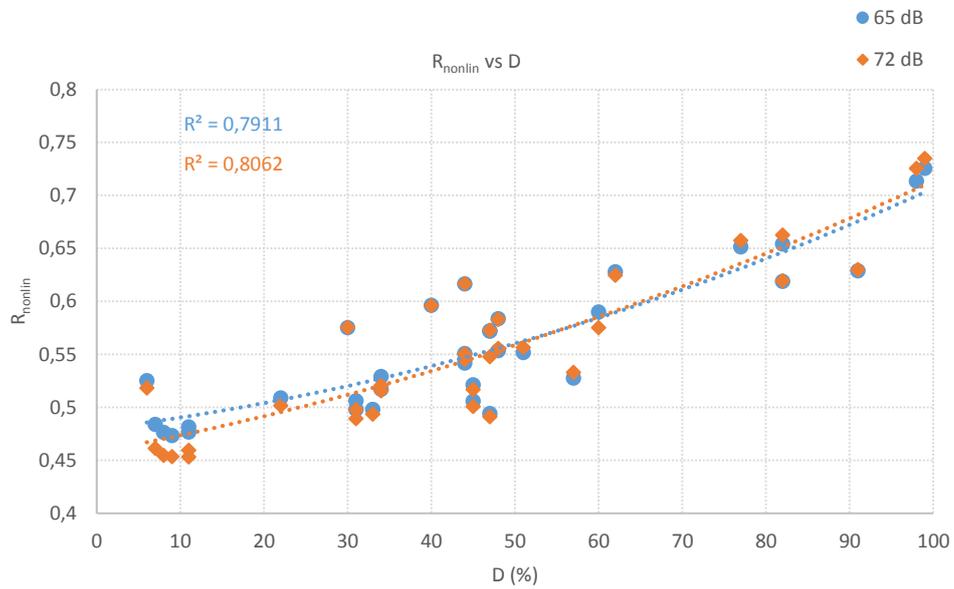
Los resultados representados son coherentes con los parámetros objetivos, si bien en nuestra opinión la representación de la métrica implementada con este formato tiene un valor añadido a la hora de establecer comparaciones entre sistemas: en primer lugar da cuenta de las diferencias cualitativas entre los sistemas, en segundo lugar no se limita a asignar un único valor a un sistema sino que muestra el rango de variación dentro del mismo, lo cual puede ser un marcador cualitativo en sí mismo. Finalmente, dadas las características de la métrica implementada, la valoración se establece en función de un evento sonoro concreto, con lo cual, pueden compararse diferentes sistemas en función de los tipos de eventos sonoros.

### 8.5.2 *Señales de voz*

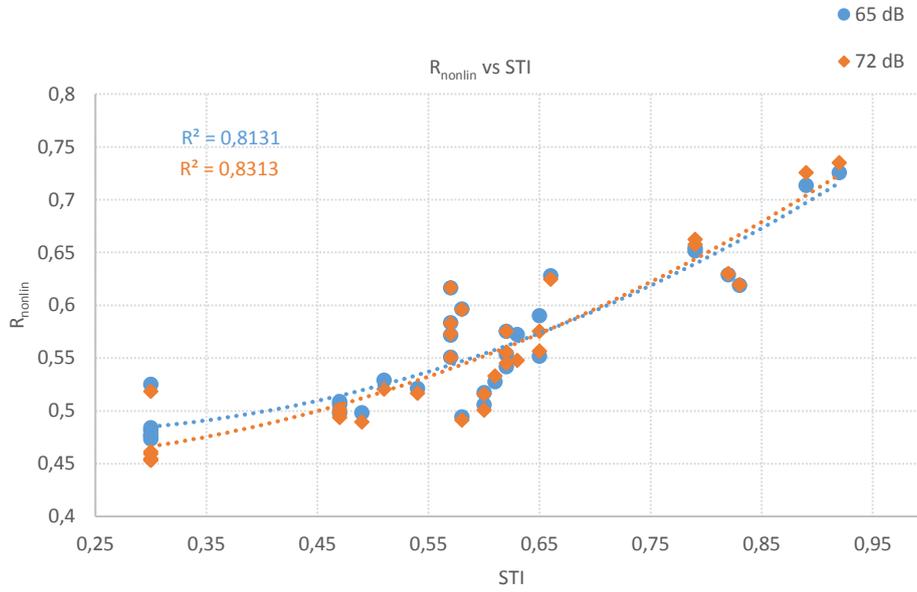
Se representan en las siguientes páginas los resultados obtenidos con las señales de voz, en primer lugar los relativos al párrafo fonéticamente equilibrado, desde la Gráfica 8.25 hasta Gráfica 8.32 para la voz masculina y femenina en relación con los parámetros EDT, D, STI y RASTI. Los resultados representados entre la Gráfica 8.33 y la 8.36 corresponden a los obtenidos empleando como señal de excitación la batería de 20 frases psicoacústicas de la Tabla 8.1. Se representan las correlaciones obtenidas respecto a los índices de transmisión STI y RASTI. Resultados adicionales con diferente nivel de excitación y relacionados con los tiempos de reverberación y la definición D pueden consultarse en el Anexo I. Finalmente, las gráficas 8.37 y 8.38 muestran la distribución estadística de los resultados obtenidos con la métrica empleando frases psicoacústicas como señal de excitación, estableciendo la comparación de resultados entre salas tanto con la voz masculina como con la femenina.



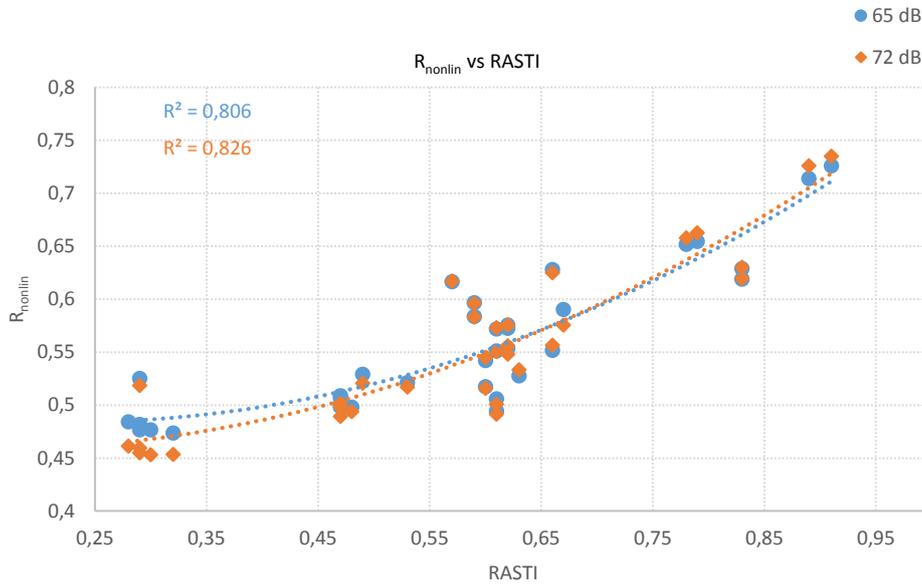
**Gráfica 8.25. Pares EDT - R<sub>nonlin</sub> para la señal de párrafo masculino**



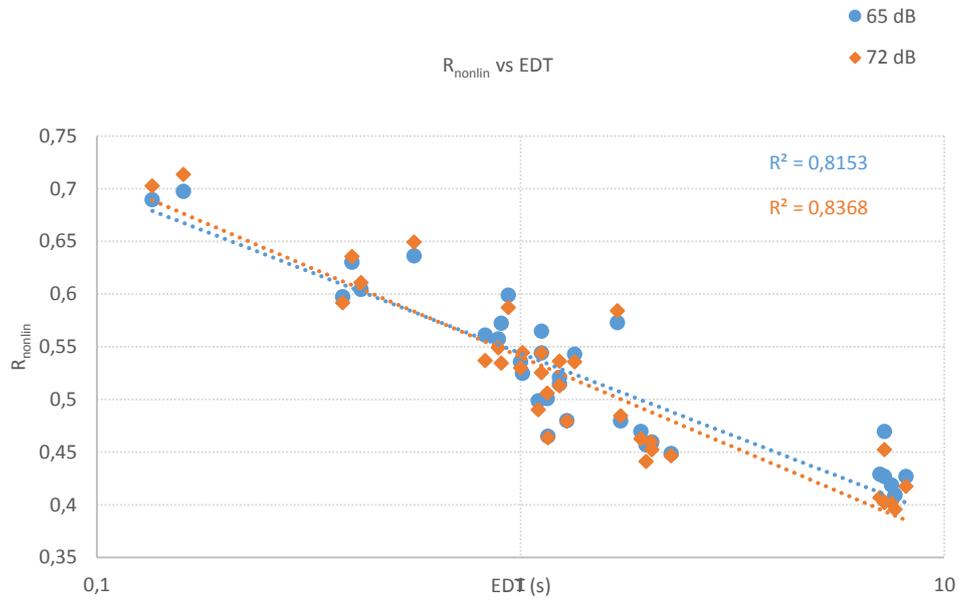
**Gráfica 8.26. Pares D - R<sub>nonlin</sub> para la señal de párrafo masculino**



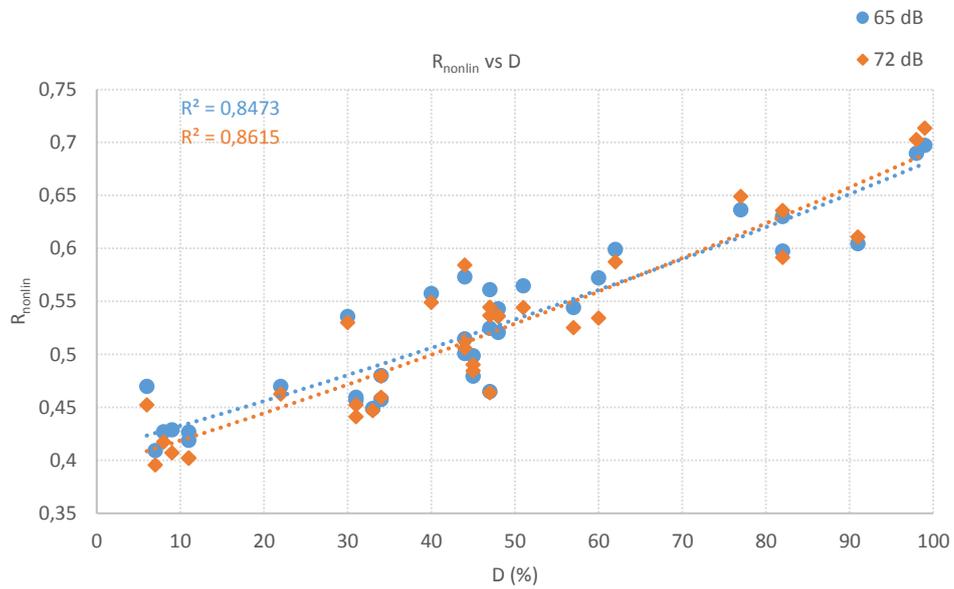
Gráfica 8.27. Pares STI - R<sub>nonlin</sub> para la señal de párrafo masculino



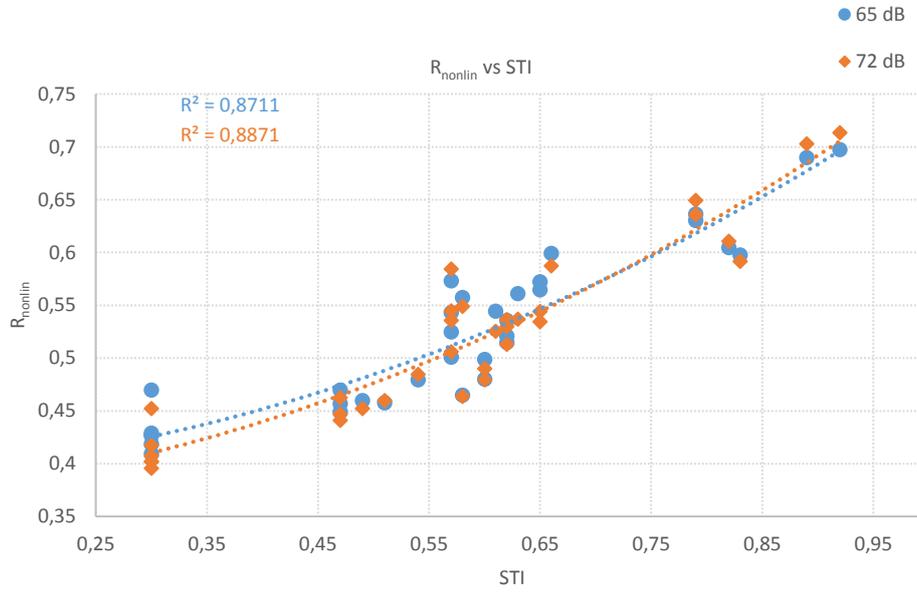
Gráfica 8.28. Pares RASTI - R<sub>nonlin</sub> para la señal de párrafo masculino



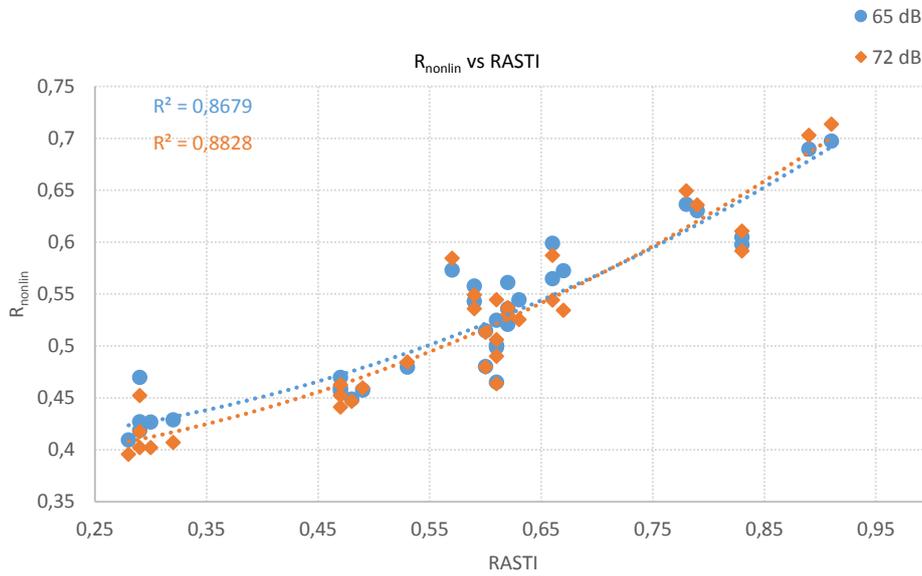
Gráfica 8.29. Pares EDT -  $R_{nonlin}$  para la señal de párrafo femenino



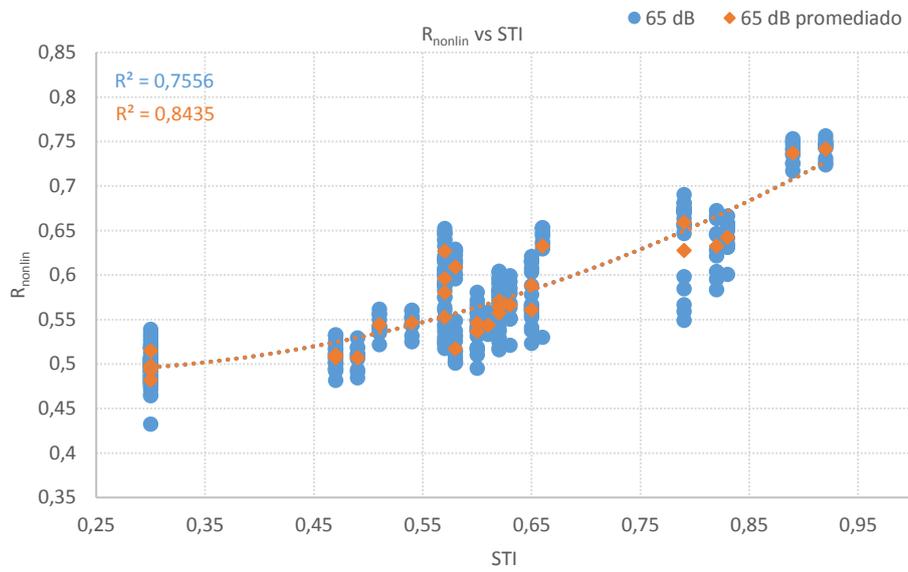
Gráfica 8.30. Pares D -  $R_{nonlin}$  para la señal de párrafo femenino



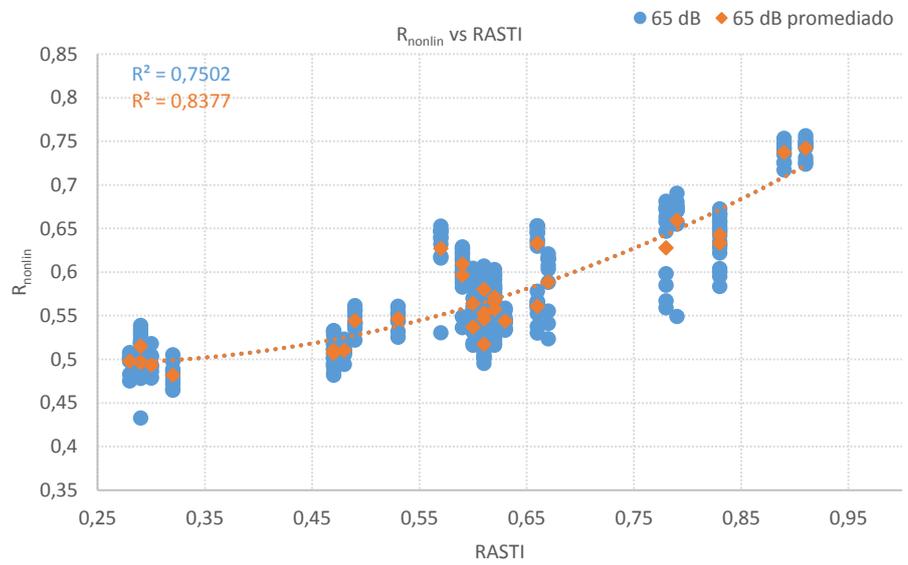
Gráfica 8.31. Pares STI - R<sub>nonlin</sub> para la señal de párrafo femenino



