

FELIPE STOLNIK BORGES

AUDIODESCRIÇÃO, PERCEPÇÃO E SIMULAÇÃO BINAURAL

Orientadora: Prof^ª. Doutora Adriana de Almeida P.V. de Sá

Co-orientador: Prof. Doutor João Manuel Marques Carrilho

Universidade Lusófona de Humanidades e Tecnologias

**Lisboa
2023**

FELIPE STOLNIK BORGES

AUDIODESCRIÇÃO, PERCEPÇÃO E SIMULAÇÃO BINAURAL

Dissertação defendida em provas públicas na Universidade Lusófona de Humanidades e Tecnologias no dia 18/04/2023 para obtenção do grau de mestre em Produção e Tecnologias de Som perante o júri nomeado pelo Despacho de Nomeação n.º: 223/2023, de 13 de Abril, com a seguinte composição:

Presidente: Prof. Doutor João Manuel Marques Carrilho

Arguente: Profª. Doutora Ana Carrasco

Orientadora: Profª. Doutora Adriana de Almeida P.V. de Sá

Universidade Lusófona de Humanidades e Tecnologias

Escola de Comunicação, Arquitetura, Artes e Tecnologias da Informação

Lisboa

2023

"O respeito pela diversidade é o sinal de uma sociedade civilizada e avançada."

- Martin Luther King Jr.

Resumo

Pretende-se que esta investigação contribua para uma reflexão sobre a acessibilidade de pessoas com deficiência visual às narrativas audiovisuais. O principal método adotado por canais de televisão, cinemas e plataformas de *streaming* para promover essa acessibilidade é a audiodescrição (AD). Consiste em uma narração verbal que se justapõe à banda sonora original, de modo a fornecer informações relevantes sobre o que a pessoa com deficiência visual não pode ver. Tipicamente, a AD é posicionada frontalmente no campo auditivo, como se o narrador estivesse à frente e muito próximo do ouvinte. A descrição dos acontecimentos visuais é executada nas pausas entre diálogos, mas tende a ofuscar os efeitos sonoros, a música e os silêncios, que também possuem significados para a narrativa. Será possível conjugar o som original e a audiodescrição sem desfavorecer a inteligibilidade das diversas componentes sonoras? São poucas as investigações que se dedicam a esta questão, que é foco central no nosso próprio trabalho.

Esta dissertação aborda os princípios que governam a agregação e segregação dos sons, e os mecanismos de localização e percepção auditiva. Debruça-se sobre espacialização sonora e apresenta um estudo de caso, que explora diversas maneiras de posicionar a audiodescrição no campo auditivo. O estudo utiliza as ciências da percepção para analisar como a simulação computacional de binauralidade pode afetar a inteligibilidade da AD e do som original do filme. Com efeito, a utilização de tecnologias binaurais abre novas perspectivas sobre o problema, mas não garante a sua resolução; a eficácia de cada estratégia dependerá sempre de um estudo aprofundado dos nossos mecanismos de percepção.

Palavras-chave: acessibilidade; audiodescrição; audiovisual; binaural; percepção auditiva.

Abstract

This investigation contributes to a reflection on how visually impaired people can be given access to audiovisual narratives. The main method adopted by television channels, cinemas and streaming platforms is audio description (AD). It consists of a verbal narration that is played-back along with the original soundtrack, in order to provide relevant information about what a blind person cannot see. Typically, the AD is positioned frontally in the auditory field, as if the narrator were in front and very close to the listener. The description of visual events is performed in the pauses between dialogues, but tends to overshadow the sound effects, music and silences, which might be equally significant in the narrative. So, is it possible to combine the original sound and the audio description without damaging the intelligibility of the different sound components? There are few investigations dedicated to this issue, which is the central focus in our own work.

This dissertation addresses the principles that govern the aggregation and segregation of sounds, and the mechanisms of localization and auditory perception. It focuses on sound spatialization and presents a case study, which explores different ways of positioning the AD in the auditory field; The study uses perceptual sciences to analyze how computer simulations of binaurality might affect the intelligibility of AD and original film sound. Indeed, the use of binaural technologies opens up new perspectives on the problem, but it does not guarantee its resolution; the effectiveness of each strategy will always depend on an in-depth study of our own perceptual mechanisms.

Keywords: accessibility; audio description; audio-visual; binaural; auditory perception.

Abreviaturas

AD	Audiodescrição
ASA	Auditory Scene Analysis
DAW	<i>Digital audio workstation</i> é o software de produção e mistura de áudio
HRTF	<i>Head related transfer function</i> é uma resposta que caracteriza como um ouvido recebe um som de um ponto no espaço. Além de incluir ITD e ILD, as características estão ligadas aos efeitos de filtragem produzidos pela orelha. É também uma forma de descrever matematicamente a influência da cabeça e torso do ouvinte.
HRIR	<i>Head related impulse-response</i> é uma é uma medição do atraso de tempo e mudanças de amplitude que ocorrem a um som quando ele atinge o ouvido de um indivíduo depois de passar por sua cabeça, pescoço e tronco.
ILD	Diferença de Nível Interaural (<i>interaural level differences</i>)- é a diferença de nível ou volume que um ouvido escuta antes do outro.
ITD	Diferença de Tempo Interaural (<i>interaural time differences</i>)- é a diferença de tempo que um ouvido escuta antes do outro.
Pan	Panorama ou <i>Panning</i> é a distribuição de sinal de som em um campo de som estéreo ou multicanal determinado por controle de configuração da posição do som.

Índice

1. Introdução	1
2. Audiodescrição	4
2.1 Princípios da Audiodescrição	6
2.2 Audiodescrição em Cinema e TV	9
2.3 Audiodescrição e Acessibilidade: Problemas e Possibilidades	13
3. Espacialização Sonora: Do Mono ao Binaural	17
3.1 Mono	18
3.2 Estéreo	19
3.3 Surround	20
3.4 Binaural	22
3.4.1 Gravação e Reprodução Binaural	27
3.4.2 Técnicas de Gravação Binaural	30
3.4.3 Protótipo de Microfone Binaural	33
3.4.4 Simulação Binaural via Plugins	35
3.4.5 Reprodução Binaural via auscultadores	38
4. Percepção Auditiva	40
4.1 A Escuta em Pessoas com Deficiência Visual	42
4.2 Análise de Cena Auditiva	44
4.3 Percepção da Origem Espacial do Som	48
4.3.1 Função de Transferência Relacionada com a Cabeça (HRTF)	50
4.3.2 Diferença de Tempo Interaural (ITD)	51
4.3.3 Diferença de Nível Interaural (ILD)	53
4.3.4 Mecanismos Perceptivos que Influenciam a Inteligibilidade da Fala	54
5. Estudo de caso: o posicionamento da audiodescrição no campo auditivo	56
5.1 Pressupostos	57
5.2 A Obra Escolhida e a Gravação da Audiodescrição	61
5.3 As Versões De Mistura	63
5.4 Entrevistas	69
5.5 Discussão De Resultados	71
Referências Bibliográficas	82
Glossário	86
Anexos	87

