

**MESTRADO**

MULTIMÉDIA - ESPECIALIZAÇÃO EM MÚSICA INTERACTIVA E DESIGN DE SOM

# **A CONTRIBUIÇÃO DA ALTURA PARA A IMPRESSÃO ESPACIAL NO CONTEXTO DA PRODUÇÃO ÁUDIO**

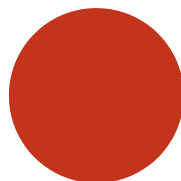
Joana Castro Rodrigues

**M**

**2019**

FACULDADES PARTICIPANTES:

**FACULDADE DE ENGENHARIA  
FACULDADE DE BELAS ARTES  
FACULDADE DE CIÊNCIAS  
FACULDADE DE ECONOMIA  
FACULDADE DE LETRAS**



**A contribuição da altura para a  
impressão espacial no contexto da  
produção áudio**

**Joana Castro Rodrigues**

Mestrado em Multimédia da Universidade do Porto

Orientador: Rui Luís Nogueira Penha (PhD)

Coorientador: Marco Paulo Barbosa da Conceição (PhD)

Julho de 2019



© Joana Rodrigues, 2019

# **A contribuição da altura para a impressão espacial no contexto da produção áudio**

**Joana Castro Rodrigues**

Mestrado em Multimédia da Universidade do Porto

Aprovado em provas públicas pelo Júri:

Presidente: Gilberto Bernardes de Almeida

Vogal Externo: Nuno Miguel da Costa Santos Fonseca

Orientador: Rui Luís Nogueira Penha



# Resumo

A presente dissertação é proposta num contexto de renovado interesse pela espacialização sonora e pela compreensão sobre as formas como esta pode ser utilizada no contexto da produção musical para formatos de reprodução multicanal. Desenvolvido em parceria com a Orquestra Jazz de Matosinhos, o presente estudo explora diferentes estratégias, sistemas e tecnologias de produção áudio de forma a compreender de que modo a adição de altifalantes de altura potencia a impressão espacial do ouvinte. Para melhor compreender estes conceitos, foi feita a avaliação comparativa entre formatos de reprodução com e sem recurso a altura através da análise subjectiva de dois formatos de captação aplicados a estes sistemas. Desta forma, este estudo pretende explorar, validar e documentar não só estratégias de produção musical mas também analisar a resposta subjectiva aos estímulos acústicos, focalizando-se na impressão espacial e na sua contribuição para a apreciação musical.



# Abstract

This dissertation is proposed in a context of renewed interest in sound spatialization and in the understanding of how it can be used in the context of musical production for multichannel reproduction formats. Developed in partnership with Orquestra Jazz de Matosinhos, the present study explores different strategies, systems and technologies of audio production in order to understand how the addition of height reproduction channels improves the spatial impression of the listener. To better understand these concepts, a paired comparison evaluation was made between reproduction setups - with and without the height layer - through the subjective assessment of two different recording techniques applied to these systems. In this way, this study intends to explore and report not only musical production strategies but also to analyze the subjective response to the acoustic stimuli, focusing on the spatial impression and its contribution to musical appreciation.



# Agradecimentos

Aos meus pais pelo carinho, paciência e confiança.

Ao meu irmão, a minha maior referência.

Aos meus orientadores, Rui Penha e Marco Conceição, por todo o apoio e incentivo.

A todas as reuniões de terça-feira pela inspiração, motivação e pelo saber.

Ao Bruno Castro pela ajuda inestimável.

Ao Gilberto Bernardes pela disponibilidade e ajuda.

A todos os meus amigos e ao Penas pelo apoio e horas de partilha.

A todos os que despenderam de um bocadinho do seu tempo para participar neste projecto.

À Orquestra de Jazz de Matozinhos

Ao CARA

À Escola Superior de Música e Artes do Espectáculo

Aos Serviços Audiovisuais – ESMAE

Joana Rodrigues

# Índice

<b>1. Introdução .....</b>	<b>1</b>
1.1 Contexto, Enquadramento e Motivação .....	2
1.2 Projecto .....	3
1.3 Questões e Objectivos de Investigação .....	3
1.4 Metodologia de Investigação .....	3
<b>2. Revisão Bibliográfica.....</b>	<b>5</b>
2.1 Psicoacústica do áudio espacial .....	5
2.1.1 Percepção e caracterização do espaço acústico .....	6
2.1.2 <i>Spaciousness</i> e Envolvência .....	7
2.1.3 Avaliação subjectiva dos atributos espaciais .....	7
2.1.4 Impressão espacial nos diferentes sistemas de reprodução.....	9
2.2 Sistemas e tecnologias de captação e reprodução multicanal .....	10
2.2.1 Estereofonia .....	11
2.2.2 Aparecimento dos formatos multicanal .....	12
2.2.3 Surround.....	13
2.2.4 Sistemas de reprodução imersivos: Auro 3D e Dolby Atmos .....	13
2.2.5 Ambisonics .....	15
2.2.6 Técnicas de captação estereofónicas para sistemas multicanal .....	16
2.2.7 <i>Soundfield</i> .....	18
2.3 Normas e Recomendações: Sistemas de reprodução e avaliação subjectiva de atributos sonoros.....	19
2.4 Trabalhos relacionados e projectos de referência .....	20
<b>3. Metodologia e Implementação.....</b>	<b>23</b>
3.1 Gravação .....	25

3.1.1	Descrição das matrizes de captação tridimensional.....	29
3.2	Edição, espacialização e ajuste de níveis .....	31
3.2.1	Plugins de descodificação, espacialização e de medição e ajuste de níveis .....	32
3.3	Sistema de reprodução .....	34
<b>4.</b>	<b>Avaliação .....</b>	<b>36</b>
4.1	Condições de teste e programa de excertos .....	36
4.2	Sujeitos de Avaliação .....	38
4.3	Configuração de Teste .....	39
4.4	Atributo .....	41
4.5	Análise estatística dos dados obtidos .....	42
4.6	Breve conclusão sobre os resultados.....	48
4.7	Comentários finais .....	49
<b>5.</b>	<b>Conclusões e Trabalho Futuro .....</b>	<b>51</b>
<b>6.</b>	<b>Referências .....</b>	<b>53</b>
<b>7.</b>	<b>Anexos .....</b>	<b>57</b>
7.1	Especificações técnicas do material de captação .....	57
7.2	Plugins de descodificação, espacialização e de medição e ajuste de níveis .....	59
7.3	Instruções e Questionário de Avaliação .....	62

# Lista de Figuras

<b>Figura 1:</b> “ <i>Proposed hierarchy of spatial attributes for use in subjective analysis (Mason, 1999)</i> ” (Rumsey, 2001, p.58)	8
<b>Figura 2:</b> Stage Plot da Orquestra de Jazz de Matosinhos nas sessões de gravação. Secção central: Bateria, contrabaixo, piano e guitarra eléctrica. O ponto “M” indica o local intermédio do maestro. O quadrado tracejado indica o posicionamento das matrizes de captação principal.	27
<b>Figura 3:</b> Array de captação OCT <i>surround</i> + 4 hipercardióides em altura e FOA <i>Soundfield</i> .	28
<b>Figura 4:</b> <i>Stage plot</i> com o array de captação OCT <i>surround</i> + 4 hipercardióides em altura e FOA <i>Soundfield</i> .	28
<b>Figura 5:</b> Matriz de captação OCT <i>surround</i> + 4 supercardioides: identificação dos pontos discretos de captação e orientação dos pontos de captação discretos (cubo superior). Matriz <i>Soundfiled</i> (cubo inferior): identificação do eixo frontal de captação e do posicionamento de em relação à matriz OCT 9.	29
<b>Figura 6:</b> Estúdio PTM Sul, Escola Superior de Música e Artes do Espectáculo	34
<b>Figura 7:</b> <i>Setup</i> para os testes de avaliação subjectiva.	39
<b>Figura 8:</b> Interface de teste. Os participantes podiam permutar entre os excertos A e B sem alteração na <i>timeline</i> musical. A interface permitia ainda fazer pausa na reprodução, reiniciar os excertos e avançar nos exemplos de forma a garantir a autonomia dos participantes durante o tempo de teste.	40
<b>Figura 9:</b> <i>Comparison Category Rating</i> (CCR) (Bech & Zacharov, 2006). Escala de avaliação utilizada para avaliação nos testes auditivos.	41
<b>Figura 10:</b> Representação gráfica das respostas dos 24 inquiridos sob as formas de <i>violin plot</i> e <i>boxplot</i> .	45
<b>Figura 11:</b> Percentagem de respostas por exemplo.	46
<b>Figura 12:</b> Distribuição percentual do total de respostas sobre a avaliação da contribuição da altura na impressão espacial a partir da técnica OCT.	47
<b>Figura 13:</b> Distribuição percentual do total de respostas sobre a avaliação da contribuição da altura na impressão espacial a partir da técnica <i>Soundfield</i> .	47
<b>Figura 14:</b> SoundField by RØDE	59

<b>Figura 15:</b> IEM Plug-in Suite: AIIRA Decoder e SimpleDecoder	60
<b>Figura 16:</b> Reaper ReaSurround (Cockos)	61
<b>Figura 17:</b> Waves Loudness Meter	61

# Lista de Tabelas

<b>Tabela 1:</b> Especificações do formato de reprodução	35
<b>Tabela 2:</b> Lista de exemplos apresentada aos sujeitos de avaliação	37
<b>Tabela 3:</b> Definição do atributo, sub-atributos e termos descritivos comuns relativos ao atributo principal “Impressão Espacial”	41
<b>Tabela 4:</b> Resultados estatísticos para cada exemplo segundo os testes t-Student e Wilcoxon	42
<b>Tabela 5:</b> Teste Shapiro-Wilk	43
<b>Tabela 6:</b> Valores estatísticos descritivos: Média e Mediana	43



# Abreviaturas e Símbolos

<b>ASW</b>	Apparent Source Width
<b>DAW</b>	Digital Audio Workstation
<b>EBU</b>	European Broadcasting Union
<b>EDT</b>	Early Decay Time
<b>FOA</b>	First-Order Ambisonic
<b>HOA</b>	High-Order Ambisonic
<b>IEM</b>	Institute of Electronic Music and Acoustics
<b>IRT</b>	Institute of Radio Technology
<b>ITU</b>	International Telecommunication Union
<b>LFE</b>	Low Frequency Effects
<b>LUFS</b>	Loudness Units, referenced to Full Scale
<b>MS</b>	Middle-Side
<b>OCT</b>	Optimum Cardioid Triangle
<b>OJM</b>	Orquestra Jazz de Matosinhos
<b>ORTF</b>	Office de Radiodiffusion-Télévision Française
<b>WFS</b>	Wave-Field Synthesis





# 1. Introdução

Até muito recentemente, a gravação e reprodução musical focou-se essencialmente em sistemas estereofónicos ou *surround*, reproduzindo apenas o campo sonoro no plano horizontal. A reprodução imersiva, ou tridimensional, faz uso de canais de altura que melhoram a percepção dos atributos espaciais trazendo ao ouvinte um ambiente de reprodução sonora que se aproxima da sua forma natural de escuta. O ouvido humano avalia as várias propriedades do campo sonoro e utiliza a interpretação das suas diversas componentes para a audição espacial. (Blauert, 1997; Theile & Wittek, 2011).

Os formatos áudio tridimensionais oferecem um grande potencial no que toca à recriação de performances musicais e à simulação de profundidade, realismo e envolvimento. Inúmeras técnicas de captação foram já desenvolvidas e testadas relativamente à reprodução do espaço sonoro em sistemas sem recurso a altura. No entanto, quando comparado ao espectro do áudio tridimensional, é de notar que poucas das técnicas desenvolvidas para estes sistemas foram formalmente testadas – quer por métodos objectivos como subjectivos –, deixando assim uma área ainda por explorar no âmbito da produção musical para sistemas imersivos (Howie, Martin, Benson, Kelly, e King, 2018; Howie, 2018).

A indústria áudio encontra-se neste momento apta para a utilização de sistemas de reprodução *surround*, seja em produção, distribuição ou até mesmo em formatos de reprodução para consumo, existindo maior aceitação por parte dos consumidores perante *setups* de reprodução com maior número de altifalantes. Apesar disso, no que toca a formatos de *home-cinema*, existem dificuldades em recriar a qualidade sonora que este tipo de sistemas é capaz de reproduzir, quer seja pelos espaços acústicos de reprodução não serem os mais adequados, como pela montagem do sistema não ser a mais apropriada (Theile & Wittek, 2012). Para além disso, o acesso a sistemas de reprodução multicanal para configurações de *home-cinema* é também ainda pouco frequente e dispendioso. No entanto, a utilização deste tipo de formato tem uma vasta área de aplicações e o áudio imersivo começa a ganhar impulso em domínios como o cinema, realidade virtual, *gaming* e até mesmo na produção musical.

O seguinte projecto de dissertação surge no contexto da exploração de métodos de espacialização sonora tridimensional aplicados a novos contextos de produção musical. A argumentação estende-se ainda ao estudo de estratégias de captação para difusão multicanal e ao entendimento sobre a contribuição de camadas de reprodução em altura na percepção do espaço e na experiência musical.

## 1.1 Contexto, Enquadramento e Motivação

Ouvir música orquestral ao vivo é uma experiência única em termos de gama dinâmica musical e pela extensão do palco sonoro criado. Estes factores têm uma influência directa na percepção da componente acústica do local do concerto e na forma como interpretamos o conteúdo musical. A música orquestral tem mantido grande popularidade, não só no contexto de concertos ao vivo, mas também pela sua utilização em bandas sonoras no contexto do cinema e jogos. A gravação deste tipo de conteúdo, ou ensemble, tem sido, até muito recentemente, principalmente explorada para sistemas convencionais estereofónicos e *surround*. Estes formatos reproduzem o campo sonoro apenas no plano horizontal, comprometendo a experiência comparativamente à prática auditiva ao vivo. Isto traduz-se numa representação do espaço acústico de forma menos imersiva e realista em termos espaciais. A reprodução sonora tridimensional, com canais de altura orientados verticalmente, traz o ouvinte para uma experiência auditiva mais próxima da real (Kim, Ko, Nagendra & Woszczyk, 2013; Howie, 2018).

A análise e compreensão sobre metodologias de produção musical para sistemas imersivos pode estender-se neste sentido para diversas áreas de aplicação. Desta forma, torna-se pertinente o estudo sobre estratégias de produção musical para sistemas imersivos, aplicados a diferentes contextos musicais, desde o momento de captação até ao sistema de reprodução final. Tendo em conta estes elementos e a investigação preliminar do estado da arte, fica clara a necessidade de desenvolver e efectuar estudos empíricos no âmbito de formatos de produção para áudio imersivo, bem como compreender de que forma o tipo de sistema de captação explorado influencia a percepção espacial dos ouvintes. Este projecto pretende compreender e documentar de que forma as diferentes técnicas se distinguem entre si (pelas suas diferentes tipologias), quais as suas características subjectivas de percepção e quais os métodos de mistura e renderização adequados. Os resultados deste estudo irão permitir obter informações sobre a implementação de matrizes de microfones tridimensionais com o intuito de recolher dados que, por sua vez, validem, esquematizem e contribuam para a compreensão da impressão espacial neste contexto.

## 1.2 Projecto

Este projecto foi desenvolvido em parceria com a Orquestra Jazz de Matosinhos, o que possibilitou a exploração de métodos de produção e especialização musical a partir da gravação de novo material e da experimentação de formatos de reprodução com e sem recurso a altura. A presente investigação torna-se relevante não só pelo seu contexto específico, mas também pela exploração e definição de metodologias aplicadas ao âmbito da produção áudio. As fases de desenvolvimento do projecto serão apresentadas na secção 1.4.

## 1.3 Questões e Objectivos de Investigação

1. De que forma a adição de camadas de altura em sistemas de reprodução multicanal influencia a percepção de espaço do ouvinte?

2. Quais as estratégias, metodologias e técnicas de produção áudio - desde o momento de captação até ao sistema de reprodução – que melhor beneficiam a reprodução sonora tridimensional?

### **Objectivos:**

São inúmeros os estudos práticos desenvolvidos no que toca à criação, reprodução e percepção do espaço sonoro em sistemas *surround*. Existe a necessidade de desenvolver e concretizar de forma igualmente extensiva estudos do mesmo âmbito aplicados a sistemas imersivos. Através da implementação prática e da análise subjectiva entre sistemas de captação e reprodução, pretendemos que este projecto possa, por um lado, contribuir para a apreciação e compreensão deste tipo de sistemas e, por outro, explorar e documentar as estratégias de produção musical aplicadas a este contexto.

## 1.4 Metodologia de Investigação

Através de uma análise das áreas de estudo envolvidas, a presente investigação parte de um processo de revisão bibliográfica de forma a aprofundar conceitos e compreender de que forma poderá contribuir para a expansão do conhecimento. Os principais focos de investigação são fundamentados após uma primeira fase de revisão da literatura onde foram definidas as metodologias de desenvolvimento e implementação do projecto.

O principal foco de pesquisa é determinado pela necessidade de expandir o conhecimento sobre a prática e promover o conhecimento a partir da mesma. Neste sentido, o *design* metodológico utilizado para o desenvolvimento do presente projecto traduz-se numa *practice-led research* a partir da qual existe a exploração do conhecimento que se traduz numa maior compreensão sobre a prática. Esta tipologia de investigação inclui a execução prática como parte integrante do seu método e enquadra-se na área de pesquisa-acção (Candy, 2006).

Após a análise e validação teórica e prática da investigação, as fases de desenvolvimento dividem-se da seguinte forma:

- Selecção das técnicas de captação tridimensionais a implementar;
- Gravações de Orquestra Jazz de Matosinhos;
- Fase de experimentação sobre metodologias de mistura e espacialização do conteúdo gravado;
- Espacialização sonora para sistema de reprodução multicanal com e sem a altura;
- Avaliação / comparação subjectiva do atributo sonoro: impressão espacial;
- Análise dos resultados obtidos;
- Conclusões.

## 2. Revisão Bibliográfica

No presente capítulo são apresentados alguns conceitos para uma melhor contextualização teórica e compreensão prática do projecto. Foi a partir da revisão literária, bem como através da análise a trabalhos relacionados, que foi definida a estratégia para o desenvolvimento e concepção do trabalho final proposto.

Este capítulo divide-se em quatro grandes secções referentes à psicoacústica do áudio espacial, aos sistemas e tecnologias de captação e reprodução multicanal, normas e recomendações sobre sistemas de reprodução e análise subjectiva de atributos áudio e, por último, trabalhos relacionados. A primeira secção é referente à cognição do áudio espacial fazendo a ponte para os formatos de produção e reprodução sonora a partir de uma breve contextualização história e apresentação de sistemas aplicados ao áudio imersivo. Após esta secção, são apresentadas as normas de implementação dos sistemas de reprodução explorados, bem como as recomendações que auxiliaram a conceptualização dos testes perceptuais. A última secção faz referência a alguns projectos relacionados que exploram as temáticas abordadas nas secções anteriores e que partem de metodologias idênticas à adoptada para o presente projecto.

### 2.1 Psicoacústica do áudio espacial

Nesta secção é feito um breve enquadramento relativo à percepção espacial e aos parâmetros de avaliação física e subjectiva dos atributos espaciais. Tendo em conta o âmbito do projecto de investigação, é também apresentada uma secção referente aos métodos de avaliação qualitativa e preferencial sobre os parâmetros de impressão espacial.

### 2.1.1 Percepção e caracterização do espaço acústico

*“If we want to understand spatiality (...), we must understand how the listener’s mental processes give rise to the experience of meaning.”* (Kendall, 2010)

O que o ouvinte interpreta no momento em que percebe o som pode ser descrito como uma série de fenómenos que ocorrem no espaço. Segundo Kendall (2010), os *sound events* e os seus constituintes são vivenciados através de múltiplas modalidades sensoriais e o entendimento destas experiências como eventos é fundamentalmente multimodal. Um *sound event* está directamente associado à ocorrência do som no mundo físico enquanto que um *auditory event* se refere à representação da percepção auditiva. A sequência ou simultaneidade de *auditory events* percebidos por um ouvinte define-se como uma *auditory scene* (Spors et al., 2013). O ouvido humano utiliza diversos mecanismos através dos quais localiza e interpreta os sons, mecanismos esses que dependem da decomposição frequencial dos eventos sonoros, de detecção de diferenças temporais e de amplitude, bem como da informação derivada da reverberação e do espaço onde estes ocorrem (Gerzon, 1974; Blauert, 1997; Rumsey, 2001). O presente estudo foca-se na percepção do conteúdo espacial e sobre de que forma o ouvinte interpreta subjectivamente essa informação.