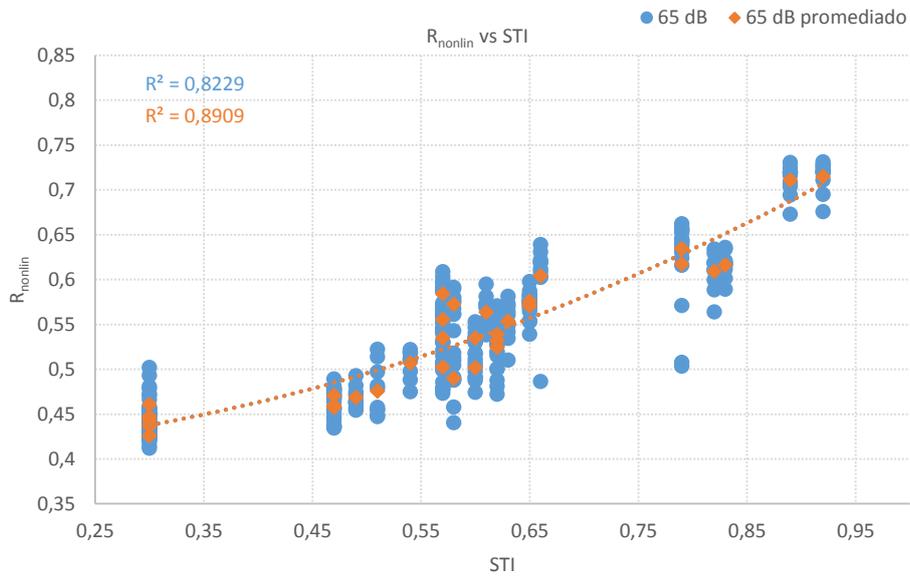
Gráfica 8.32. Pares RASTI - R<sub>nonlin</sub> para la señal de párrafo femenino



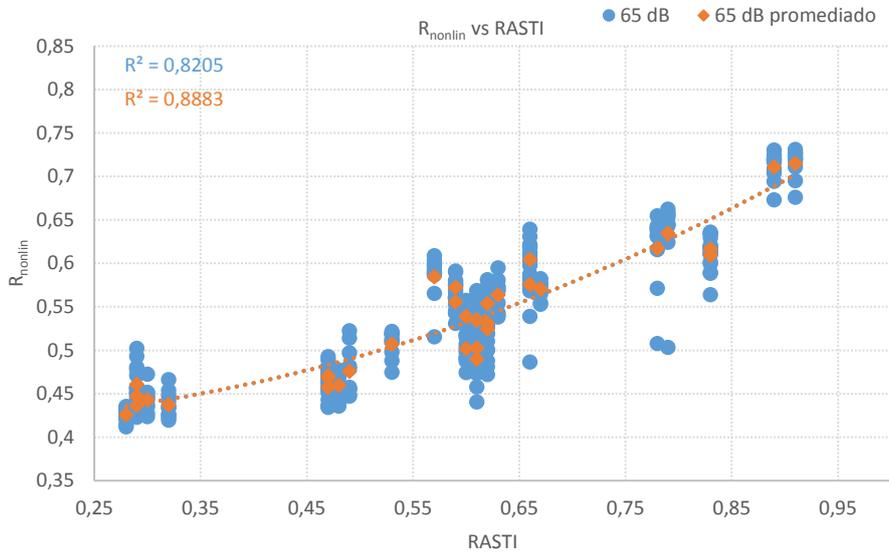
Gráfica 8.33. Pares STI - R<sub>nonlin</sub> frases psicoacústicas voz masculina



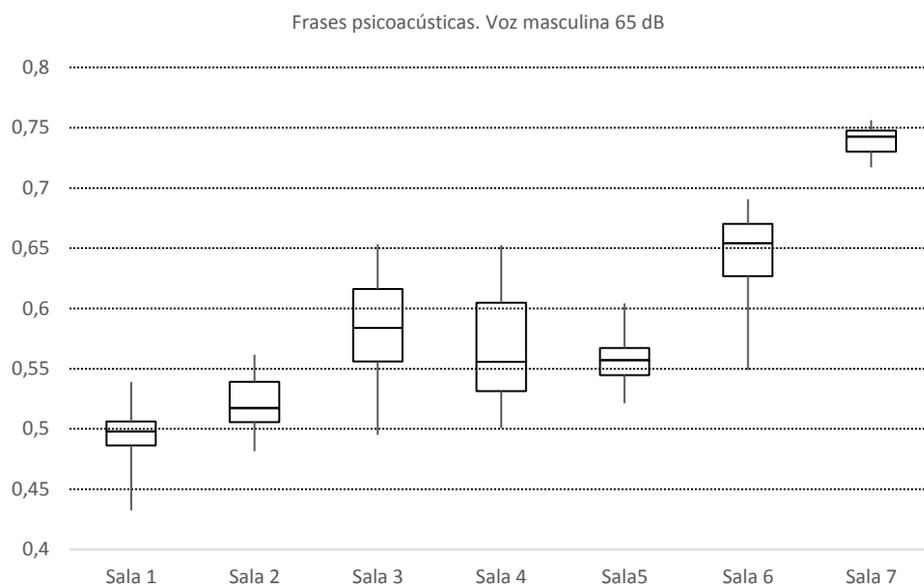
Gráfica 8.34. Pares RASTI - R<sub>nonlin</sub> frases psicoacústicas voz masculina



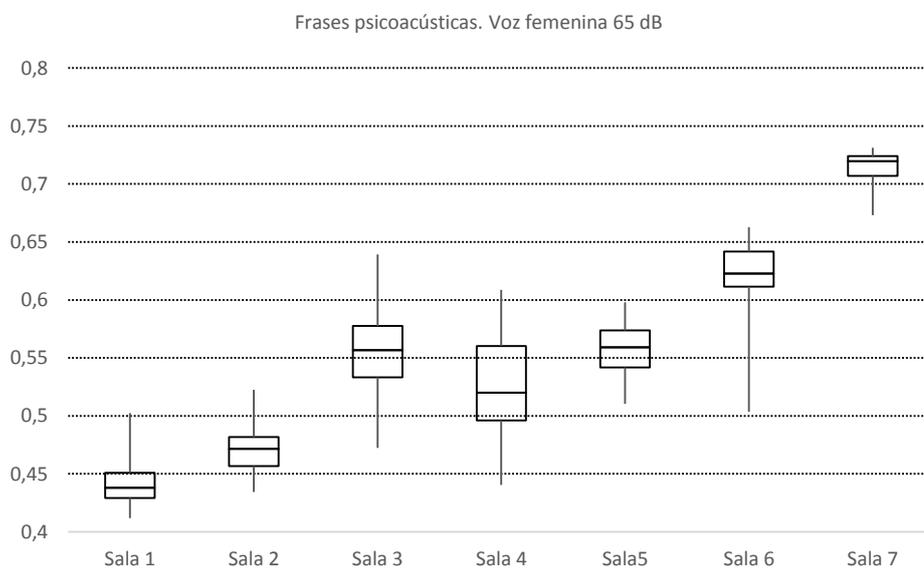
Gráfica 8.35. Pares STI - R<sub>nonlin</sub> frases psicoacústicas voz femenina



Gráfica 8.36. Pares RASTI - R<sub>nonlin</sub> frases psicoacústicas voz femenina



**Gráfica 8.37. Resultados frases psicoacústicas voz masculina por salas**



**Gráfica 8.38. Resultados frases psicoacústicas voz femenina por salas.**

#### *8.5.2.1 Discusión de resultados*

Las representaciones se han realizado de forma análoga al caso de las señales musicales, en forma de gráficas de dispersión de puntos formados por pares del valor obtenido por la métrica y el valor del parámetro especificado medido en cada punto. Los resultados obtenidos al procesar las señales de voz se han comparado con el tiempo de reverberación EDT, la definición D y los índices de transmisión STI y RASTI.

Comentaremos en primer lugar los resultados obtenidos con el párrafo de voz masculina. En el caso del párrafo masculino se han obtenido valores de  $R_{\text{nonlin}}$  en un intervalo de resultados comprendido entre 0,45 para el caso más desfavorable y cerca de 0,74 para el más favorable. Relacionándolo con el tiempo de reverberación EDT se produce una traslación entre este intervalo de valores y el rango de tiempos EDT obtenidos mediante medición directa, que varían entre 0,14 s y 7,6 s. De nuevo se puede establecer un ajuste de los valores a una función logarítmica equivalente a la de la ecuación 8.2, si bien el índice de correlación obtenido en el caso de la voz masculina es algo inferior al obtenido con las señales musicales (exceptuando el caso ya discutido de la señal  $S_2$ ). Así, si los índices de correlación respecto a las funciones de ajuste en cuanto a los tiempos de reverberación en las señales musicales estaban en torno a 0,85, para el párrafo de voz masculina desciende a 0,77 para el nivel más alto de excitación de la sala.

Al igual que en el caso de las señales musicales, las nubes de puntos más aisladas correspondientes a las salas con características más extremas son claramente distinguibles mientras que en las salas intermedias existe una mayor dispersión entre el parámetro objetivo y el resultado de la métrica. También, de la misma forma que en el caso de las señales musicales, existen variaciones de comportamiento en función del nivel de presión sonora entregado por la fuente de excitación de la sala. Esta variación es de nuevo más manifiesta en el caso de los recintos con parámetros más extremos. Se repite de nuevo el fenómeno correspondiente a la obtención de un valor de  $R_{\text{nonlin}}$  menor al incrementar el nivel de presión de la señal de test en el caso de las salas con más reverberación (sala 1 y 2). Esto también ocurría con las señales musicales, aunque en el caso de la señal con el párrafo masculino el efecto es más destacado.

Al incrementar el nivel de presión sonora se incrementa también la energía reverberada y en un recinto muy reverberante la degradación de la señal es mayor y consecuentemente es peor el resultado obtenido al comparar la señal registrada con la de referencia. Además de esta diferencia en la zona de mayor reverberación el efecto se invierte en la zona más anecoica: al excitar la sala con mayor nivel se obtiene una mayor correlación al comparar señales si el entorno es muy absorbente, como es el caso de la sala 7. A la vista de las gráficas este efecto se evidencia únicamente en los dos puntos de medición de la sala 7, con un EDT entre 0,14 y 0,16. Para valores de reverberación intermedios, entre 0,4 y 1,6 s, las diferencias de la métrica en función del nivel de excitación de la sala no son significativas.

Al comparar estos resultados respecto a la definición D obtenemos que en el caso de la señal de voz, especialmente la masculina, la dispersión de puntos se ajusta algo mejor a una función polinómica de segundo orden, que a una función lineal, como habíamos empleado en el caso de las señales musicales (ecuación 8.4). Esto ocurre fundamentalmente con los niveles de excitación más bajos y debido al efecto comentado en el párrafo anterior relativo a las variaciones de la métrica en función del nivel. Para niveles de excitación bajos obtendremos valores ligeramente más altos de  $R_{nonlin}$  en recintos muy reverberantes y ligeramente más bajos en recintos muy anecoicos, con lo cual el ajuste se alejará de una recta ideal en los extremos. Si incrementamos en nivel de presión, en el extremo reverberante la métrica responderá ofreciendo un valor inferior, mientras que en el extremo anecoico se obtendrá un valor más elevado, por lo que el ajuste tenderá a ser más lineal.

En el Anexo II se muestran en una tabla todas las funciones de correlación obtenidas en este experimento. Para el caso de las señales de voz se recopilan las funciones resultantes realizando un ajuste lineal y un ajuste polinómico para los dos niveles de excitación. Puede allí corroborarse que los índices de correlación obtenidos realizando un ajuste lineal son más altos para los niveles de presión sonora de la señal de excitación más elevados. En el rango de definición D comprendido entre 35 – 70 % las diferencias respecto al nivel de excitación no son relevantes.

Otros parámetros objetivos de interés a la hora de relacionarlos con la métrica implementada para la evaluación de señales de voz son el índice de transmisión del habla STI y su versión reducida aplicada a salas, RASTI. La Gráfica 8.27 y la Gráfica 8.28 establecen esta relación para la señal del párrafo de voz masculina. En el caso del índice STI se ha trabajado en un rango de valores experimentales comprendido en 0,3 y 0,92, mientras que en el caso del RASTI este rango ha variado entre 0,28 y 0,91. Como se ha explicado en la sección 4.7.3, la medición de RASTI en el ámbito de aplicación para el que fue diseñado ofrece resultados muy ajustados a los obtenidos por STI. Esto puede corroborarse a la vista de las dos gráficas citadas comparando en los puntos la coordenada correspondiente al parámetro objetivo (STI / RASTI). Por otra parte, si retomamos la Figura 4.13 constataremos que el intervalo estudiado abarca prácticamente todo el rango cualitativo aplicable a la inteligibilidad del habla. Tanto STI como RASTI se ajustan a la métrica implementada relacionándola con una función polinómica de segundo orden, de forma que se tiene

$$STI = a \cdot (R_{nonlin})^2 + b \cdot R_{nonlin} + c \quad (8.5)$$

$$RASTI = a \cdot (R_{nonlin})^2 + b \cdot R_{nonlin} + c \quad (8.6)$$

Obteniendo índices de correlación entre 0,81 y 0,83 según el nivel de excitación de la sala. En este sentido, en función del nivel de excitación, las variaciones descritas anteriormente en las salas con características más extremas acentúan el carácter polinómico del ajuste si bien no ocurre como en el caso de la definición, en la que la diferencia con un ajuste lineal era escasa principalmente con el nivel más alto.

Veamos ahora los resultados obtenidos con la señal del párrafo hablado con voz femenina. Para esta señal se han obtenido resultados de  $R_{\text{nonlin}}$  en una horquilla comprendida entre 0,40 y 0,71, es decir, un rango de variación de similar magnitud al caso del párrafo de voz masculina – en torno a 0,3 - pero con valores absolutos de  $R_{\text{nonlin}}$  alrededor de 0,05 puntos inferiores de forma promediada. El valor absoluto de la métrica por sí sólo no significa nada, si no está vinculado a una misma señal de referencia. Que los valores obtenidos con la voz femenina sean más bajos que los obtenidos con la masculina puede justificarse por el propio contenido espectral de ambas señales (ver Figura 8.4). De hecho, los ajustes de las dispersiones obtenidas con el párrafo de voz femenina obtienen mejores índices de correlación que con la voz masculina. Así, con respecto al tiempo de reverberación EDT se obtiene un índice de correlación alrededor del 7% superior que en el caso de la voz masculina. Esto ocurre igualmente en el caso del ajuste realizado respecto a la definición, siendo este último bastante más lineal que el relativo a la voz masculina (comparar Gráfica 8.26 y Gráfica 8.30, ver Anexo II).

Respecto a las variaciones experimentadas en relación a la diferencia de niveles, se repite igualmente el efecto en los extremos de las gráficas, al igual que en todos los demás casos. Hay una particularidad relativa a la diferencia de resultados con respecto al nivel de excitación en el caso de esta señal. Al observar los resultados representados desde la Gráfica 8.30 a la Gráfica 8.32, a diferencia de con la voz masculina, se constata que en las zonas centrales de las gráficas existen puntos cuyos valores  $R_{\text{nonlin}}$  son claramente superiores para la señal de excitación de menor nivel. Normalmente, en ese rango de las curvas, las diferencias de valores en función del nivel no son tan significativas. A excepción de uno de esos puntos, todos los demás ocurren en la sala 5, cuyos resultados ya hemos comentado en varias ocasiones precedentes. Parece ser que para esta señal de excitación se reproduce, únicamente en esta sala, el efecto que tiene lugar con recintos mucho más reverberantes, lo cual, a priori, no es de esperar que ocurra en una sala de estas dimensiones.

Una posible justificación de este resultado lo podemos encontrar en los resultados de RT y EDT obtenidos en este recinto. Remitimos al lector a comparar los valores de estos parámetros para la octava de 2000 Hz entre la sala 5 y la sala 3 – ver para ello el Anexo III - . Mientras que la sala 3 tiene un volumen más de 10 veces superior al de la 5, los EDT de ambas salas en la octava de 2000 Hz son del mismo orden de magnitud<sup>26</sup>. Esta característica tonal de la sala 5 puede estar influyendo en el resultado comentado, especialmente en el caso de la voz femenina, cuyo espectro está más desplazado hacia la alta frecuencia que la masculina (Figura 8.4).

Respecto de los ajustes obtenidos con esta señal de test relacionados con los índices STI y RASTI lo más destacable que cabe mencionar son los altos índices de correlación obtenidos al ajustarlos a una función polinómica. A pesar de los muy similares resultados

---

<sup>26</sup> En la sala 3, los valores EDT obtenidos están en un intervalo entre 0,71 y 0,99 s. En la sala 5 el intervalo es entre 0,71 y 0,94 s. Todo ello en la octava de 2000 Hz.

obtenidos con STI y con RASTI, tanto para el caso de la voz masculina como para la femenina el ajuste siempre es ligeramente más preciso en relación con STI.

Finalmente comentaremos los resultados obtenidos utilizando como señales de excitación para palabra la batería de 20 frases psicoacústicas que se muestra en la Tabla 8.1. En todos los puntos de medición se han registrado las señales correspondientes a estas veinte frases tanto para voz masculina como femenina y con dos niveles de excitación. Las gráficas relativas a estas señales que se han mostrado en las páginas precedentes (gráficas de la 8.33 a 8.38) corresponden al nivel de excitación más bajo ( $L_{eq}=65$  dB) y son las que establecen la relación entre la métrica  $R_{nonlin}$  y los índices STI y RASTI. Las gráficas 8.33 y 8.34 son las relativas a la voz masculina, mientras que las 8.35 y 8.36 se refieren a la voz femenina. Finalmente, la Gráfica 8.37 y la Gráfica 8.38 representan, en forma de diagrama de cajas y bigotes el intervalo de resultados y la distribución de éstos por cuartiles obtenidos en cada sala. Esta representación es especialmente útil en el caso de las frases psicoacústicas dado que en las salas en las que se han utilizado 6 puntos de medición (salas 1, 2, 3 y 4) se han obtenido series con 120 valores de la métrica, 80 valores en las salas con cuatro puntos de medición (salas 5 y 6) y 40 en la sala con dos puntos (sala 7).

La gran cantidad de datos obtenidos con las frases psicoacústicas nos ha llevado a optar por representar los resultados básicos relativos a la relación con STI y RASTI y a los valores absolutos comparados entre salas, como se ha hecho en las páginas precedentes, mientras que en el Anexo I se incluyen el resto de resultados obtenidos, como los ajustes relacionados con los índices de transmisión de la palabra con mayor nivel de la señal de excitación, así como los ajustes relativos al tiempo de reverberación EDT y a la definición D con ambos niveles de excitación, tanto para la voz masculina como para la femenina.