Índice de Figuras

Figura 1 - Exemplo de audiodescrição	7
Figura 2 - Guião de audiodescrição	12
Figura 3 - Reprodução dos sistemas de áudio	17
Figura 4 - Sistema Surround 5.1	21
Figura 5 - Ângulo azimute e elevação no espectro de audição em 360°	23
Figura 6 - Oscar, o primeiro dummy head microphone	24
Figura 7 - Configuração do posicionamento estéreo via altifalantes	25
Figura 8 - Plano médio, plano frontal e plano horizontal	28
Figura 9 - Exemplo de localização sonora desempenhada pelo ouvido	29
Figura 10 - 3DIO Omni Binaural Microphone / Neumann KU100	30
Figura 11 - Microfone binaural Roland CS10-EM	31
Figura 12 - Fluxo de funcionamento do microfone binaural Neumann KU100	32
Figura 13 - Construção do protótipo do microfone binaural (Orelhas PVC + Behringer C2)	33
Figura 14 - Protótipo do microfone binaural (Orelhas silicone + Microfones lapelas)	34
Figura 15 - The Anaglyph binaural audio engine	36
Figura 16 - DearVR Pro Binaural Plugin	37
Figura 17 - Anatomia do canal auditivo	41
Figura 18 - Comparativo de segregação primitiva	46
Figura 19 - O ouvido mais próximo da fonte sonora recebe o estímulo sonoro antes do outro	52
Figura 20 - Fontes sonoras de um lado são mais altas no ouvido próximo devido ao mascaramento da cabeça	53
Figura 21 - Análise espectral do som da curta-metragem	61
Figura 22 - Mistura #1 - espectro (Hz) e amplitude(dB) da AD e da banda sonora	64
Figura 23 - Mistura #2 - espectro (Hz) e amplitude(dB) da AD e da banda sonora	65
Figura 24 - Plugin de simulação binaural DearVR	66
Figura 25 - Análise de espectro de frequências e nível (dB) da AD binaural à direita	67
Figura 26: Plugin de simulação binaural Ambeo	68
Figura 27 - Análise de espectro de frequências e nível (dB) da AD binaural com elevação	69
Figura 28 - Espectrograma com onda sonora	72

Índice de Tabelas

Tabela 1 - Ordem de acertos sobre inteligibilidade da obra das versões de audiodescrição	71
Tabela 2 - Análise de frequências das quatro misturas em relação a banda sonora original	73
Tabela 3 - Análise espectral de frequências das quatro versões de AD	75
Tabela 4 - Análise espectral de frequências da Mistura #1 de AD	76
Tabela 5 - Análise espectral de frequências da Mistura #2 de AD	76
Tabela 6 - Análise espectral de frequências da Mistura #3 de AD	77
Tabela 7 - Análise espectral de frequências da Mistura #4 de AD	78

1. Introdução

A audiodescrição (AD) serve como um meio de acessibilidade para que pessoas que possuem baixa visão e cegueira compreendam detalhes visuais sobre as obras audiovisuais. A modalidade consiste em uma narração descritiva inserida nas pausas entre diálogos para informar sobre algum acontecimento que não seria possível de compreender somente ao ouvir. Apontam-se acontecimentos relevantes como o surgimento de personagens em cena ou a apresentação de um novo lugar. O processo de realização da audiodescrição de um filme inicia-se quando os audiodescritores adquirem acesso à obra. Em seguida, estrutura-se um guião que é gravado e inserido numa nova exportação de arquivo sonoro para ser enviado à distribuição digital.

Embora a narração audiodescritiva seja reconhecida e esteja disponível em cinemas, plataformas de *streaming* e canais de televisão, é alvo de severas críticas relacionadas com a falta de tempo para descrever a narrativa. Devido a ser inserida nas pausas entre diálogos, a descrição é limitada ao tempo da pausa. Entretanto, este curto intervalo costuma possuir algum efeito sonoro ou música, que também são elementos que podem contribuir para a narrativa fílmica.

Devido aos criadores do filme na maioria das vezes não serem os responsáveis pelo guião audiodescritivo, este é escrito por guionistas que não estiveram envolvidos nas obras. A audiodescrição costuma ser inserida por empresas contratadas após o processo de finalização das obras. Ressalta-se uma falta de interesse em incluir a escuta acessível por parte das produtoras audiovisuais, pois tornam a permitir a livre interpretação de um guionista de AD que não participou da obra, o que pode fazer com que ela não se adapte tão bem à mistura final de som (e.g. Lopez, 2020).

Um problema fundamental diz respeito à justaposição da audiodescrição e do som original do filme. Quando múltiplos sons são simultaneamente produzidos pela mesma fonte sonora, podem-se tornar menos inteligíveis do que quando reproduzidos por duas fontes distintas; acontece, por exemplo, quando se sobrepõem sons na mesma gama de frequências (e.g. Bregman, 1994).

Nosso estudo buscou utilizar as possibilidades do som binaural para investigar sobre um posicionamento da audiodescrição no campo auditivo que favoreça a inteligibilidade de escuta. Para isso, esta dissertação foi dividida em capítulos sobre audiodescrição, espacialização sonora, percepção auditiva e estudo prático.

No primeiro capítulo, aprofundamos o que foi dito nos parágrafos anteriores, para perceber quais são os princípios e as limitações do cinema inclusivo para as pessoas com deficiência visual. Lopez (2020) sugere que as opções de acessibilidade devem ser mais variadas, pois nos dias de hoje o ouvinte com deficiência visual apenas conta com a modalidade da audiodescrição como referência inclusiva no cinema.

No capítulo seguinte, dedicado à espacialização, abordamos as noções de mono, estéreo e *surround*, para depois especificarmos a trajetória, as técnicas de reprodução, gravação e simulação de áudio binaural. Mono consiste num único sinal sonoro, estéreo resulta de dois sinais distintos, e tipicamente, o termo '*Surround*' implica múltiplos sinais sonoros. Assim como o estéreo, o som binaural é composto por duas ondas sonoras mas costuma ser reproduzido via auscultadores para que maximize-se a informação espacial, reproduzindo as condições físicas de um ambiente e o modo como o ouvido humano capta as frequências e reflexões de som, tendo em conta o torso, ombros e face. Um microfone binaural de alta qualidade é muito dispendioso, e gostaríamos de incentivar pequenas empresas a produzir o seu próprio material de acessibilidade. Experimentamos criar um protótipo próprio, cujas limitações foram instrutivas. Alternativamente, pode-se utilizar um modelo computacional para criar uma simulação binaural.