Enquanto os parâmetros físicos para a caracterização de espaços acústicos se dedicam à definição de propriedades objectivas, a caracterização subjectiva da percepção destes parâmetros estuda a nossa resposta perceptual aos estímulos acústicos. São vários os parâmetros objectivos<sup>1</sup> - ou físicos - que nos permitem avaliar e caracterizar um espaço acústico. No entanto, as principais características e atributos subjectivos de um campo sonoro resultam da percepção do ouvinte relativamente a esse mesmo espaço (Henrique, 2011). A percepção e cognição do áudio espacial podem ser caracterizadas a partir de diversos atributos psicoacústicos que permitem a caracterização subjectiva do espaço. Estes parâmetros de caracterização do campo sonoro são aplicados à avaliação qualitativa e/ou preferencial do conteúdo espacial e focam-se na cognição da experiência sonora. Autores como Rumsey (2001, 2002) Berg e Rumsey (2003), Toole (2008), Kendall (2010) e Henrique (2011) identificam e descrevem alguns destes atributos. Apesar das diferentes contribuições, a terminologia utilizada para a descrição dos parâmetros de percepção subjectiva dos espaços acústicos não é sistemática.

---

<sup>1</sup> Tempo de reverberação (T60), tempo de decaimento inicial (EDT), clareza objectiva (*early-to-late sound index - C80*), initial-time delay gap (ITDG), intensidade, bass ratio (BR), coeficiente de intercorrelação binaural (*interaural cross-correlation coefficient - IACC*) e o método RASTI (*Rapid Speech Transmission Index*) (L. Henrique, 2011, p. 788-790)

## 2.1.2 *Spaciousness* e Envolvência

A percepção espacial baseia-se em diferentes atributos dos eventos sonoros, incluindo dois principais: reverberância e *auditory spaciousness*. A reverberância é o sentido da reverberação percebido num espaço ou gravação, dependendo do tempo de reverberação e do EDT (Henrique, 2011). A *auditory spaciousness* relaciona-se com a dispersão dos eventos auditivos, sendo este efeito essencialmente fruto das *early lateral reflections* (Conceição, 2015). Os termos *spaciousness* e envolvência surgem recorrentemente para descrever as propriedades espaciais de sistemas de reprodução sonora. Estes estão primariamente relacionados à impressão espacial do ambiente e são em grande parte resultado do som refletido<sup>2</sup> (Rumsey, 2001; Toole, 2008).

Segundo Rumsey (2001), *Spaciousness* é a expressão utilizada com mais frequência para descrever “*the sense of open space or room*” (p.38) no qual o ouvinte está localizado, relacionando-se com a percepção do espaço sonoro à volta da cabeça, não se restringindo a uma região próxima da mesma. Segundo o autor, envolvência difere deste termo por se relacionar com “*the sense of immersivity and involvement in a (reverberant) soundfield*” (p.38). Desta forma, se existir uma boa envolvência, o ouvinte sente que a quantidade de energia proveniente de todas as direcções é perceptualmente igual. Por outro lado, Toole (2008) refere *spaciousness* e envolvência como sinónimos, na medida em que ambas se referem à percepção de se estar rodeado por um espaço amplo e envolvente. O autor refere ainda que *spaciousness*, ou impressão espacial, é composta por duas componentes perceptuais principais, nomeadamente largura aparente da fonte (ASW)<sup>3</sup> e envolvência<sup>4</sup>.

Torna-se importante referir que a impressão espacial numa sala de concerto ao vivo poderá - e provavelmente irá - diferir da impressão espacial de um sistema de reprodução sonora.

## 2.1.3 Avaliação subjectiva dos atributos espaciais

Apesar da existência de atributos subjectivos para a impressão espacial, a definição e verbalização por parte do ouvinte torna-se por vezes bastante complexa pelo seu carácter subjectivo e sensorial. A este problema acresce o facto de diferentes ouvintes utilizarem

---

<sup>2</sup> Several of the measures related to pleasurable perceptions in concert halls are related to the angles of incidence of reflected sounds relative to the direct sounds from the stage. In particular, those reflections arriving from the sides have been found to be especially useful contributors to what was originally called “spatial impression.” (Toole, 2008, p. 50)

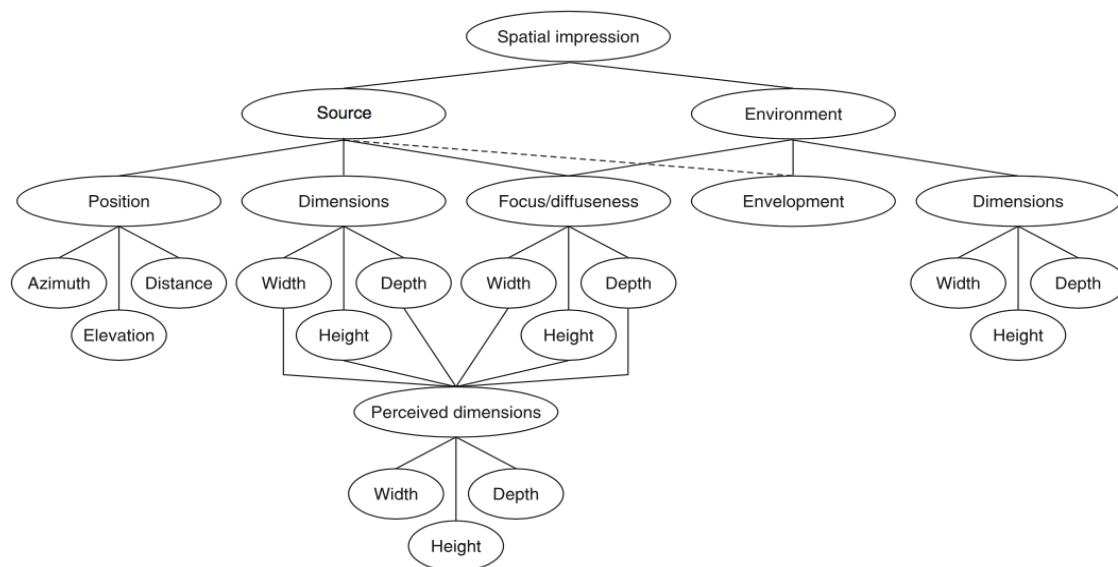
<sup>3</sup> “*Apparent Source Width (ASW), a measure of perceived broadening of a sound image whose location is defined by direct sound. In live performances, it is the auditory illusion of a sound source that is wider than the visible sources;*” (Toole, 2008, p. 34)..

<sup>4</sup> “*Listener envelopment (LEV) is a sense of being in a large space, of being surrounded by a diffuse array of sounds not associated with any localizable sound images. This is regarded as perhaps the more important component of spaciousness (...)*” (Toole, 2008, p. 34).



terminologia diferente para a designação dos mesmos eventos. Como já referido na secção anterior, são vários os autores que identificam terminologia para avaliar a impressão espacial através da definição de diversos atributos. Apesar de existir uma extensa investigação ao longo dos anos relativamente à impressão subjectiva de espaços acústicos, os termos utilizados para descrever estes fenómenos ainda não estão formalmente sistematizados, existindo variações na nomenclatura apresentada em vários projectos deste âmbito.

Contudo, numa análise subjectiva das características do som reproduzido, pode ser vantajoso criar uma organização hierárquica para os diferentes atributos espaciais (Rumsey, 2001). Esta organização medeia a comunicação dos significados concedidos aos diferentes atributos de forma a facilitar o estudo sobre os atributos físicos e a sua percepção subjectiva. Mason (1999) apresenta uma proposta de organização hierárquica destes termos descritivos para os atributos sonoros que poderá servir como um ponto de partida.



**Figura 1:** “Proposed hierarchy of spatial attributes for use in subjective analysis (Mason, 1999)” (Rumsey, 2001, p.58)

Na secção 2.3 serão apresentadas duas recomendações protocolares sobre metodologias de avaliação subjectiva de atributos sonoros. A partir das mesmas é feita a descrição pormenorizada do atributo impressão espacial (através da definição de sub-atributos e termos descritivos comuns) e da sua aplicação prática no contexto de testes perceptuais subjectivos.

#### 2.1.4 Impressão espacial nos diferentes sistemas de reprodução

No caso dos formatos de reprodução estereofónicos de dois canais, a impressão espacial tem de ser necessariamente recriada a partir de uma perspectiva bidimensional entre os dois altifalantes frontais. Já no caso do formato estereofónico 3/2, o ambiente acústico pode ser moldado através do uso de altifalantes *surround* adicionais - o mesmo se aplica ao sistema *surround 5.1*. Adicionar estes novos canais permite a reprodução das *early lateral reflections*, o que se traduz numa percepção mais realista da distância e profundidade do campo sonoro. Desta forma, a partir da reprodução das reflexões e reverberação por via dos altifalantes que se encontram fora da imagem frontal estereofónica, é possível conseguir uma impressão espacial mais natural. Estes são colocados em torno do ouvinte, lateralmente ou atrás do mesmo (Theile, 2001).

Um sistema de reprodução tridimensional procura criar um campo sonoro natural que se aproxime da nossa percepção sonora real. Um dos principais aspectos que molda a nossa consciência e apreciação dos sons consiste nas reflexões tridimensionais que se tornam cruciais para o nosso cérebro interpretar e analisar o campo sonoro. Esta componente é essencial para a percepção de vários atributos sonoros, nomeadamente o timbre. Existem bastantes diferenças relativamente à perspectiva psicoacústica entre as dimensões vertical e horizontal, tais como a eficácia ou acuidade com que o ouvinte consegue avaliar a localização espacial de uma fonte sonora. O ouvido consegue compreender com grande facilidade e resolução as variações na localização das fontes sonoras no eixo horizontal, o mesmo não acontece no eixo vertical. (Blauert, 1997; Kendall, 1995)

The related psychoacoustic principles should be understood as phenomena of spatial hearing governed by specific laws and thus requiring suitable types, configurations and locations of microphones, as well as distinct handling of delay, interchannel correlation and level balancing of direct / indirect sound. (Theile, 2001, p.4)

Tendo em conta o contexto prático do presente projecto, o conceito de impressão espacial e a descrição detalhada dos principais componentes perceptuais relativos a este atributo subjectivo serão descritos em maior detalhe no capítulo 3.

## 2.2 Sistemas e tecnologias de captação e reprodução multicanal

Existem três métodos para a reprodução de ambientes imersivos, nomeadamente *channel-based*, *object-based* e *scene-based*. Cada um destes conceitos tem diferentes características, diferenciando-se essencialmente na forma como o conteúdo áudio é transmitido.

O método *channel-based* é uma forma de representação na qual os diversos elementos sonoros são misturados para um número pré-definido de canais de reprodução. Cada canal é reproduzido através da sua associação a um altifalante com uma posição estática e pré-definida. Este formato é dos mais explorados na indústria da transmissão áudio pela predominância de formatos como o estéreo e *surround*. A adaptação de conteúdo representado através deste método, implica a aplicação de técnicas de *up* e *down mixing*<sup>5</sup>. No método *object-based* os componentes individuais são tratados separadamente e acompanhados por metadados que descrevem a correlação entre os mesmos. Cada fonte é tratada como uma entidade individual à qual se dá o nome de *object*. Este sistema permite, a partir de um processo de renderização dedicado e tendo em conta o número de canais de reprodução existentes, definir quais serão utilizados para a recriação do *object* sonoro, o que torna esta tecnologia independente de um formato de reprodução fixo. Os metadados podem variar ao longo do tempo para, por exemplo, alterar a posição espacial de um elemento. O sistema Dolby Atmos é um exemplo de uma tecnologia que aplica este método. Já o áudio *scene-based* faz a representação do conteúdo através de um conjunto de sinais de coeficiente. Este método aproxima-se de uma combinação dos dois anteriormente referidos. Semelhante ao método *channel-based*, a mistura é feita para um número de canais áudio fixo. Pelo contrário, e neste sentido aproximando-se do método *object-based*, os sinais não estão directamente associados a um formato de reprodução específico. A partir deste sistema, o conteúdo é transmitido sobre a forma de harmónicos esféricos que descrevem o campo sonoro e as suas propriedades temporais e espaciais, tornando o conteúdo independente do número e da posição dos canais de reprodução. Um exemplo de áudio baseado em cenas é o HOA (Peters, et al., 2015; International Telecommunication Union [ITU], 2018).

O presente projecto consiste na exploração e comparação de uma técnica de captação esterofónica multicanal e de um microfone *Soundfield*, baseando-se, respectivamente, em dois dos métodos anteriormente referidos: *channel-based* e *scene-based*.

De forma a melhor contextualizar os formatos e fundamentos explorados na fase de concepção e implementação prática, a secção 2.2 faz a apresentação e enquadramento dos diferentes sistemas de captação e reprodução sonora.

---

5

### 2.2.1 Estereofonia

Os sistemas de reprodução estereofónicos de dois canais são o formato mais simples e comum para a transmissão de conteúdo espacial tanto na gravação como na reprodução sonora. Os primeiros trabalhos sobre reprodução direccional no Bell Labs, no início da década de 30, produziram avanços significativos na captação e reprodução estereofónicas. Alan Blumlein mostrou que, a partir da introdução de diferenças de amplitude entre um par de altifalantes, seria possível criar diferenças de fase entre as orelhas semelhantes às da nossa audição natural. As técnicas de captação estereofónicas são introduzidas por Blumlein, em 1931, a partir dos primeiros sistemas de captação estereofónicos de cápsula coincidente. Estas resultam da aproximação das diferenças de fase das fontes originais nos ouvidos quando reproduzidas em altifalantes. Estas técnicas baseiam-se na capacidade de interpretação e na relação de distância entre os dois ouvidos humanos relativamente à discriminação e localização dos diferentes sons provenientes de várias direcções. Na mesma década, um sistema estereofónico de três canais foi apresentado por Steinberg e Snow em 1933, partindo da simplificação prática de um sistema teórico composto por um número infinito de canais. Este sistema envolveu tentativas de aproximar a representação de uma frente de onda sonora a partir de um número reduzido de canais. Steinberg e Snow revelaram que a partir da utilização de três canais obtinham-se resultados bastante convincentes e que, ao reduzir o número de canais de três para dois, as fontes centrais pareciam recuar em direcção à parte de trás do palco, resultando num aumento da largura do palco sonoro reproduzido. O trabalho de Steinberg e Snow foi principalmente destinado à reprodução sonora de grandes palcos sonoros. É interessante notar que, embora a aplicação de três canais frontais não tenha sido muito utilizada em sistemas de reprodução para formatos de consumo, são hoje em dia a norma na reprodução sonora para cinema, em parte, devido à grande variedade de posições dos espectadores. Neste sentido, o canal central tem o capacidade de estabilizar a imagem frontal, importante para os ouvintes descentralizados, isto é, fora da zona ideal de escuta. (Streicher & Everest, 1998; Rumsey, 2001; Penha, 2014).

Blumlein desenvolve os primeiros sistemas de captação estereofónicos com microfones coincidentes (1931). De forma a reconstruir o palco sonoro estereofónico, a utilização destas técnicas explora a captação das diferenças de intensidade das fontes sonoras no momento em que chegam às cápsulas. Estas técnicas partem da intenção de criar uma ilusão acústica virtual entre dois altifalantes e, por vezes, tentar estender esta ilusão para além do posicionamento e extensão física dos mesmos, sendo capaz de produzir *phantom sources*<sup>6</sup>. As diferenças de intensidade captadas por cada cápsula tornam-se importantes na criação de pistas interaurais para a percepção da ilusão estereofónica, a partir das quais é possível captar e reproduzir as diferenças de cada fonte sonora colocando-as desta forma no palco sonoro reproduzido. O

---

<sup>6</sup> Percepcionamos a existência de uma fonte sonora virtual num ponto onde não existe um canal de reprodução físico. Percebemos tais fontes fantasma pelas relações de nível e tempo entre os diferentes canais de reprodução.

primeiro sistema de captação proposto pelo autor é conhecido como *Blumlein* e consiste na utilização de dois microfones bidireccionais sobrepostos de forma coincidente formando um ângulo de 90° entre eles. Na mesma patente, o autor apresenta a técnica *Mid-side*, composta pela sobreposição de um microfone bidireccional, apontado para os lados, e um omnidireccional ou direccional em direcção à fonte sonora. A técnica *Mid-side* pressupõe a utilização de uma matriz de descodificação onde é depois feita a combinação dos diferentes sinais. Uma das técnicas estereofónicas coincidentes mais conhecidas e comuns é a XY, composta por dois microfones cardióide dispostos num ângulo entre 90° a 135°.

Os sistemas de captação estereofónicos passam também pela exploração de técnicas não-coincidentes ou de pares espaçados. Estes formatos captam e reproduzem não só diferenças de intensidade mas também diferenças temporais. O formato mais comum neste âmbito é o AB, tipicamente composto por dois microfones omnidireccionais em que o espaçamento entre os mesmos poderá variar consoante a largura do palco sonoro. Quanto maior o espaçamento entre microfones, maior será o extensão do palco estéreo. A técnica Decca Tree parte do formato AB à qual é adicionada um microfone central (Streicher & Everest, 1998; Rumsey, 2001; Owsinski, 2009; Conceição, 2015; Penha, 2014).

A escolha da técnica de captação, bem como o padrão de polaridade, são de grande importância tanto para a reprodução da imagem estereofónica como para a impressão espacial obtida do campo sonoro. A compreensão destes princípios e técnicas torna-se necessária para o entendimento dos novos formatos e tecnologias aplicados aos sistemas multicanal.

## 2.2.2 Aparecimento dos formatos multicanal

Em 1941, é adicionado ao sistema de três canais frontais um canal traseiro dedicado a efeitos sonoros. Este formato de quatro canais, "*Fantasound*", utilizado pela Disney no famoso filme *Fantasia*, não foi muito adoptado até à década de 60 quando a Dolby Stereo se torna o formato *surround standard*. Tal como o sistema anteriormente referido, este também utiliza quatro canais - *left*, *center*, *right* e *surround* mono (LCRS). A vantagem deste formato é a sua adaptabilidade aos sistemas estéreo 2.0 e mono se nenhum descodificador estiver presente. Por volta da década de 70, surge associado particularmente à indústria musical, o sistema quadrifónico de reprodução. A quadrifonia nasce como um formato natural de extensão do estéreo 2.0, prometendo melhorar este mesmo a partir da adição de dois canais traseiros. Apesar da tentativa, o sistema acaba por se tornar obsoleto pelas suas fragilidades psicoacústicas. Com o aparecimento dos formatos de distribuição digital nos anos 80, o número de canais traseiros aumentou para dois, aos quais é também adicionado um canal de efeitos de baixa frequência (LFE). Assim, é constituído o formato *standard 5.1* que rapidamente se torna o formato padrão

para cinema e que se estende até aos dias de hoje, sendo também amplamente utilizado em sistemas de *home cinema*, televisão digital e para reprodução musical. (Owsinski, 2006)

### 2.2.3 Surround

O standard ITU-R BS.775-1 é lançado em 1992 e traz novas características: um sistema de reprodução que se aproxima de um ambiente acústico natural a partir do aumento da área de reprodução e, conseqüentemente, de audição. Este formato melhora a estabilidade da imagem estereofónica central e, dentro de algumas limitações, permite a criação de um ambiente acústico realista pela adição de dois pontos de reprodução traseiros - no caso do 7.1 com o acréscimo de dois canais laterais em relação ao ouvinte. Nos últimos anos, foram introduzidas melhorias e sistemas inovadores que vão desde vários formatos 7.1 até aos sistemas HOA e WFS<sup>7</sup>. No entanto, o aparecimento de novos formatos exige novos paradigmas, novo hardware e atenção especial no momento da gravação para esta tipologia de sistemas de reprodução (Theile & Wittek, 2012). É importante compreender que todos estes formatos multicanal surgem a partir da estereofonia ao tentar transferir algumas das suas mais importantes características, tal como a criação de *phantom sources* entre dois altifalantes adjacentes para a geração de imagens virtuais no campo sonoro. Tal como nos sistemas estereofónicos, a direcção destas fontes sonoras depende da posição de escuta - *sweet spot*. Neste sentido, acrescentar mais canais no plano horizontal ajuda a ampliar a zona de escuta, proporcionando uma resolução direccional mais homogénea e estável. No entanto, existem formas de utilizar estes canais adicionais fora do plano horizontal, nomeadamente, quando os colocarmos verticalmente em relação à cabeça do ouvinte.

### 2.2.4 Sistemas de reprodução imersivos: Auro 3D e Dolby Atmos

O método utilizado para a reprodução de um ambiente tridimensional pode ser tecnicamente bastante variado. Desde sistemas *channel-based* – como os formatos originais Auro 3D –, *object-based* – como o Dolby Atmos –, ou ainda *scene-based* – como a tecnologia Ambisonics (Peters, et al., 2015; Auro Technologies, 2016; I.T.U., 2018). Sinais *channel-based* requerem uma configuração de reprodução onde o número e as posições dos altifalantes estão bem definidos. Sinais *object-based*, bem como sinais *scene-based*, podem ser reproduzidos

---

<sup>7</sup> A Wave Field Synthesis é capaz de criar um área de escuta consideravelmente grande através da síntese ou recriação de frentes de onda. As propriedades espaciais da cena acústica podem ser percebidas corretamente por um número arbitrariamente grande de ouvintes, independentemente da sua posição dentro dessa área. Devido ao elevado número de altifalantes necessários, a WFS torna-se muitas vezes impraticável.

através de sistemas de reprodução configurados para canais *channel-based* ou através de matrizes compostas por altifalantes adicionais (ITU, 2018).