Remitimos al lector en primer lugar a las gráficas 8.33 y 8.34. Se representan los valores de  $R_{nonlin}$  obtenidos para las frases psicoacústicas con voz masculina con todas las frases empleadas con un nivel de excitación de referencia de  $L_{eq}=65$  dB en el punto de medición más cercano a la fuente de señal. Como para cada uno de los 34 puntos de medición se han obtenido 20 resultados de la métrica empleada, éstos se representan como una única serie – serie de puntos azules - relacionada con el parámetro objetivo (en estos casos, STI y RASTI) medido en ese punto. Así, para un mismo valor de STI y RASTI se tendrán como mínimo 20 valores de  $R_{nonlin}$ . En el caso de que varios puntos de medición obtengan el mismo valor STI o RASTI se multiplicará el número de resultados de  $R_{nonlin}$  asociados a ese valor objetivo. Esto ocurre, por ejemplo, en la sala 1, en la que los 6 puntos de medición obtienen un  $STI=0,3$ , con lo cual, los 120 valores de  $R_{nonlin}$  obtenidos en esa sala quedan asociados a ese valor STI. La cercanía entre los valores hace que los puntos representados se solapen entre ellos pareciendo que está representada una menor cantidad de puntos, pero, siguiendo el mismo ejemplo anterior, en la Gráfica 8.33, a lo largo del eje  $STI=0,3$  hay representados (en azul) 120 valores de  $R_{nonlin}$ . Esto permite ver también la dispersión de valores obtenidos con la métrica implementada para un mismo

punto o en diferentes puntos en los que se obtiene un mismo valor de algún parámetro de los denominados objetivos.

Por otra parte, dado que a nivel de distribución del contenido fonético de las frases psicoacústicas, lo que está equilibrado fonéticamente es el conjunto de las frases, no cada una de ellas, hemos estimado conveniente añadir a estas representaciones el promedio de los valores  $R_{\text{nonlin}}$  obtenidos en cada punto de registro, de forma que a cada punto de medición se le asocia también un único valor de la métrica implementada, obteniendo así un conjunto de 34 pares (STI,  $R_{\text{nonlin}}$ ) o (RASTI,  $R_{\text{nonlin}}$ ) que son los representados en color anaranjado en las citadas gráficas.

El rango de valores de la métrica obtenido para el conjunto de frases psicoacústicas en todos los registros realizados, considerando los dos niveles de excitación oscila entre unos valores absolutos delimitados entre 0,42 y 0,76, estableciéndose un rango de variación aproximado de 0,34. En términos absolutos, los valores obtenidos por las frases individualmente suelen ser entre 0,01 y 0,03 puntos superiores a los obtenidos por el párrafo. En este extremo, dado que el contenido espectral es muy similar en el caso de las frases que en el párrafo, dado que se trata de la misma voz, la explicación a este menor nivel absoluto del valor de la métrica en el párrafo la podemos encontrar en el hecho de que el éste incluye un determinado número de pausas y silencios, al tratarse de un texto escrito. Puesto que las correlaciones cruzadas de la métrica se establecen en *frames* de 30 ms, es de esperar que un número no despreciable de ellas se produzcan en ventanas temporales que no incluyan contenido espectral de la voz (especialmente las pausas y silencios) con lo cual las correlaciones se establecerán entre el ruido de fondo de la grabación anecoica (a efectos prácticos despreciable) y el ruido de fondo en el punto de registro de la sala. Son de esperar correlaciones cruzadas muy bajas en estas ventanas temporales, que influirán en el promediado del valor absoluto final.

Evidentemente, la característica antes citada relativa a la cantidad de valores  $R_{\text{nonlin}}$  asociados a un valor STI o RASTI determina la particularidad de estas gráficas, que distribuyen de forma axial los valores obtenidos utilizando la métrica perceptual. Aún con esta característica dispersiva, se obtienen índices de correlación respecto de los parámetros STI y RASTI aceptables, en torno a 0,75. Esto viene a significar, ejemplificando un caso extremo, que en el sistema implementado para la realización del experimento, podría determinarse el parámetro STI o RASTI con un acierto del 75% utilizando una única frase de las mostradas en la Tabla 8.1. Lógicamente, al utilizar el promediado de los resultados de las frases en cada punto, el índice de correlación con respecto a STI y RASTI mejora ostensiblemente, llegándose a valores alrededor de 0,84. De igual manera que sucedía con los párrafos, con las frases también se obtienen índices de correlación ligeramente superiores en las correlaciones con STI, al compararlas con las RASTI.

La función que mejor se ajusta en ambos casos a las distribuciones de puntos es la representada en las ecuaciones 8.5 y 8.6, teniendo esto también su justificación en la

fenomenología que tiene lugar en los extremos, tanto en la parte más reverberante como en la más anecoica. Se constata de nuevo, un descenso de los valores arrojados por la métrica en entornos muy reverberantes al incrementar el nivel de excitación de la sala – aunque en este caso más pronunciado únicamente en la sala 1 – así como un tenue incremento del valor absoluto de  $R_{\text{nonlin}}$  en el caso de la sala 7 también al incrementar el nivel de presión sonora entregado por la fuente. Esto produce una ligera mejor correlación de datos en el caso de la excitación de las salas con mayor nivel – ver gráficas I.1, I.2 en el Anexo I, obteniéndose una mejora del índice de correlación alrededor del 1%.

Respecto a los resultados con la señal de test utilizando frases psicoacústicas con voz femenina, éstos se representan en las gráficas 8.35 y 8.36 en relación a los índices STI y RASTI. Los resultados con mayor nivel de excitación y en relación con otros parámetros pueden encontrarse también en el Anexo I. Puede observarse que la distribución es muy similar a la obtenida en el caso de la voz masculina, si bien de nuevo, en términos de valor absoluto, los valores entregados por la métrica son ligeramente inferiores comparados con la de la voz masculina. Así, el rango de variación de valores está delimitado entre 0,39 y 0,74 considerando los dos niveles de excitación. Generalmente, en todos los casos, el valor mínimo del rango siempre se obtiene en la sala 1 con el nivel de excitación más alto. De nuevo se obtienen índices de correlación considerablemente mejores en el caso de la voz femenina, de manera que éstos llegan a ser de 0,82 ya la nube de puntos formada por todas las frases psicoacústicas por separado (un 7% mejor que con la voz masculina). Al promediar el resultado de la métrica en cada punto se llega a índices de correlación de 0,89 para el nivel de excitación  $L_{\text{eq}}=65$  dB y de 0,9 para el nivel de excitación  $L_{\text{eq}}=72$  dB (ver Gráfica I.7 e I.8 en el Anexo I).

Entrando ahora en una valoración de la métrica para su utilidad en la evaluación cualitativa de canales de transmisión para voz, sin relacionarla con otros parámetros de uso habitual, los datos obtenidos con las frases psicoacústicas en los diferentes recintos que han formado parte del experimento se representan en la Gráfica 8.37 para la voz masculina y en la Gráfica 8.38 para la femenina. En estas representaciones está contenida la distribución de los 120, 80 o 40 resultados de cada sala, según se ha explicado más arriba. En estas gráficas podemos visualizar las diferencias entre recintos, así como la distribución de resultados para cada uno de éstos.

Vemos en primer lugar que existe una graduación por recintos que se repite aproximadamente en todos los casos. En el extremo inferior son claramente diferenciables las dos salas más reverberantes, cuya distribución de resultados se ajusta bastante a un valor central ofreciendo menor dispersión, lo cual parece bastante coherente si consideramos que entornos muy reverberantes tienden a tener un campo acústico difuso y tendiente a la homogeneidad. En la parte derecha de las gráficas, relativas a las salas 6 y 7 estas también se diferencian claramente, siendo esta última la que menor dispersión de valores ofrece en todos los casos, lo cual es lógico teniendo en cuenta sus reducidas dimensiones. En la parte central, en lo relativo a las dos salas más

comparables entre sí en cuanto a sus dimensiones, la sala 3 obtiene siempre mejores resultados que la 4, observándose en estos dos casos una mayor dispersión de los valores obtenidos. Las dimensiones y características de estos dos recintos justifican la mayor variabilidad de los resultados obtenidos en cada punto, también más distantes entre sí. Esto ocurre igualmente con las técnicas de medición de parámetros objetivos clásicos (puede consultarse el Anexo III en este sentido).

De nuevo, el único recinto que presenta problemas de graduación de resultados y variabilidad de éstos en función de la señal de excitación y del nivel de ésta es la sala 5. A juzgar por los resultados del índice STI y RASTI medidos en esta sala, ésta presenta algunas deficiencias ya que a priori son esperables mejores índices de transmisión en un recinto tan reducido. Ya se ha comentado que son valores idénticos a los de la sala 3, que tiene un volumen más de 10 veces superior. Con lo cual, los resultados aportados por la métrica perceptual implementada reflejan también esta anomalía, y esta se refleja además en la variación de resultados en función del nivel de excitación. En el caso de la voz masculina se produce en esta sala un ligero descenso de la distribución absoluta de resultados al incrementar el nivel de señal y en el caso de la voz femenina, este descenso es más remarcado (ver Gráfica I.14) lo cual, a priori, sólo es esperable en un entorno muy reverberante como la sala 1.



## 9 Conclusiones y futuras líneas de trabajo

Empezábamos esta memoria haciendo alusión a la dificultad para enmarcar el término calidad en una teoría general dentro del ámbito de la acústica. Hemos constatado, a lo largo de estas páginas, el carácter multidisciplinar implícito en cualquier aproximación que persiga una evaluación cualitativa de un sistema. Para la experimentación con la métrica, se ha recurrido a los fundamentos teóricos que la sustentan para proceder a su desarrollo e implementación. Éstos conjugan tanto conocimientos de carácter fisiológico y de comportamiento del sistema auditivo humano obtenidos a lo largo del último siglo por la psicoacústica, como técnicas de procesado de señal necesarias para el modelado matemático de ese comportamiento. Por otra parte, dado que el ámbito de aplicación de la métrica se realiza sobre sistemas no evaluados anteriormente con este tipo de procedimientos se hace necesaria una revisión de los métodos tradicionalmente usados para la evaluación de estos sistemas, con tal de tener referencias sobre las que comparar los resultados obtenidos. Por ello, se ha realizado un estudio de las principales métricas utilizadas en los dos tipos de canales de transmisión utilizados.

Sin haber entrado en valoraciones cualitativas en sentido estricto, las cuales se realizan mediante experimentación psicoacústica, se ha realizado la implementación de una métrica basada en consideraciones de carácter perceptivo con el objetivo de evaluar su viabilidad en canales de transmisión de naturaleza distinta a los inicialmente previstos para este tipo de métricas.

Aun no tratándose de una evaluación cualitativa como se considera en sentido tradicional, la aproximación con un modelo perceptual puede proporcionarnos una evaluación a priori relacionada en mayor medida con aspectos que pueden resultar relevantes en una valoración cualitativa: los elementos perceptivos implícitos en esta. La principal limitación de los modelos perceptivos estriba en que aun realizando una representación de la imagen auditiva interna, evidentemente, las funciones cognitivas quedan fuera del modelo y el rol desempeñado por estas últimas en determinadas ocasiones puede resultar no despreciable.

Siendo conscientes de estas limitaciones, la métrica elegida, en su concepción original (Moore, Tan, Zacharov, & Ville-Veikko, 2004) se propuso como un modelo de predicción de calidad subjetiva, obteniéndose un ajuste basado en distorsiones generadas artificialmente. Este ajuste - Figura 6.14 - es válido únicamente para el tipo de distorsiones utilizado en el trabajo referenciado. Lo que es importante destacar aquí, es que la métrica está concebida para poder llegar a ser utilizada como una herramienta de predicción cualitativa, en última instancia capaz de sustituir a las propias encuestas de valoración subjetiva. Para ello sería necesario establecer una correlación entre los resultados obtenidos al implementar la métrica en los canales de transmisión utilizados en este trabajo y las valoraciones cualitativas realizadas mediante una amplia experimentación psicoacústica. El objeto experimental de este trabajo se ha situado fuera de este ámbito, centrándonos en la implementación del algoritmo por un lado y en la aplicación de éste a señales propagadas en los canales de transmisión estudiados, con la finalidad de evaluar la viabilidad del método trabajando con dispositivos y recintos reales.

Queda pues para futuros trabajos la experimentación psicoacústica necesaria para establecer las correlaciones cualitativas con los órdenes de magnitud obtenidos por la métrica en los dos tipos de canales estudiados, lo cual permitiría obtener unas curvas, que, como las de la Figura 6.14, posibilitaran la predicción cualitativa de una señal registrada previamente según los procedimientos experimentales realizados en este trabajo. El uso práctico de  $R_{\text{nonlin}}$  como herramienta de predicción requeriría de un conjunto de valores de ajuste de curvas predeterminado para una variedad de tipos de señales, de forma que no se tuviera que recurrir al ajuste con datos procedentes de un test de evaluación en todas las ocasiones.

Resumiremos a continuación las principales conclusiones obtenidas a partir de la experimentación con la métrica perceptual llevada a cabo en este trabajo, en primer lugar en lo referente a la propia implementación de la métrica. En segundo lugar, se

enumerarán las conclusiones obtenidas en la evaluación realizada sobre transductores y finalmente en la evaluación sobre salas y recintos. Finalmente se establecerán unas futuras líneas de trabajo y propuestas de mejora sobre algunos de los procedimientos llevados a cabo.

### **9.1 Implementación de la métrica**

La implementación del algoritmo no ha estado exenta de dificultad, debido por una parte a su descripción sintética en el trabajo de referencia, lo cual ha hecho necesario un trabajo de profundización de las técnicas y procedimientos disponibles para su desarrollo que se ha resumido en el capítulo 6. De su uso práctico en el procedimiento experimental podemos concluir lo siguiente:

- a) Uno de los aspectos del procesado que resulta crucial es el alineado previo de señales. Dado que las comparaciones se establecen correlando una ventana temporal de 30 ms con otra de 90 ms, sin un correcto alineado previo de las señales la comparación entre éstas no tendría coherencia. Esto es especialmente importante en el caso de señales registradas en recintos, donde los tiempos de propagación y de llegada del sonido directo son mayores, en comparación con los realizados en transductores – en la misma garganta del motor de compresión o muy cerca de la superficie de radiación -. Estos retardos deben ajustarse antes de iniciar el procesado
- b) Dado que el análisis de las señales se realiza en el dominio temporal para las 40 bandas de frecuencia que representan el filtrado del oído interno, el algoritmo implementado utiliza un tiempo de procesado considerable. En el apartado de futuras líneas se resumen algunas propuestas de optimización del algoritmo.
- c) La presencia de silencios en las señales de test tenderá a ofrecer valores absolutos de la métrica más bajos que separando los eventos sonoros en señales diferentes. Esto no es relevante si la métrica se utiliza para comparar sistemas, en cuyo caso lo importante son las diferencias relativas, pero la presencia de silencios debe ajustarse como máximo a las propias de las pausas en el habla o musicales.

### **9.2 Resultados en transductores**

Se han realizado medidas en dos tipos de transductores diferenciados. Enumeraremos en primer lugar las conclusiones comunes a ambos tipos de dispositivos.

- a) Se ha observado una significativa diferencia de comportamiento de la métrica en función de las señales de test utilizadas, especialmente comparando entre señales de test estacionarias o complejas, llegando a la conclusión de que las señales estacionarias son inadecuadas para su uso en esta métrica. La explicación a este resultado puede radicar en la poca o casi nula información que contienen estas señales en un gran número de filtros del banco de filtros auditivos, como se ha explicado en el capítulo 7.

- b) Por otra parte, se ha constatado que los resultados obtenidos con señales complejas muestran un comportamiento coherente con lo que cabe esperar de la métrica, valores de  $R_{\text{nonlin}}$  ajustados a una dispersión limitada y cercanos a la unidad. En el caso de los transductores es esperable este comportamiento dada la distancia a la que se produce el registro de la señal y que éste se produce en un entorno anecoico, con lo que no hay degradación de señal introducida por presencia de reverberación.
- c) Se aprecian diferencias significativas en el uso de señales complejas según la estructura dinámica de la señal utilizada, de forma que señales con más componentes transitorios obtienen mejores resultados. Esto sugiere que los umbrales de percepción de la distorsión pueden ser más elevados en caso de utilizar señales con estructuras temporales en las que predominan los transitorios. Este resultado coincide con el obtenido por en (Schmitt, 1995) mediante métodos conceptualmente diferentes sin considerar supuestos psicoacústicos.
- d) En lo que respecta a la posibilidad de uso de la métrica para la comparación entre dispositivos, la repetición del mismo patrón de resultados en todas las pruebas lleva a pensar en la viabilidad de la métrica en la comparación de dispositivos reales, si bien hay que tener en cuenta el papel determinante de la señal utilizada en la evaluación, lo cual sugiere, en principio, el uso de diversos tipos de señales para comparar dispositivos.
- e) Dadas las diferencias obtenidas según la estructura de la señal de test se ha planteado la posible validez de un parámetro como descriptor de calidad, independientemente de que la métrica también sea utilizada en su vertiente predictiva. En este sentido, utilizando dos señales con estructura dinámica bien diferenciada se ha visto que la distancia entre los valores de  $R_{\text{nonlin}}$  para ambas señales obtenidos en cada transductor podría funcionar como un indicador de calidad de éste. Este extremo tendría que ser corroborado más ampliamente de forma experimental.
- f) El comportamiento de la distorsión, o más bien de la percepción de ésta en función del nivel de entrada al sistema está también vinculado a las características de la señal que se esté evaluando.
- g) Al contrario que las métricas clásicas o de otras basadas en supuestos perceptuales como la GedLee, que presupone unas características intrínsecas al dispositivo independientes de las señales que éste procese, en esta métrica, la valoración está fuertemente vinculada con la señal de test. Esto, a priori, puede parecer un inconveniente a la hora de comparar objetivamente entre dispositivos, pero al parecer, nuestra forma de percibir la distorsión también resulta estar vinculada al tipo y características de la señal en que ésta se presenta. Esta particularidad hace esta métrica especialmente interesante para estudios previos de aplicaciones concretas, en las que se conozcan las señales o los tipos de señales que van predominar en la futura aplicación.

Las siguientes conclusiones se refieren a diferencias de resultados observadas entre los dos tipos de transductores o a comportamientos observados en una de las tipologías de transductores utilizadas en el experimento.

- a) Los transductores de radiación directa obtienen valores absolutos superiores a los de radiación indirecta, lo cual es coherente con el grado de distorsión que se les presupone a los diferentes tipos de dispositivos.
- b) La degradación de la señal (en relación con el incremento del nivel de presión sonora) medida en los transductores de radiación directa para el caso de las señales estacionarias es menos acusada que la obtenida en los motores de compresión. También es cierto que los niveles de presión sonora generados por los motores de compresión son extremadamente altos y, por tanto, más propicios a la aparición de no linealidades, que quedan reflejadas en los resultados.
- c) De los resultados obtenidos con las señales complejas puede deducirse que los motores de compresión, por sí solos, es decir, sin estar acoplados a una bocina son capaces de generar distorsión que sea perceptible.