O terceiro capítulo aprofunda conhecimentos sobre a percepção auditiva, debruçando-se sobre as particularidades da escuta em invisuais, sobre a agregação e segregação dos sons, e a identificação das suas posições relativas no espaço. Para aprimorar o conhecimento sobre a tecnologia binaural, investigamos como o percurso que um som faz até atingir os ouvidos se reflete em diferenças de tempo e de nível (i.e. volume) interaural. Para melhor compreendermos as vicissitudes da audiodescrição, estudamos os mecanismos que influenciam a inteligibilidade da fala.

Finalmente, apresentamos um estudo de caso. Para investigar a inteligibilidade de diferentes posicionamentos da narração, foram testadas quatro versões de AD: mono, estéreo, estéreo binaural e estéreo binaural com elevação. Utilizámos um

plugin binaural para simular um ambiente completo de reflexão, reverb, distância e posição manipuláveis dentro do campo de escuta. Para analisar resultados servimo-nos de diversas investigações dedicadas à análise da cena auditiva (e.g. Bregman, 1994), bem como de um inquérito a participantes com deficiência visual, sobre a compreensão da narrativa audiodescritiva.

2. Audiodescrição

Esta seção descreve a trajetória histórica da audiodescrição (AD), além de suas regras e aplicações. Este capítulo traça os princípios da audiodescrição, isto é, como ela é percebida pelos desenvolvedores e usuários. Apresentamos a aplicação da AD no seu primórdio (1981) e no tempo presente (2022), incluindo reflexão seguida de críticas pelos investigadores de audiodescrição na atualidade.

Franco e Silva (2010) descrevem a audiodescrição como um processo de transformação de imagens em palavras que possam informar alguém que tenha nula ou baixa visibilidade sobre algum acontecimento que é considerado relevante para dar o melhor entendimento do que está a acontecer na imagem. A narração audiodescritiva pode ser utilizada em peças de teatro, filmes, programas de TV e espetáculos de dança. Esse formato de descrição está sendo reconhecido no mundo inteiro, embora muitas pessoas com deficiência visual ainda não conheçam esta opção devido a falta de informação sobre o assunto. Bourne (2008) caracteriza a AD como uma informação verbal inserida entre os diálogos que visa auxiliar as pessoas com deficiência visual a apreciar o que está passando na tela. Franco e Silva (2010) definem que “a audiodescrição consiste na transformação de imagens em palavras para que informações-chave transmitidas visualmente não passem despercebidas e possam também ser acessadas por pessoas cegas ou com baixa visão”. (Franco e Silva, 2010, p.23)

De acordo com Snyder (2014), o conceito da AD foi estruturado entre os anos 1960 e 1980, mas ganhou popularidade entre pessoas com deficiência visual somente na década seguinte. Em 1981, Margaret Rockwell e seu marido Cody Pfanstiehl, que trabalhavam com pessoas com deficiência visual, receberam incentivo do governo para promover a audiodescrição em uma peça de teatro transmitida por rádio em Washington, chamada Major Barbara. No ano seguinte, eles promoveram a audiodescrição para a televisão, durante a série American Playhouse, usando o canal pré-gravado via satélite Second Audio Program (SAP *channel*) para transmitir a descrição pelo rádio.

Em Portugal, segundo o censo de 2011, cerca de 900 mil pessoas têm algum tipo de deficiência visual, e cerca de 28 mil têm cegueira total. De acordo com Neves

(2011), as primeiras narrativas de audiodescrição em Portugal ocorreram em jogos de futebol, seguidos por radionovelas, sem se considerar o fato de que estavam servindo como ajuda para pessoas com deficiência visual. A autora afirma que é difícil determinar a origem precisa da história da audiodescrição em Portugal, uma vez que ela surgiu de forma espontânea em círculos restritos. Na televisão, houve o registro de dezembro de 2003 com o filme *A Menina da Rádio* (1944), transmitido pela RTP. Em 2004, essa emissora transmitiu mais filmes com audiodescrição. A autora cita sobre a deliberação em vigor desde 2009 pela RTP:

Em cumprimento da Deliberação 5/OUT-TV/2009, aprovada pelo Conselho Regulador da Entidade Reguladora para a Comunicação Social (ERC) a 28 de Abril de 2009, definindo “o conjunto de obrigações dos operadores estatais e privados que permitam o acompanhamento das emissões por pessoas com necessidades especiais”, a RTP passou a oferecer com maior regularidade audiodescrição em séries de produção nacional. (Neves, 2011, p.71)

Embora a emissora portuguesa RTP conte com alguns programas com legendagem audiodescritiva, o catálogo ainda é muito escasso e não são todos os dias que são transmitidos programas audiodescritos. A emissora conta com a opção da audiodescrição sair por rádio numa frequência pré-definida ou pela própria televisão (TDT¹).

Por outro lado, no Brasil há uma recente Lei 14.009/20, que tinha prazo de 1º de janeiro de 2021 para que todas as salas de cinema brasileiras oferecessem recursos de acessibilidade para pessoas com deficiência visual ou auditiva, como audiodescrição e legendagem descritiva. Nos Estados Unidos uma lei semelhante já está em vigor desde 2016. Já no universo do teatro e nas artes performáticas também é pouco utilizado o recurso da AD. Segundo Neves (2011), pode ser pelo facto das organizações possuírem pouco conhecimento a respeito do assunto ou pelo grau de dificuldade em realizar tal descrição ao vivo.

Afundo, buscando um pouco da representação da importância de acessibilidade, de acordo com uma pesquisa de Mestrado da PUC-SP por Camila Neves, com uma amostra de 30 invisuais residentes de Belém no estado do Pará no Brasil, foi observado que 97% destes indivíduos consome televisão, sendo quase a metade (47%) entre 2 a 4 horas diárias. Com isso, foi possível perceber que os meios

¹ TDT - Televisão digital terrestre.

de inclusão e acessibilidade ainda são escassos, pois 93% do público afirmou que a televisão ainda precisa aprimorar os recursos auditivos para que haja a melhor compreensão deste grupo. Ainda, apenas 52% destes conheciam o conceito de audiodescrição, sendo que a maioria que conhece a utiliza para consumir material audiovisual como filmes e séries.