Os formatos *surround* diferem dos formatos imersivos no número de dimensões de reprodução sonora. Na prática, os formatos *surround* estão posicionados no plano horizontal em torno do ouvinte, permitindo que o som se propague lateralmente como também no eixo frente/trás. Os formatos imersivos adicionam uma terceira dimensão: a altura. Desta forma, a adição de altifalantes no eixo vertical complementa e enriquece a área espacial, permitindo a criação de um ambiente sonoro tridimensional. O primeiro sistema de reprodução perifónica<sup>8</sup> a ser anunciado pertence a G. Cooper, consistindo num sistema tetraédrico de quatro canais (Gerzon, 1973). Gerzon (1973) apresenta alguns dos primeiros formatos perifónicos de reprodução e captação sonora que antecedem e influenciam a exploração de algumas das mais recentes tecnologias aplicadas aos sistemas imersivos. É importante notar que um sistema de reprodução tridimensional não é apenas criado a partir do som directo proveniente dos altifalantes de altura. Este tipo de formato requer uma exploração da informação espacial criada pelas reflexões em torno do ouvinte. Este tipo de informação não pode ser apenas reproduzida ou recriada através da colocação de altifalantes de altura, mas a partir da sua combinação com as restantes camadas de reprodução.

Em 2006, Wilfried Van Baelen introduz o formato Auro 9.1, desenvolvido a partir do formato *standard* 5.1 ao qual acrescenta quatro canais para informação de altura. Estes quatro altifalantes superiores estão posicionados nos pontos correspondentes em altura dos dois canais traseiros *surround* – LS/RS –, e dos dois canais frontais – L/R. É claro que esta extensão do formato pode ser transposta para configurações semelhantes, tal como o 7.1 *surround*. Esta disposição cubóide dos altifalantes possibilita a inclusão de uma nova dimensão do espaço sonoro, permitindo tanto a reprodução das primeiras reflexões provenientes do tecto como a inclusão da difusão espacial da reverberação. O Auro 9.1 estabelece um novo paradigma para diversos parâmetros tais como envolvimento, impressão espacial e profundidade (Theile & Wittek, 2012). A combinação das configurações de reprodução com a tecnologia Auro-3D permite a utilização de formatos desde o Auro 8.0 até ao Auro 13.1. O sistema AuroMax adiciona a tecnologia *object-based* e explora sistemas com mais canais de reprodução (Auro Technologies, 2015).

A tecnologia Dolby Atmos foi anunciada em Abril de 2012 e rapidamente se tornou um formato *standard* no som para cinema, sendo aplicável a todas as configurações de reprodução anteriormente referidas. Este sistema, por se basear no método *object-based*, torna mais flexível a adição de objectos áudio e a colocação de sons no plano superior de reprodução. Esta tecnologia permite que cada fonte seja tratada como uma entidade individual à qual se dá o nome de objectos. O processador do sistema Dolby Atmos determina, dentro do número de

---

<sup>8</sup> Sistema de reprodução com recurso aos eixos vertical e horizontal em torno do ouvinte.

canais de reprodução existentes, quais serão utilizados para a recriação do objecto sonoro, tornando esta tecnologia independente de um formato de reprodução fixo. A partir de um mecanismo dedicado de renderização flexível, este sistema otimiza a qualidade áudio e os efeitos *surround* de acordo com o formato de reprodução e as características da sala. Para além disso, o Dolby Atmos foi projectado desde o início para manter a compatibilidade com versões anteriores e minimizar o impacto nos fluxos de trabalho atuais de produção e distribuição (Dolby Laboratories, 2016).

### 2.2.5 Ambisonics

Uma vez que uma das técnicas de captação implementadas no presente projecto é um microfone *Soundfield* de primeira ordem, a presente secção introduz em maior detalhe a tecnologia adoptada para a abordagem de espacialização do conteúdo gravado.

A tecnologia Ambisonics surge no início dos anos 70 quando Michael Gerzon (Gerzon, 1973) apresenta pela primeira vez o formato de primeira-ordem conhecido como *B-format*. Este, codifica a informação direccional de um espaço sonoro tridimensional – *soundfield* – em quatro canais discretos – W, X, Y e Z –, aplicando os conceitos explorados por Blumlein agora para o áudio tridimensional. Cada canal associa-se, respectivamente, à informação omnidireccional, *front-back*, lateral e elevação vertical, sendo as três últimas caracterizadas por informação direccional no plano horizontal.

Em meados de 1990, a tecnologia Ambisonics estende-se para *high-orders*, reflectindo-se na discretização do campo sonoro em mais harmónicos esféricos<sup>9</sup>, aumentando assim não só o tamanho do *sweet spot* no qual o campo sonoro é reproduzido, mas também a qualidade da localização das fontes. Com o aumento da ordem de resolução, novos canais são introduzidos, o que se traduz na necessidade de aumentar o número de altifalantes necessários para a reprodução rigorosa do campo sonoro. No entanto, é importante compreender que os sinais áudio codificados são independentes do sistema de reprodução onde o campo sonoro é posteriormente descodificado e reproduzido. Isto acontece porque os diferentes sinais não alimentam directamente nenhum canal de reprodução dedicado<sup>10</sup>, mas transportam a informação direccional do campo sonoro. “*An Ambisonic decoder is therefore always designed for a specific speaker layout, and an Ambisonic-encoded soundfield can be reproduced on any Ambisonic decoding system*” (Hollerweger, F., 2008, p.2).

---

<sup>9</sup> Cada *beam* da esfera harmónica contém informação sobre uma secção do campo sonoro.

<sup>10</sup> Como acontece em sistemas de tradicionais *channel-based* que não se baseiam em considerações físicas.



## 2.2.6 Técnicas de captação estereofónicas para sistemas multicanal

These microphone techniques tend to split into two main groups: those that are based on a single array of microphones in reasonably close proximity to each other, and those that treat the front and rear channels separately. The former are usually based on some theory that attempts to generate phantom images with different degrees of accuracy around the full 360° in the horizontal plane. The latter usually have a front array providing reasonably accurate phantom images in the front, coupled with a separate means of capturing the ambient sound of the recording space. (Rumsey, 2001, p.188)

Para reproduzir a qualidade sonora e espacial de um grande coro, uma orquestra ou qualquer fonte sonora de grande dimensão, torna-se ineficiente a captação através de fontes pontuais. Desta forma, com o aparecimento dos formatos de reprodução multicanal, surgem também técnicas de captação optimizadas para os mesmos. A maior parte das técnicas de captação deste tipo de formato parte da exploração de pares coincidentes e espaçados de microfones, explorando os mesmos princípios do formato estéreo 2.0 em que os pontos de captação destas matrizes alimentam directamente os canais discretos dos sistemas de reprodução. Wittek e Theile (2017) defendem que as técnicas de captação multicanal estereofónicas conseguem resultados de alta qualidade sonora e eficiência na distribuição dos canais por explorarem os mesmos requisitos observados no *design* de matrizes de captação estéreo. Os autores definem os três critérios essenciais:

- Signal separation among all channels in order to avoid comb filtering: No one signal should be present at significant levels in more than two channels;
  - Level and/or arrival time differences between adjacent channels to achieve the desired imaging characteristics;
  - Decorrelation of diffuse-field sound for optimal envelopment and sound quality.
- (Wittek & Theile, 2017, p.4)

Segundo os autores, um sistema de captação estereofónico é sempre preferível por facilmente criar fontes fantasma em várias direcções sem introduzir artefactos consideráveis, nomeadamente coloração. Gribben (2018) refere também neste sentido que “*Interchannel decorrelation is necessary to maintain or achieve the spatial impression that is desired from multichannel reproduction. Without some decorrelation of signals, it is likely that the sound scene would sound very unnatural and unstable (...).*” (p.24).

O áudio tridimensional potencia a percepção espacial quando comparado a sistemas de reprodução *surround*. Acrescentar a componente de altura no momento da captação torna-se preferível quando comparada à simulação dessa componente espacial. A componente vertical

## Revisão Bibliográfica

ajuda não só na percepção da elevação das fontes sonoras reproduzidas mas, principalmente, a conseguir melhor resultados no que toca a envolvimento, naturalidade e precisão tímbrica. Para além destes pontos, torna-se evidente as melhorias obtidas quando se fala do espaço e campo sonoro. Depois da introdução do sistema *surround* 5.1, a inclusão de canais de altura foi o passo seguinte na produção musical “realista”. Para isso, são necessárias técnicas dedicadas para controlar os quatro principais atributos de uma gravação tridimensional: largura e direcção das fontes, profundidade, impressão espacial e envolvimento (Theile & Wittek, 2012).

As técnicas de captação estereofónicas podem ser categorizadas como coincidentes – XY, *Middle-Side* e Blumlein –, quase-coincidentes – ORTF –, e técnicas espaçadas – AB e Decca Tree (Owsinsky, 2009; Conceição, 2014). Tal como no estéreo tradicional e nos sistemas *surround* 5.1, as técnicas de gravação para áudio 3D podem ser divididas a partir destas três categorias (Howie et al., 2018). Vários autores nomeiam técnicas como OCT, ORTF, IRT Cross, Hamaski Square e duplo *Mid Side* como algumas das principais matrizes de captação para formatos *surround*.

A técnica OCT é uma variação do formato Decca Tree onde as cápsulas omnidireccionais são substituídas por três microfones direccionais organizados num triângulo. Ao adicionar dois cardióides traseiros à matriz de captação frontal, é criada a extensão desta técnica para o formato *surround*. A IRT Cross parte também da adaptação de uma outra técnica, a ORTF. O sistema ORTF utiliza dois cardióides, dispostos de forma quase coincidente, com um ângulo de 110 graus e um espaçamento de cerca 17 centímetros entre cada cápsula. Tal como na matriz OCT, a extensão desta técnica para captação multicanal é feita através da duplicação da disposição frontal para a captação da informação traseira. Já a configuração IRT, semelhante à técnica ORTF, é composta por quatro cardióides dispostos num quadrado, onde os diferentes pontos de captação têm um ângulo de 90° entre si. Para compensar o ângulo mais estreito em comparação ao formato ORTF, a distância entre as cápsulas dos microfones é maior. A configuração Hamasaki Square aproxima-se da técnica IRT, onde é feita a substituição das cápsulas direccionais por figuras-de-oito. Os pontos nulos de captação são direccionados frontalmente para o palco sonoro, tornando este sistema relativamente insensível à captação de som directo (ideal para captações de ambiência). Por último, a matriz *Middle-Side* utiliza uma configuração MS padrão com a adição de um microfone cardióide voltado para trás.

Tal como a maior parte dos sistemas de captação para *surround* partem da extensão de técnicas estereofónicas, muitos dos métodos de captação para formatos tridimensionais partem também da adaptação de matrizes *surround*. A partir das mesmas, surgem formatos como OCT 9, ORTF 3D, IRT cross 5.0.4 e triplo *Mid Side*, sendo estas normalmente caracterizados pela duplicação da disposição da camada inferior para a captação da componente em altura. Tal como nas técnicas estereofónicas e *surround* anteriormente referidas, as técnicas de microfone tridimensional espaçadas captam e permitem a reprodução da informação espacial através das

diferenças de tempo e de intensidade entre os microfones (Rumsey, 2001; Theile, 2001; Schoeps, 2006; Owsinski, 2009; Wittek & Theile, 2012; Wittek & Theile, 2017).

### **2.2.6.1 Main Microphone Vs. Ambiente Microphone**

Uma técnica de captação de ambiente pode funcionar como uma técnica de captação principal. A diferença está na localização do espaço e das fontes sonoras, encontrando-se em torno do ouvinte em 360° em vez de se focalizar na parte frontal do palco sonoro (Wittek, & Theile, 2017). Neste sentido, uma matriz de captação ambiente não privilegia a captação do som directo frontal mas sim a composição do espaço sonoro onde a energia é uniformemente distribuída por todo o campo de reprodução. Segundo Theile (2001), as matrizes principais devem combinar dois princípios básicos da psicoacústica: a imagem direccional e a imagem espacial, referindo-se à importância da captação do som directo, das reflexões naturais e reverberação do campo sonoro<sup>11</sup>.

Algumas técnicas de captação principal mais populares são a AB, Decca Tree e XY – para sistema estereofónicos de dois canais – e a OCT, IRT Cross, ORTF e Decca Tree para sistemas multicanal sem altura. Já no áudio tridimensional, a OCT 9 é indicada como uma técnica de captação principal enquanto a ORTF 3D como uma técnica de captação ambiente. No entanto, dependendo do contexto prático, as aplicações como técnica principal ou ambiente podem variar.

### **2.2.7 Soundfield**

A informação das diferentes coordenadas Ambisonics pode ser captada a partir da combinação de um microfone omnidireccional (W) e três figuras-de-oito (X, Y e Z), permitindo uma captação FOA do espaço sonoro real (Hollerweger, 2008). Uma vez que a criação de um *array* de microfones perfeitamente coincidente é fisicamente impossível, Gerzon (1975) introduz o microfone *Soundfield*. Composto por quatro cápsulas cardióide ou sub-cardióide dispostas num formato tetraédrico de forma coincidente, a composição do microfone *Soundfield* é baseada na decomposição do campo sonoro tridimensional em esféricos harmónicos. O *output* gerado a partir deste microfone classifica-se como *A-format*, sendo a partir do processo de codificação do sinal que se obtém o *B-format*. (Conceição, 2015; Zhang et al., 2017).

---

<sup>11</sup> “*Directional imaging: Picking up the prime sound of a source or group of sources which forms the “first wave front” during reproduction (direct sound). Spatial imaging: Picking up natural reflections and reverberation (indirect sound).*” (Theile, 2001, p.26).

O acesso a microfones *soundfield* com alta resolução espacial é ainda um obstáculo. Wittek e Theile (2017), defendem que a resolução dos microfones *Soundfield* de primeira ordem são apenas adequados para sistemas *surround 5.1*, revelando-se insuficientes na resolução espacial para áudio tridimensional. Os autores defendem ainda que, quando comparada a esta tipologia de captação, a utilização de matrizes de captação estereofónicas é superior pela facilidade na criação de fontes fantasma sem a adição de artefactos consideráveis<sup>12</sup>. “*This becomes evident in their low interchannel signal separation as well as the insufficient quality of their reproduced spatiality.*” (Wittek & Theile, 2017, p.3). O argumento dos autores baseia-se também nos critérios apresentados na secção 2.2.6.

### **2.3 Normas e Recomendações: Sistemas de reprodução e avaliação subjectiva de atributos sonoros**

As Recomendações ITU-R constituem um conjunto de normas técnicas internacionais desenvolvidas pelo sector de radiocomunicações da International Telecommunication Union. A implementação destas recomendações não é obrigatória. No entanto, como são desenvolvidos por peritos de inúmeras áreas do áudio, são metodologias adoptadas e reconhecidas em todas as áreas de investigação abrangidas. A partir destas recomendações, são sugeridas várias normas e metodologias que vão desde a implementação de sistemas de reprodução até à apresentação de directrizes para a realização de testes perceptuais auditivos. Neste sentido, passaremos a referir algumas destas referências que vão ao encontro do âmbito do presente projecto.

A norma ITU-R BS.775 apresenta detalhadamente um sistema multicanal estereofónico universal composto por três canais frontais e dois canais traseiros/laterais, juntamente com um canal LFE opcional, constituindo o tão conhecido formato *surround 5.1*. Já a norma ITU-R BS.2051-2 referencia sistemas de som avançados, isto é, formatos de reprodução com configurações para além das referidas na norma ITU-R BS.775. Estas especificações incluem requisitos para vários formatos de reprodução, nomeadamente *layouts* com camadas de reprodução em altura. Para garantir que a configuração de reprodução a partir de sistemas multicanal avançados tem uma definição e implementação consistente, a norma 2051-2 define um conjunto de parâmetros que detalham várias configurações alusivas a estes formatos, especificando cada ponto de reprodução, a sua posição e valores de azimute e elevação.

Estas duas recomendações são de grande importância pela sua aplicação nos formatos de reprodução adoptados para o desenvolvimento deste projecto, nomeadamente *surround 5.1* e

---

<sup>12</sup> “*No time-of-arrival differences exist between any of the channels. When deriving surround-sound channels from a single-point array, 1st-order polar patterns overlap excessively, as compared to a surround-sound array with some degree of spacing between microphones.*” (Bowels, 2015, p.4)

Auro 9.1. As mesmas, apresentam especificações detalhadas acerca da implementação destes sistemas de reprodução. Não obstante, as recomendações alusivas às avaliações perceptuais não revelam indicações acerca das condições de teste a serem avaliadas. Elas apenas sugerem metodologias para a elaboração desta tipologia de teste.

A recomendação 1284-2 constitui a referência principal que conduziu à elaboração da metodologia adoptada na fase dos testes de avaliação subjectiva. Esta recomendação descreve métodos gerais para a avaliação subjectiva de qualidade sonora, definindo uma série de atributos sonoros para a descrição da qualidade percebida. Esta referência parte da recomendação 1116-3, sendo esta última a base para várias recomendações que possam conter condições especiais adicionais, ou, pelo contrário, que não exijam requisitos tão rigorosos de implementação prática. A 1116-3 destina-se a ser utilizada na avaliação de sistemas que introduzam pequenas disparidades ao ponto de serem indetectáveis sem o controlo rigoroso das condições experimentais. Já a recomendação 1284-2 pretende orientar a avaliação geral da qualidade sonora, definindo directrizes para avaliações gerais que envolvam diferenças perceptuais mais evidentes e, portanto, que não exijam um controlo tão rigoroso dos parâmetros do teste. Por se basear na avaliação qualitativa de parâmetros áudio a partir de diferenças claramente perceptíveis, a utilização de um sinal de referência<sup>13</sup> poderá não ser necessária. Deste modo, é possível a comparação entre várias condições de teste, com diferentes características, tendo em conta o atributo sonoro sobre avaliação. Esta recomendação elenca e descreve os diferentes processos para a conceptualização e desenvolvimento de um teste de avaliação subjectiva, tais como a selecção do painel de participantes, métodos e escalas de avaliação, programa de excertos, condições de escuta, entre outros parâmetros. A aplicação desta recomendação será apresentada no capítulo 4. Avaliação.

É importante reforçar que apesar da implementação destas referências não ser obrigatória, elas constituem uma base sólida para o desenvolvimento desta tipologia de testes

## 2.4 Trabalhos relacionados e projectos de referência

São vários os trabalhos relacionados com a avaliação subjectiva ou qualitativa de sistemas de reprodução multicanal bem como os estudos realizados para a compreensão detalhada de técnicas de captação aplicadas a esses mesmos sistemas. Nesta secção são apresentados alguns estudos de referência para a presente investigação.

Marui, Kamekawa e Irimaji (2006) apresentam uma investigação sobre a percepção subjectiva de sete técnicas de captação *surround* a partir da audição de duas peças musicais.

---

<sup>13</sup> Excerto de teste reproduzido sem processamento, usado como base de comparação para um teste de imparidade.

## Revisão Bibliográfica

Para os testes de avaliação perceptual foi aplicado um método comparativo. Este estudo fez também a comparação entre diferentes formatos de reprodução, nomeadamente o estéreo 2.0 e o formato *surround* 5.1. Os atributos para a avaliação comparativa entre as técnicas focaram-se essencialmente em avaliações preferenciais e não na avaliação da percepção da eficiência das técnicas perante os estímulos acústicos. A partir das análises, observou-se que as matrizes de microfones apresentaram diferenças significativas nas respostas subjectivas dos ouvintes. Relativamente à sensação de potência sonora e largura de palco, as técnicas Decca Tree e Omni Square Mid foram as favoritas. Já a técnica Duplo MS revelou resultados preferenciais inferiores em relação às outras matrizes de microfones. As restantes técnicas revelaram diferenças significativas no programas de excertos, o que influenciou a impressão subjectiva. Relativamente à comparação entre os dois formatos de reprodução, os sujeitos de avaliação apontaram que o sistema multicanal apresenta maior largura horizontal e envolvimento. As conclusões apontam também que as avaliações perceptuais não se alteraram consoante as diferentes amostras de participantes.

Kim, Ko, Nagendra e Woszczyk (2013) apresentam um estudo sobre a percepção do ouvinte perante a adição de camadas superiores de altura a formatos de reprodução multicanal *standard*. Este estudo teve origem no facto da maior parte das investigações se centrarem em factores essencialmente relacionados com o posicionamento de altifalantes de altura e com a percepção de altifalantes de altura frontais. Considerando estes factores, o foco de estudo centrou-se nos efeitos perceptuais dos canais de altura traseiros que, até ao momento, não tinham sido um base central de investigação.