### **9.3 Resultados en recintos**

Se enumeran en este apartado las conclusiones obtenidas de la aplicación experimental en recintos.

- a) La introducción de la reverberación como agente de degradación de la señal procesada produce un descenso significativo de los valores absolutos ofrecidos por la métrica. Si en el caso de las distorsiones artificiales, estudiadas en (Moore, Tan, Zacharov, & Ville-Veikko, 2004) el rango de uso de la métrica oscilaba entre 0,85-1, hemos visto que este rango es similar en transductores. Con la introducción del efecto sala, los valores absolutos de la métrica se obtienen en unos rangos entre 0,4 y 0,8 para señales complejas tanto de voz como musicales.
- b) El resultado anterior hace suponer que a nivel perceptual, la degradación introducida por el efecto sala puede enmascarar efectos de distorsión en transductores que sí pueden ser detectados en un entorno anecoico. Por otra parte, en la consideración de un sistema formado por el conjunto transductor más sala, en condiciones normales de funcionamiento del transductor es de esperar que el efecto de la sala sea predominante en el resultado de la métrica. Este es un resultado a tener en cuenta en caso de realizar evaluaciones de salas con sistemas de refuerzo sonoro.
- c) En el caso de las salas no se han utilizado señales deterministas como fuente de excitación sino señales musicales complejas, con diferentes tipos de contenido espectral y estructura temporal. Se ha observado una diferencia de

comportamiento de la métrica en función de ambas características descriptivas de las señales.

- d) Para señales con un mismo o similar ancho de banda, como ocurre con  $S_3$  y  $S_4$  se obtendrán valores absolutos más altos para aquellas con una estructura temporal más estacionaria. Así, se produce una fuerte degradación de los transitorios muy pronunciados proporcional a la reverberación del recinto. Este resultado es inverso al obtenido con los transductores, dada la ausencia de campo reverberante en las señales registradas con aquéllos.
- e) El contenido espectral, y más concretamente, la densidad espectral de potencia de la señal puede invalidar los resultados obtenidos, si ésta no se ajusta a la respuesta del conjunto del sistema, como ha ocurrido con la señal de test  $S_2$ . Se necesita un nivel mínimo de excitación del recinto en las bandas de frecuencia más significativas del rango audible. Por ello, es recomendable realizar la medida de presión sonora generada por la señal de test en dB(A).
- f) La respuesta en términos de valor absoluto de la métrica en función del nivel de excitación del recinto –con los niveles SPL empleados en este experimento- es sólo significativa en entornos muy reverberantes, en los que un incremento de la presión sonora se traducirá en un menor valor absoluto de la métrica.
- g) También se observa un muy ligero incremento de los valores arrojados por la métrica al incrementar el nivel de excitación para cualquier tiempo de reverberación o relación energética al utilizar señales con estructuras temporales cuasi estacionarias, como ocurre exclusivamente con la señal  $S_4$ .
- h) Con las señales musicales complejas utilizadas se han obtenido funciones de ajuste con parámetros clásicos de evaluación de salas, obteniéndose índices de correlación entre la métrica implementada y estos parámetros entre 0,81 y 0,89<sup>27</sup>. La función de ajuste de la métrica con los parámetros clásicos es logarítmica para los tiempos de reverberación RT y EDT, polinómica para la claridad  $C_{80}$  y lineal para la definición D.
- i) A la vista de los resultados mostrados en representaciones como las de las Gráfica 8.21, Gráfica 8.23 y Gráfica 8.24, se confirma la posibilidad de uso de la métrica para la comparación cualitativa entre recintos, obteniéndose patrones similares de graduación entre recintos que pueden presentar ligeras variaciones en función de la señal de test utilizada. La representación de los resultados obtenidos mediante diagramas de cajas resulta conveniente a la hora de establecer comparaciones entre sistemas: en primer lugar da cuenta de las diferencias cualitativas entre los sistemas, en segundo lugar no se limita a asignar un único valor a un sistema sino que muestra el rango de variación dentro del mismo, lo cual puede ser un marcador cualitativo en sí mismo o de comparación de comportamiento entre recintos.
- j) Los resultados obtenidos con las señales de voz son inferiores, en términos de valor absoluto, al utilizar la señal del párrafo escrito, respecto a las frases

---

<sup>27</sup> Exceptuando el ya discutido caso de la señal  $S_2$ .

psicoacústicas. Esto se debe a la presencia de las pausas propias del lenguaje, fragmentos en los cuales la baja correlación entre señales obtenida contribuye a un descenso del valor absoluto establecido en el promediado. En términos relativos, especialmente en lo referente a comparación entre recintos, los resultados obtenidos con ambas señales de test son equiparables.

- k) Una conclusión similar a la anterior puede obtenerse al comparar los resultados obtenidos con la voz masculina y la femenina. Los valores absolutos son superiores en el caso de la voz masculina, resultado que podría justificarse por el mayor contenido espectral de ésta pero que habría que ratificar estudiando en mayor profundidad la respuesta de la métrica en función del contenido espectral, aspecto que se enumera en las futuras líneas de trabajo. En lo que respecta a las diferencias relativas, los resultados obtenidos con ambas voces también son equiparables.
- l) En el caso de las señales de voz, la respuesta en términos de valor absoluto de la métrica en función del nivel de excitación del recinto –con los niveles SPL empleados en este experimento- es, a diferencia de la obtenida con las señales musicales, significativa tanto en entornos muy reverberantes como en recintos con muy baja reverberación. De esta forma, un incremento de nivel en un recinto excesivamente reverberante se traducirá en un menor valor absoluto de la métrica. En el extremo más anecoico, un incremento del nivel supondrá un mayor valor absoluto del resultado. En el rango medio de tiempos de reverberación y relaciones
- m) Con las señales de voz utilizadas se han obtenido funciones de ajuste con parámetros clásicos de evaluación de la palabra en salas, obteniéndose índices de correlación entre la métrica implementada y los índices de transmisión de la palabra STI y RASTI entre 0,81 y 0,90<sup>28</sup>. Debido al efecto observado en la conclusión anterior la función de ajuste de esta correlación es de tipo polinómico.
- n) En los ajustes realizados entre los resultados obtenidos con las señales de voz y los parámetros clásicos de tiempo de reverberación EDT y definición D se han obtenido unos índices de correlación entre 0,74 y 0,88. En general se obtienen mejores correlaciones con niveles de excitación más altos y con la voz femenina. La función de ajuste con el tiempo de reverberación EDT es logarítmica, mientras que en el caso de la definición D es de tipo polinómica si bien tiende a la linealidad al incrementar el nivel de excitación de la sala.
- o) En general, se obtienen índices de correlación ligeramente superiores respecto a los parámetros clásicos con el promediado de valores obtenidos en las frases psicoacústicas que con la señal del párrafo. El uso de las frases psicoacústicas reduce también el tiempo de procesado y permite establecer una dispersión de valores en un mismo punto de medida en función del contenido fonético de cada una de las frases.

---

<sup>28</sup> Considerando la señal del párrafo de texto y los promediados del conjunto de frases psicoacústicas..

- p) Dados los resultados mostrados en representaciones como las de las Gráfica 8.37 y Gráfica 8.38, por ejemplo, puede concluirse la viabilidad de la métrica para uso como herramienta de comparación entre recintos también en el caso de señales de palabra o voz. Las dispersiones representadas en estos diagramas de cajas incluyen tanto la variabilidad entre puntos de medición como las variaciones en función de los contenidos de cada frase, siendo representativos de series de 40 y 120 valores en el experimento llevado a cabo.
- q) La métrica implementada es de aplicación directa en situaciones de sala ocupada en el caso de salas en las que se haga uso de un sistema de refuerzo sonoro. En éstas, generalmente se dispone de la señal que se envía al sistema de refuerzo sonoro, que constituiría la señal de referencia.

#### 9.4 Futuras líneas de trabajo

Como ya se ha adelantado en la introducción de estas conclusiones, la principal línea de trabajo que queda pendiente tras la implementación y aplicación de la métrica perceptual es la investigación acerca de sus capacidades predictivas para la tipología de canales de transmisión estudiados. Esta tarea se enmarca en el ámbito de la experimentación psicoacústica y conllevaría el establecimiento de una correlación entre valoraciones cualitativas obtenidas mediante procedimientos como los vistos en la sección 2.2.1 utilizando los registros empleados para la aplicación de la métrica en este experimento y los propios resultados ofrecidos por ésta. Una vez establecida esta relación, podría evaluarse la capacidad de predicción de la métrica a partir de registros efectuados en cualquier recinto.

Otro aspecto a profundizar sería un análisis en mayor detalle del comportamiento de la métrica en función del contenido espectral de las señales de test empleadas, con la finalidad de determinar en qué medida afectan a los promediados finales las fuertes atenuaciones de señal que se aplican en determinadas bandas en la emulación de la respuesta del oído externo y medio. Un análisis de resultados en función de la distribución de potencia a lo largo del espectro de la señal de test podría esclarecer aspectos de que evitarían casos como el de la señal  $S_2$  en este experimento.

En el caso de los transductores se ha observado que la distancia entre  $R_{\text{nonlin}}$  obtenidos con diferentes tipologías de señales para un mismo dispositivo (expuesto en la conclusión e) referida a los transductores) podría ser de relevancia como un indicador de calidad. Sería necesaria más experimentación sobre un mayor número de dispositivos para corroborar este extremo.

En lo referente a la implementación del algoritmo, algunas propuestas de mejora para la optimización del cálculo podrían centrarse en el cálculo previo de la matriz de niveles con anterioridad al cálculo de las correlaciones cruzadas, de manera que para aquellos *frames* que no vayan a contribuir en la matriz de ponderación no se calcule  $X_{\text{max}}$ . Esto ocurre con bastante asiduidad en la baja y alta frecuencia. Eliminar el cálculo de estas correlaciones mejoraría considerablemente el tiempo de procesado. Por otra parte, una

mejora del algoritmo enfocada a descartar el procesado en franjas de silencio también sería deseable. Estos aspectos son de implementación relativamente sencilla y podrían ser el inicio de una tarea encaminada a poder implementar un análisis en tiempo real de las señales o al menos producir una muy significativa reducción de los tiempos de procesado.

Por último, se podría profundizar en la relación entre  $R_{\text{nonlin}}$  y los diferentes parámetros de calidad estudiados, medidos y correlados. Dado el buen ajuste obtenido a priori, podría proponerse la obtención de parámetros de calidad de salas a través de la medición de  $R_{\text{nonlin}}$ . Esto requiere ampliar el estudio a un mayor número de salas y obtener un ajuste más amplio en diferentes entornos.



# 10 Referencias bibliográficas

- AES-11D-1991 r 2003 standards and information documents. (1991). *Plane Wave Tubes: Designs and Practices*. AES.
- Aguilar, L. (1991). *Propuesta de un test de evaluación segmental del habla para el castellano: el Test de Rimas Modificado*. Departamento de Filología Española. Universidad Autónoma de Barcelona.
- Alías, F., & Triviño, M. (2007). A Phonetically balanced modified rhyme test for evaluating catalan speech intelligibility. *16th International Congress of Phonetic Sciences*. Saarbrücken.
- Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small room acoustics. *J. Acoust. soc. Am.*, 65(4), 943-950.
- Barron, M. (1981). Spatial impression due to early lateral reflections in concert halls. *J. Sound and Vibration*, 77(2), 211-232.

- Barron, M. (1988). Subjective study of British symphony concert halls. *Acustica*, 66(1), 1-14.
- Barron, M. (1993). *Auditorium Acoustics and Architectural Design*. E&FN Span.
- Barron, M., & Marshal, M. (1981). Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. *Journal of Sound and Vibration*, 77(2), 211-231.
- Beaton, R., Beerends, J., Keyhl, M., & Treurniet, W. (1996). Objective Perceptual Measurement of Audio Quality. *Colected Papers on Digital Audio Bit Rate Reduction*, 126-152.
- Bech, S., & Zacharov, N. (2006). *Perceptual Audio Evaluation. Theory, Method and Application*. Wiley.
- Belcher, R. (mayo de 1978). A new distortion measurement. *Wireless World*.
- Beraneck, L., & Mellow, T. (2012). *Acoustics: Sound Fields and Transducers*. Academic Press.
- Beranek, L. (1996). *Concert and Opera Halls -How they Sound*. Acoustical society of America.
- Beranek, L. L. (1962). *Music, Acoustics and Architecture*. John Wiley and sons.
- Berkhout, A. J., de Vries, D., & Boone, M. M. (1980). A new method to acquire impulse responses in concert halls. *J. Acoust. Soc. Am.*, 68(1), 179-183.
- Berkhout, A. J., de Vries, D., Baan, J., & van den Oetelaar, B. W. (1999). A wave field extrapolation approach to acoustical modelling in enclosed spaces. *J. Acoust. Soc. Am.*, 105(3), 1725-1733.
- Berkhout, A., de Vries, D., & Boone, M. (1980). A new method to acquire impulse responses in concert halls. *J. Acoust. Soc. Am.*, 68(1), 179-183.
- Blauert, J. (2012). A perceptionist view on psychoacoustics. *Archives of Acoustics*, 37(3), 365-371.

- Blauert, J. (2013). Conceptual aspects regarding the qualification of spaces for aural performances. *Acta acustica united with acustica*, 99(1), 1-13.
- Blauert, J. (2013). *The technology of binaural listening*. Springer.
- Blauert, J., & Dominicus, R. (2013). The things, feeling and thoughts of the aural world – an epistemological analysis regarding quality judgment. En W. G. Schmidt, *The Fascination of Sound: Anthropology, Mediality, and Cultural Practice*. De Gruyter.
- Blauert, J., & Guski, R. (2009). Critique of Pure Psychoacoustics. *International Conference on Acoustics; NAG-DAGA*. Berlin.
- Blauert, J., & Jekosh, U. (1997). Sound quality evaluation: a multilayered problem. *Acta Acustica united with Acustica*, 85(5), 747 -753.
- Blauert, J., & Jekosh, U. (2012). A layer model of sound quality. *J. Audio Eng. Soc.*, 60(1), 4-12.
- Borwick, J. (2001). *Loudspeaker and headphone handbook*. Taylor & Francis US.
- Bradley, J. (2011). Review of objective room acoustics measures and future needs. *Applied Acoustics*, 72, 713–720.
- Bruyninckx, M., Harmegnies, B., Llisterri, J., & Poch, D. (1994). Language-Induced voice quality variability in bilinguals. *Journal of Phonetics*, 22(1), 19-31.
- Buus, S. (1985). Release from Masking Caused by Envelopment Fluctuations. *Journal of the Acoustical Society of America*, 78, 1958-1965.
- Cagigas, A. (2001). An busca de la supraconciencia: Gustav Theodor Fechner. *Revista de Historia de la Psicología*, 22(3-4), 283-287.
- Carrión, A. (1998). *Diseño acústico de espacios arquitectónicos*. Edicions UPC.
- Cooke, M. (1993). *Modelling Auditory Processing and Organization*. New York: Cambridge University Press.

- Craggs, A. (1994). A finite element method for the free vibration of air in ducts and rooms with absorbing walls. *Journal of Sound and Vibration*, 173(16), 568-576.
- Cruaños, J., Francés, J., Ramis, J., & Alba, J. (2007). Application of the Geddes-Lee Model for subjective evaluation of sound emitted by loudspeakers. *19th International Congress on Acoustics*. Madrid.
- Cruaños, J., Francés, J., Ramis, J., & Alba, J. (2008). Psychoacoustic assessment of compression drivers. *Acoustics'08*. Paris.
- Cunha, I.B., Smiderle, R., & Bertoli, S. (2013). Influence of sound reinforcement system on acoustic performance in a Catholic Church. *International Symposium on Room Acoustics*. Toronto.
- Czewrwiniski, E., & Voishvillo, A. (Noviembre de 2001). Multitone Testing of Sound system Components. Some results and conclusions. *Journal of the Audio Engineering Society*, 1011-1048.
- Davis, D. (1979). The role of the initial time delay gap in the acoustic design of control rooms for recording and reinforcing systems. *64th Audio Eng. Soc. Convention*.
- Davis, D., & Patronis, E. (2006). *Sound system engineering*. Taylor & Francis.
- de Boer, E., & Jong, H. (1978). On cochlear encoding: Potentialities and limitations of the reverse-correlation technique. *Journal of the Acoustical Society of America*, 115-135.
- de Santis, E., & Henin, S. (2007). *Perception & Thresholds of Nonlinear Distortion using Complex Signals*. Institute of Electronic Systems. Aalborg University.
- Dickason, V. (2006). *Loudspeaker design cookbook*. Audio Press.
- Dietsch, L., & Kraak, W. (1986). Ein objektives Kriterium zur Erfassung von Echostörungen bei Musik- und Sprachdarbietungen (An objective criterion for capturing echo disturbances with music and speech performances). *Acustica*, 60, 205.

- Dominicus, R. (2009). *The subject theory of radical constructivism*. Diplomatica-Verlag.
- Ergan, J. P. (1948). Articulation Testing Methods. *Laryngoscope*, 58, 955-991.
- Eyring, C. F. (1930). Reverberation time in dead rooms. *J. Acoust. Soc. Am.*, 1, 217-241.
- Fairbanks, G. (1958). Test of Phonemic Differentiation: The Rhyme Test. *J. Acoust. Soc. Am.*, 30(7), 596-600.
- Farina, A. (2001). Acoustic quality of theatres: correlations between experimental measures and subjective evaluations. *Applied Acoustics*, 62(8), 889-916.
- Farina, A. (2007). Advancements in impulse response measurements by sine sweeps. *Proceedings of the 122nd AES Convention*. Viena.
- Fechner, G. (1860). *Elemente der Psychophysic*. Leipzig: Breitkopf und Härtel.
- Fletcher, H. (1922). The nature of speech and its interpretation. *Bell System Technical Journal*, 129-144.
- Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Physics*, 12, 47-65.
- Fletcher, H., & Munson, W. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical society of America*, 5, 82-108.
- Franz, M., & Scholkopf, B. (2006). A Unifying View of Wiener and Volterra Theory and Polynomial Kernel Regression. *Neural computation*, 18(12), 3097-3118.
- French, N., & Steinberg, J. (1947). Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.*, 19(1), 90-119.
- Galiana, M., Llinares, C., & Llinares, J. (2014). Does Architecture affect Acoustic perception in music halls? *International Conference on Kansei Engineering and Emotion Research*. Linköping.
- Geddes, E., & Clark, D. (1995). Computer simulation of horn loaded compression drivers. *J. Audio Eng.Soc.*, 431-436.