2.1 Princípios da Audiodescrição

De acordo com Neves (2011), existem diferentes modalidades de aplicação da audiodescrição. Ela pode ser utilizada para nichos como cinema, televisão, web, teatro, espetáculos de música e dança, museus, espaços públicos e outras aplicações. Embora o enfoque da dissertação seja a audiodescrição em peças audiovisuais, é considerado relevante apresentar um pouco deste segmento para aumentar o conhecimento sobre o tema. Assim, a autora define a audiodescrição ou áudio-narrativa como:

A arte de traduzir, através de uma narrativa descritiva ou outras técnicas verbais, mensagens visuais não perceptíveis apenas através dos sinais acústicos presentes em textos (áudio) visuais (filmes, produtos multimídia, Web, ...); a arte de descrever imagens, objectos, realidades com valor comunicativo essencialmente visualista (ex. paisagens, património construído, peças de museu). (Neves, 2011, p.11)

Neves (2011) ainda descreve alguns tipos de audiodescrição. O primeiro é a descrição de imagens estáticas, encontrada em museus, na web ou em outros contextos em que uma imagem é objeto de contemplação visual, como fotografias, pinturas, esculturas ou paisagens. Em seguida, a descrição de imagens em movimento inclui filmes e programas de televisão. Outro tipo de audiodescrição mencionado pela autora é a descrição de exploração, que consiste em ajudar a orientação de mobilidade para pessoas com deficiência visual. A descrição de orientação é outra modalidade de audiodescrição, sistemas aprimorados de GPS com foco em mobilidade para ajudar as pessoas com deficiência visual a se sentirem mais confortáveis nas ruas.

Para que um audiodescritor execute bem o seu trabalho, Neves (2011) recomenda que ele tenha conhecimentos de psicofisiologia da visão, de cegueira e

dos processos sócio-cognitivos da visão. Além disso, é preciso conhecer as técnicas de oralidade, escrita e locução, incluindo noções de ritmo, pronúncia, dicção, respiração, etc. É possível que a pessoa que escreve o texto não seja a mesma que o fala, pois cada produtora de pós-produção trabalha de acordo com suas preferências, sem uma regulamentação de regras a seguir.

Como esta dissertação se concentra na audiodescrição audiovisual, um exemplo simples da sua aplicação é apresentado. Na Figura 1, busca-se uma explicação simples do contexto da cena. Desta forma, questões que o invisual não conseguiria captar são relatadas. A narração poderia ser algo como: ‘Jack e Rose estão à deriva no oceano, apoiados por uma peça de madeira’.



Figura 1 - Exemplo de audiodescrição

Fonte: Titanic (1997) James Cameron. Paramount Pictures

Cada país tem a sua própria forma de conduzir a audiodescrição. Por exemplo, tanto no Brasil como em Portugal não há leis que estabeleçam uma padronização para os guiões de AD, existindo apenas guias sugestivos disponibilizados, respectivamente, pelo Ministério da Cultura e pela Secretaria do Audiovisual. Esses guiões têm como objetivo orientar a elaboração da audiodescrição em questões técnicas, linguísticas e de tradução. Nem sempre o guionista é o narrador da audiodescrição, por isso é importante manter um padrão na estrutura do guião, incluindo o tempo inicial e final das inserções da AD, bem como a última fala da personagem antes da entrada da AD.

Quanto ao momento de inserção, o guia sugere que as narrações sejam colocadas entre os diálogos sonoros, isto é, quando não há falas.

Pode ser ligeiramente adiantada ou atrasada em relação à cena para dar informações necessárias ao andamento da narrativa, desde que não antecipe factos ou faça versões do que está previsto. O audiodescritor precisa decidir se a informação é importante e se vale a pena operar esse tipo de deslocamento. (Naves, Mauch, Alves Araujo, 2016)

Ainda sobre as questões relacionadas à narração, é importante que seja fluida e não monótona, ajustada à narrativa. Por exemplo, em programas infantis, sugere-se que a narração seja lúdica e animada. De acordo com Naves, Mauch, Alves, Araújo (2016), a narração/audiodescrição não é um elemento que participe da construção do significado na elaboração de uma obra, mas quando inserida, torna-se um elemento da composição do significado para quem a utiliza. É aconselhável que a audiodescrição não se sobreponha aos diálogos e demais sons importantes para a narrativa, incluindo a banda sonora. No entanto, nem sempre é possível dar prioridade à música, já que os diálogos ainda são a principal forma de transmitir informações.

De acordo com Routledge (2016), o Royal National Institute of Blind People (RNIB) na Inglaterra estabeleceu algumas regras para a audiodescrição infantil com base nas opiniões de pais e terapeutas ao longo dos anos. Ao lidar com crianças com deficiência visual, é necessário ter cuidado para não incluir informações irrelevantes ou que alterem o contexto da obra. Além disso, a linguagem utilizada deve ser mais lenta e a entonação deve ser dinâmica. No entanto, estas não são regras universais, já que cada país pode escolher seguir ou não algum guia de referência para a elaboração da audiodescrição. Por exemplo, o guia brasileiro não inclui uma seção específica para audiodescrição infantil, mas mantém as mesmas recomendações de manter diálogos claros e evitar sobreposições. No entanto, esse guia destaca a importância de se adaptar ao tom emocional da obra.

2.2 Audiodescrição em Cinema e TV

De acordo com Neves (2011), a AD audiovisual inclui a AD de cinema, de televisão, DVD e DVB/web. Em sessões especiais inclusivas, a AD pode ser apresentada em sinal aberto, isto é, está inserida para todos os ouvintes da sala a escutarem. Alternativamente, a AD pode ser apresentada em sinal fechado, com uso de equipamento específico - normalmente um receptor com auscultador. Esta segunda opção requer mais investimento, mas de acordo com a autora é a melhor para oferecer acessibilidade à pessoa invisual numa sala de cinema, e assim fazer coabitar formas diferentes de acesso ao texto fílmico. A experiência de um indivíduo visual e um invisual sobre algum filme é diferente, mas devido a AD, estes dois indivíduos poderão discutir sobre algum mesmo filme.

Por outro lado, ainda segundo Neves (2011), a *web* é um dos meios mais convenientes para a audiodescrição. Devido à personalização e à possibilidade de inserir ilimitadas camadas de áudio a serem tocadas simultaneamente, juntamente com diversos *plugins* e *softwares* que são capazes de sincronizar o tempo de dois canais de áudio ao mesmo tempo, pode-se afirmar que o maior portal com maior variedade de catálogos acessíveis é a *web*.