Lee e Gribben (2014) focaram-se na condução de avaliações subjectivas sobre *arrays* de captação tridimensional. Os testes tinham como objectivo compreender de que modo diferentes espaçamentos entre camadas principais e de altura influenciam a impressão espacial através da avaliação preferencial de diversos participantes. A técnica de captação utilizada no desenvolvimento deste estudo foi a extensão do formato OCT *surround* – OCT 9 –, como proposto por Theile e Wittek, a partir de gravações de instrumentos a solo. Os resultados indicaram de uma forma geral que o efeito do espaçamento entre as camadas principal e de altura na impressão espacial não foi significativo, dependendo do tipo de fonte sonora.

Ton e Meister (2018) publicaram um artigo relativo à captação de ambiência como parte fundamental na criação de ambientes e cenografias acústicas. A aplicação desta tipologia de captação é estudada no âmbito do cinema, jogos e realidade virtual. O estudo parte da experimentação de seis microfones ou *arrays* de microfones a partir dos quais são feitos três testes auditivos onde os participantes fizeram uma avaliação preferencial comparativa entre as diferentes técnicas. Todos os três testes resultaram numa classificação preferencial que tendeu essencialmente para o AH-Array (configuração desenvolvida pelos autores), ORTF 3D e IRT Cross. Os microfones Ambisonics de primeira ordem foram percebidos como piores.

Em 2017, Wittek apresenta uma aplicação móvel que permite ouvir gravações da banda jazz da SCHOEPS - Wrong and Strong - feita a partir da implementação de uma nova técnica

para áudio 3D e realidade virtual, a ORTF 3D. A aplicação permite ainda a comparação, feita pelo próprio utilizador, entre a técnica desenvolvida e a captação feita a partir de um microfone FOA *Soundfield*. Wittek e Theile (2017), apresentam posteriormente um artigo que apresenta a nova técnica como um “*ambience arrangement*” (p.5).

Howie (2018) apresenta na sua tese de doutoramento o desenvolvimento, implementação e avaliação de uma técnica de captação para música orquestral aplicada a sistemas de reprodução tridimensionais. A investigação parte da exploração da técnica 22.2 Multichannel Sound, ou Hamasaki 22.2, sendo este estudo complementado através da realização de testes qualitativos focalizados na imersividade e coesão na representação do cenário acústico. A partir desta análise, são apresentadas avaliações preferenciais de diferentes padrões de polaridade para a captação das camadas superiores de reprodução.

Howie, Martin, Benson, Kelly, e King (2018) conduziram um estudo comparativo entre quatro técnicas de gravação distintas (otimizadas para a gravação de um piano solo) para a reprodução tridimensional a partir de 9 canais. Tal como no presente projecto, as diferentes técnicas variaram na tipologia de matriz, tendo sido utilizadas configurações espaçadas, quase-coincidentes e coincidentes. Este estudo teve como objectivo investigar as possíveis diferenças perceptuais entre as várias técnicas de gravação através da análise objectiva e subjectiva das mesmas. As conclusões mostram que os dois sistemas de gravação espaçados obtiveram resultados superiores para todos os atributos subjectivos sob investigação, nomeadamente envolvimento, “naturalidade da cena sonora”, naturalidade tímbrica e tamanho das fontes. A técnica quase-coincidente (OCT 9) apresentou resultados mais moderados, enquanto a matriz coincidente foi classificada como inferior. Neste estudo é reforçada a necessidade de continuar a análise comparativa formal entre as várias técnicas de captação já desenvolvidas para sistemas tridimensionais de reprodução.

As referências citadas demonstram, simultaneamente, métodos de implementação prática a partir da exploração de tecnologias, sistemas e conceitos semelhantes aos propostos nas secções seguintes. A análise destas referências e o cruzamento da informação obtida serviu de base para desenvolver a metodologia prática implementada, para identificar as variáveis a ter em consideração para o desenvolvimento de testes de avaliação perceptual, e elucidou o modo como a presente investigação poderia contribuir para a expansão do conhecimento nesta área.

### 3. Metodologia e Implementação

A reprodução da experiência acústica é vista como o padrão máximo que os sistemas de captação e reprodução tentam recriar. Geralmente, as técnicas mais tradicionais de gravação e reprodução sonora não representam a performance como ela é percebida, não só por normalmente partirem da mistura de gravações próximas com captações do espaço, mas também pela reprodução do campo sonoro acontecer ou entre dois altifalantes frontais ou apenas no eixo horizontal. Neste sentido a questão que se coloca é de que forma o sistema de captação e reprodução têm a capacidade de registar e reproduzir as principais variáveis perceptuais que contribuem para as experiências auditivas ao vivo. É neste sentido que a exploração de sistemas multicanal pode ser vantajosa quando comparada a formatos mais tradicionais de produção e reprodução sonora.

Quais as variáveis que temos de ter em atenção para a reconstrução de um campo sonoro da forma mais idêntica à nossa percepção em experiências de audição ao vivo?

Aqui, podemos definir três pontos centrais: direcção, distância e impressão espacial. A direcção está directamente relacionada com a nossa capacidade de localização das fontes sonoras. Mais canais de reprodução permitem localizações mais discretas, sendo isto principalmente evidente no plano horizontal. Relativamente ao factor distância, quando uma fonte sonora - captada de forma próxima - é enviada para um único altifalante, ela é percebida à distância desse mesmo ponto de reprodução. Se reflexões simuladas ou reais dessa mesma fonte forem adicionadas, é possível criar a ilusão de maior distância. Por último, mas não menos importante, a impressão espacial refere-se à percepção relativa à escuta num espaço, associada essencialmente a dois principais componentes perceptuais: a largura aparente das fontes sonoras e envolvimento. A ASW refere-se à medida de percepção da largura das fontes sonoras e resulta da fusão temporal e espacial da largura da imagem sonora com a informação direccional da mesma. Já a envolvimento refere-se à sensação de se estar envolvido por um campo difuso sem privilegiar qualquer informação direccional ou fonte localizável.



Estes dois factores estão directamente relacionados com o conteúdo reflectido. Reflexões mais intensas têm a capacidade de deslocar a posição aparente de uma fonte na direcção da reflexão e/ou fazer a fonte parecer maior, bem como contribuir para a sensação de envolvimento (Toole, 2008). Uma vez que a ilusão espacial requer que sons reflectidos de baixo nível sejam audíveis, a impressão espacial é extremamente dependente dos níveis amplitude. Várias das medidas relacionadas com a percepção de salas de concerto estão relacionadas com os ângulos de incidência dos sons reflectidos em relação à componente direccional proveniente do palco. Em particular, as reflexões laterais foram consideradas contribuições especialmente úteis para o que foi originalmente chamado de impressão espacial. Uma vez que a adição de canais de reprodução em altura permite a inclusão de toda a área superior do campo sonoro, torna-se possível a inclusão de reflexões iniciais nesse plano e a reprodução apropriada da difusão espacial.

Como já referido anteriormente, inúmeras técnicas de captação foram já desenvolvidas e testadas no âmbito da reprodução sonora em sistemas sem recurso a altura. Quando comparado ao espectro do áudio tridimensional, é de notar que as metodologias e sistemas dedicados necessitam de continuar a ser formalmente testados relativamente à forma como as percebemos. Contudo, é de referir que vários estudos têm vindo a ser realizados neste sentido. Apesar de ser incontestável a influência directa da adição de camadas superiores de reprodução na impressão espacial (Kim, et al., 2013; Howie, 2018), a maior parte das referências analisadas foca-se maioritariamente na análise da variação de critérios específicos dos sistemas de captação e reprodução que possam influenciar a nossa percepção espacial. Embora a maior parte destas investigações valide o foco de estudo através de testes de avaliação perceptual, a exploração da impressão espacial e a sua percepção acontece maioritariamente em âmbitos essencialmente técnicos a nível de implementação. Neste sentido, poucas técnicas desenvolvidas para a gravação musical tridimensional foram submetidas a avaliações comparativas formais, quer através de meios subjectivos ou objectivos (Howie, et al., 2018).

De forma a ampliar e expandir os contextos de implementação das tecnologias existentes dedicadas ao áudio imersivo, este projecto tem como objectivo compreender e analisar a resposta subjectiva aos estímulos acústicos, focalizando-se na impressão espacial no contexto da experiência musical, e na influência da adição de camadas de reprodução em altura na impressão espacial. Para melhor compreender estes conceitos, foi feita a avaliação quantitativa e análise preferencial de duas técnicas de captação aplicadas a sistemas de reprodução com e sem recurso a altura. A metodologia prática adoptada foi dividida em três fases principais: captação, espacialização e testes de avaliação subjectiva do atributo espacial. A partir da implementação de duas técnicas de captação foram aplicadas metodologias de espacialização e calibração para os testes perceptuais consequentes. A partir destas primeiras fases de exploração prática, o presente estudo pretendeu compreender de que forma é que o ouvinte se posiciona perante estes sistemas, indo ao encontro da necessidade de formalizar este tipo de investigação no âmbito da produção musical. Neste sentido, a última etapa é referente ao desenvolvimento e realização dos

testes auditivos onde foi feita a avaliação subjectiva da impressão espacial a partir dos resultados práticos obtidos nas fases anteriores.

Tendo em conta a conceptualização e desenvolvimento do processo prático, a próxima secção apresenta e documenta as estratégias e metodologias de produção musical adoptadas.

### 3.1 Gravação

Em colaboração com a Orquestra de Jazz de Matosinhos, foi feita a captação de várias peças de repertório, tais como *Farol e Peça em Peças*, interpretadas pela orquestra residente. As gravações decorreram no Centro de Alto Rendimento Artístico, em Matosinhos. Desde Setembro de 2018, ano de inauguração, o novo espaço tem recebido diversos concertos, ensaios, gravações e iniciativas do serviço educativo. Através de um processo colaborativo, a partir do qual foi possível a exploração de técnicas adaptadas ao espaço e à formação musical, esta parceria nasceu no sentido de dar forma ao presente projecto, indo de encontro aos objectivos de investigação propostos. As gravações decorreram nas instalações do CARA, em simultâneo com as gravações da OJM, com o apoio dos assistentes técnicos residentes.

Uma das tarefas mais sofisticadas no momento da gravação é registar o campo sonoro de forma "realista". Para tal, é necessário o uso de técnicas eficientes de captação que permitam controlar os três atributos principais da gravação tridimensional – direcção profundidade/distância e impressão espacial. A gravação e reprodução multicanal em três dimensões exigem que se tome especial atenção às reflexões provenientes do tecto, levando a adoptar técnicas de captação e produção musical que tenham em conta estes detalhes.

Ao procurar uma técnica de gravação adequada para um sistema de reprodução multicanal, o objectivo da gravação desempenha um papel fundamental na determinação da técnica a implementar. Existem técnicas indicadas para a captação das características do som directo em combinação com a informação espacial - matrizes de captação principal –, e outras optimizadas para a combinação com microfones pontuais – matrizes de captação ambiente. No entanto, investigações práticas são essenciais para poder verificar, aperfeiçoar e adaptar estas técnicas para situações individuais de gravação e intenções estéticas. Conseguir a separação entre canais a partir de uma matriz de captação ambiente torna-se significativamente mais difícil por existir maior probabilidade de *crosstalk* indesejado, isto é, conteúdos correlacionados estarem presentes em mais do que um canal de reprodução. Ao procurar uma técnica de gravação adequada aos sistemas de reprodução a serem explorados – *surround* 5.1 e Auro 9.1 –, torna-se difícil colocar os microfones de forma a evitar interferências entre nove canais. Tendo em conta a intenção de explorar técnicas de captação principal, a utilização de matrizes já optimizadas

para o formato *surround 5.1*, constituiu um ponto de partida para a escolha das técnicas a serem aplicadas. Neste sentido, as matrizes escolhidas partiram também da compatibilidade na sua aplicação tanto num sistema de reprodução bidimensional como num formato tridimensional, levando à implementação de técnicas comprovadamente eficientes na impressão espacial para o sistema sem recurso a altura a ser testado. Desta forma, é possível focar as avaliações subjectivas na forma como a adaptação e descodificação de cada técnica para um formato tridimensional potencia a impressão espacial, percebendo directamente a contribuição da adição da altura nas mesmas. Para além destes factores, a escolha das técnicas de captação partiu de duas condições principais: a implementação de técnicas de pares espaçados e coincidentes – de forma a explorar a influência desse factor na impressão espacial –, e a aplicação das mesmas em novos contextos musicais. Considerando todos estes pontos, este projecto parte da exploração de duas técnicas principais de captação: a OCT 9 (Theile & Wittek, 2012), em que os pontos de captação destas matrizes alimentam directamente os canais discretos dos sistemas de reprodução, e o *Soundfield FOA*, onde a espacialização sonora é feita através da descodificação da informação captada para o formato de reprodução a ser utilizado.

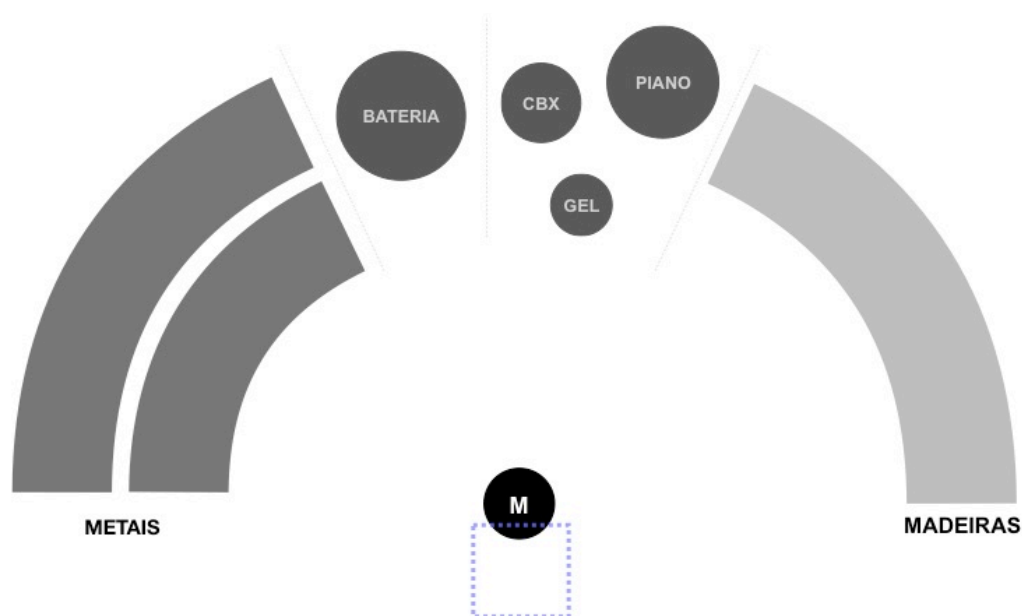
Todos estes factores, em combinação com o contexto prático de gravação, implicaram que a abordagem das técnicas utilizadas fosse adaptada nesse sentido, trazendo, desta forma, uma metodologia de captação moldada ao contexto.

A abordagem de captação adoptada fez a combinação de dois *arrays* de captação – compostos por duas matrizes de microfones posicionadas num único ponto da sala – e microfones pontuais para os diversos instrumentos que compunham a orquestra. As técnicas escolhidas têm como ponto de partida a aplicação de duas técnicas de captação principais com disposições de captação que permitissem explorar a influência da proximidade e afastamento dos pontos individuais de captação na impressão espacial. O contexto prático de gravação implicou que a intenção de implementação das técnicas como sistemas de captação principais fosse moldada ao espaço onde as gravações decorreram. O principal objectivo desta tipologia de técnica, factor mais fundamental no caso do OCT 9, é captar tanto a informação direccionada da performance com as características acústicas do espaço. A intenção inicial era a utilização destas técnicas como matrizes de captação principal, onde a captação do som directo seria feita de forma equilibrada com o conteúdo difuso, de forma a não necessitar de qualquer reforço pontual. A aplicação de um microfone principal parece ser vantajosa se forem dadas condições de gravação adequadas e a localização correcta do microfone puder ser encontrada para garantir a captação da imagem direccional de forma adequada. Apesar da técnica OCT 9 se inserir na tipologia de matrizes de captação principal, o posicionamento da mesma não permitiu assegurar o balanço adequado de som directo para todas as fontes discretas. Desta forma, a estabilidade direccionada da imagem frontal estereofónica (ver secção central indicada na imagem 2) através dos canais LCR ficou comprometida e o canal central poderia ser dispensado no momento da gravação. Neste caso, microfones pontuais foram adicionados para realizar correções em relação

## Metodologia e Implementação

ao equilíbrio de intensidade, coloração e timbre. Mais pormenores sobre este ponto serão discutidos na secção 3.2.

Para além dos dois *arrays* de captação tridimensional, foram também utilizados vários microfones de reforço direccionados para instrumentos em particular ou para secções instrumentais. O intuito deste *setup* de captação é, numa fase posterior de mistura, combinar as diferentes abordagens de captação – as matrizes multicanal e as captações pontuais - para a recriação do palco sonoro original. Uma vez que o objectivo deste estudo é também analisar o desempenho de cada técnica na impressão espacial percebida, esta abordagem tinha como objectivo a captação mais natural e realista possível em termos espaciais, isto é, sem implicar a necessidade de processamento adicional numa fase posterior de espacialização. Vários autores e referências defendem que, quando comparada ao *Soundfield* de primeira ordem, a utilização de matrizes de captação estereofónicas é superior pela facilidade na criação de fontes imaginárias sem a adição de artefactos consideráveis. Apesar dos estudos já efectuados, o principal objectivo da utilização destas técnicas não se prende apenas com a avaliação subjectiva das mesmas no que toca à impressão espacial obtida. Considerando estes factores, este estudo pretende também avaliar e compreender de que forma cada técnica, quando descodificada para um sistema de reprodução com altura, melhora a impressão geral obtida em comparação com a sua própria aplicação a um sistema sem recurso a fontes de reprodução em altura.



**Figura 2:** Stage Plot da Orquestra de Jazz de Matosinhos nas sessões de gravação. Secção central: Bateria, contrabaixo, piano e guitarra eléctrica. O ponto “M” indica o local intermédio do maestro. O quadrado tracejado indica o posicionamento das matrizes de captação principal.

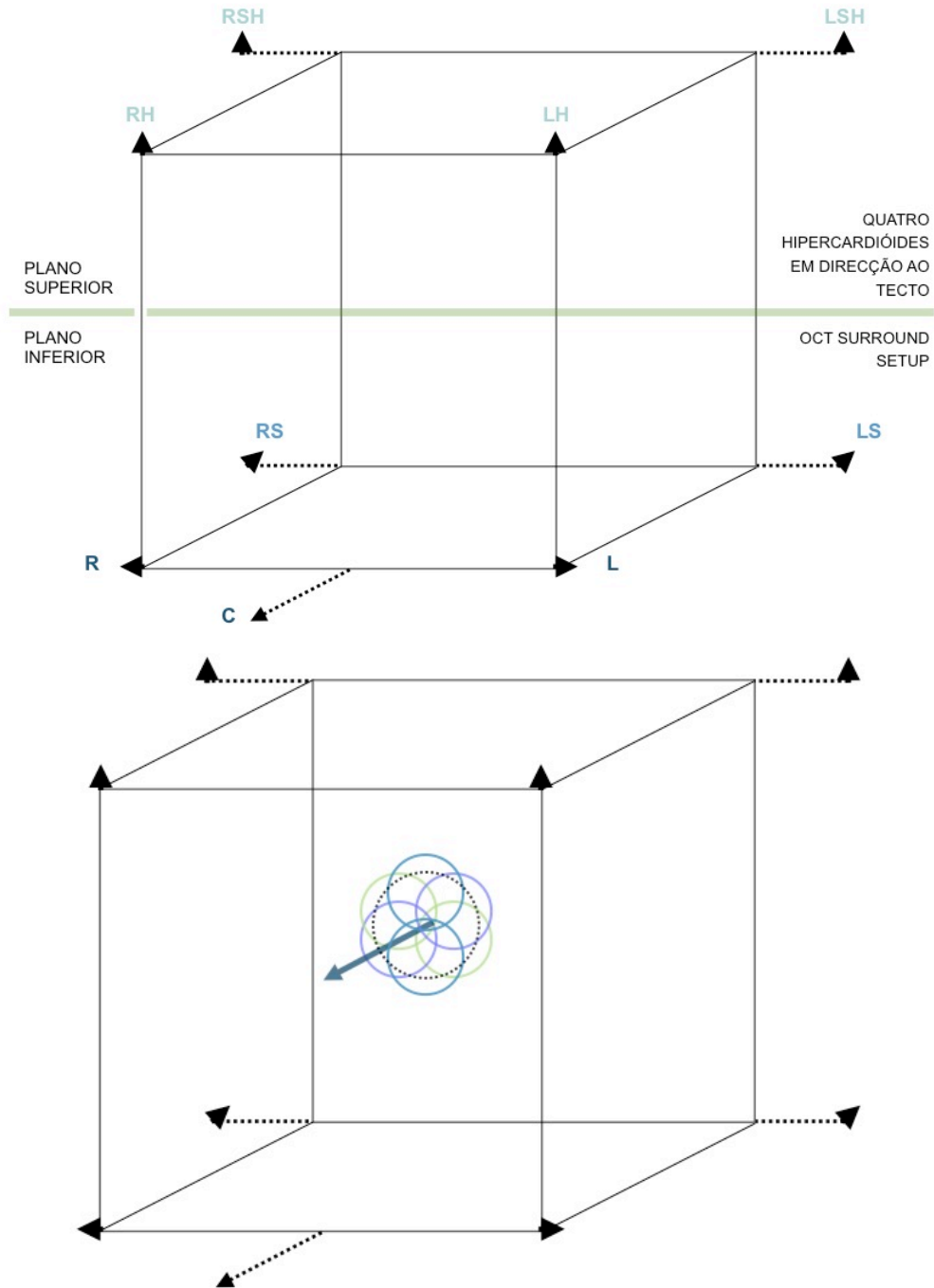


**Figura 3:** Array de captação OCT *surround* + 4 hipercardióides em altura e FOA *Soundfield*.



**Figura 4:** *Stage plot* com o array de captação OCT *surround* + 4 hipercardióides em altura e FOA *Soundfield*.

### 3.1.1 Descrição das matrizes de captação tridimensional



**Figura 5:** Matriz de captação OCT *surround* + 4 supercardioides: identificação dos pontos discretos de captação e orientação dos pontos de captação discretos (cubo superior). Matriz *Soundfiled* (cubo inferior): identificação do eixo frontal de captação e do posicionamento de em relação à matriz OCT 9.