- Geddes, E., & Lee, L. (2003). Auditory perception of nonlinear distortion. New York: 115th AES Convention.
- Geddes, E., & Lee, L. (2003). Auditory perception of nonlinear distortion - Theory. New York: 115th AES Convention.
- Geddes, E., & Lee, L. (2006). Audibility of linear distortion with variations in sound pressure level and group delay. San Francisco: 121st AES Convention.
- Geddes, E., Lee, L., & Magalotti, R. (2005). Subjective testing of compression drivers. *JAES*, 1152-1157.
- Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filters shapes from notched noise data. *Hearing Research*, 47, 103-138.
- Glasberg, B., & Moore, B. (2002). A model of loudness applicable to time-varying sounds. *J. Audio Engineering Society*, 331-342.
- Glasberg, B., & Moore, B. (2006). Prediction of absolute thresholds and equal-loudness contours using a modified loudness model. *Journal of the Acoustical Society of America*, 120, 585-588.
- Guski, R., Felscher-Suhr, U., & Schuemer, R. (1999). The concept of noise annoyance: How international experts see it. *J. Sound Vibr.*, 223(4), 513-527.
- Hartmann, W. (1997). *Signals, sound and sensation*. Woodbury: American Institute of Physics.
- Heckbert, P. S., & Hanrahan, P. (1984). Beam tracing polygonal objects. *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*.
- Henricksen, C. (1988). Loudspeakers, enclosures and headphones. En G. Ballou, *Handbook for sound engineers* (págs. 405-488). Howard W. Sams & CO.
- Hernández, F. (2012). Mejora de la difusión acústica de la cámara reverberante del Campus de Gandía de la UPV. *TFM. Master en Ingeniería Acústica. Universitat Politècnica de València*.

- Hidalgo, J., Sommerhoff, J., & Rosas, C. (2011). Evaluación de un corpus de logatomos fonéticamente balanceado para la evaluación de la inteligibilidad de la palabra en español. *Síntesis Tecnológica*, 4(2), 37-49.
- Hilliard, J. K. (1941). Distortion Test by the Intermodulation Method. *Proceedings of the IRE*, 614-620.
- Hougast, T., & Steeneken, H. (1984). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.*, 77(3), 1069-177.
- House, A., Williams, C., Hecker, M., & Kryter, K. (1965). Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set. *J. Acoust. Soc. Am.*, 37(1), 158-166.
- IEC-60268-16. (2011). Sound system equipment. Part16: Objective rating of speech intelligibility by speech transmission index.
- ISO 3382. (1997). Measurement of the Reverberation Time of Rooms with Reference to other acoustical Parameters“, Annex B: "Binaural Auditorium Measures derived from Impulse Responses“. *International Standard Organization*.
- ISO 9921. (2003). Ergonomics-Assessment of speech communication. *International Standards Organization*.
- ITU BS.1387-1. (2001). Method for Objective Measurement of Perceived Audio Quality.
- ITU-R BS.1534-1. (2003). Method for the subjective assessment of the intermediate quality level of coding systems. Recommendation. *International Telecommunication Union*.
- ITU-T P.800. (1996). Methods for subjective determination of transmission quality.

- ITU-T P.835. (2003). *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm. Recommendation*. Geneva: International Telecommunication Union.
- ITU-T P.862. (2001). Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to End Speech Quality Assessment of Narrow Band Telephone Networks and Codecs.
- Jekosh, U. (2004). Basic concepts and terms of “quality”, reconsidered in the context of product-sound quality. *Acta Acustica united with Acustica*, 90(6), 999–1006.
- Jekosh, U. (2005). Assigning Meaning to Sounds – Semiotics in the Context of Product-Sound Design . En J. Blauert, *Communication Acoustics*. Springer.
- Johannesma, P. (1972). The per-response stimulus ensemble of neurons in cochlear nucleus. *Hearing Theory*, 58-69.
- Jordan, V. (1970). Acoustical criteria for auditoriums and their relation to model techniques. *J. Acoust. Soc. Am.*, 47(2), 408-412.
- Jordan, V. (1980). *Acoustical design of concert halls and theaters*. Applied Sciences Publishers.
- Jordan, V. L. (1982). Acoustical criteria for concert hall stages. *Applied Acoustics*, 15(5), 321-328.
- Kaizer, A. (1986). The Modelling of the Nonlinear Response of an Electrodinamic Loudspeaker by a Volterra Series Expansion. Montreux: 80th AES Convention.
- Kates, J. (1992). On using coherence to measure distortion in hearing aids. *Journal of the Acoustical Society of America*, 91(4), 2335-2244.
- Kinsler, L., Frey, A., Coppens, A., & Sanders, J. (1991). *Fundamentos de acústica*. Noriega Limusa.

- Kirkegaard, L., & Gulsrud, T. (2011). In search of a new paradigm: How do our parameters and measurement techniques constrain approaches to concert hall design? *Acoustics Today*, 7(1), 7-14.
- Kleiner, M. (1989). A New Way of Measuring Lateral Energy Fractions. *Applied Acoustics*, 27, 321-332.
- Klippel, W. (1990). Dynamic measurement and interpretation of the nonlinear parameters of electrodynamic loudspeakers. *J. Audio Eng. Soc.*, 35, 421-433.
- Klippel, W. (1992). The non linear large signal behaviour of electrodynamic loudspeaker at low frequencies. *J. Audio Eng. Soc.*, 40, 483-496.
- Klippel, W. (1996). Modelling the nonlinearities in horn loudspeakers. *Journal of the audio engineering society*, 44(8), 470-480.
- Klippel, W. (1996). Nonlinear System Identification for Horn Loudspeakers. *Journal of the Audio Engineering Society*, 44(10), 811-820.
- Klippel, W. (2007). *Loudspeakers Nonlinearities. Causes, Parameters, symptoms.* Klippel GmbH, Dresden.
- Krokstad, A., Strom, S., & Sorsdal, S. (1968). Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1), 118-125.
- Kürer, R. (1971). A simple measuring procedure for determining the central time of room acoustical impulse responses. *7th International Congress on Acoustics.* Busapest.
- Lacatis, R., Giménez, A., Barba, A., Cerdá, S., Romero, J., & Cibrián, R. (2008). Historical and chronological evolution of the concert hall acoustical parameters. *Proceedings of Acoustics '08, Paris*, (págs. 2151-2156).
- Le Callet, P., Moller, S., & Perkins, A. (2012). *Qualinet White Paper on Definitions of Quality of Experience.* Qualinet.

- Lehmann, P. (1976). *Über die Ermittlung raumakustischer Kriterien und deren Zusammenhang mit subjektiven Beurteilungen der Hörsamkeit (On the ascertainment of room acoustical criteria and correlation of the same with subjective assessments of the acoustic overall impression)*. Dissertation TU Berlin.
- Llisterri, J., & Poch, D. (1991). Caracterización fonética del bilingüismo, análisis acústico del habla espontánea y evaluación de sistemas de síntesis del habla. *Simposio de la lengua española. Ciencia y Tecnología*. Barcelona.
- Lokki, T. (2014). Tasting music like wine: sensory evaluation of concert halls. *Physics Today*, 67(1), 27-32.
- Lokki, T., Vertanen, H., Kuusinen, A., Pätynen, J., & Tervo, S. (2010). Auditorium acoustics assessment with sensory evaluation methods. *Proceedings of the International Symposium on Room Acoustics*. Melbourne.
- Magalotti, R., Zuccatti, C., & Pasini, P. (1999). Building a plane wave tube: experimental and theoretical aspects. *Journal of the Audio engineering Society*, 47, 596-601.
- Maggiolo, D. (2014). Apuntes de acústica musical. Obtenido de [www.eumus.edu.uy](http://www.eumus.edu.uy)
- Marquis-Favre, C., Premat, E., & Aubree, D. (2005b). Noise and its effects: a review on qualitative aspects of sound. Part II: noise and annoyance. *Acta Acustica united with Acustica*, 91(4), 626-642.
- Marquis-Favre, C., Premat, E., Aubree, D., & Vallet, M. (2005a). Noise and its effects. A review on qualitative aspects of sound. Part 1: Notions and acoustic ratings. *Acta Acustica united with Acustica*, 91(4), 613-625.
- Martínez, J., Cruaños, J., & Ramis, J. (2009). Comparison of different methods for the subjective sound quality evaluation of compression drivers. Munich: 126 AES Convention.

- Meddis, R., & López Poveda, E. (2010). *Computational models of the auditory system*. Springer.
- Metzler, B. (1993). *Audio Measurement Handbook*. Beaverton: Audio Precision.
- Millington, G. (1932). A modified formula for reverberation. *J. Acous. Soc. Am.*, 4, 69-82.
- Möller, S. (2005). Quality of Transmitted Speech for Humans and Machines. En J. Blauert, *Communication Acoustics*. Springer.
- Moore, B. (1995). Masking in the human auditory system. (AES, Ed.) *Collected Papers on Digital Audio Bit-Rate Reduction*.
- Moore, B. C. (2007). *Cochlear hearing loss*. John Wiley & Sons Ltd.
- Moore, B. C., Tan, C., & Zacharov, N. ( November de 2003). The Effect of Nonlinear Distortion on the Perceived Quality of Music and Speech Signals. *J. Audio Eng. Soc.*, 51(11).
- Moore, B. C., Tan, C., Zacharov, N., & Ville-Veikko, M. (Diciembre de 2004). Measuring and Predicting the Perceived Quality of Music and Speech Subjected to Combined Linear and Nonlinear Distortion. *Journal of the Audio Engineering Society*, 52(12), 1228-1244.
- Moore, B., & Glasberg, R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hearing Research*, 28, 209-225.
- Moore, B., Glasberg, B., & Gaer, T. (1997). A model for the prediction of thresholds, loudness and partial loudness. *Journal of the Audio engineering Society*, 224-240.
- Morabito, C., & Della Roca, M. (2010). Epistemological Models in Psychoacoustics: historical overview. *26th Annual Meeting of the International Society for Psychophysics. Fechner Day*.

- Neuhoff, J. (2004). *Ecological Psychoacoustics*. Elsevier Academic Press.
- Nowak, J., Liebetrau, J., & Sporer, T. (2013). On the perception of apparent sound width and listener envelopment. *Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*.
- Ogunfunmi, T. (2007). *Adaptive Nonlinear System Identification. The Volterra and Wiever model approaches*. Springer.
- Ortega, J., Gonzalez, J., & Marrero, V. (2000). AHUMADA: A large speech corpus in Spanish for speaker characterization and identification. *Speech Communication*, 31, 255-264.
- Patterson, R. (1976). Auditory Filter Shapes derived with Noise Stimuli. *Journal of the Acoustical Society of America*, 640-654.
- Patterson, R., & Henning, G. (1977). Stimulus variability and auditory filter shape. *Journal of the Acoustical Society of America*, 62, 649-664.
- Patterson, R., & Nimmo-Smith, I. (1980). Off-Frequency listening and Auditory Filter Assymetry. *journal of the Acoustical Society of America*, 67, 229-245.
- Patterson, R., Nimmo-Smith, I., Holdsworth, J., & Rice, P. (1987). An efficient auditory filter bank based on the gammatone function. *Speech-Group meeting of the Institute of Acoustics on Auditory Modelling*. Malvern.
- Patterson, R., Nimmo-Smith, I., Weber, D., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram and speech threshold. *J. Acoust. Soc. Am.*(72), 1788-1803.
- Pätynen, J., Tervo, S., Robinson, P., & Lokki, T. (2014). Concert halls with strong lateral reflections enhance musical dynamics. *Proc. Nat. Acad. Sci.*, 111(12), 4409-4414.
- Pérez Miñana, J. (1969). *Compendio práctico de acústica*. Barcelona: Labor.

- Peutz, V. (1971). Articulatory loss of consonants as a criterion for speech transmission in a room. *J. Audio Eng. Soc.*, 19, 915-919.
- Preiss, D., & Gregg, R. (2000). Coherence based, wide band, signal to distortion ratio versus total harmonic distortion of slightly non-linear audio systems. Los Angeles: 109th AES Convention.
- Pulkki, V., & Karjalainen, M. (2015). *Communication Acoustics: an introduction to speech, audio and psychoacoustics*. Wiley.
- Reichardt, W., Abdel Alim, O., & Schmidt, W. (1974). Abhängigkeit der Grenzen zwischen brauchbarer und unbrauchbarer Durchsichtigkeit von der Art des Musikmotivs, der Nachhallzeit und der Nachhalleinsatzzzeit. *Applied Acoustics*, 7, 243–264.
- Reichardt, W., Abdel Alim, O., & Schmidt, W. (1975). Definition und Messgrundlage eines objektiven Masses zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit beim Musikdarbietung. *Acustica*, 32, 126–137.
- Rife, D. D., & Vanderkooy, J. (1989). Transfer-function measurement with Maximum-Length Sequences. *J. Aud. Eng. Soc.*, 37(6), 419-444.
- Robles, L., & Ruggero, M. (2001). Mechanics of the mammalian cochlea. *Physiological Reviews*, 81, 1305–1352.
- Robles, L., Ruggero, M., & Rich, N. (1986). Basilar membrane mechanics at the base of the chinchilla cochlea. Input-Output functions, tuning curves and response phases. *Journal of the Acoustical Society of America*, 80, 1364-1374.
- Rosas, C., & Sommerhoff, J. (2008). Inteligibilidad acústica en español: una propuesta para su medición. *Estudios Filológicos*(43), 179-190.
- Rugh, W. J. (1981). *Nonlinear System Theory. The Volterra and Wiener approach*. The Johns Hopkins University Press.

- Rumsey, F., Zielinski, S., Kassier, R., & Bech, S. (2005). On the relative importance of spacial and timbral fidelities in judgments of degraded multichannel audio quality. *J. Acoust. Soc. Am.*, 118(2), 968-976.
- Sabine, W. C. (1922). *Collected Papers on Acoustics*. Harvard University Press.
- Schetzen, M. (1989). *The Volterra & Wiener theories of nonlinear systems*. Malabar, Florida: Krieger publishing company, inc.
- Schmitt, R. (1995). Audibility of nonlinear loudspeaker distortions. Paris: 98th AES Convention.
- Schmitt, R. (1995). Causes of nonlinear compression driver distortions and their audibility. New York: 99th AES Convention.
- Schroeder, M. (1965). New method for measuring reverberation time. *J. Acoust. Soc. Am.*, 37(3), 409-412.
- Schroeder, M., Gottlob, D., & Siebrasse, K. (1974). Comparative study of european concert halls: correlation of subjective preference with geometric and acoustic parameters. *J. Acoust. Soc. Am.*, 56(4), 1196-1200.
- Schurer, H., Berkhoff, A., Slump, C., & Herrmann, O. (1995). Modeling and compensation of nonlinear distortion in horn loudspeakers. *Journal of the Audio engineering Society*, 43(7/8).
- Sepúlveda, C., & Cortez, H. (2013). Modelo electroacústico análogo del oído. Obtenido de [www.mediatools.cl](http://www.mediatools.cl)
- Seron, D. (2011). The Fechner-Brentano Controversy on the Measurement of Sensation. *Revue roumaine de philosophie*, 55(1), 87-102.
- Shannon, E. (1948). Mathematical theory of communication. *Bell system Technical Journal*(27), 379-423.

- Shaw, E. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *Journal of the Acoustical Society of America*, 56, 1848–1861.
- Shorter, D. (1950). The influence of high order products on nonlinear distortion. *Electronic engineering*.
- Slaney, M. (1993). *An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank*. Apple Computer Technical Report #35.
- Sommerhoff, J., & Rosas, C. (2007). Evaluación de la inteligibilidad del habla en español. *Estudios filológicos*, 42, 215-225.
- Spiegel, M. (1981). Thresholds for tones in maskers of various bandwidths and for signals of various bandwidths as a function of signal frequency. *Journal of the Acoustical Society of America*, 69, 791-795.
- Sporer, T., Liebetrau, J., & Scheneider, S. (2009). Statistics of MUSHRA revisited. *Audio Eng. Soc. Convention 127*.
- Stan, G., Embrechts, J., & Archambeau, D. (2002). Comparison of Different Impulse Response Measurement Techniques. *J. Aud. Eng. Soc.*, 50(4), 249-262.
- Steeneken, H., & Houtgast, T. (2002). Validation of the revised STI method. *Speech Communication*, 38(3), 413-425.
- Steeneken, H., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *J. Acoust. Soc. Am.*, 67(1), 318-326.
- Steeneken, H., van Wijngaarden, S., & Verhave, J. (2011). The evolution of speech transmission index. *Aud. Eng. Soc. Convention 130th* . AES.
- Stevens, S. (1961). To Honor Fechner and Repeal His Law. *Science*, 133(3446), 80-86.
- Stevens, S. (1966). Operation or words? A reply to Savage. *Psychological Monographs*(80), 33-38.
- Stevens, S. (1975). *Psychophysics*. Wiley.

- Tan, C., Moore, B. C., Zacharov, N., & Ville-Veikko, M. (2004). Predicting the Perceived Quality of Nonlinearly Distorted Music and Speech Signals. *Journal of the Audio Engineering Society*, 52(7/8), 699-711.
- Tarnóczy, T. (1986). Noise Interference with Oral Communications. En L. Sáenz, & R. Stephens, *Noise Pollution* (págs. 249-263). John Wiley and Sons Ltd.
- Temme, S., & Brunet, P. (2008). A New Method for Measuring Distortion Using a Multitone Stimulus and Noncoherence. *Journal of the Audio engineering Society*, 56(3), 176-188.
- Thiele, R. (1953). Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen. *Acustica*, 3, 291-302.
- Tortosa, F., & Civera, C. (2006). *Historis de la Psicología*. McGraw-Hill.
- Totzek, U., & Preis, D. (1987). How to measure and Interpret Coherence Loss in Magnetic Recording. *JAES*, 35(11).
- Toyota, Y., Oguchi, K., & Nagata, M. (1988). A study of the characteristics of early reflections in concert halls. *2nd Joint Meeting: Acoustical Societies of America and Japan. J. Acous. Soc. Am.* 84 (S1).
- UNE EN ISO 3382:1. (2010). Medición de parámetros acústicos en recintos. Parte I: salas de espectáculos.
- Unoki, M., Irino, T., Glasberg, B., Moore, B., & Patterson, R. (2006). Comparison of the roex and gammachirp filters as representations of the auditory filter. *Journal of the Acoustical Society of America*, 120, 1474-1492.
- Valero, A. (1991). *El corpus de las frases psicoacústicas de Harvard: una adaptación al castellano*. Departamento de filología española. Universidad Autónoma de Barcelona.
- van Dorp Schuitman, J. (2011). *Auditory modelling for assessing room acoustics*. PhD Thesis. Universiteit Delft.