De acordo com Snyder (2014), embora o acervo de AD televisivo seja menor em comparação à internet, a televisão é o meio mais conhecido e utilizado na audiodescrição, mas por motivos econômicos e técnicos, o catálogo de filmes com AD em canais de televisão e no cinema é pequeno. Por outro lado, por mais que o mercado de DVD nos dias de hoje esteja em declínio, a nível global contribuiu muito para a expansão do conceito da AD, pois oferecia diversos idiomas e não custava muito gravar uma pista adicional para ser colocada no menu do disco (e.g. Snyder, 2014).

Seja qual for o meio audiovisual utilizado, a AD de um filme narrativo é um dos tipos mais complexos de se produzir, visto que é preciso “abordar a imagem nas suas relações sequenciais - desenrolar da ‘história’ - e nas suas relações simultâneas com todos os elementos composicionais com que interage, e em particular, com a banda sonora” (Neves, 2011, p.31). A autora menciona o som e suas diretrizes - efeitos sonoros e música - como por ser necessário a busca pela harmonia entre a narração e

o restante dos elementos sonoros. Afinal, existem diversos exemplos e situações dentro das narrativas filmicas com que o narrador deverá lidar:

Tomemos, a título de exemplo, os filmes musicais que partilham muitas das condicionantes da AD das artes performativas; comédias com efeitos visuais cômicos; filmes de ação, densamente condicionados pelos efeitos sonoros e pelo ritmo; filmes de terror, com enorme impacto emocional e tensão cumulativa; filmes com imagens de sexo explícito ou violência gráfica; ou, ainda, as séries e telenovelas, produtos altamente convencionados, que se estendem por inúmeros episódios, com um elevado número de personagens e enredos, que se deslocam num igualmente diversificado espaço físico e social; os documentários com narração em off, com uma profusão de imagens e bandas sonoras densamente preenchidas com efeitos sonoros, música e a voz do narrador e eventuais participantes; eventos desportivos com imagens em diversos ângulos; obras em língua estrangeira com legendas na língua de chegada; programas infantis, com grande carga pictórica e imagética; ou até mesmo a publicidade (Neves, 2011, p.32).

Nos dias de hoje, plataformas de Streaming com um vasto catálogo de gêneros filmicos com material audiodescritivo. Assim como a Netflix disponibiliza a opção do idioma com audiodescrição em seus títulos originais a HBO, Amazon Prime, Disney + e Hulu também contam com um acervo de AD.

Dentro das obras audiovisuais da Netflix, o guião da AD pode ser diferente entre a versão brasileira e a versão feita nos Estados Unidos. Ao comparar as ADs da série *Super Drags*², pode-se perceber que a quantidade de detalhes e informações na versão brasileira é maior do que na estadunidense. No entanto, devido à narração descritiva brasileira ser mais executada durante o decorrer de algum episódio da obra, ela é reproduzida simultaneamente ao som original, interferindo por cima de diálogos. Na versão norte-americana (EUA), os diálogos nunca são invadidos pela AD, tornando a descrição narrativa menos frequente.

Para Neves (2011), o processo para a elaboração da audiodescrição de imagens em movimento inicia-se com a leitura e a interpretação da obra. Em seguida, é preciso que o guionista de AD compreenda o enredo e outros aspectos relevantes sobre a narrativa, tal como os personagens. Após, inicia-se o processo de elaboração do guião de AD. “Não existem estratégias pré-definidas, pelo que se verifica que cada profissional segue a sua própria intuição e utiliza as ferramentas técnicas e linguísticas

² Dir. Fernando Mendonça. *Combo Estúdio*. 2018

que transporta de experiências anteriores enquanto guionista ou tradutor/tradaptador.” (Neves, 2011, p.38)

O guia da Netflix sobre a produção de audiodescrição sugere que tanto o texto como a narração sejam adaptados ao gênero da obra.

Ao descrever ações, nem todos os elementos precisam ser incluídos o tempo todo. Determine o que é mais relevante para a história fluir sem impactar negativamente a experiência dos espectadores. Evite a sobrecarga de informações quando não forem relevantes ou quando os mesmos detalhes puderem ser discernidos do diálogo/música. (Netflix, 2022)³

De fato, é preciso que sejam mantidas algumas questões formais para a elaboração de guiões de AD. Neves (2011) explica que nem sempre os ouvintes invisuais se interessam pelo mesmo nível de detalhamento dos acontecimentos. Algumas pessoas podem preferir mais ou menos explicações sobre o contexto de uma cena fílmica. Ainda, é recomendado que o audiodescritor não faça interpretações pessoais, para que se torne mais fácil a compreensão por parte dos ouvintes. Deve-se deixar questões e linguagens pessoais de fora do texto, e não generalizar com base na ideia de deficiência, ou tentar compensar com informações que não estão disponíveis em cena. Por exemplo, se uma personagem transmitir uma emoção por meio da expressão facial, deve-se levar em conta que a interpretação deste sentimento pode ser subjetiva, portanto deve-se tentar ser o mais neutro possível nas descrições.

Embora não exista um padrão para a elaboração dos guiões audiodescritivos, provavelmente não faltarão indicações quanto ao tempo - os milissegundos do intervalo em que a descrição vai ser inserida na obra (*TCR - Timecode Reader*), juntamente com a descrição a ser gravada pela voz do narrador. A Figura 2 exemplifica um simples roteiro áudio descritivo com o tempo de inserção para cada descrição.

³ Texto original: *When describing actions, not all elements need to be included at all times. Determine what is most relevant for the story to flow without negatively impacting the viewers' experience. Avoid information overload when not relevant or when the same details can be discerned from dialogue/music.* Disponível em <<https://partnerhelp.netflixstudios.com/hc/en-us/articles/215510667-Audio-Description-Style-Guide-v2-3>>

Nº ordem	TCR		Descrição
7	00:00:46,613 00:00:51,651	=>	Uma rodovia de mão dupla que corta o sertão é vista de dentro de um veículo em movimento.
8	00:00:52,826 00:00:55,948	=>	Às margens da rodovia uma vegetação seca.
68	00:15:50,981 00:16:02,377	=>	Lentamente a câmera percorre o quintal, onde há uma pequena planta de folhas finas e verdes, um bode, chão seco, uma cerca de pau-a-pique e uma árvore com poucas folhagens.
231	01:05:23,185 01:05:36,052	=>	Sertão. A paisagem é seca e há duas árvores ao fundo. Atrás delas há uma cerca de pau-a-pique. Bodes caminham perto das árvores. Ao longe, o pai tange os animais.