Uma matriz OCT é composta por três microfones direccionais cujos sinais são enviados discretamente para os canais de reprodução esquerdo, central e direito – tendo em conta os formatos de reprodução utilizados no projecto prático. Os dois microfones de captação lateral são por norma hipercardióides e o central cardióide. A extensão desta técnica para captação *surround* permite a captação discreta de cinco canais. O sistema frontal da OCT é complementado por dois cardióides apontados para trás para evitar a captação do som direto. As diferenças temporais e de nível entre o par cardióide e hipercardióide de cada lado produzem uma representação estereofónica dos sons laterais para complementar a imagem estereofónica frontal. As reflexões laterais, importantes para a percepção da sala, são desta forma reproduzidas corretamente, recriando uma perspectiva espacial mais realista. Theile e Wittek (2012) apresentam uma proposta de adaptação do formato *surround* para um sistema de reprodução tridimensional a partir da adição de quatro hipercardióides posicionados e apontados verticalmente de forma a captar as reflexões e o campo sonoro difuso proveniente do tecto. Por oposição a esta técnica, foi também utilizado um microfone FOA *Soundfield* – descrito na secção 2.2.7.

As especificações técnicas do material, referentes aos modelos de microfone utilizados na captação, podem ser consultadas no Anexo 1, secção 7.1. Especificações técnicas do material de captação.

### **3.1.1.1 Posicionamento e dimensão dos *arrays* de captação tridimensional**

- Posicionamento: Ligeiramente atrás do maestro (ver figura 2);
- Aresta do cubo: 50 centímetros;
- Largura da aresta LS/RS e LSH/RSH: 70 centímetros;
- Altura do plano inferior: 1,71 metros;
- Altura total: 2,21 metros;
- Os pontos de captação indicados por uma linha a tracejado - no casos C/LS/RS/LSH/RSH – indicam o acréscimo de 10 centímetros ao comprimento inicial da aresta na direcção indicada pela linha. As setas indicam a orientação de cada cápsula na matriz OCT 9;
- O microfone Soundfield encontrava-se posicionado entre as duas camadas de captação da matriz OCT 9, rigorosamente no ponto central.

### 3.2 Edição, espacialização e ajuste de níveis

A partir das gravações realizadas, foram trabalhadas duas obras, *Peça em Peças* e *Farol*. A fase de edição passou pela selecção de vários excertos de cada peça para a sua utilização posterior nos testes de avaliação subjectiva. A partir das duas obras foram escolhidos excertos que explorassem um carácter musical variado em termos de dinâmica, traduzindo-se tanto na escolha de excertos com secção de grande potencia sonora, com todos os elementos da orquestra a tocar em simultâneo, como em momentos dinâmicos mais *piano*. Neste sentido, a selecção dos excertos a serem trabalhados partiu das variações dinâmicas, de duração e da frase musical de forma a tornar as sequências musicais o mais variadas possível entre si. Dependendo da passagem seleccionada, cada excerto teve a duração de entre 10 a 20 segundos de forma a preservar a frase musical e evitar criar cortes abruptos que pudessem interferir na experiência auditiva durante os testes. Tendo em conta a possibilidade de reprodução em modo repetição, todos os excertos tinham *fade-ins* e *fade-outs* breves.

Uma vez que o posicionamento dos *arrays* de captação principal não permitiu assegurar o balanço adequado na captação de som directo da imagem frontal do palco sonoro, a estabilidade direccional frontal, captada através dos canais LCR, ficou comprometida. Por conseguinte, foi necessário recorrer às captações dos microfones pontuais para realizar correções de equilíbrio de intensidade e de informação direccional de alguns instrumentos em falta no campo sonoro. As maiores falhas foram notadas na representação da secção central da frente do palco onde foi necessário reforçar a secção rítmica da orquestra, nomeadamente a bateria, contrabaixo, piano e guitarra (ver figura 2). As secções de sopro – madeiras e metais – não necessitaram de reforço directo pelo seu posicionamento lateral no campo sonoro, tendo os *arrays* de captação tridimensional – devido à proximidade no seu posicionamento – sido suficientes para a captação da informação sonora direccional destes elementos. A aplicação das matrizes tridimensionais, devido ao seu posicionamento, cumprem essencialmente os requisitos de um microfone de sala, não privilegiando a captação do som directo frontal, mas antes, a composição e distribuição uniforme do espaço sonoro no campo de reprodução. Neste sentido, os microfones pontuais permitiram criar uma mistura completa das fontes no palco frontal que atende a todos os requisitos em relação ao equilíbrio direccional e sonoro.

A metodologia de mistura adoptada teve sempre em vista as condicionantes implícitas na realização dos testes perceptuais a partir dos resultados obtidos. Isto exigiu que a abordagem fosse o mais objectiva possível de forma a evitar a adição de variáveis subjectivas às gravações originais. Desta forma, as gravações feitas a partir dos *arrays* tridimensionais foram tratadas essencialmente na sua distribuição espacial, não tendo sido feito qualquer ajuste em termos frequenciais, de nível ou de processamento de sinal. No caso das gravações pontuais, foram apenas feitos alguns ajustes para assegurar coerência tímbrica mas nenhum processamento adicional de espaço foi acrescentado. Esta abordagem tem em conta a necessidade de preservar



as condições espaciais representadas através das técnicas de captação principais - as quais foram objecto de avaliação nos testes subjectivos - de forma a não trazer condicionantes espaciais adicionais que pudessem influenciar, na fase posterior do projecto, a avaliação da impressão espacial.

A intenção relativa à recriação do palco sonoro era colocar o ouvinte num ponto próximo da orquestra, aproximadamente no local do maestro (ver Figura 2), de forma a preservar a sensação de envolvimento e espaço e a disposição do palco sonoro original. A espacialização das captações pontuais manteve-se igual para as duas configurações de captação principal e para os dois formatos de reprodução a serem utilizados. A distribuição espacial das mesmas foi feita no plano de reprodução inferior – *surround* 5.1 – e manteve-se igual quando reproduzido no formato de reprodução com altura, uma vez que a localização e elevação das fontes sonoras não era uma variável de estudo neste projecto. A distribuição panorâmica das fontes pontuais foi feita entre os canais de reprodução LCR onde foi respeitado o posicionamento original da instrumentação no palco sonoro. Relativamente à abordagem de espacialização das matrizes tridimensionais, no caso das gravações obtidas através da OCT 9 cada ponto de captação corresponde a um ponto no sistema de reprodução que lhe é simétrico. Neste sentido, o processo de espacialização passou pelo envio discreto e directo de cada ponto de captação para o ponto de reprodução correspondente. Já no caso das gravações do microfone FOA *Soundfield*, a espacialização do conteúdo dependeu da descodificação independente para os dois formatos de reprodução em utilização. As metodologias adoptadas para este processo serão apresentadas na secção 3.2.1. Plugins de descodificação, espacialização e de medição e ajuste níveis.

### **3.2.1 Plugins de descodificação, espacialização e de medição e ajuste de níveis**

Para a descodificação das gravações obtidas a partir do microfone *Soundfield*, foi necessária a conversão do formato B FuMa para AmbiX<sup>14</sup> pela utilização posterior de *plugins* de descodificação que utilizam esta norma. O SoundField by RØDE oferece uma abordagem ao processamento Ambisonics intuitiva e acessível. Este *plugin* permite a descodificação de três tipos de input – B-Format FuMa, B-Format AmbiX e A-Format NT-SF1<sup>15</sup> – para vários formatos de reprodução *standard* – indo desde mono até 7.1.4 –, permitindo também a conversão entre os formatos B AmbiX e FuMa. O *plugin* permite ainda controlar em pós-produção o padrão de polaridade dos microfones virtuais e definir a angulação dos mesmos (ver

---

<sup>14</sup> Existem duas convenções dentro do formato B Ambisonics: AmbiX e FuMa. Apesar de semelhantes, diferem pela sequência na qual os quatro canais estão dispostos. O formato AmbiX é organizado como WYXZ, enquanto no formato FuMa a ordem dos canais é WXYZ.

<sup>15</sup> FOA *Soundfield* da RØDE.

## Metodologia e Implementação

Figura 14, Anexo 7.2). Este processo de conversão entre *B-formats* foi necessário para a criação de dois modelos de descodificação para os diferentes formatos de reprodução a partir do *plugin* AIIRA Decoder (ver figura 15, Anexo 7.2). O AIIRA Decoder da IEM permite a criação de raiz de descodificadores Ambisonics para qualquer formato de reprodução desejado, permitindo a descodificação do campo sonoro até sétima ordem. A vantagem em utilizar esta ferramenta está na possibilidade de controlar parâmetros costumizáveis para o formato de reprodução em utilização. Para criar um *template* de descodificação, é necessário definir as coordenadas correspondentes - azimute e elevação - e o número de canal para cada altifalante físico. O decodificador resultante é aplicado automaticamente no canal de entrada seleccionado para que possa ser directamente reproduzido no formato desejado. Os pontos de reprodução que se encontram ao mesmo nível da cabeça do ouvinte devem ser especificados por um ângulo de elevação de 0°. A existência de canais imaginários permite a triangulação de um sistema de reprodução de forma não ambígua no caso de, por exemplo, se tratar de um sistema de reprodução sem recurso a elevação. A partir deste *plugin* foram criados dois *templates* de descodificação a serem importados para o SimpleDecoder da IEM (ver figura 15, Anexo 7.2) que fez a leitura das configurações de reprodução definidas (*surround* 5.1 e Auro 9.1, segundo a configuração física do estúdio).

Para a espacialização das fontes pontuais foi utilizado o *plugin* ReaSurround. A ferramenta nativa da DAW Reaper é um *panner* multicanal que suporta qualquer número de canais de entrada e qualquer número de altifalantes. Permite a utilização de sistemas de reprodução padrão predefinidos – tal como a técnica *standard* 5.1 - ou adicionar fontes de reprodução manualmente em qualquer espaço do campo sonoro. Este *plugin* foi utilizado para a espacialização e manipulação de vários pontos individuais como uma fonte sonora única, ou grupo, conforme pode ser observado na Figura 16, do Anexo 7.2.

Após a espacialização dos diferentes elementos, foi necessário fazer o ajuste de níveis de forma a assegurar a mesma intensidade sonora entre todos os excertos. Este procedimento é crucial para garantir que esta variável não influencia ou introduz condicionantes na avaliação dos exemplos e para que as passagens entre os mesmos sejam feitas da forma mais coerente possível. O *plugin* Waves WLM Plus Loudness Meter (ver figura 17, Anexo 7.2) permite a medição precisa da sensação de intensidade. Esta ferramenta é totalmente compatível com todas as especificações ITU e EBU actuais, permitindo leituras de parâmetros intensidade Momentary, Long Term, Short Term e de True Peak<sup>16</sup>. Um medidor de *Loudness* mede a sensação de intensidade percebida através de uma série de técnicas de filtragem e avaliação média, permitindo a tradução destes valores através de escalas numéricas. A utilizada neste

---

<sup>16</sup> Momentary, Short e Long term loudness: escalas utilizadas – com diferentes janelas temporais e ponderação de valores – para determinar se um excerto áudio excede o limite de tolerância de intensidade superior ao limite definido. True Peak: exhibe o pico absoluto da forma de onda como será ouvido no domínio analógico. Refere-se ao valor máximo da forma de onda do sinal áudio no domínio contínuo de tempo. (European Broadcast Union [E.B.U.], 2014)

projecto foi a LUFs (E.B.U., 2014), a partir da qual foram medidos e nivelados todos os excertos de forma a terem o mesmo valor de sensação de intensidade. A predefinição aplicada foi a WLM Plus Multichannel 5.1, permitindo a medição dos valores de intensidade nas duas camadas de reprodução a partir da duplicação da instância do *plugin*.

### 3.3 Sistema de reprodução



**Figura 6:** Estúdio PTM Sul, Escola Superior de Música e Artes do Espectáculo

O estúdio PTM Sul (apresentado na figura 6) no qual decorreram as fases de mistura, espacialização e realização dos testes de avaliação subjectiva, está desenhado especificamente para a exploração de sistemas de reprodução multicanal. As dimensões deste espaço são de 6,4 metros de comprimento, 3,6 metros de largura e 5,1 metros de altura. O estúdio apresenta três camadas de reprodução com diferentes alturas, nomeadamente quatro monitores de chão, oito monitores ao nível intermédio dispostos circularmente, quatro monitores em altura - dois frontais e dois traseiros - e um LFE. A camada de reprodução intermédia tem a possibilidade de ser configurada para dois formatos de reprodução *standard*, nomeadamente 5.1 e 7.1. Os formatos utilizados para este projecto prático foram o *surround* 5.1, de acordo com a norma

## Metodologia e Implementação

*standard* ITU-R BS.775.1, e o formato Auro 9.1 – a ao qual são adicionados 4 altifalantes em altura ao formato *standard* 5.1 –, de acordo com a recomendação ITU-R BS.2051-2.

CANAL		POSIÇÃO MONITOR	
ID	NOME	AZIMUTE	ELEVAÇÃO
L	Left	+30°	0°
C	Centre	0°	0°
R	Right	-30°	0°
LFE	Low Frequency Effects	0°	0°
LS	Left Surround	+105°	0°
RS	Right Surround	-105°	0°
LH	Left Height	+30°	Aprox. +30°
RH	Right Height	-30°	Aprox. +30°
LSH	Left Surround Height	+105°	Aprox. +30°
RSH	Right Surround Height	-105°	Aprox. +30°

**Tabela 1:** Especificações do formato de reprodução

As reflexões laterais são consideradas contribuições especialmente úteis para a impressão do espaço e desempenham um papel fundamental na audição espacial. A relação entre o som directo e indirecto determina os atributos espaciais de um evento sonoro e permitem a reprodução do espaço de gravação. Nesta fase, a componente do som reflectido merece especial atenção, pois afecta criticamente atributos como a distância, profundidade e impressão espacial. A principal característica do formato de reprodução com altura utilizado neste projecto - Auro 9.1 - é a disposição em forma de cubo de oito altifalantes. O som resultante da reprodução em altura permite desta forma a inclusão de todo o espaço superior para a reprodução de primeiras reflexões e para a reprodução apropriada da difusão espacial subjectiva da parte da reverberação. Para além de permitir a recriação da componente espacial de todo o espaço superior do campo sonoro original, permite também a comutação rápida entre os sistemas de reprodução com e sem altura. Desta forma, assegura-se a configuração correcta de ambos os formatos de reprodução 5.1 e 9.1 segundo as recomendações anteriormente mencionadas.

## 4. Avaliação

De forma a compreender de que modo a adição de altifalantes de altura potencia a impressão espacial do ouvinte, o presente estudo explora diferentes estratégias e tecnologias de produção musical aplicadas a sistemas áudio imersivos. Para melhor compreender estes conceitos, foi feita a avaliação comparativa entre formatos de reprodução com e sem recurso a altura através da análise subjectiva de dois formatos de captação aplicados a estes sistemas. Desta forma, este teste tem como objectivo compreender e analisar a resposta subjectiva aos estímulos acústicos, focalizando-se na impressão espacial geral e na contribuição da altura na impressão subjectiva do espaço e na experiência auditiva. Este teste foi desenvolvido a partir de duas recomendações, nomeadamente a ITU-R BS. 1284-2 e a ITU-R BS. 1116-3. A primeira descreve métodos gerais para a avaliação subjectiva da qualidade sonora e baseia-se na segunda recomendação referida que especifica métodos para a avaliação subjectiva de pequenos artefactos em sistemas de áudio. A apresentação destas normas pode ser revista na secção 2.3 Normas e Recomendações: Sistemas de reprodução e avaliação subjectiva de atributos sonoros.

### 4.1 Condições de teste e programa de excertos

A tipologia de teste adoptada para o presente estudo foi uma comparação em pares envolvendo duas condições para cada exemplo de avaliação. O teste parte então da existência de quatro condições iniciais a serem avaliadas subjectivamente através de uma série de comparações.

As 4 condições de teste são apresentadas na seguinte lista:

- I. Gravação OCT *surround* para formato de reprodução *surround* 5.1;
- II. Gravação OCT 9 para formato de reprodução Auro 9.1;
- III. Gravação *Soundfield* FOA descodificado para o formato de reprodução *surround* 5.1;
- IV. Gravação *Soundfield* FOA descodificado para o formato de reprodução Auro 9.1.

## Avaliação

A partir das obras *Peça em Peças* e *Farol*, interpretadas pela Orquestra de Jazz de Matosinhos, foram retirados no total 7 excertos musicais a serem utilizados nos 9 exemplos de teste. Cada excerto teve uma duração máxima de entre 10 a 20 segundos e foi ajustado de forma a assegurar um nível de intensidade constante entre os diferentes exemplos. A tabela apresentada em seguida indica os diferentes exemplos a partir dos quais foram feitas as avaliações subjectivas do atributo sonoro. Desta forma, são apresentadas as diversas relações estabelecidas entre as condições de teste anteriormente referidas, resultantes nos 9 exemplos de avaliação:

EXEMPLO	CONDIÇÃO A	CONDIÇÃO B	EXCERTO
1	SF 5.1	OCT 5.1	1. Peça em Peças
2	OCT 5.1	OCT 9.1	2. Farol
3	SF 5.1	SF 9.1	2. Farol
4	SF 9.1	OCT 9.1	3. Peça em Peças
5	OCT 9.1	OCT 5.1	4. Peça em Peças
6	SF 5.1	SF 9.1	4. Peça em Peças
7	OCT 9.1	SF 9.1	5. Farol
8	OCT 5.1	OCT 5.1	6. Peça em Peças
9	SF 5.1	SF 5.1	7. Farol

**Tabela 2:** Lista de exemplos apresentada aos sujeitos de avaliação

### Legenda de Tabela:

- I. Exemplos marcados com o mesmo código de cor indicam que as condições sobre teste (A e B) se repetem mas que existe a alteração do excerto musical e, em alguns casos, da ordem das condições.
- II. A repetição das mesmas condições de teste com diferentes excertos musicais prende-se com a necessidade de descartar avaliações preferenciais do próprio conteúdo musical ao invés do parâmetro de teste a ser avaliado – ver Exemplos 2-5 e 3-6.
- III. É trocada a ordem das condições (A e B) em exemplos que se repetem de forma a verificar a precisão nas avaliações sobre o atributo em questão.

- IV. A repetição de excertos entre condições de teste diferentes prende-se com a necessidade de manter coesão entre a avaliação das próprias técnicas de captação de forma a minimizar as diferenças entre as mesmas – ver Exemplos 2-3 e 5-6.
- V. Os exemplos 8 e 9 apresentam condições iguais – excertos de controlo – de forma a verificar a precisão nas avaliações tendo em conta o parâmetro sobre avaliação.