- van Wijngaarden, S., Verhave, J., & Steeneken, H. (2012). The speech transmission index after four decades of development. Technical note. *Acoustics Australia*, 40(2), 134-138.
- Vilkamo, J., Lokki, T., & Pulkki, V. (2009). Directional audio Coding: Virtual microphone-based synthesis and subjective evaluation. *J. Aud. Eng. Soc.*, 57(9), 709-724.
- Voishvillo, A. (2004). Nonlinearity in Horn Drivers and Engineering Aspects of Horn Driver Design. *ALMA Winter Symposium, January 7, 2004*. Gibson Labs Technology Group, Inc., Sunnyvale, CA, USA.
- Voishvillo, A. (2006). Assesment of nonlineraity in transducers and sound systems. San Francisco: 126 AES convention.
- Voishvillo, A., Terekov, A., & Czerwinski, E. (Abril de 2004). Graphing, Interpretation and Comparison of Results of Loudspeaker Nonlinear distortion Measurements. *Journal of the audio engineering society*, 52(4), 332-357.
- von Békésy, G. (1960). *Experiments on hearing*. New York: McGraw-Hill.
- Watanabe, K. (2008). Objective Perceptual Audio Quality Measurement Methods. *Boadcast Technology*(35).
- Yoo, J.-C. (2009). Fast Normalized Cross Correlation. *Circuits Syst Signal Process*, 819–843.
- Zwicker, E. (1961). Subdivision of the Audible Frequency Range into Critical Bands. *Journal of the Acoustical Society of America*, 33(2), 248.
- Zwicker, E. (1965). Temporal Effects in Simultaneous Masking by White-Noise Bursts. *Journal of the Acoustical Society of America*, 37, 653-663.
- Zwicker, E. (1965). Temporal Effects in Simultaneous Masking and Loudness. *Journal of the Acoustical Society of America*, 38, 132-141.

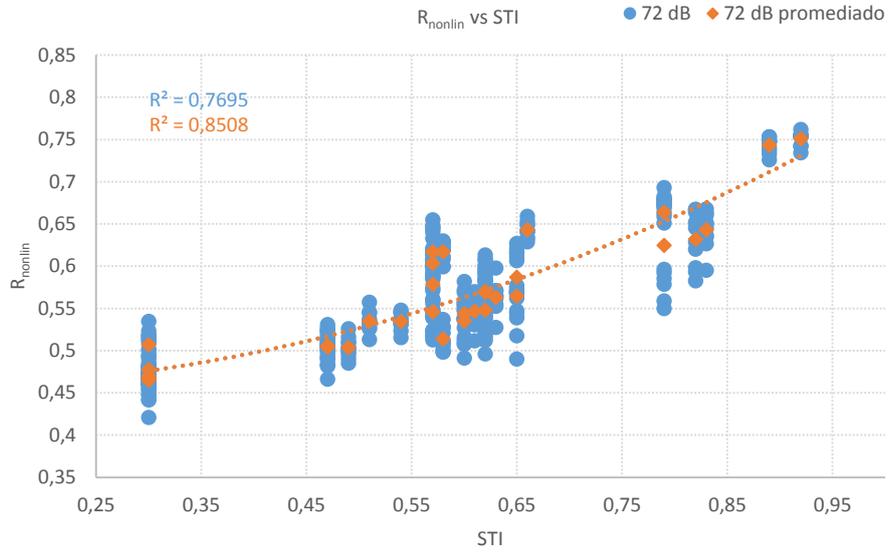
Zwicker, E., & Fastl, H. (1972). On the Development of the Critical Band. *Journal of the Acoustical Society of America*, 52, 699-702.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics. Facts and models*. Berlin: Springer-Verlag.

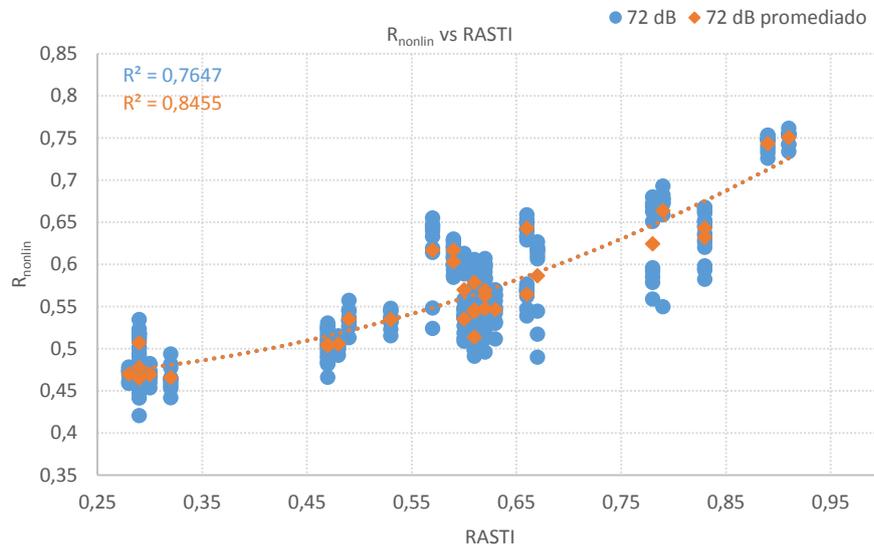
# Anexo I

**Resultados adicionales obtenidos con las frases psicoacústicas**

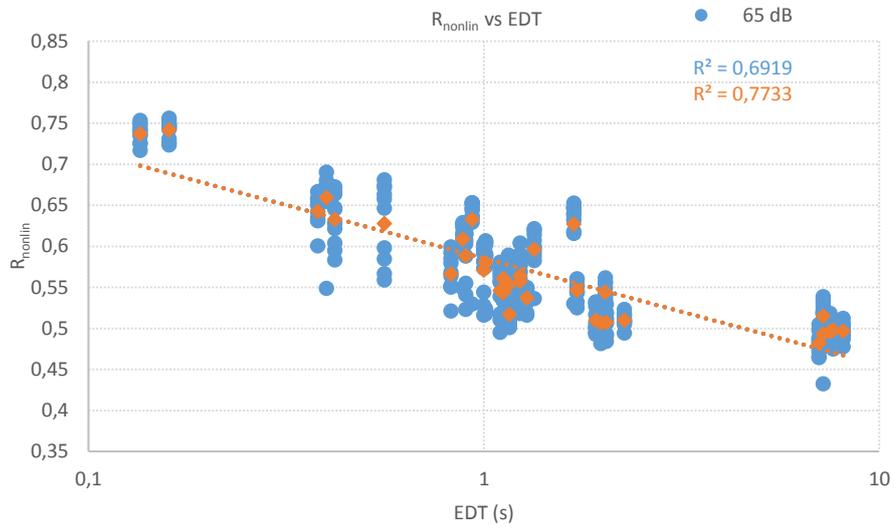




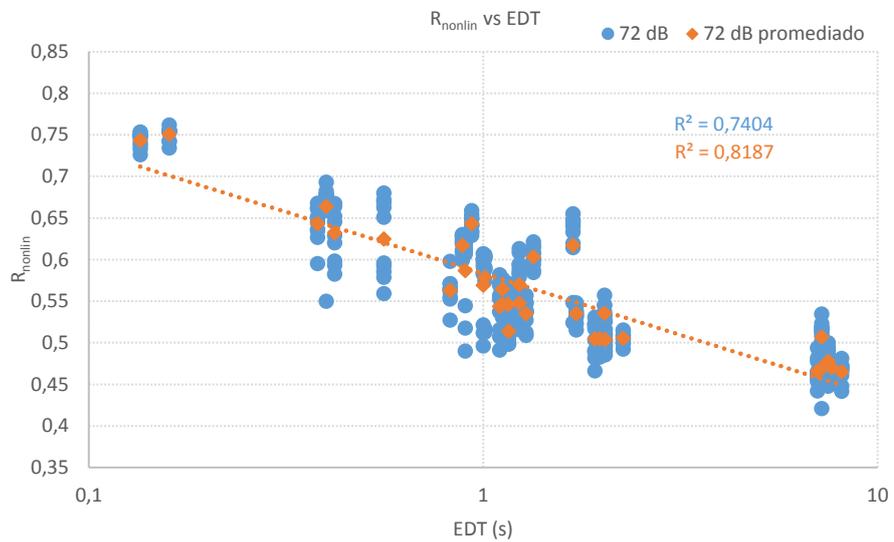
**Gráfica I.1 Pares STI - R<sub>nonlin</sub> frases psicoacústicas voz masculina (72 dB)**



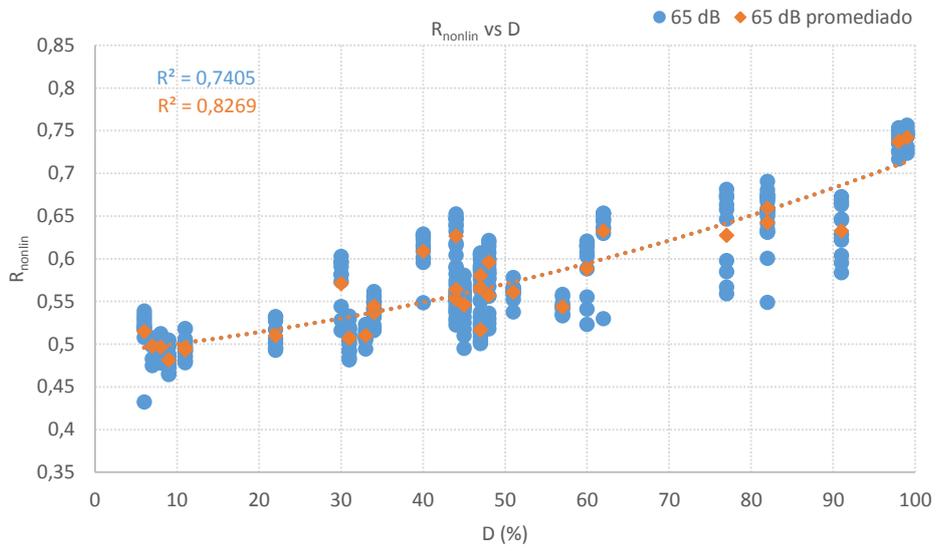
**Gráfica I.2 Pares RASTI - R<sub>nonlin</sub> frases psicoacústicas voz masculina (72 dB)**



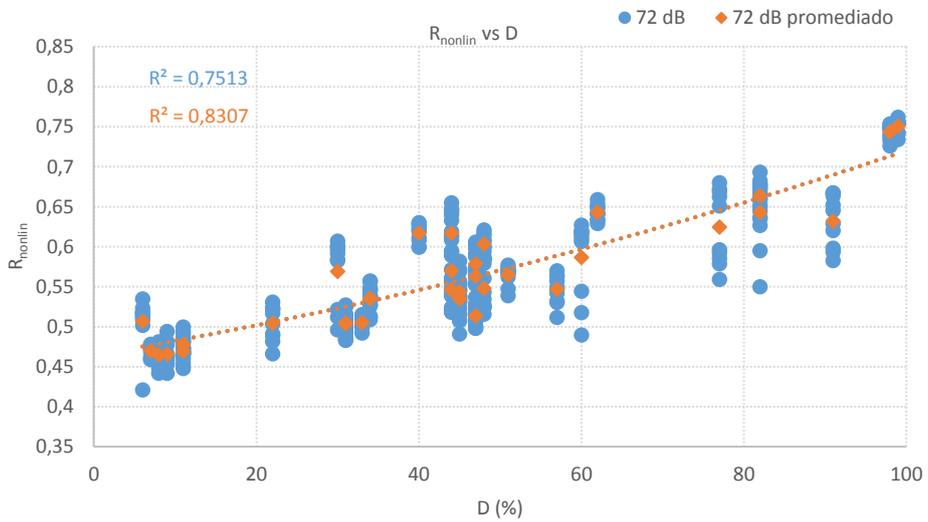
Gráfica I.3 Pares EDT - R<sub>nonlin</sub> frases psicoacústicas voz masculina (65 dB)



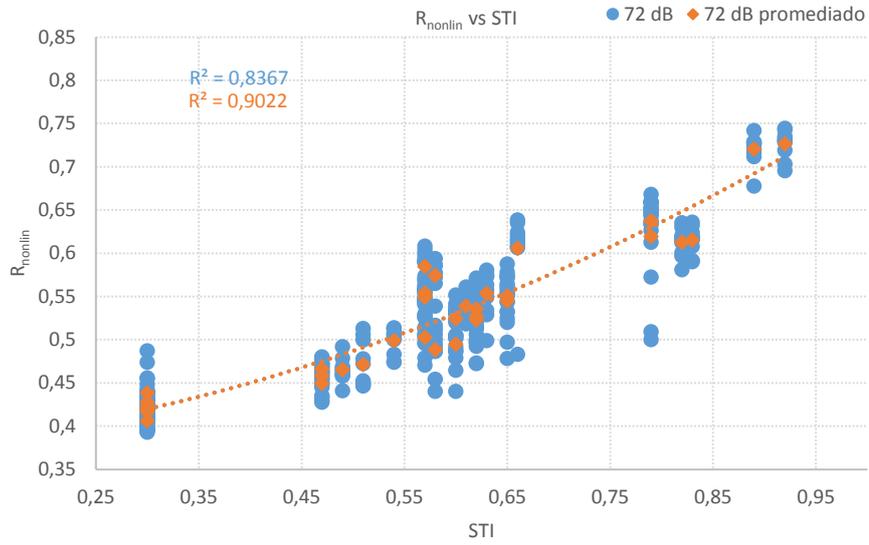
Gráfica I.4 Pares EDT - R<sub>nonlin</sub> frases psicoacústicas voz masculina (72 dB)



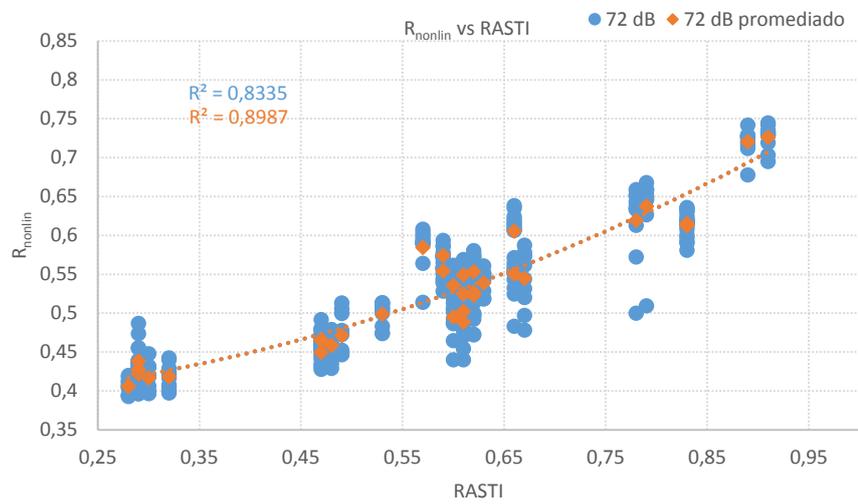
Gráfica I.5 Pares D -  $R_{nonlin}$  frases psicoacústicas voz masculina (65 dB)



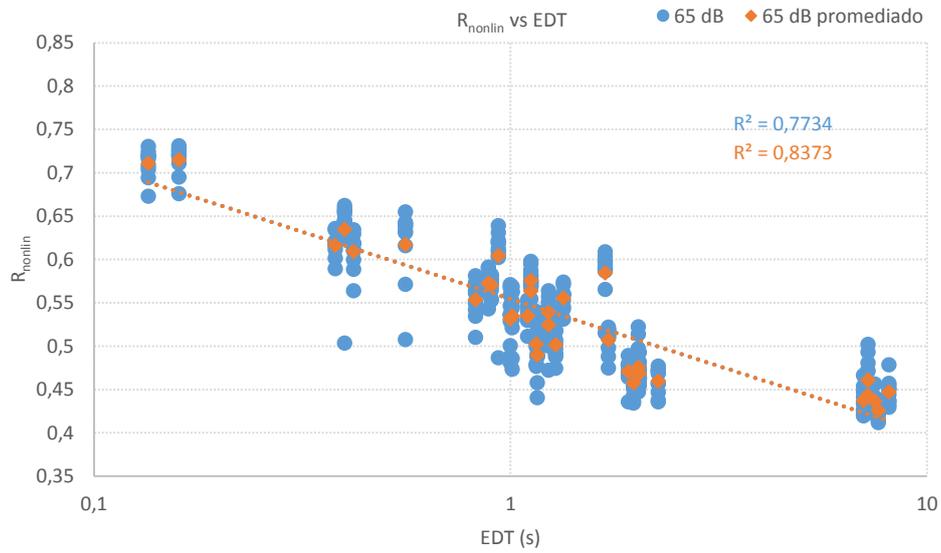
Gráfica I.6 Pares D -  $R_{nonlin}$  frases psicoacústicas voz masculina (72 dB)



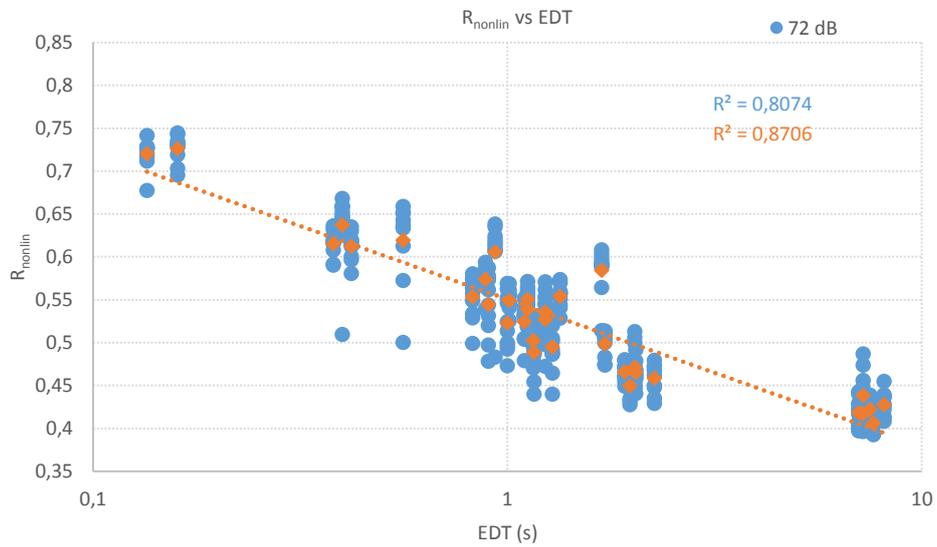
**Gráfica I.7 Pares STI - R<sub>nonlin</sub> frases psicoacústicas voz femenina (72 dB)**