Figura 2 - Guião de audiodescrição

Fonte: Araujo, Vera. 2011 ⁴

É possível perceber que os narradores costumam alterar o tom de suas vozes de acordo com o gênero do filme que está sendo gravado. Por exemplo, na Netflix, pode-se comparar a AD de *Nasce Uma Rainha*⁵ com *Reality Z*⁶, ambas as séries em português. O primeiro é um reality show de comédia, em que a narração audiodescritiva possui um tom sorridente, embora as frases sejam curtas e rápidas, já a segunda possui um tom de voz mais neutro, visto que a série é um drama de terror.

⁴ Disponível em: <<https://scielo.br/j/tla/a/mFsPFb8FBvGFwRgXtNh7qDm/?lang=pt>>

⁵ Gustavo Mello, Mari Nunes. Original Netflix, 2020

⁶ Cláudio Torres, Charlie Brooker. Conspiração Filmes e Netflix, 2020.

2.3 Audiodescrição e Acessibilidade: Problemas e Possibilidades

A experiência cinematográfica de um indivíduo invisual não é a mesma que a de um indivíduo visual: o que define cinema é a imagem. “Se, como dissemos, é a imagem que, ontologicamente, define o cinema, aquilo que, em contrapartida, marca a diferença entre cinema e televisão não é tanto a especificidade da imagem, mas a posição diferente que nela ocupa o som.” (Chion, 2011, p.125)

A percepção é a interpretação do que foi captado pelos sentidos, sendo influenciada por fatores fisiológicos e psicológicos. A integração multissensorial faz parte dos seus aspetos automáticos e inconscientes. O termo refere-se aos processos pelos quais os diversos sentidos se influenciam reciprocamente, originando uma experiência unificada do mundo (e.g. Sá, 2013). As chamadas ‘ilusões multisensoriais’ exemplificam como a informação auditiva pode alterar a percepção visual e vice-versa.

O termo *valor acrescentado* foi introduzido na teoria audiovisual por Michel Chion (2011), para designar o modo como a combinação entre som e imagem produz significados e efeitos que não estão contidos na própria imagem ou no próprio som. O efeito total é diferente da soma das partes. Chion afirma que o som pode ser um auxílio à imagem que está sendo projetada. Para o autor, um aspecto importante a considerar é a linguagem: ele postula que os filmes costumam centrar-se na voz, que é o aspecto mais significativo para a comunicação. Além disso, o som dá contexto às imagens apresentadas, orientando o espectador. O autor comenta que os sons influenciam na temporalidade. Por *animação temporal da imagem*, refere-se à capacidade do som de mudar a percepção do espectador sobre a imagem. A *linearização temporal dos planos* trata do modo como o som providencia uma lógica abrangente. O som também possui a capacidade de guiar as cenas. Chion usa o termo *vetorização* para descrever uma orientação ou expectativa que leva o espectador a um objetivo final. Segundo o autor, diferentes condições se aplicam a diferentes situações fílmicas, e diferentes métodos podem ser aplicados para utilizar as possibilidades de som e de temporalidade. Basicamente, quando o som aparenta ser natural, quase como desnecessário, é porque está a funcionar: “O que ouvimos é o que não tivemos tempo de ver.” (Chion, 2011, p.61). Portanto, promover a mesma experiência para a pessoa visual e a invisual sobre uma obra cinematográfica não é possível:

- Se a imagem é descrita por palavras, perde-se o ‘valor acrescentado’, que emerge da relação audiovisual.
- Um indivíduo com visão tende a focar mais a atenção na imagem do que no som, que por sua vez, costuma servir como apenas um auxílio ao que está apresentado na imagem. O invisual possui apenas os ouvidos para perceber a história e os acontecimentos principais.
- A banda sonora e os efeitos sonoros podem perder funções dramatúrgicas quando se sobrepõe uma AD. Aquilo em que focamos a atenção corresponde à informação que está a ser processada na memória a curto prazo, e esta tem capacidade limitada em qualquer humano. (e.g. Sá, 2013) Logo, os estímulos recebidos passam por um processo de competição. Apenas algumas memórias passam a ser de longo prazo, ou seja, memoráveis. Tendo em conta que estímulos como efeitos sonoros ocorrem normalmente em um rápido intervalo de tempo, em termos narrativos o áudio passa a ser pouco lembrado pelo ouvinte.

De acordo com uma investigação⁷ que contou com uma amostra de 80 pessoas invisuais ou com baixa visão realizada em Portugal, constatou-se que 85% dos participantes do teste nunca assistiram séries ou filmes com audiodescrição na emissora RTP. Sabe-se que no Brasil há uma média de 6,5 milhões de deficientes visuais (IBGE), sendo 500 mil com perda total, e em Portugal há uma média de 900 mil (Censo 2011), sendo 28 mil com perda total.

Sendo a AD o único modo de acessibilidade reconhecido na atualidade, é fundamental perceber quais são os seus problemas. A experiência de uma pessoa invisual é diferente da vivência de uma pessoa visual, mas não tem que ser inferior.

⁷ Disponível em:

[<https://acessibilidade.gov.pt/wp-content/uploads/2020/07/2016_estudo_audiodescricao.docx>](https://acessibilidade.gov.pt/wp-content/uploads/2020/07/2016_estudo_audiodescricao.docx)

Um dos problemas apontados por Neves (2011) e Fryer (2016) sobre a AD é a dificuldade de adaptação ao tempo, pois são nas pausas entre os diálogos que a narração audiodescritiva precisa ser reproduzida. É preciso encurtar e resumir as descrições. Embora os guionistas de AD procurem informar sobre tudo o que for relevante para a compreensão da narrativa, como movimentos de personagens ou trocas de cenários, muitos podem acabar por focar-se apenas nos diálogos, e inserir as narrações descritivas durante alguma parte importante da banda sonora ou de um efeito sonoro, que muitas vezes são essenciais para a experiência e a emoção que a obra tenta transmitir.

Mariana Lopez, investigadora do projeto Enhacing Audiodescription⁸, afirma que a AD não consegue transmitir 100% dos acontecimentos da história de uma obra audiovisual à pessoa invisual, fazendo com que ela possa perder um tanto da compreensão da narrativa. Além disso, ela observa que a audiodescrição ainda não está apurada. De acordo com Lopez (2018), normalmente, os filmes são feitos por produtoras, mas a audiodescrição é feita em outro lugar, ou seja, o que temos é uma interpretação do filme feita pelos audiodescritores, o que pode não condizer com o objetivo principal da obra. De acordo com a investigadora: "Em um grupo focal, quando uma mulher descobre que a audiodescrição não é feita pelos criadores do filme, eu poderia simplesmente ter um amigo a me contar os acontecimentos ao meu ouvido. Afinal, por que os audiodescritores saberiam mais do que qualquer outra pessoa?"⁹

Ainda de acordo com Lopez (2018), outra desvantagem da AD é que ela é muito focada em transmitir informação, o que não significa entretenimento. A AD mascara muita informação sonora que está presente na música e nos demais efeitos sons. Afirma-se que não há diversidade no setor de acessibilidade. A audiodescrição é o único modo reconhecido de acessibilidade audiovisual, e é apenas isso que a pessoa invisual vai consumir das emissoras de televisão. A autora conclui: "Por que assumir que todo mundo com algum tipo de deficiência é igual e quer consumir acessibilidade da mesma maneira?"