## 4.2 Sujeitos de Avaliação

A contribuição de cada participante centrou-se na avaliação comparativa de uma série de excertos musicais focando-se na avaliação da impressão espacial. Os excertos puderam ser ouvidos as vezes necessárias até o ouvinte se sentir confortável e seguro da sua avaliação do atributo em questão. Os testes decorreram num período de 3 dias, contando com um total de 24 participantes. Todos os sujeitos de avaliação eram alunos e/ou professores da Escola Superior de Música e Artes do Espectáculo e da Faculdade de Engenharia da Universidade do Porto, do Mestrado em Multimédia. Todos os participantes apresentaram interesse ou formação nas áreas do áudio, composição, tecnologias da música, gravação e reprodução musical, tendo uma parte significativa da amostra de participantes experiência prévia em testes auditivos.

Nesta tipologia de teste, relativa à avaliação global de qualidade tal como qualidade básica de áudio, *selected assessors* ou *expert assessors* (participantes que têm uma compreensão comum dos atributos a serem avaliados e que podem avaliar e classificar de forma objectiva os estímulos auditivos) são a categoria de participantes indicada. Os sujeitos que participaram na presente experiência podem, neste sentido, ser considerados como *selected* ou *expert assessors* (Bech & Zacharov, 2006), o que significa que cada participante foi escolhido pela sua capacidade de realizar um teste sensorial e que, através do seu conhecimento ou experiência, tem competência para dar uma opinião no campo sobre o qual é consultado. Esta classificação não está directamente relacionada com a qualificação académica e/ou profissional do mesmo. O painel foi composto por participantes considerados membros da amostra desejada, com acuidade auditiva normal, sensíveis às características de qualidade de áudio e que demonstram a capacidade de avaliar repetidamente os estímulos de forma consistente. Neste sentido, era expectável que o painel de avaliadores produzisse resultados fiáveis.

Nenhum dos sujeitos reportou problemas auditivos.

### 4.3 Configuração de Teste



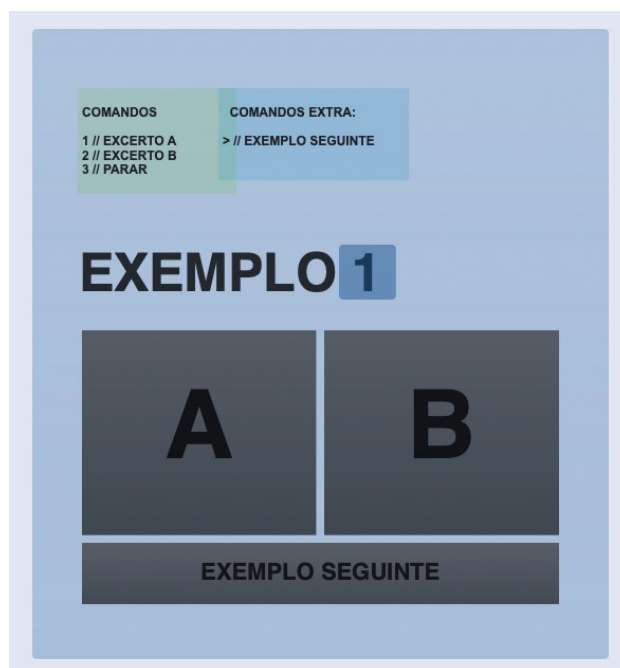
**Figura 7:** *Setup* para os testes de avaliação subjectiva.

Os testes decorreram no estúdio PTM Sul na Escola Superior Música e Artes do Espectáculo. As configurações de reprodução utilizadas foram as já referidas e a partir das quais foram desenvolvidas as fases de mistura e espacialização. Este estúdio foi escolhido pela disponibilidade do sistema de reprodução multicanal com recurso a altura e também pelo seu ambiente controlado de reprodução.

A reprodução da lista de exemplos (Tabela 2) foi feita de forma aleatória entre testes. A ordem de reprodução variável teve como objectivo evitar que os participantes se influenciassem entre si ao revelarem as suas respostas, e também evitar a formulação de avaliações prévias no caso do participante estar presente no momento de teste de outro sujeito. Os exemplos foram apresentados no software Max Msp, através do qual foi criada uma *interface* de teste que permitiu a alternância entre os excertos A e B e também a atribuição aleatória da ordem de reprodução. Esta plataforma de programação permitiu ainda a comutação simples entre os dois formatos de reprodução sobre teste. As comparações foram feitas entre pares de excertos, tendo sido pedido a cada participante para se concentrar na qualidade da impressão espacial do excerto B quando comparado com o A.



O teste apresentou 9 exemplos de avaliação, cada um composto por 2 excertos, A e B (ver Tabela 2). Dado que cada participante podia repetir cada exemplo as vezes necessárias até se sentir confortável para atribuir uma avaliação, o tempo de teste foi variável. No entanto, o tempo máximo de teste nunca excedeu os 15 minutos.



**Figura 8:** Interface de teste. Os participantes podiam permutar entre os excertos A e B sem alteração na *timeline* musical. A interface permitia ainda fazer pausa na reprodução, reiniciar os excertos e avançar nos exemplos de forma a garantir a autonomia dos participantes durante o tempo de teste.

Para testes de comparação entre pares, um método baseado na comparação discreta bipolar de cinco ou sete graus ou uma escala discreta baseada em diferenças numéricas a partir de escalas de cinco graus pode ser usada. Em geral, estas não são equivalentes e podem não dar os mesmos resultados. Quando são utilizados termos descritivos ou atributos de avaliação de parâmetros específicos, qualquer um dos métodos referidos pode ser utilizado para teste. Neste sentido, foi pedido aos participantes que avaliassem cada exemplo utilizando uma escala bipolar de 7 níveis discretos, sendo a avaliação, em termos de impressão espacial percebida, feita em relação ao excerto B quando comparado com A. O questionário de teste pode ser consultado no anexo 7.3. Instruções e Questionário de Avaliação. No mesmo, é feita uma breve contextualização do projecto, descrição do questionário, do atributo sonoro e do método de avaliação.

## Avaliação

- 3	- 2	- 1	0	1	2	3
Bastante menos espacial	Menos espacial	Ligeiramente menos espacial	Igual	Ligeiramente mais espacial	Mais espacial	Bastante mais espacial

**Figura 9:** *Comparison Category Rating (CCR)* (Bech & Zacharov, 2006). Escala de avaliação utilizada para avaliação nos testes auditivos.

### 4.4 Atributo

Dependendo dos objectivos de teste, diferentes números e tipos de atributos podem ser utilizados para qualificar os mesmos. Quaisquer atributos utilizados durante os testes devem ser claramente definidos. Tendo em conta a subjectividade entre os participantes na interpretação prática do conceito “impressão espacial”, foi necessário definir o mesmo de forma a uniformizar o significado atribuído. Neste seguimento, a recomendação ITU-R BS.1284-2 indica alguns atributos, determinando o interpretativo no momento de avaliação. Para melhor compreender o foco de avaliação perceptual do teste, foi apresentada uma tabela que descreve o atributo principal, sub-atributos e termos descritivos comuns para a avaliação exacta e detalhada do mesmo. Para além dos conceitos apresentados na seguinte tabela, a recomendação estudada descreve em detalhe a definição dos sub-atributos relacionados com o principal.

#### IMPRESSÃO ESPACIAL

<b>ATRIBUTO PRINCIPAL</b>	A performance parece estar a ocorrer num ambiente espacial apropriado.
<b>SUB - ATRIBUTOS</b>	Homogeneidade do som espacial Reverberância Tamanho aparente da sala Equilíbrio acústico Perspectiva de profundidade
<b>TERMOS DESCRITIVOS COMUNS</b>	Sala reverberante / Sala seca Som directo / Som indirecto Sala grande / Sala pequena

**Tabela 3:** Definição do atributo, sub-atributos e termos descritivos comuns relativos ao atributo principal “Impressão Espacial”

## 4.5 Análise estatística dos dados obtidos

Na presente secção, apresentam-se os resultados dos testes efectuados para cada um dos exemplos através da análise estatística dos dados obtidos<sup>17</sup>. Foram aplicadas duas tipologias de teste diferentes: um teste de t-Student, onde se assume a normalidade dos dados e homogeneidade de variâncias, e o teste de Wilcoxon, um teste não paramétrico para uma amostra.

One Sample T-Test					
	Test	Statistic	df	p	Location Parameter
Exemplo 1	Student	5.290	23	< .001	1.458
	Wilcoxon	273.000		< .001	1.500
Exemplo 2	Student	7.446	23	< .001	1.792
	Wilcoxon	290.000		< .001	2.000
Exemplo 3	Student	6.224	23	< .001	1.333
	Wilcoxon	237.500		< .001	1.500
Exemplo 4	Student	17.944	23	< .001	2.333
	Wilcoxon	300.000		< .001	2.500
Exemplo 5	Student	-4.399	23	< .001	-1.042
	Wilcoxon	26.000		< .001	-1.500
Exemplo 6	Student	7.400	23	< .001	1.542
	Wilcoxon	261.500		< .001	1.500
Exemplo 7	Student	-3.358	23	0.003	-0.958
	Wilcoxon	49.500		0.006	-1.500
Exemplo 8	Student	1.813	23	0.083	0.125
	Wilcoxon	6.000		0.149	1.000
Exemplo 9	Student	-0.296	23	0.770	-0.042
	Wilcoxon	6.000		0.783	-0.500

Note. For the Student t-test, location parameter is given by mean difference  $d$ ; for the Wilcoxon test, location parameter is given by the Hodges-Lehmann estimate.

**Tabela 4:** Resultados estatísticos para cada exemplo segundo os testes t-Student e Wilcoxon

Quando  $p < 0.05$ , considera-se que a avaliação dos inquiridos quanto à impressão espacial foi significativamente diferente, superior ou inferior, de 0. Todos os exemplos avaliados, à excepção do 8 e 9, encontram um valor de  $p$  entre  $< 0.001$  e  $0.006$ , revelando desta forma dados estatísticos com bastante significância. Considerando estes factores, os dados mostraram resultados significativos para as diferentes técnicas de captação e reprodução multicanal tendo em conta o impacto sobre a impressão espacial percebida. No caso dos exemplos 8 e 9, sendo estes os exemplos de controlo, o valor de  $p$  apresentou valores que revelam um impacto não significativo na impressão espacial. A tabela seguinte demonstra que em nenhum caso foi cumprido o pressuposto da normalidade da distribuição dos dados, condição requerida para poder ser aplicado o teste t-Student. Como tal, é mais seguro confiar no teste de Wilcoxon, onde

<sup>17</sup> Para a análise estatística dos dados pudemos contar com o auxílio do professor Bruno Castro do Departamento de Biologia & Centro de Biologia Molecular e Ambiental (CBMA) da Universidade do Minho, Escola de Ciências.

## Avaliação

nos concentramos nos resultados da mediana e não na média de respostas obtida para cada exemplo.

Test of Normality (Shapiro-Wilk)		
	W	p
Exemplo 1	0.858	0.003
Exemplo 2	0.836	0.001
Exemplo 3	0.842	0.002
Exemplo 4	0.768	< .001
Exemplo 5	0.858	0.003
Exemplo 6	0.764	< .001
Exemplo 7	0.835	0.001
Exemplo 8	0.393	< .001
Exemplo 9	0.587	< .001

*Note.* Significant results suggest a deviation from normality.

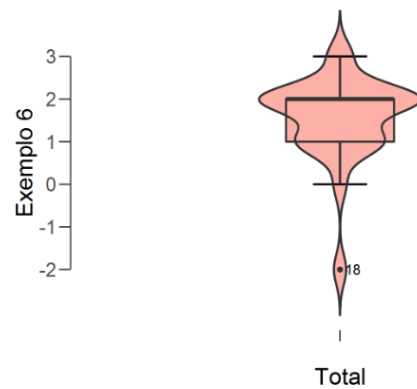
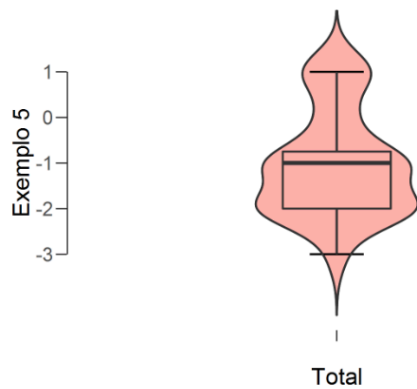
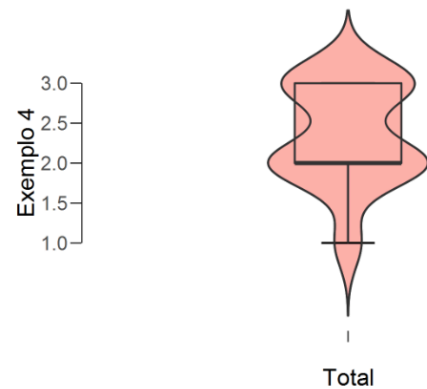
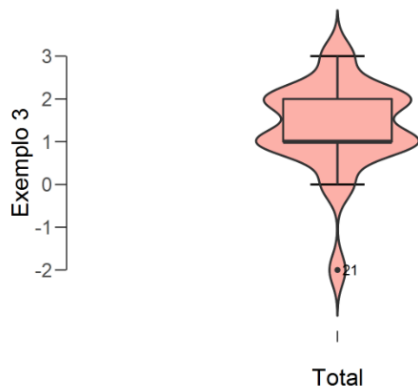
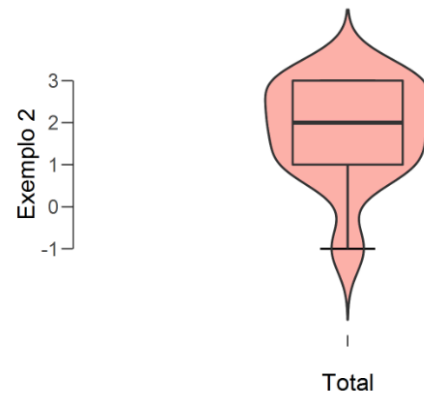
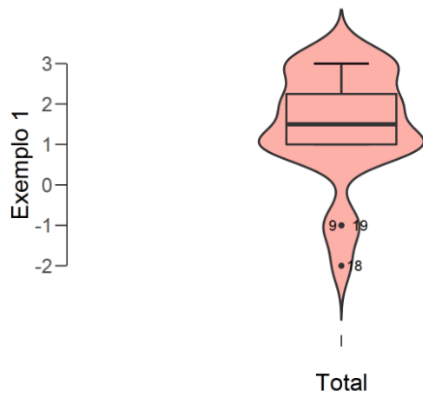
**Tabela 5:** Teste Shapiro-Wilk

Os valores de significância (Tabela 4) demonstram que a percepção subjectiva foi considerada positiva para todos os exemplos, excepto para os exemplos de controlo, cumprindo os requisitos pressupostos. Independentemente do teste aplicado para o tratamento estatístico dos dados, apenas os exemplos 8 e 9 foram considerados equivalentes em termos da classificação especial (mediana da classificação espacial = 0). No caso dos exemplos 5 e 7, a percepção dos inquiridos foi que a impressão espacial diminuiu significativamente do primeiro para o segundo excerto (ver mediana < 0 na tabela abaixo). No caso dos exemplos 1, 2, 3, 4 e 6, a percepção dos inquiridos foi de um aumento significativo de "espacialidade" (ver mediana > 0 na tabela abaixo).

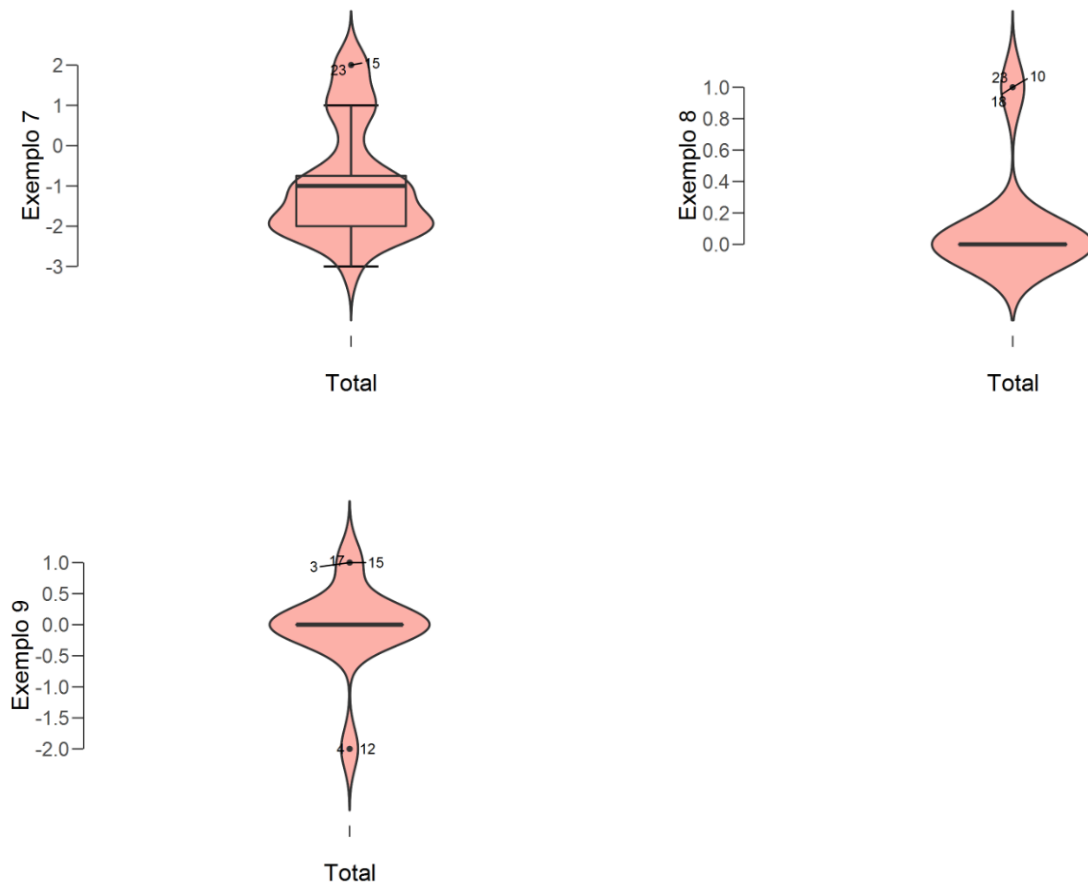
Descriptive Statistics									
	Exemplo 1	Exemplo 2	Exemplo 3	Exemplo 4	Exemplo 5	Exemplo 6	Exemplo 7	Exemplo 8	Exemplo 9
Valid	24	24	24	24	24	24	24	24	24
Missing	0	0	0	0	0	0	0	0	0
Mean	1.458	1.792	1.333	2.333	-1.042	1.542	-0.9583	0.1250	-0.04167
Median	1.500	2.000	1.000	2.000	-1.000	2.000	-1.000	0.000	0.000

**Tabela 6:** Valores estatísticos descritivos: Média e Mediana

De seguida, apresentam-se os gráficos com as respostas dos 24 inquiridos sob a forma de uma *violin plot*. Este método permite a representação gráfica da concentração das respostas, indicando que quanto mais larga a figura, maior a frequência de respostas no valor de avaliação correspondente - indicado na escala à esquerda. As respostas para cada exemplo são também representados sob a forma de um *boxplot* que indica a mediana, o intervalo inter-quartis e possíveis *outliers*.