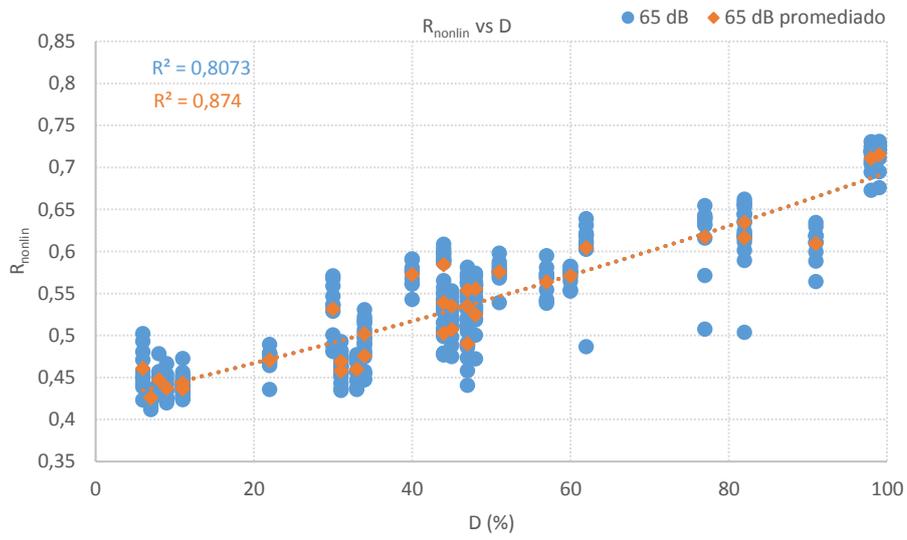
**Gráfica I.8 Pares RASTI - R<sub>nonlin</sub> frases psicoacústicas voz femenina (72 dB)**



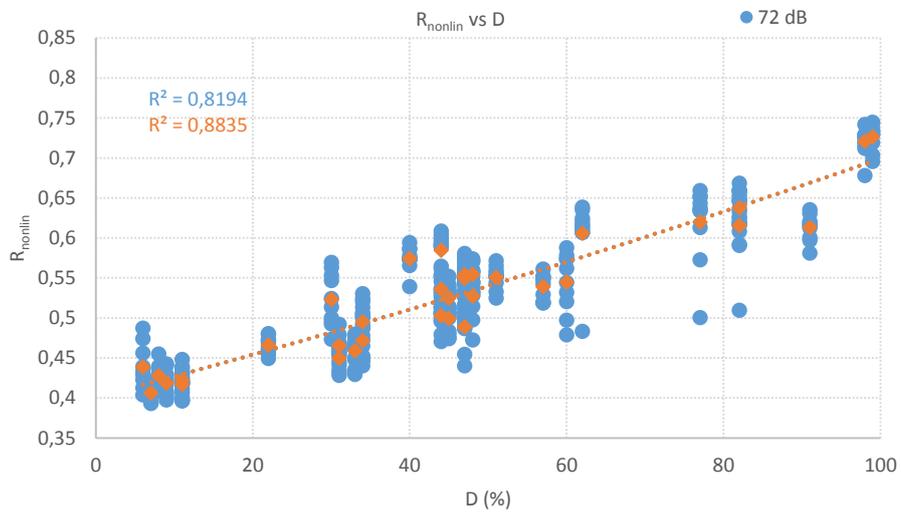
Gráfica I.9 Pares EDT -  $R_{nonlin}$  frases psicoacústicas voz femenina (65 dB)



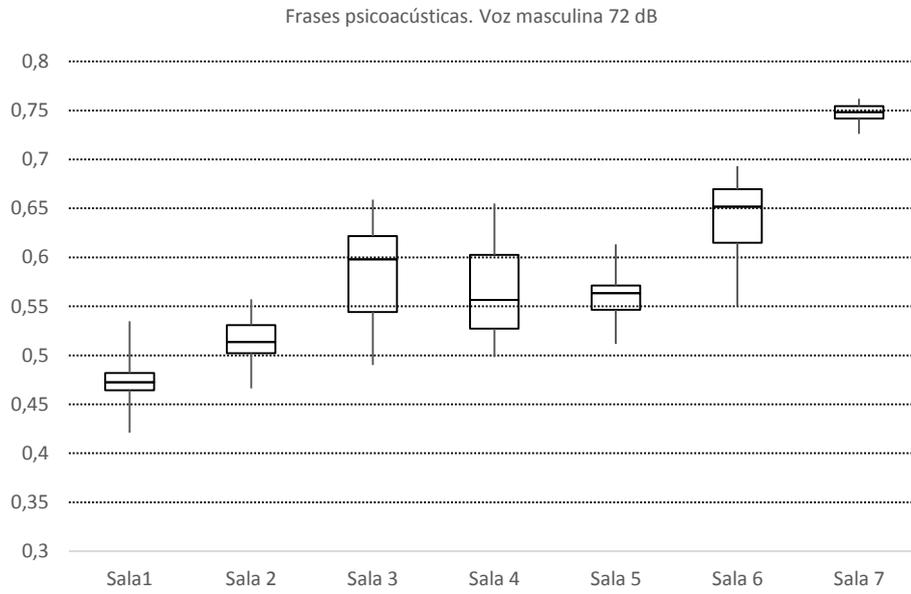
Gráfica I.10 Pares EDT -  $R_{nonlin}$  frases psicoacústicas voz femenina (72 dB)



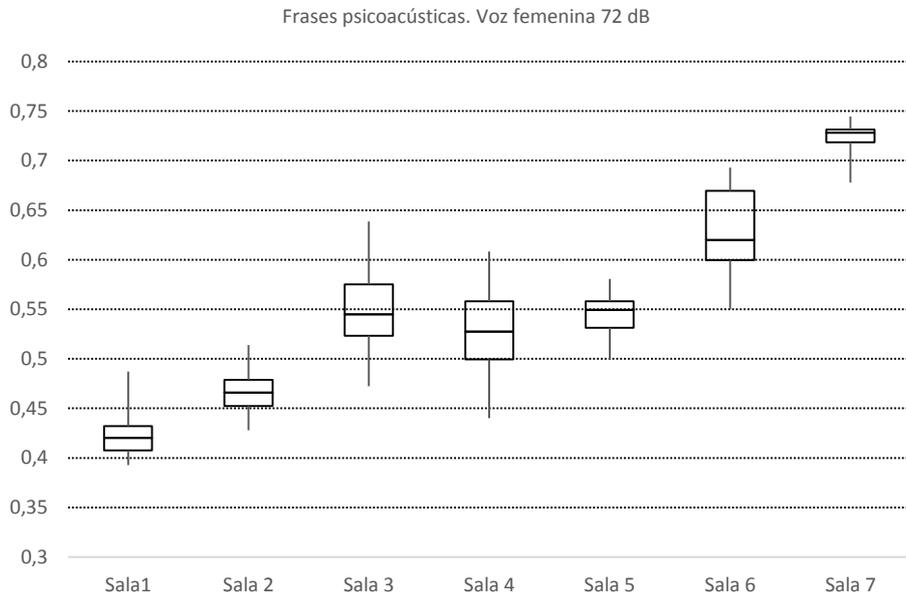
**Gráfica I.11 Pares D - R<sub>nonlin</sub> frases psicoacústicas voz femenina (65 dB)**



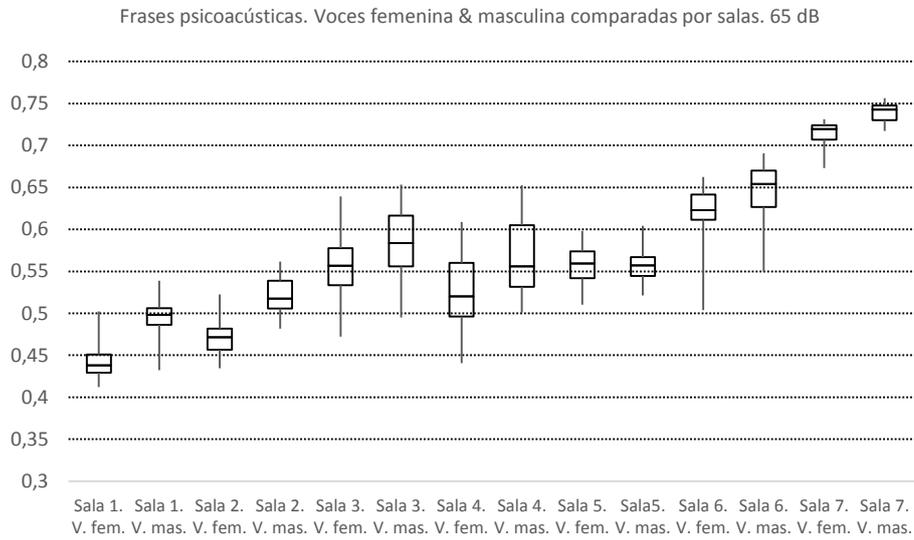
**Gráfica I.12 Pares D - R<sub>nonlin</sub> frases psicoacústicas voz femenina (72 dB)**



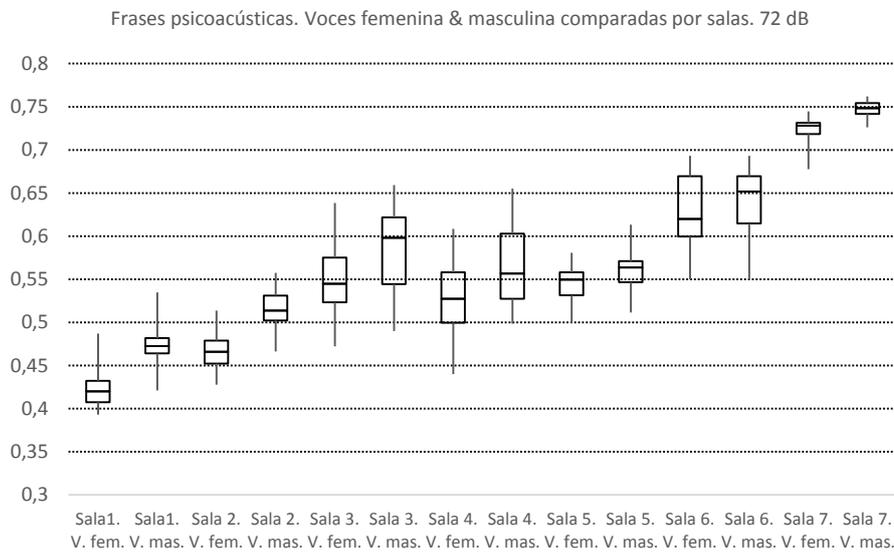
**Gráfica I.13 Resultados frases psicoacústicas voz masculina por salas (72 dB)**



**Gráfica I.14 Resultados frases psicoacústicas voz femenina por salas (72 dB)**



**Gráfica I.15 Resultados frases psicoacústicas voces masculina/femenina por salas (65 dB)**



**Gráfica I.16 Resultados frases psicoacústicas voces masculina/femenina por salas (65 dB)**

# Anexo II

**Tabla de funciones de ajuste de datos experimentales**



Señal	Nivel	Función	R <sup>2</sup>
S <sub>1</sub>	L <sub>1</sub> (70 dB)	$RT = -0.04Ln(R_{nonlin}) + 0.6519$	0.8060
		$EDT = -0.04Ln(R_{nonlin}) + 0.6536$	0.8188
		$C_{80} = -7 \cdot 10^{-5} \cdot (R_{nonlin})^2 + 0.0059 \cdot R_{nonlin} + 0.6306$	0.8534
		$D = 0,0017R_{nonlin} + 0.5666$	0.8632
S <sub>1</sub>	L <sub>2</sub> (80 dB)	$RT = -0.047Ln(R_{nonlin}) + 0.6462$	0.8648
		$EDT = -0.047Ln(R_{nonlin}) + 0.6481$	0.8663
		$C_{80} = -0.0001 \cdot (R_{nonlin})^2 + 0.0073 \cdot R_{nonlin} + 0.6231$	0.8807
		$D = 0.0019R_{nonlin} + 0.5487$	0,8737
S <sub>2</sub>	L <sub>1</sub> (65 dB)	$RT = -0.014Ln(R_{nonlin}) + 0.5423$	0.5422
		$EDT = -0.014Ln(R_{nonlin}) + 0.5427$	0.5031
		$C_{80} = -0.00004 \cdot (R_{nonlin})^2 + 0.0022 \cdot R_{nonlin} + 0.5355$	0.5557
		$D = 0.0006R_{nonlin} + 0.5123$	0,5810
S <sub>2</sub>	L <sub>2</sub> (72 dB)	$RT = -0.024Ln(R_{nonlin}) + 0.556$	0.7062
		$EDT = -0.023Ln(R_{nonlin}) + 0.5568$	0.6806
		$C_{80} = -0.00007 \cdot (R_{nonlin})^2 + 0.0038 \cdot R_{nonlin} + 0.5453$	0.6885
		$D = 0.0009R_{nonlin} + 0.509$	0.6559
S <sub>3</sub>	L <sub>1</sub> (70 dB)	$RT = -0.085Ln(R_{nonlin}) + 0.568$	0.8675
		$EDT = -0.085 \cdot Ln(R_{nonlin}) + 0.5714$	0.8633
		$C_{80} = -0.0001 \cdot (R_{nonlin})^2 + 0.0122 \cdot R_{nonlin} + 0.5244$	0.8484
		$D = 0.0033R_{nonlin} + 0.3948$	0.8474
S <sub>3</sub>	L <sub>2</sub> (78 dB)	$RT = -0.081Ln(R_{nonlin}) + 0.571$	0.8593
		$EDT = -0.081Ln(R_{nonlin}) + 0.5742$	0.8577
		$C_{80} = -0.0001 \cdot (R_{nonlin})^2 + 0.0117 \cdot R_{nonlin} + 0.5294$	0.8441
		$D = 0.0032R_{nonlin} + 0.4063$	0.8327
S <sub>4</sub>	L <sub>1</sub> (70 dB)	$RT = -0.09Ln(R_{nonlin}) + 0.6799$	0.8872
		$EDT = -0.09Ln(R_{nonlin}) + 0.6833$	0.8782
		$C_{80} = -0.0003 \cdot (R_{nonlin})^2 + 0.0147 \cdot R_{nonlin} + 0.6405$	0.8458
		$D = 0.0034R_{nonlin} + 0.5025$	0.7957
S <sub>4</sub>	L <sub>2</sub> (78 dB)	$RT = -0.092Ln(R_{nonlin}) + 0.6859$	0.8778
		$EDT = -0.092Ln(R_{nonlin}) + 0.6896$	0.8793
		$C_{80} = -0.0003 \cdot (R_{nonlin})^2 + 0.0149 \cdot R_{nonlin} + 0.6452$	0.8679

		$D = 0.0035R_{nonlin} + 0.5043$	0.8064
		$EDT = -0.056Ln(R_{nonlin}) + 0.575$	0.7416
Párrafo Voz masculina	L <sub>1</sub> (65 dB)	$D = 0.0023 \cdot R_{nonlin} + 0.455$	0.7676
		$D = -0.00001 \cdot (R_{nonlin})^2 + 0.001 \cdot R_{nonlin} + 0.4795$	0.7911
		$STI = 0.4447 \cdot (R_{nonlin})^2 - 0.1701 \cdot R_{nonlin} + 0.4961$	0.8131
		$RASTI = 0.487 \cdot (R_{nonlin})^2 - 0.221 \cdot R_{nonlin} + 0.5089$	0.8066
		$EDT = -0.064Ln(R_{nonlin}) + 0.5727$	0.7782
Párrafo Voz masculina	L <sub>2</sub> (72 dB)	$D = 0.0026 \cdot R_{nonlin} + 0.4376$	0.7983
		$D = -0.00001 \cdot (R_{nonlin})^2 + 0.0147 \cdot R_{nonlin} + 0.4576$	0.8062
		$STI = 0.4065 \cdot (R_{nonlin})^2 - 0.0808 \cdot R_{nonlin} + 0.454$	0.8313
		$RASTI = 0.4541 \cdot (R_{nonlin})^2 - 0.1383 \cdot R_{nonlin} + 0.4686$	0.8260
Párrafo Voz femenina	L <sub>1</sub> (65 dB)	$EDT = -0.068Ln(R_{nonlin}) + 0.5437$	0.8153
		$D = 0.0027 \cdot R_{nonlin} + 0.3997$	0.8440
		$D = 0.000006 \cdot (R_{nonlin})^2 + 0.0022 \cdot R_{nonlin} + 0.4102$	0.8473
		$STI = 0.3365 \cdot (R_{nonlin})^2 + 0.0261 \cdot R_{nonlin} + 0.3875$	0.8711
		$RASTI = 0.3847 \cdot (R_{nonlin})^2 - 0.0324 \cdot R_{nonlin} + 0.4024$	0.8679
Párrafo Voz femenina	L <sub>2</sub> (72 dB)	$EDT = -0.074Ln(R_{nonlin}) + 0.5405$	0.8368
		$D = 0.003 \cdot R_{nonlin} + 0.3833$	0.8585
		$D = 0.000006 \cdot (R_{nonlin})^2 + 0.0024 \cdot R_{nonlin} + 0.394$	0.8615
		$STI = 0.3471 \cdot (R_{nonlin})^2 + 0.0527 \cdot R_{nonlin} + 0.3633$	0.8871
		$RASTI = 0.3991 \cdot (R_{nonlin})^2 - 0.0108 \cdot R_{nonlin} + 0.3796$	0.8828
Frases Voz masculina	L <sub>1</sub> (65 dB)	$EDT = -0.056Ln(R_{nonlin}) + 0.575$	0.6919
		$D = 0.0023 \cdot R_{nonlin} + 0.4649$	0.7181
		$D = 0.00001 \cdot (R_{nonlin})^2 + 0.001 \cdot R_{nonlin} + 0.4897$	0.7405
		$STI = 0.459 \cdot (R_{nonlin})^2 - 0.1872 \cdot R_{nonlin} + 0.5111$	0.7556
		$RASTI = 0.5033 \cdot (R_{nonlin})^2 - 0.2401 \cdot R_{nonlin} + 0.5244$	0.7502
Frases promediadas Voz masculina	L <sub>1</sub> (65 dB)	$EDT = -0.056Ln(R_{nonlin}) + 0.5852$	0.7733
		$D = 0.0023 \cdot R_{nonlin} + 0.4645$	0.8018
		$D = 0.00001 \cdot (R_{nonlin})^2 + 0.001 \cdot R_{nonlin} + 0.4898$	0.8269
		$STI = 0.4579 \cdot (R_{nonlin})^2 - 0.1865 \cdot R_{nonlin} + 0.5111$	0.8435
		$RASTI = 0.5033 \cdot (R_{nonlin})^2 - 0.2395 \cdot R_{nonlin} + 0.5243$	0.8377
Frases	L <sub>2</sub> (72 dB)	$EDT = -0.064Ln(R_{nonlin}) + 0.5833$	0.7404