⁸ Disponível em <<http://enhancingaudiodescription.com>>

⁹ Disponível em 12'03' <<https://youtu.be/13pablzz8zl>> Texto original: *Once in a focus group, one lady, when she found out that, well, I might just well have a friend telling me in my ear what's happening, cuz, why would the audio describers know more than anybody else?*

As sugestões da investigadora para aprimorar audiodescrição incluem o maior uso de efeitos sonoros, menos descrições verbais, mais uso de áudio espacial - binaural - e personalização dos acontecimentos. Ainda, pode-se fazer com que as vozes das personagens circulem no espectro estéreo, mas isso requer romper a mistura convencional que é realizada nos dias de hoje. Afinal, os efeitos que são mais espacializados são específicos, como normalmente para acentuar espaçonaves, carros e demais sons de vertente ficção científica.

3. Espacialização Sonora: Do Mono ao Binaural

Esta seção apresenta as características e aplicações dos diferentes sistemas de som, popularmente conhecidos como mono, estéreo e *surround*. Para esclarecer as particularidades de um sistema binaural, é preciso que sejam esclarecidas as diferenças entre estes principais sistemas de captação e reprodução na atualidade.

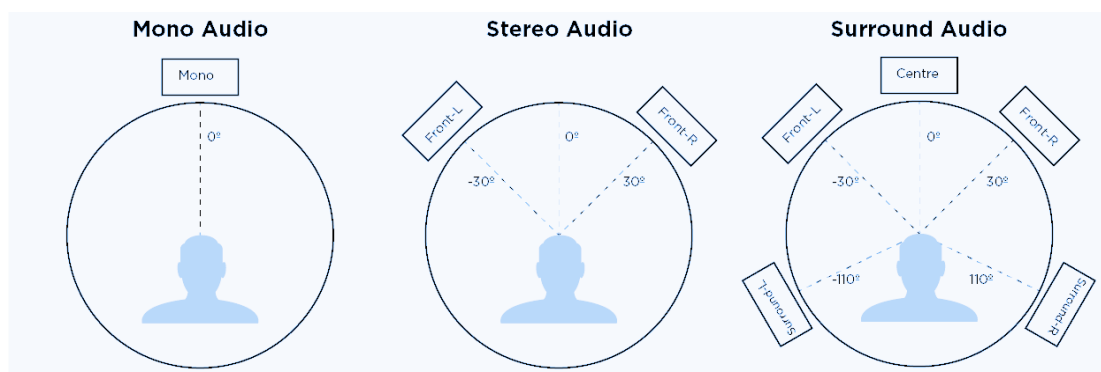


Figura 3 - Reprodução dos sistemas de áudio

Fonte: <<https://www.frozenmountain.com/developers/blog/archive/spatial-audio-in-vr-why-and-how>>

A espacialização sonora refere-se à produção de um efeito auditivo tridimensional perceptível. Nela, mecanismos psicoacústicos inerentes à percepção da localização, do envolvimento por um ambiente acústico. Várias técnicas de processamento de sinais podem ser empregadas para forjar efeitos desta natureza, sem necessariamente utilizar um modelo formal de propagação de frentes de onda (e.g. Yost, 2008). A Figura 3 apresenta a diferença que o acréscimo de fontes de energia sonora nos modelos padrão de reprodução sonora.

A invenção da gravação de áudio no século XIX era dada como espaço de dimensão zero, o que significa que era representada por um sinal monofônico. Com o passar dos anos, a possibilidade de captação de som se tornou estereofônica. Após isso, a Bell Labs¹⁰ registrou captações bidimensionais (*surround*), até a

¹⁰ Empresa de investigação e tecnologia situada em Nova Jersey, EUA.

tridimensionalidade dos dias de hoje, que é apresentada em vários formatos, como binaural e *ambisonics*.

3.1 Mono

O termo 'mono' refere-se a um único ponto de transmissão de som. Um sistema monofônico é constituído por um único sinal sonoro, e basta um altifalante para transmitir esse sinal na integralidade. Roginska (2018) explica que quando o ouvinte está de frente a um único altifalante, o sinal de propagação na direção do ouvido esquerdo e direito é idêntico. Com o uso de auscultador, esta reprodução causa a sensação de estar - no centro da cabeça, por ser igual nos lados esquerdo e direito. As reproduções monofônicas são utilizadas em redes telefônicas, aparelhos auditivos ou em sistemas de transmissão de audiodescrição no cinema, assim como na amplificação de instrumentos musicais e na locução audiovisual ou rádio.

O uso de áudio mono remonta aos primórdios da gravação e transmissão de áudio, onde recursos tecnológicos limitados dificultavam a produção e distribuição de áudio estéreo. Ainda hoje, o áudio mono continua a ser usado em determinados contextos, como em sistemas ou tecnologias mais antigas que não suportam a reprodução de áudio estéreo. Pode também ser usado por motivos de custo ou recursos, pois requer menos componentes e menos poder de processamento do que o áudio estéreo. Devido ao mono possuir apenas um canal para trabalhar, isso pode resultar em uma experiência de áudio mais clara e consistente, especialmente em ambientes ruidosos ou acusticamente desafiadores. Além disso, o áudio mono pode ser mais acessível para pessoas com deficiência auditiva, pois elimina a necessidade de distinguir entre os canais esquerdo e direito (e.g. Watkinson, 1998).

Chion (2011) complementa que o som mono cria um campo sonoro unificado e centralizado, que pode ser usado para direcionar a atenção do ouvinte para elementos específicos da mixagem de áudio. O autor enfatiza a simplicidade e a pureza do áudio mono, que pode transmitir uma sensação de clareza e objetividade em comparação com o som multicanal.

Sinais mono podem ser reproduzidos em sistemas estéreo devido a uma conversão que reproduz no lado esquerdo e direito os canais igualmente, resultando

em uma imagem sonora central que não possui informação espacial. (e.g. Watkinson, 1998)

3.2 Estéreo

Um sistema estereofônico consiste na combinação de dois sinais de áudio. Portanto, utilizam-se dois sinais microfonados para gravação, e dois altifalantes ou um auscultador com dois sinais de reprodução. Segundo Geluso (2018), sistemas estereofônicos são concebidos para criar a ilusão de uma cena auditiva com fontes sonoras direcionadas, que se localizam entre dois ou mais altifalantes posicionados à frente do ouvinte.” (e.g. Geluso, 2018, p.63).