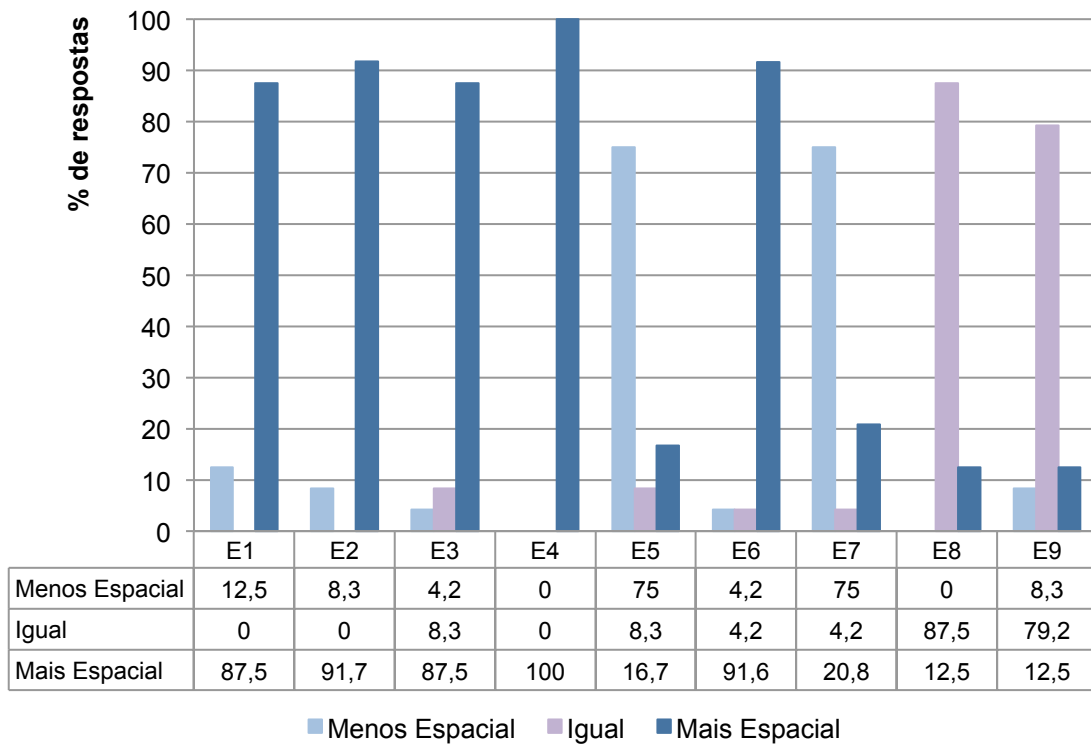


## Avaliação



**Figura 10:** Representação gráfica das respostas dos 24 inquiridos sob as formas de *violin plot* e *boxplot*.

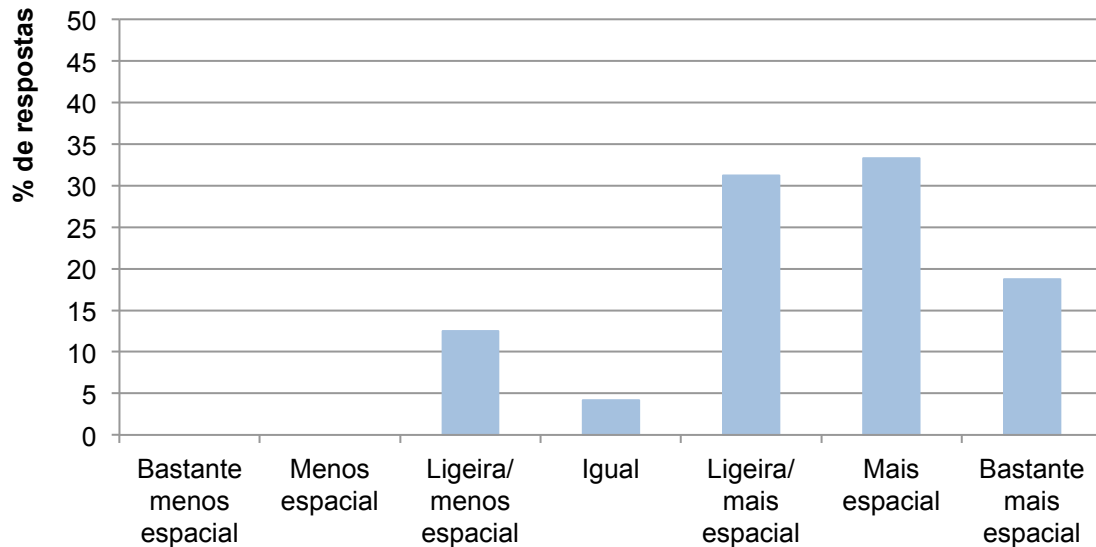
A partir dos resultados obtidos, foi possível desenhar um gráfico com a distribuição percentual do total de respostas obtidas (n = 216), que expõe a percentagem de respostas sobre a impressão espacial para os nove exemplos de teste.



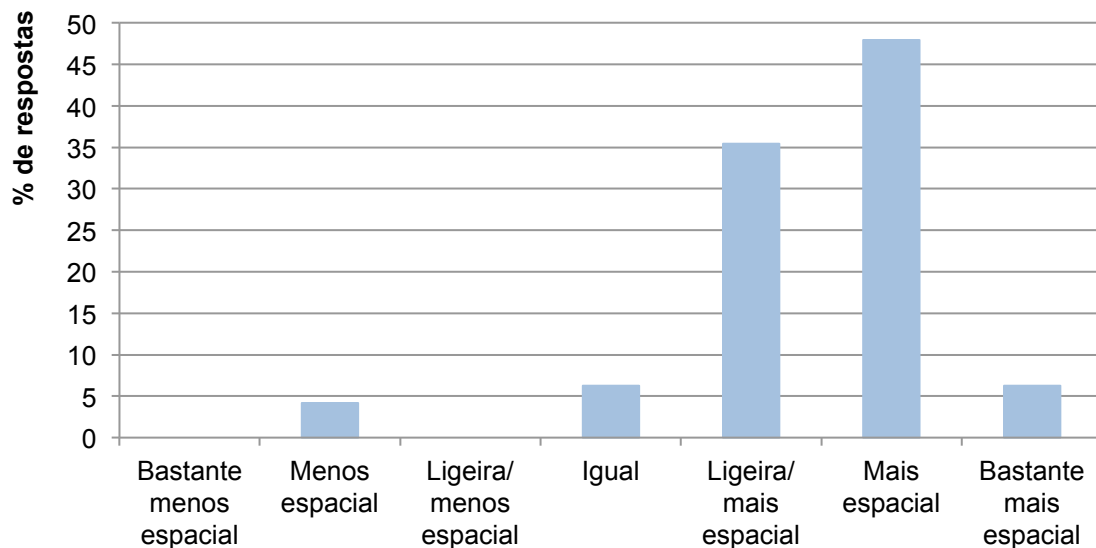
**Figura 11:** Percentagem de respostas por exemplo.

## Avaliação

As figuras 12 e 13 fazem, respectivamente, a representação gráfica percentual das respostas obtidas a partir da avaliação da contribuição da adição da camada de altura na técnica de captação OCT *surround* (ver exemplos 2 e 5 da Tabela 2) e *Soundfield* (ver exemplos 3 e 6 da Tabela 2).



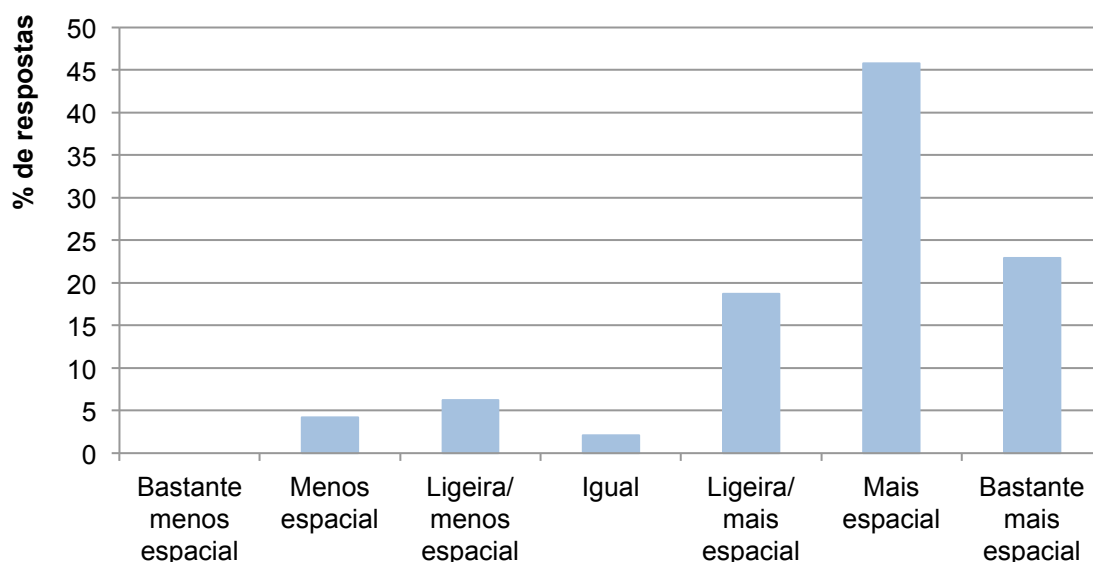
**Figura 12:** Distribuição percentual do total de respostas sobre a avaliação da contribuição da altura na impressão espacial a partir da técnica OCT.



**Figura 13:** Distribuição percentual do total de respostas sobre a avaliação da contribuição da altura na impressão espacial a partir da técnica *Soundfield*.



Ao avaliar a comparativamente as duas técnicas de captação com a camada de reprodução de altura (ver exemplos 4 e 7 da Tabela 2), obtiveram-se as seguintes respostas relativas à técnica OCT 9 quando comparada com a matriz *Soundfield* descodificada para a reprodução em 9.1:



## 4.6 Breve conclusão sobre os resultados

Sucintamente:

- Nos exemplos 8 e 9, os dois excertos foram considerados equivalentes pelos inquiridos;
- No caso dos exemplos 5 e 7, a percepção dos inquiridos foi que a impressão espacial diminuiu significativamente;
- No caso dos exemplos 1, 2, 3, 4 e 6, a percepção dos inquiridos foi que a impressão espacial aumentou;
- A adição de altura, no caso dos exemplos 2, 3, 5 e 6, traduziu-se num aumento significativo da impressão espacial;

Os resultados obtidos corresponderam positivamente às expectativas elaboradas durante a fase de concepção prática. A aplicação e descodificação das diferentes técnicas para um formato de reprodução tridimensional influenciou directamente a apreciação das mesmas em termos de desempenho espacial. O formato OCT 9, quando comparado com a própria configuração para *surround*, demonstrou ter um influência significativamente positiva em termos de impressão

espacial. O mesmo se pode concluir na comparação da técnica *Soundfield* quando descodificada para os dois formatos de reprodução. Comparando as duas técnicas aplicadas ao sistema de reprodução com altura, existiu uma avaliação preferencial evidente que tendeu para a matriz de captação OCT 9. Tendo em conta os dados obtidos, podemos concluir que a utilização da técnica de captação multicanal estereofónica consegue resultados com maior eficiência na recriação do campo sonoro e na impressão espacial recriada, indo ao encontro dos critérios defendidos por Theile e Wittek (2017) referidos na secção 2.2.6. Estes critérios referem-se à separação do sinal captado entre canais de forma a evitar comb filtering e diferenças de nível e de tempo entre pontos de captação adjacentes, para desse modo recriar as características de imagem desejadas e descorrelação do campo sonoro difuso de forma a otimizar a sensação de envolvimento. A influência destes parâmetros é clara nos comentários feitos pelos inquiridos, referidos na secção seguinte.

### 4.7 Comentários finais

No final de cada teste, a grande maioria dos participantes fez alguns comentários de forma a analisar as expectativas criadas antes da experiência e as associações feitas ao longo da mesma. Grande parte dos comentários focou-se na descrição das sensações associadas à escuta, auto-análise de percepção e comentários de cariz preferencial.

Um dos grandes desafios na elaboração de avaliações que partem da subjectividade da interpretação do participante é definir da forma mais objectiva possível, o parâmetro em que o mesmo se deve focar. Foi quase transversal a subjectividade imposta ao atributo sonoro em avaliação, notando-se, entre os diferentes participantes, uma clara atribuição de significado pessoal ao próprio tributo. Esta interpretação do mesmo dividia-se essencialmente na associação da impressão espacial à percepção de um espaço acústico em contexto de concerto - em que as fontes sonoras se concentram na parte frontal de reprodução - ou à percepção de um espaço acústico circundante ao ouvinte. Neste sentido, foi necessário clarificar o âmbito do próprio atributo de forma a assegurar que os diferentes participantes se focaram no mesmo conceito para que o seu significado e interpretação fossem o mais idênticos possível.

Uma das observações mais recorrentes entre os diferentes participantes foi a expectativa de sentir altura na reprodução, o que na maior parte dos casos não se confirmou depois da experiência auditiva. Uma grande parte dos inquiridos relatou não conseguir perceber a adição da camada superior de reprodução, isto é, perceber a directividade das colunas em altura. No mesmo sentido, ficou claro que não existiu a alteração do palco sonoro reproduzido em altura, mas que existiu uma alteração clara no campo sonoro em termos espaciais. Na maior parte dos casos, os sujeitos de avaliação não conseguiram associar se este fenómeno acontecia devido à adição de camadas de reprodução em altura ou pela alternância entre as diferentes

técnicas de captação que estavam a ser testadas. Os participantes que revelaram maior sensibilidade à alteração do campo sonoro em relação à impressão espacial estabeleceram um paralelismo sensorial auditivo semelhante à passagem entre um sistema estéreo 2.0 e *surround* 5.1, de forma a tornar mais claro que a adição de novos planos de reprodução potencia a sensação de espaço, sendo a alteração na impressão espacial bastante evidente em grande parte dos casos.

A sensação de estarem perante a orquestra e a estabilidade e realismo da imagem foram algumas das características mais mencionadas pelos participantes. A maior parte dos comentários reforçam a naturalidade, realismo na representação do palco sonoro, envolvimento, profundidade e clareza como termos descritivos comuns para os exemplos mais espaciais. É de referir que, mesmo quando fora do *sweet spot* ou da área total de reprodução, a alteração do campo sonoro em termos espaciais era bastante clara quando se adicionava a altura. Nestes casos, a sensação de estarem perante a orquestra num contexto de espectadores tornava-se mais evidente por não se encontrarem no centro do palco sonoro mas na perspectiva ou posição mais comum de um espectador numa posição de plateia. Foi referido que, quando em passagens de dinâmica forte em *tutti*, o aumento espacial era muito mais notório quando comparado com passagens de carácter musical mais silencioso, não se notando tão significativamente a alteração de espaço neste último caso. Isto pode ser justificado pela excitação do espaço acústico não ser tão dominante quando comparado com excertos de maior massa sonora. Desta forma, a captação de energia proveniente do tecto não tem a intensidade necessária para ser significativa na sua percepção. Relativamente às técnicas de captação exploradas, os participantes com um *background* associado às áreas do áudio e tecnologias do som destacaram a forte influência do timbre na impressão espacial. Foi referido que, na passagem entre as duas técnicas de captação tridimensional, as diferenças tímbricas foram também um factor de avaliação na impressão espacial pelas claras diferenças na naturalidade e equilíbrio acústico de representação do palco sonoro. Para além disso, foi também referido que este factor tímbrico pode induzir em erro no momento de avaliação da impressão espacial percebida por, em alguns casos, serem diferenças bastante notórias.

De um modo geral, a experiência foi muito bem recebida por todos os participantes. Na perspectiva de apreciação do conteúdo musical – abstraindo-se da componente associada à avaliação dos estímulos acústicos – a sensação de se estar perante um sistema tridimensional de reprodução, e a eficácia do mesmo para a recriação da experiência musical mais realista, revelou um grande impacto positivo na experiência auditiva geral. Apesar da sensação de incremento espacial não ter sido igualmente evidente em todos os exemplos apresentados, a adição da camada superior de reprodução revelou ter uma influência significativa na apreciação musical quando comparada a sistemas sem recurso a esta terceira dimensão.

## 5. Conclusões e Trabalho Futuro

Através de gravações dedicadas para a reprodução em sistemas imersivos, foi feita a análise subjectiva da característica perceptual da impressão de espaço para campos sonoros reconstruídos. A compreensão sobre como o espaço pode ser controlado e a forma como o processo de produção musical (motivado pela composição e manipulação de sensações auditivas) influencia a experiência musical no contexto da reprodução sonora, foram as principais motivações deste projecto. O principal foco de pesquisa foi determinado pela necessidade de expandir o conhecimento sobre a prática e promover o conhecimento a partir da mesma, contribuindo para a expansão dos contextos de aplicação prática destes sistemas e para a compreensão mais detalhada sobre a resposta subjectiva aos estímulos espaciais. Neste sentido, foram exploradas estratégias de produção áudio aplicadas a sistemas imersivos de forma a compreender de que modo a adição de camadas de reprodução em altura influencia a impressão espacial. Este estudo pretendeu, por um lado, ampliar e expandir os contextos de implementação de várias tecnologias de forma a validar e documentar os métodos de aplicação prática no âmbito da produção musical, e por outro, compreender de que forma o ouvinte percebe, através dessas tecnologias, o espaço e a experiência de escuta musical. Para melhor compreender o conceito de impressão espacial no contexto de sistemas áudio imersivos, foi feita a avaliação quantitativa e análise preferencial de duas técnicas de captação aplicadas a dois sistemas de reprodução com e sem recurso a altura.

A partir das gravações de duas peças interpretadas pela Orquestra de Jazz de Matosinhos, foi possível explorar duas matrizes de captação optimizadas para a reprodução tridimensional. As técnicas adoptadas tiveram em vista a realização posterior de testes de avaliação subjectiva sobre a impressão espacial recriada a partir das mesmas. Contando com a colaboração de 24 participantes, foi feita não só a avaliação subjectiva da influência da altura na impressão espacial, mas também a análise preferencial deste atributo através da apreciação comparativa entre as duas técnicas de captação aplicadas a dois sistemas de reprodução – *surround 5.1* e *Auro 9.1*.

Esta tese apresentou várias respostas relacionadas com o campo da percepção do espaço auditivo a partir da reconstrução de campos sonoros em sistemas de reprodução multicanal. No

entanto, há também várias questões que foram levantadas pelo trabalho realizado que poderão ser abordadas em trabalhos futuros.

- Realizar esta tipologia de testes em vários ambientes de escuta de forma a compreender se o próprio ambiente de escuta introduz artefactos na qualidade da impressão espacial percebida.
- Compreender de que forma, num contexto de consumidor final comum, a optimização deste tipo de tecnologia de captação pode ser vantajoso para sistemas de escuta como *headphones*.
- Compreender de que modo a alteração do posicionamento das técnicas de captação no espaço podem ter uma influência directa na impressão espacial e localização das fontes. Neste caso, a exploração das duas técnicas escolhidas como tipologia de matriz de captação principal VS. matriz de captação ambiente em combinação com captações pontuais para compreender os contextos ideais de aplicação das mesmas.
- A aplicação dos sistemas de captação como matriz ambiente poderá ser projectado e colocado além da distância crítica para capturar, principalmente, o som difuso. Para isso, um arranjo de microfones de ambiente semelhante à técnica “Hamasaki Square” - composto por quatro microfones figura-de-oito, dispostos numa formação quadrada – poderá ser transposto para a técnica OCT 9, onde se abdica da formação frontal LCR para captação do som directo.

O estudo apresentado nesta dissertação contribui para a expansão das investigações práticas sobre a aplicação de métodos de produção musical dedicados ao contexto do áudio imersivo. Esta tipologia de estudo é ainda necessária para poder verificar, aperfeiçoar e adaptar estes formatos de captação a situações particulares de gravação e contextos e intenções musicais específicas. Esta exploração prática permite aprofundar os contextos de aplicação destes sistemas, contribuindo para uma melhor e mais extensa compreensão dos mesmos. Para concluir, a aplicação deste tipo de metodologias de produção áudio pode potenciar a experiência sonora em várias áreas, não só num contexto de produção musical, mas também no cinema, jogos e realidade virtual. Daí, a relevância em continuar a explorar a contribuição da altura e as técnicas e sistemas que estimulam a nossa percepção espacial, no sentido de tornar a experiência auditiva uma prática cada vez mais imersiva e realista.