Voz masculina		$D = 0.0026 \cdot R_{nonlin} + 0.4485$	0.7432
		$D = 0.000009 \cdot (R_{nonlin})^2 + 0.0017 \cdot R_{nonlin} + 0.4648$	0.7513
		$STI = 0.3753 \cdot (R_{nonlin})^2 - 0.0474 \cdot R_{nonlin} + 0.4564$	0.7695
		$RASTI = 0.4225 \cdot (R_{nonlin})^2 - 0.1046 \cdot R_{nonlin} + 0.471$	0.7647
		$EDT = -0.064Ln(R_{nonlin}) + 0.5833$	0.8187
Frases promediadas Voz masculina	L <sub>2</sub> (72 dB)	$D = 0.0026 \cdot R_{nonlin} + 0.4485$	0.8217
		$D = 0.000009 \cdot (R_{nonlin})^2 + 0.0017 \cdot R_{nonlin} + 0.4648$	0.8307
		$STI = 0.3753 \cdot (R_{nonlin})^2 - 0.0474 \cdot R_{nonlin} + 0.4564$	0.8508
		$RASTI = 0.4225 \cdot (R_{nonlin})^2 - 0.1046 \cdot R_{nonlin} + 0.471$	0.8455
		$EDT = -0.067Ln(R_{nonlin}) + 0.5545$	0.7734
Frases Voz femenina	L <sub>1</sub> (65 dB)	$D = 0.0027 \cdot R_{nonlin} + 0.4109$	0.8042
		$D = 0.000006 \cdot (R_{nonlin})^2 + 0.022 \cdot R_{nonlin} + 0.4214$	0.8073
		$STI = 0.3445 \cdot (R_{nonlin})^2 + 0.0135 \cdot R_{nonlin} + 0.4027$	0.8229
		$RASTI = 0.3949 \cdot (R_{nonlin})^2 - 0.0473 \cdot R_{nonlin} + 0.4182$	0.8205
		$EDT = -0.067Ln(R_{nonlin}) + 0.5544$	0.8373
Frases promediadas Voz femenina	L <sub>1</sub> (65 dB)	$D = 0.0027 \cdot R_{nonlin} + 0.411$	0.8707
		$D = 0.000006 \cdot (R_{nonlin})^2 + 0.022 \cdot R_{nonlin} + 0.4214$	0.8740
		$STI = 0.3444 \cdot (R_{nonlin})^2 + 0.0136 \cdot R_{nonlin} + 0.4027$	0.8909
		$RASTI = 0.3948 \cdot (R_{nonlin})^2 - 0.0472 \cdot R_{nonlin} + 0.4182$	0.8883
		$EDT = -0.074Ln(R_{nonlin}) + 0.5505$	0.8074
Frases Voz femenina	L <sub>2</sub> (72 dB)	$D = 0.003 \cdot R_{nonlin} + 0.3934$	0.8180
		$D = 0.000004 \cdot (R_{nonlin})^2 + 0.0026 \cdot R_{nonlin} + 0.401$	0.8194
		$STI = 0.3215 \cdot (R_{nonlin})^2 - 0.0794 \cdot R_{nonlin} + 0.367$	0.8367
		$RASTI = 0.3737 \cdot (R_{nonlin})^2 - 0.0156 \cdot R_{nonlin} + 0.3834$	0.8335
		$EDT = -0.074Ln(R_{nonlin}) + 0.5505$	0.8706
Frases promediadas Voz femenina	L <sub>2</sub> (72 dB)	$D = 0.003 \cdot R_{nonlin} + 0.3934$	0.8820
		$D = 0.000004 \cdot (R_{nonlin})^2 + 0.0026 \cdot R_{nonlin} + 0.401$	0.8835
		$STI = 0.3215 \cdot (R_{nonlin})^2 - 0.0794 \cdot R_{nonlin} + 0.367$	0.9022
		$RASTI = 0.3737 \cdot (R_{nonlin})^2 - 0.0156 \cdot R_{nonlin} + 0.3834$	0.8987



# Anexo III

**Mediciones de parámetros en salas**



Sala 1. P1							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	6,72	7,03	7,82	6,7	5,27	3,64	1,99
EDT (s)	3,76	6,81	8,12	6,88	4,69	3,69	1,85
C <sub>80</sub> (dB)	-6,5	-8,8	-7,6	-7,7	-7,8	-4,7	-0,4
D (%)	7	7	8	11	8	17	35
STI	0,3						
RASTI	0,29						
Sala 1. P2							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	6,49	7,29	7,75	6,65	5,37	3,69	1,97
EDT (s)	5,21	6,2	7,8	6,5	4,98	3,7	1,89
C <sub>80</sub> (dB)	-5,7	-6,3	-7,8	-9,5	-8,7	-4,8	0,5
D (%)	16	10	9	6	7	16	39
STI	0,3						
RASTI	0,29						
Sala 1. P3							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	6,98	6,77	7,39	6,67	5,41	3,7	2
EDT (s)	5,32	9,32	8,03	7,27	5,28	3,63	2,1
C <sub>80</sub> (dB)	-5	-6	-8,6	-7,9	-6,1	-4,8	-0,5
D (%)	14	7	8	7	13	14	38
STI	0,3						
RASTI	0,28						
Sala 1. P4							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	6,24	6,92	7,56	6,66	5,43	3,78	2
EDT (s)	6,74	7,3	6,87	7,22	5,17	3,76	1,49
C <sub>80</sub> (dB)	-8	-8,9	-7,5	-7	-6,5	-4,2	1
D (%)	12	5	7	9	11	17	41
STI	0,3						
RASTI	0,32						
Sala 1. P5							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	6,95	8,43	7,61	6,61	5,3	3,79	2
EDT (s)	6,77	7,36	8,37	7,85	5,83	3,51	2,06
C <sub>80</sub> (dB)	-6,2	-8,7	-10,3	-8,1	-6	-4,6	0,3
D (%)	15	7	5	8	11	16	33
STI	0,3						
RASTI	0,29						
Sala 1. P6							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	6,67	7,12	7,7	6,78	5,26	3,67	2
EDT (s)	8,12	7,08	7,8	6,63	5,29	3,7	1,96
C <sub>80</sub> (dB)	-10,6	-8,3	-8,8	-6,5	-4,4	-4,1	-0,1
D (%)	6	6	7	11	19	17	35
STI	0,3						
RASTI	0,3						

Sala 2. P1							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	2,32	1,69	1,99	2	0,93	0,94	0,94
EDT (s)	1,92	1,87	2,38	1,68	2,08	1,58	1,08
C <sub>80</sub> (dB)	0,2	1,1	-1,3	-0,1	0	1,2	3,5
D (%)	28	46	32	34	36	47	53
STI	0,51						
RASTI	0,49						
Sala 2. P2							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	2,06	1,78	1,99	1,92	1,9	1,61	1,26
EDT (s)	1,99	1,48	0,94	1,5	1,69	1,27	1,13
C <sub>80</sub> (dB)	1,9	1,3	1,6	0,9	0,6	2	3,6
D (%)	50	51	49	45	44	49	60
STI	0,53						
RASTI	0,54						
Sala 2. P3							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,72	1,76	1,98	1,95	1,87	1,61	1,28
EDT (s)	2,48	3,12	2,16	1,79	1,72	1,69	1,07
C <sub>80</sub> (dB)	-1,7	-2,3	-2,3	-1,2	-1,7	-0,6	3,7
D (%)	16	30	23	31	25	32	60
STI	0,47						
RASTI	0,47						
Sala 2. P4							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,98	1,81	2,02	1,96	1,89	1,63	1,28
EDT (s)	2,47	2,01	1,99	1,85	1,89	1,72	1,3
C <sub>80</sub> (dB)	-3,9	-1,2	-3,1	-3,4	-1,9	-1	3,8
D (%)	13	18	22	22	23	30	61
STI	0,47						
RASTI	0,47						
Sala 2. P5							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,5	1,82	2,03	1,9	1,89	1,64	1,28
EDT (s)	0,93	1,63	1,78	2,32	1,88	1,59	1,34
C <sub>80</sub> (dB)	2,6	0,7	-1,4	-1,5	-1,2	-0,9	2,8
D (%)	29	34	28	31	30	31	5
STI	0,49						
RASTI	0,27						
Sala 2. P6							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,74	1,85	2,07	1,88	1,86	1,62	1,31
EDT (s)	2,59	2,33	2,38	2,15	2	1,63	1,5
C <sub>80</sub> (dB)	-2,1	-1,3	-1,1	-0,9	-1,5	0,6	0,4
D (%)	20	15	32	33	26	37	37
STI	0,47						
RASTI	0,48						

Sala 3. P1							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,69	1,15	1,14	1,14	0,94	0,75	0,52
EDT (s)	2,34	1,98	0,98	0,83	0,71	0,79	0,51
C <sub>80</sub> (dB)	-2,2	-2,1	-1,8	2,5	2,7	4,3	10,4
D (%)	20	29	26	40	40	42	66
STI	0,58						
RASTI	0,59						
Sala 3. P2							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,41	1,13	1,07	1,1	0,82	0,72	0,57
EDT (s)	0,95	1,2	1,17	0,7	0,74	0,78	0,6
C <sub>80</sub> (dB)	2	0,3	2,8	5	6,9	6,7	11,6
D (%)	46	42	53	62	75	68	88
STI	0,66						
RASTI	0,66						
Sala 3. P3							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,31	1,12	1,04	1,09	0,94	0,73	0,57
EDT (s)	2,78	1,32	1,31	0,69	0,99	0,84	0,57
C <sub>80</sub> (dB)	2	0,6	-0,1	3,3	3,2	4,1	12
D (%)	20	39	37	30	34	46	53
STI	0,62						
RASTI	0,62						
Sala 3. P4							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,47	0,97	0,9	0,9	0,86	0,71	0,6
EDT (s)	2,14	1,17	1	0,8	0,86	0,9	0,66
C <sub>80</sub> (dB)	-0,2	2,4	4,6	6,1	6,3	6,6	10,1
D (%)	23	51	49	60	68	68	81
STI	0,65						
RASTI	0,67						
Sala 3. P5							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,37	1,2	1	1,02	0,9	0,75	0,57
EDT (s)	1,62	1,38	1,24	0,96	0,77	0,74	0,7
C <sub>80</sub> (dB)	-3,2	-1,7	1	2,8	2,4	5,2	9,3
D (%)	12	26	43	45	42	56	82
STI	0,60						
RASTI	0,61						
Sala 3. P6							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,5	1,1	1,06	1,02	0,86	0,7	0,55
EDT (s)	2,25	1,59	1,25	1,22	0,89	0,8	0,58
C <sub>80</sub> (dB)	-2	1,2	1,3	4,5	3,4	4,4	10,5
D (%)	10	23	32	48	40	57	83
STI	0,62						
RASTI	0,62						

Sala 4. P1							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,27	1,26	1,29	1,24	1,27	1,24	1,05
EDT (s)	2,1	1,74	1,4	1,28	1,31	1,25	1,11
C <sub>80</sub> (dB)	-2,6	1	1,3	1,9	2,1	2,3	5,5
D (%)	28	45	49	48	51	52	70
STI	0,57						
RASTI	0,59						
Sala 4. P2							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,25	1,32	1,2	1,27	1,31	1,35	1,05
EDT (s)	3,23	1,08	1,71	1,67	1,2	1,28	0,83
C <sub>80</sub> (dB)	2,5	1,4	1	2,4	0,8	1,8	8,9
D (%)	56	40	32	44	36	41	80
STI	0,57						
RASTI	0,57						
Sala 4. P3							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,35	1,32	1,22	1,27	1,33	1,28	1,03
EDT (s)	1,59	1,6	1,08	1,23	1,23	1,31	1,03
C <sub>80</sub> (dB)	-2,4	3,4	3,6	1,1	1,3	2,1	4,5
D (%)	32	42	55	44	36	44	68
STI	0,57						
RASTI	0,61						
Sala 4. P4							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,48	1,39	1,22	1,2	1,24	1,14	1,02
EDT (s)	2,44	1,76	0,86	1,16	1,25	1,2	1,11
C <sub>80</sub> (dB)	-0,8	0,9	4,3	3,1	2	1,7	4,5
D (%)	23	47	58	47	41	44	60
STI	0,57						
RASTI	0,61						
Sala 4. P5							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,73	1,24	1,22	1,24	1,3	1,34	1,11
EDT (s)	2,95	1,56	1,13	1,19	1,24	1,12	1,11
C <sub>80</sub> (dB)	0,5	1	3,2	2,8	2,6	3,4	4,8
D (%)	15	33	38	47	47	48	55
STI	0,58						
RASTI	0,61						
Sala 4. P6							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	1,41	1,43	1,26	1,39	1,49	1,49	1,16
EDT (s)	1,87	1,2	1,18	1,39	1,28	1,22	1,12
C <sub>80</sub> (dB)	2,6	2,3	2,6	1,5	2,5	3,4	6,6
D (%)	58	37	50	34	46	49	75
STI	0,6						
RASTI	0,6						

Sala 5. P1							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	0,75	0,81	0,96	0,8	0,73	0,65	0,6
EDT (s)	0,95	1,06	0,87	0,78	0,71	0,56	0,47
C <sub>80</sub> (dB)	7,3	3,8	3,7	2,8	5,5	7,3	10,4
D (%)	51	56	56	47	60	70	85
STI	0,63						
RASTI	0,62						
Sala 5. P2							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	0,64	0,86	0,85	0,84	0,71	0,64	0,58
EDT (s)	1,75	1,58	1,47	1	0,95	0,78	0,34
C <sub>80</sub> (dB)	-0,6	1,4	-2	3,5	2,4	5,1	13,8
D (%)	41	28	29	44	42	57	83
STI	0,62						
RASTI	0,60						
Sala 5. P3							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	0,61	0,8	0,85	0,81	0,7	0,65	0,64
EDT (s)	1,38	0,84	1,43	0,81	0,79	0,66	0,58
C <sub>80</sub> (dB)	3,4	3,6	3,8	4	6	6,4	7,3
D (%)	48	56	60	57	65	68	68
STI	0,61						
RASTI	0,63						
Sala 5. P4							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
RT (s)	0,89	0,81	0,83	0,8	0,69	0,63	0,6
EDT (s)	1,16	1,23	1,32	0,78	0,74	0,67	0,64
C <sub>80</sub> (dB)	-0,3	1,7	3,9	4,2	5,7	6	8
D (%)	42	32	61	60	66	71	79
STI	0,65						
RASTI	0,66						

<b>Sala 6. P1</b>							
	<b>125 Hz</b>	<b>250 Hz</b>	<b>500 Hz</b>	<b>1 kHz</b>	<b>2 kHz</b>	<b>4 kHz</b>	<b>8 kHz</b>
<b>RT (s)</b>	0,65	0,39	0,41	0,31	0,37	0,43	0,29
<b>EDT (s)</b>	0,52	0,53	0,39	0,41	0,34	0,34	0,21
<b>C<sub>80</sub> (dB)</b>	5,8	9,4	9,7	11,6	12,9	13,7	21,4
<b>D (%)</b>	51	79	78	82	87	86	97
<b>STI</b>	0,79						
<b>RASTI</b>	0,79						
<b>Sala 6. P2</b>							
	<b>125 Hz</b>	<b>250 Hz</b>	<b>500 Hz</b>	<b>1 kHz</b>	<b>2 kHz</b>	<b>4 kHz</b>	<b>8 kHz</b>
<b>RT (s)</b>	0,41	0,5	0,4	0,36	0,35	0,48	0,31
<b>EDT (s)</b>	0,76	0,59	0,75	0,37	0,36	0,36	0,18
<b>C<sub>80</sub> (dB)</b>	7,1	7,6	7,9	10,8	11,8	12,9	22,8
<b>D (%)</b>	36	64	72	77	80	80	95
<b>STI</b>	0,79						
<b>RASTI</b>	0,78						
<b>Sala 6. P3</b>							
	<b>125 Hz</b>	<b>250 Hz</b>	<b>500 Hz</b>	<b>1 kHz</b>	<b>2 kHz</b>	<b>4 kHz</b>	<b>8 kHz</b>
<b>RT (s)</b>	0,52	0,53	0,4	0,36	0,35	0,41	0,39
<b>EDT (s)</b>	0,62	0,24	0,42	0,34	0,36	0,24	0,17
<b>C<sub>80</sub> (dB)</b>	8,9	11,8	14,4	13,3	15,3	15,2	19
<b>D (%)</b>	67	85	91	82	89	90	9
<b>STI</b>	0,83						
<b>RASTI</b>	0,83						
<b>Sala 6. P4</b>							
	<b>125 Hz</b>	<b>250 Hz</b>	<b>500 Hz</b>	<b>1 kHz</b>	<b>2 kHz</b>	<b>4 kHz</b>	<b>8 kHz</b>
<b>RT (s)</b>	0,61	0,42	0,41	0,39	0,36	0,43	0,38
<b>EDT (s)</b>	0,84	0,33	0,59	0,25	0,32	0,32	0,27
<b>C<sub>80</sub> (dB)</b>	8,8	9,2	12,8	15,4	14,7	13,8	17,1
<b>D (%)</b>	78	80	88	91	88	88	93
<b>STI</b>	0,82						
<b>RASTI</b>	0,83						

<b>Sala 7. P1</b>							
	<b>125 Hz</b>	<b>250 Hz</b>	<b>500 Hz</b>	<b>1 kHz</b>	<b>2 kHz</b>	<b>4 kHz</b>	<b>8 kHz</b>
<b>RT (s)</b>	0,35	0,28	0,17	0,16	0,22	0,19	0,13
<b>EDT (s)</b>	0,29	0,1	0,08	0,19	0,29	0,23	0,12
<b>C<sub>80</sub> (dB)</b>	14,9	16,9	26,4	29	22,4	23	34,2
<b>D (%)</b>	88	96	97	98	96	97	100
<b>STI</b>				0,89			
<b>RASTI</b>				0,89			

<b>Sala 7. P2</b>							
	<b>125 Hz</b>	<b>250 Hz</b>	<b>500 Hz</b>	<b>1 kHz</b>	<b>2 kHz</b>	<b>4 kHz</b>	<b>8 kHz</b>
<b>RT (s)</b>	0,39	0,23	0,16	0,14	0,18	0,17	0,15
<b>EDT (s)</b>	0,15	0,28	0,12	0,2	0,16	0,14	0,11
<b>C<sub>80</sub> (dB)</b>	17,9	20,8	29,5	31	26,7	27,6	35,5
<b>D (%)</b>	95	92	99	99	98	99	100
<b>STI</b>				0,92			
<b>RASTI</b>				0,91			