## 6. Referências

- Auro Technologies. (2015). *Auro 3D Home Theater Setup - Guidelines and Background*. Retirado de: [https://www.auro-3d.com/wp-content/uploads/documents/Auro-3D-Home-Theater-Setup-Guidelines\\_lores.pdf](https://www.auro-3d.com/wp-content/uploads/documents/Auro-3D-Home-Theater-Setup-Guidelines_lores.pdf)
- Auro Technologies. (2016). *Auro 3D Audio: Combined Auro-3D and Dolby Atmos Setup - Guidelines and Background*. Auro Technologies NV: Publisher.
- Berg, J., & Rumsey, F. (2003). Systematic evaluation of perceived spatial quality. In *AES 24th International Conference on Multichannel*, Banff, Alberta, Canada.
- Bech, S., & Zacharov, N. (2006). *Perceptual audio evaluation: theory, method and application*. New York, NY, USA: John Wiley & Sons.
- Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.
- Bowles, D. (2015). A Microphone Array for Recording Music in Surround-Sound with Height Channels. In *Audio Engineering Society Convention 139*. Audio Engineering Society.
- Candy, L. (2006). Practice based research: A guide. *CCS Report, 1*, 1-19.
- Conceição, M. (2015). Spaciousness Control for Sound Field Recording and Reconstruction (Doctoral dissertation). Trinity College Dublin

- Dolby Laboratories (2016). *Dolby Atmos Next-Generation Audio for Cinema*. [White Paper]. Retirado de: <https://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-next-generation-audio-for-cinema-white-paper.pdf>
- European Broadcast Union. (2014). *Loudness normalisation and permitted maximum level of audio signals* (EBU R128). Retirado de: <https://tech.ebu.ch/publications/r128s1>
- Gerzon, M. A. (1973). Periphony: With-height sound reproduction. *Journal of the Audio Engineering Society*, 21(1), 2-10.
- Gerzon, M. A. (1974). Surround-sound psychoacoustics. *Wireless World*, 80(1468), 483-486.
- Gerzon, M. A. (1975, March). The design of precisely coincident microphone arrays for stereo and surround sound. In *Audio Engineering Society Convention 50*. Audio Engineering Society.
- Gribben, C. (2018). *Investigations into the Perception of Vertical Interchannel Decorrelation in 3D Surround Sound Reproduction* (Doctoral dissertation). University of Huddersfield).
- Henrique, L. (2011). *Acústica Musical*. Lisboa: Fundação Galouste Gulbenkian
- Hollerweger, F. (2008). An Introduction to Higher Order Ambisonic. *Florian Hollerweger's Website*
- Howie, W. (2018). Capturing orchestral music for three-dimensional audio playback (Doctoral dissertation). Department of Music Research, Schulich School of Music McGill University, Montreal
- Howie, W., Martin, D., Benson, D. H., Kelly, J., & King, R. (2018). Subjective and objective evaluation of 9ch three-dimensional acoustic music recording techniques. In *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*. Audio Engineering Society.
- International Telecommunication Union. (2012). *Multichannel stereophonic sound system with and without accompanying picture* (ITU-R BS, 775-3). Retirado de: <https://www.itu.int/rec/R-REC-BS/en>
- International Telecommunication Union. (2015). *Method for the subjective assessment of small impairments in audio systems including multichannel sound systems* (ITU-R BS, 1116-3). Retirado de: <https://www.itu.int/rec/R-REC-BS/en>
- International Telecommunication Union. (2018). *Advanced sound system for programme production* (ITU-R BS, 2051-2). Retirado de: <https://www.itu.int/rec/R-REC-BS/en>

## Referências

- International Telecommunication Union. (2019). *General methods for the subjective assessment of sound quality* (ITU-R BS, 1284-2). Retirado de: <https://www.itu.int/rec/R-REC-BS/en>
- Kendall, G. S. (1995). A 3D sound primer: directional hearing and stereo reproduction. *Computer music journal*, 19(4), 23-46.
- Kendall, G. S. (2010). Spatial perception and cognition in multichannel audio for electroacoustic music. *Organised Sound*, 15(3), 228-238.
- Kim, S., Ko, D., Nagendra, A., & Woszczyk, W. (2013). Subjective evaluation of multichannel sound with surround-height channels. In *Audio Engineering Society Convention 135*. Audio Engineering Society.
- Lee, H., & Gribben, C. (2014). Effect of vertical microphone layer spacing for a 3D microphone array. *Journal of the Audio Engineering Society*, 62(12), 870-884.
- Marui, A., Kamekawa, T., & Irimaji, H. (2006). Subjective impression of surround sound microphone arrays. In *Audio Engineering Society Surround Experiments Project Report*
- Nakayama, T., Miura, T., Kosaka, O., Okamoto, M., & Shiga, T. (1971). Subjective assessment of multichannel reproduction. *Journal of the Audio Engineering Society*, 19(9), 744-751.
- Owsinski, B. (1999). *The mixing engineer's handbook*. Vallejo.
- Owsinski, B. (2009). *The recording engineer's handbook* (2nd edition). Nelson Education.
- Penha, R. (2014). Modelos de Especialização: Integração no Pensamento Composicional (Doctoral dissertation). Universidade de Aveiro, Portugal.
- Peters, N., Sen, D., Kim, M. Y., Wuebbolt, O., & Weiss, S. M. (2015). Scene-based audio implemented with higher order ambisonics (HOA). In *SMPTE 2015 Annual Technical Conference and Exhibition* (pp. 1-13). SMPTE.
- Rumsey, F., & Berg, J. (2001). Verification and correlation of attributes used for describing the spatial quality of reproduced sound. In *AES 19th International Conference: Surround Sound-Techniques, Technology, and Perception*. Audio Engineering Society.
- Rumsey, F. (2001). *Spatial audio*. Taylor & Francis.
- Rumsey, F. (2002). Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society*, 50(9), 651-666.



- Schoeps Mikrofone (2006). *Surround Recording Techniques*. [White Paper]. Retirado de:  
[https://schoeps.de/fileadmin/user\\_upload/user\\_upload/Downloads/Kataloge\\_und\\_Broschueren/Anwenderbroschueren/SCHOEPS\\_surround-brochure.pdf](https://schoeps.de/fileadmin/user_upload/user_upload/Downloads/Kataloge_und_Broschueren/Anwenderbroschueren/SCHOEPS_surround-brochure.pdf)
- Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., & Zotter, F. (2013). Spatial sound with loudspeakers and its perception: A review of the current state. *Proceedings of the IEEE*, 101(9), 1920-1938.
- Streicher, R., & Everest, F. A. (1998). *The new stereo soundbook* (2nd edition). Audio Engineering Associates.
- Theile, G. (2001). Multichannel natural music recording based on psychoacoustic principles. In AES 19th International Conference.
- Theile, G., & Wittek, H. (2011). Principles in surround recordings with height. In *Audio Engineering Society Convention 130*. Audio Engineering Society.
- Theile, G., & Wittek, H. (2012). 3D Audio Natural Recording. In *27th Tonmeistertagung - VDT Internation Convention*.
- Ton & Meister (2018). *3D-Audio Ambience Recording Techniques*. Retirado de <http://www.hauptmikrofon.de/3d/tonundmeister-2>
- Toole, F. E. (2008). *Sound reproduction. Loudspeakers and rooms*. Focal Press.
- Wittek, H., & Theile, G. (2017). Development and application of a stereophonic multichannel recording technique for 3D Audio and VR. In *Audio Engineering Society Convention 143*. Audio Engineering Society.
- Wittek, H. (2017) *ORTD-3D on VR glasses*. Retirado de: <http://www.hauptmikrofon.de/3d/vrapp>
- Zhang, W., Samarasinghe, P., Chen, H., & Abhayapala, T. (2017). Surround by sound: A review of spatial audio recording and reproduction. *Applied Sciences*, 7(5), 532.

## 7. Anexos

### 7.1 Especificações técnicas do material de captação

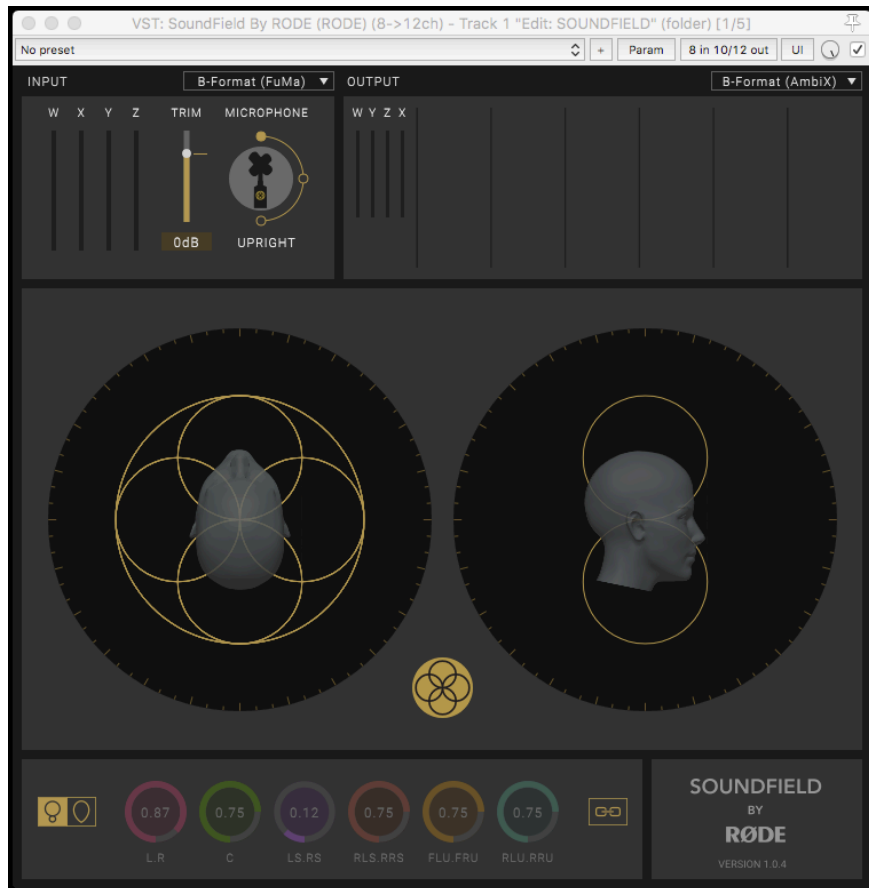
- Lista e descrição de vias

Via	Técnica	ID	DESCRIÇÃO	Observações:
1	OCT	L	1 * Schoeps CCM 41 L (compact microphone supercardioid)	Schoeps Surround Set
2	OCT	C	1 * Schoeps CCM 4 L (compact microphone cardioid)	Schoeps Surround Set
3	OCT	R	1 * Schoeps CCM 41 L (compact microphone supercardioid)	Schoeps Surround Set
4	OCT	LS	1 * Schoeps CCM 4 L (compact microphone cardioid)	Schoeps Surround Set
5	OCT	RS	1 * Schoeps CCM 4 L (compact microphone cardioid)	Schoeps Surround Set
6	OCT	LH	1 * Oktava MK-012 supercardioid	
7	OCT	RH	1 * Oktava MK-012 supercardioid	
8	OCT	LSH	1 * Oktava MK-012 supercardioid	
9	OCT	RSH	1 * Oktava MK-012 supercardioid	
10	Soundfield	LF-1	Soundfield MKV - W	MKV Processor
11	Soundfield	RF-2	Soundfield MKV - X	MKV Processor
12	Soundfield	LB-3	Soundfield MKV - Y	MKV Processor
13	Soundfield	RB-4	Soundfield MKV - Z	MKV Processor

### **MKV Processor:**

O microfone *SoundField* emprega uma matriz de quatro elementos formando um padrão tetraédrico que pode ser controlado eletronicamente a partir deste processador, que é conectado através de um cabo multicore. O controlador *SoundField* é na verdade um processador multifuncional que combina um pré-amplificador de microfone e a funcionalidades necessárias para controlar vários parâmetros. O painel frontal tem uma secção que contém um controlo do *Master Gain*, que fornece até 30 dB de ganho (discretizados em vários níveis de 10dB) e um controle *Fine Gain*. Este painel tem também botões de solo para as cápsulas individuais, bem como interruptores que compensam a orientação do microfone se o mesmo for usado numa posição invertida ou *end-fire*. O processador MKV contém também uma secção de saída estereofónica que consiste no controlo dos padrões de polaridade que variam entre omnidireccional e figura-de-oito. Para além disso, é também possível o controlo da orientação angular, que permite apontar eletronicamente as cápsulas na mesma direcção (0°) ou na direcção oposta (180°). Nesta secção é possível também retirar um *output* para *headphones* com controle de ganho dedicado. Para além destas características, existem também interruptores para a activação de um filtro passa-altos a partir dos 40 Hz, outro dedicado para a descodificação MS, bem como um conjunto de quatro barras LED que podem ser comutados para representar a saída em formato B ou a saída estéreo.

## 7.2 Plugins de descodificação, espacialização e de medição e ajuste de níveis



**Figura 14:** SoundField by RØDE

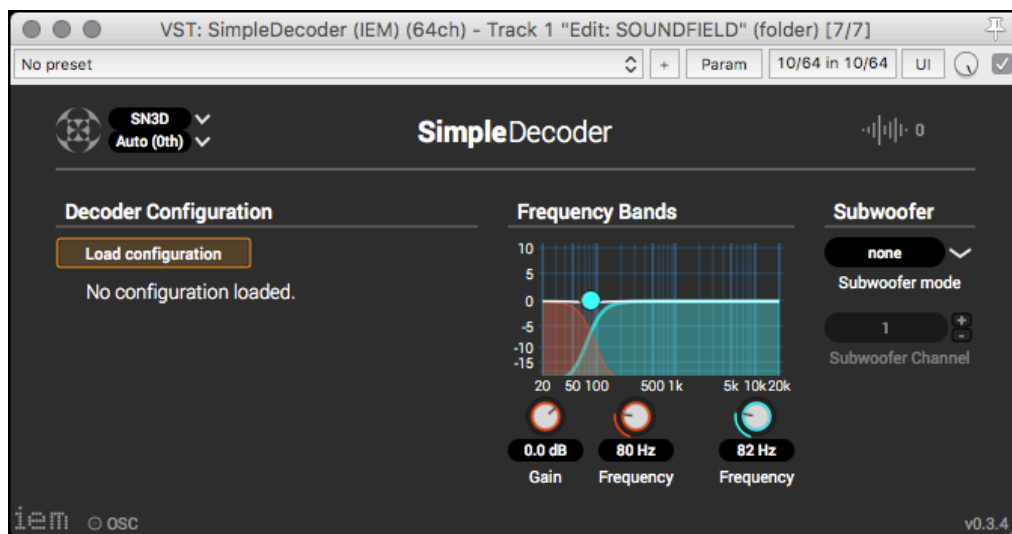
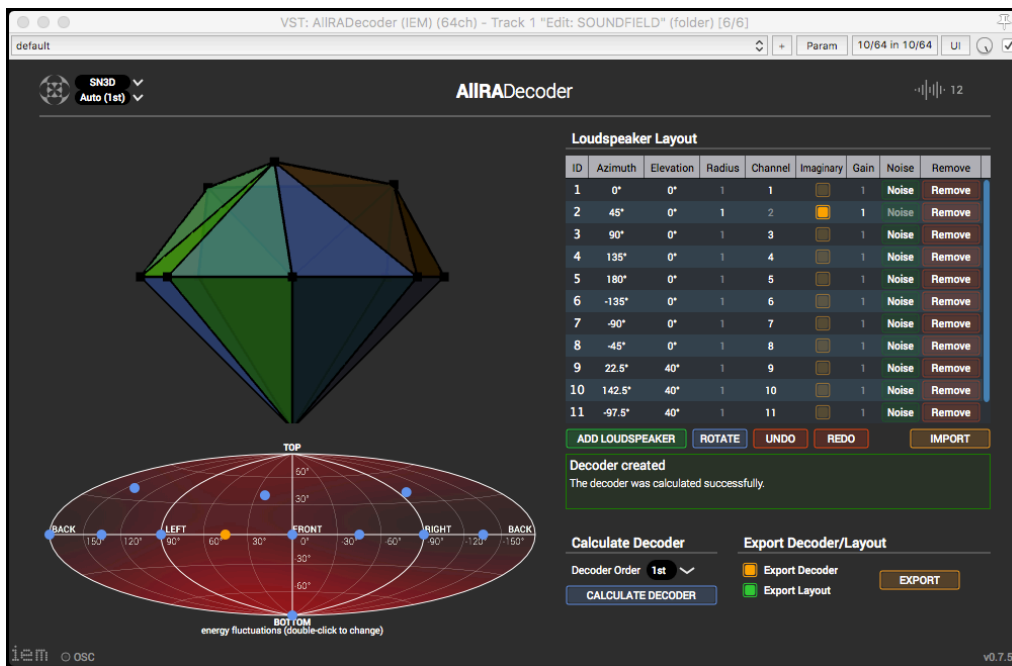
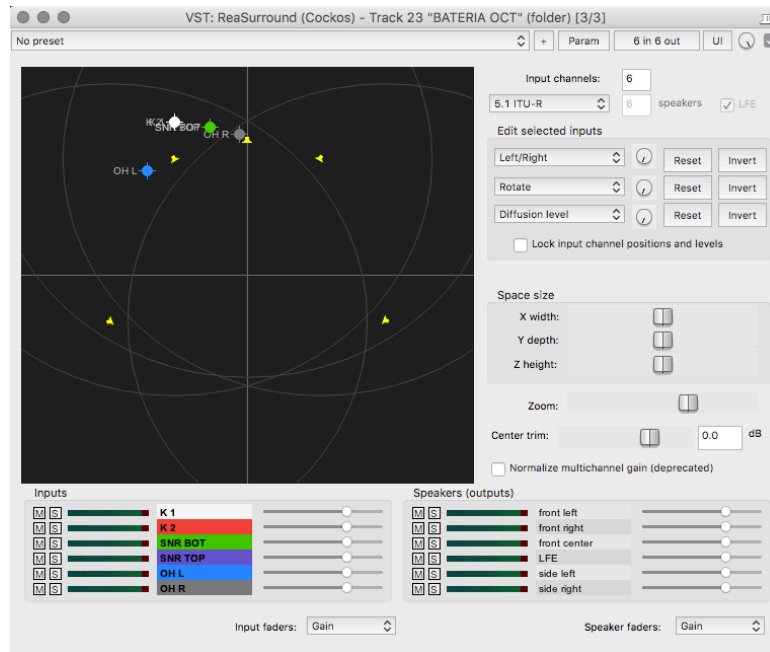


Figura 15: IEM Plug-in Suite: AIIRA Decoder e SimpleDecoder

## Referências



**Figura 16:** Reaper ReaSurround (Cockos)



**Figura 17:** Waves Loudness Meter

## 7.3 Instruções e Questionário de Avaliação

### A CONTRIBUIÇÃO DA ALTURA PARA A IMPRESSÃO ESPACIAL NO CONTEXTO DA PRODUÇÃO ÁUDIO

#### QUESTIONÁRIO DE AVALIAÇÃO SUBJETIVA

---

#### BREVE CONTEXTUALIZAÇÃO

De forma a compreender de que modo a adição de altifalantes de altura potencia a **impressão espacial** do ouvinte, o presente estudo explora diferentes estratégias, sistemas e tecnologias de produção musical para áudio imersivo.

Para melhor compreender estes conceitos, será feita a avaliação comparativa entre formatos de reprodução **com** e **sem** recurso a altura através da análise subjetiva de **dois formatos de captação** aplicados a estes sistemas. Desta forma, este teste tem como objectivo compreender e analisar a resposta subjetiva aos estímulos acústicos, focalizando-se na **impressão espacial** percebida.

#### DESCRIÇÃO DO QUESTIONÁRIO

**Em primeiro lugar, obrigada pelo interesse e pela participação nesta experiência.**

O teste será composto por **9** exemplos de avaliação, cada um composto por **2** excertos, **A** e **B**. A contribuição de cada participante centra-se na avaliação comparativa dos excertos musicais apresentados. Cada participante deve avaliar cada exemplo apresentado apenas em termos da **impressão espacial** percebida. As comparações serão feitas entre pares de excertos devendo concentrar-se na a **qualidade da impressão espacial** do excerto **B** quando comparado com **A**.

A pergunta que se propõe em cada momento de avaliação é a seguinte:  
Quão mais espacial é o excerto **B** em relação ao **A**?

Cada um dos exemplos poderá ser ouvido as vezes necessárias para que possa fazer a sua avaliação subjetiva do atributo em questão.

## Referências

### ATRIBUTO DE AVALIAÇÃO: IMPRESSÃO ESPACIAL

Para melhor compreender o foco de avaliação perceptual do teste, é importante considerar alguns sub-atributos e termos descritivos para a avaliação exacta e detalhada do atributo principal em questão.

#### IMPRESSÃO ESPACIAL

<b>ATRIBUTO PRINCIPAL</b>	A performance parece estar a ocorrer num ambiente espacial apropriado.
<b>SUB - ATRIBUTOS</b>	Homogeneidade do som espacial Reverberância Tamanho aparente da sala Equilíbrio acústico Perspectiva de profundidade
<b>TERMOS DESCRITIVOS COMUNS</b>	Sala reverberante / Sala seca Som directo / Som indirecto Sala grande / Sala pequena

### MÉTODO DE AVALIAÇÃO

Para a avaliação, é pedido que avalie cada exemplo segundo a seguinte escala:

- 3	- 2	- 1	0	1	2	3
Bastante menos espacial	Menos espacial	Ligeiramente menos espacial	Igual	Ligeiramente mais espacial	Mais espacial	Bastante mais espacial

Para cada exemplo apresentado, deverá assinalar na **Folha de Respostas** (página 3) o número correspondente à sua avaliação.

Se ainda restarem dúvidas, sinta-se à vontade para esclarecê-las antes de dar início ao teste. **Obrigada novamente pelo seu tempo!**



## FOLHA DE RESPOSTAS

Assinale as suas respostas de forma clara **indicando com uma cruz o número** correspondente à sua avaliação na **Tabela de Respostas** apresentada de seguida.

A avaliação deverá ser feita em relação ao excerto **B**.

- 3	- 2	- 1	0	1	2	3
Bastante menos espacial	Menos espacial	Ligeiramente menos espacial	Igual	Ligeiramente mais espacial	Mais espacial	Bastante mais espacial

## TABELA DE RESPOSTAS

<b>EXEMPLO 1</b>	- 3	- 2	- 1	0	1	2	3
<b>EXEMPLO 2</b>	- 3	- 2	- 1	0	1	2	3
<b>EXEMPLO 3</b>	- 3	- 2	- 1	0	1	2	3
<b>EXEMPLO 4</b>	- 3	- 2	- 1	0	1	2	3
<b>EXEMPLO 5</b>	- 3	- 2	- 1	0	1	2	3
<b>EXEMPLO 6</b>	- 3	- 2	- 1	0	1	2	3
<b>EXEMPLO 7</b>	- 3	- 2	- 1	0	1	2	3