Moylan (2007) aponta que a história do áudio estéreo remonta ao final do século XX , quando foram realizadas as primeiras experiências em gravação e reprodução de som. No entanto, foi somente nas décadas de 1950 e 1960 que o áudio estéreo começou a ser amplamente adotado, pois os avanços na tecnologia de gravação e reprodução tornaram possível produzir e distribuir gravações de áudio estéreo de alta qualidade. Hoje, o áudio estéreo é um padrão da indústria, com quase todos os produtos e sistemas de áudio comerciais sendo capazes de suportar a reprodução de áudio estéreo. Ainda de acordo com o autor, embora o áudio estéreo forneça uma experiência de som envolvente e direcional, limita-se por requerer um processamento mais complexo e um número maior de componentes do que o áudio mono, o que pode resultar em custos mais altos e maiores demandas técnicas.

A reprodução estéreo fornece uma perspectiva auditiva de 180° e promove mais espacialização sonora na reprodução de arquivos de áudio que foram gravados e misturados desta maneira. Um dos responsáveis pelo desenvolvimento do sistema de reprodução estéreo foi Alan Blumlein na década de 1930, mas foi somente na década de 1950 que o formato se popularizou para as salas de cinema. Geluso (2018) afirma que a esta altura, pensava-se no som já não apenas em termos de um sinal monofônico que acompanhava a imagem em movimento, mas antes em termos de um elemento que ajudava na condução da história para o espectador.

Uma possibilidade do estéreo no cinema é citada por Chion, quando define o supercampo (*superfield*) como um espaço que co-existe além do ecrã. Desde a criação do estéreo, os sons inseridos e produzidos nos altifalantes que, ficavam fora da área de visão, construíram narrativas independentes devido a explorarem sons

ambientes de natureza ou cidade, música. Ainda, o uso destas narrativas tornaram por auxiliar o espectador para quando fosse necessário introduzir um novo espaço na narrativa.

(...) na planificação, o plano que mostrava todo o cenário era um elemento estratégico com grande força dramática e visual, uma vez que, quer fosse colocado no início, no meio ou no fim de determinada cena, dava logo a sentir o espaço ambiente; lembrava-o e, ao mesmo tempo, situava as personagens no quadro e conferia uma ressonância particular ao momento em que intervinha. (Chion, 2011, p.119)

Ainda de acordo com Chion (2011), o som estéreo no cinema iniciou pela distribuição de elementos musicais, utilizando do supercampo para separar instrumentos no plano. Tanto estéreo como *surround* são sistemas que promovem espacialidade.

3.3 Surround

Rumsey (2018) define '*Surround*' como um termo que é frequentemente usado para descrever qualquer forma de reprodução de som de altifalante que envolve mais de dois altifalantes e tenta envolver o ouvinte com sons a serem reproduzidos por várias direções. (e.g. Roginska, 2018, p.180) Tal como o som estéreo, o *surround* foi concebido por Blumlein na década de 1930. O *Surround 5.1* ficou conhecido por causa do lançamento do DVD na década de 1990. O sistema resulta da configuração de seis sinais áudio, emitidos por cinco altifalantes: três à frente - esquerdo, centro e direito, mais um *subwoofer*¹¹, e dois altifalantes por trás do *sweet spot*¹².

Tanto '*surround*' como 'multicanal' são termos amplamente utilizados para descrever sistemas de áudio que usam mais de duas fontes de som, mas alguns autores fazem uma distinção entre eles, com 'multicanal' sendo usado como um termo genérico que inclui todas as formas de áudio que usam mais de duas fontes de som,

¹¹ É um altifalante particular para as frequências graves.

¹² *Sweet Spot* ou *Reference Listening Point* é um termo usado por engenheiros de gravação para descrever o ponto focal entre dois altifalantes, onde um indivíduo é totalmente capaz de ouvir a mistura de áudio estéreo da maneira que deveria ser ouvida pelo mixer.

enquanto 'surround' se refere a sistemas específicos de áudio que produzem uma sensação de som ambiente em todo o espaço. (e.g. Roginska, 2018)

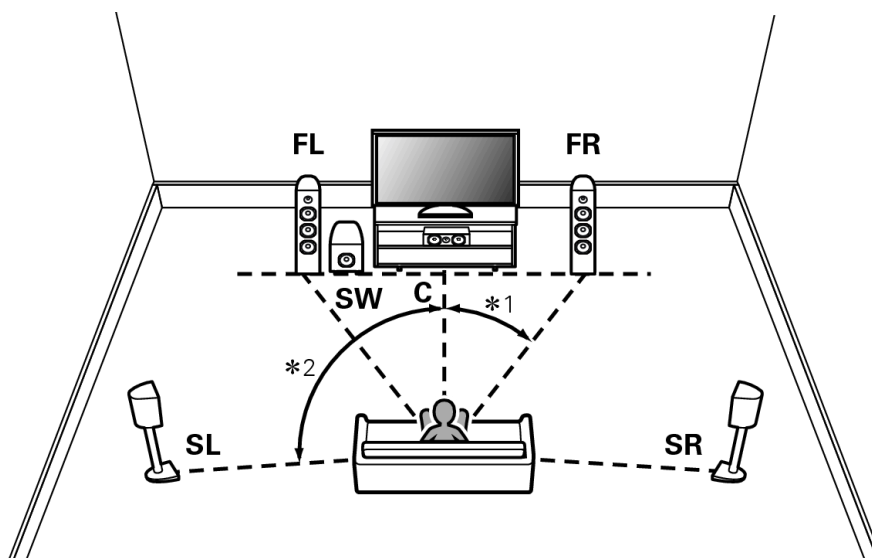


Figura 4 - Sistema *Surround* 5.1

Fonte: <<https://manuals.denon.com/AVRX6400H/NA/EN/DRDZSYpvilxnds.php>>

A Figura 4 apresenta um exemplo de filme com distribuição sonora *Surround* 5.1. Nela, podem-se distribuir os elementos que compõem o desenho de som, para que estes sejam claramente percebidos com mais espacialidade. Por exemplo, pode-se inserir o áudio da AD nas colunas SL e SR, e a banda sonora na FL E FR. Quanto mais fontes de som, mais imersividade é proposta, como o exemplo de 7.1, que é o padrão de espacialidade normalmente produzido em salas de cinema (e.g. Moylan, 2007).

Chion (2011) observa que o som *surround* pode ser usado para criar uma sensação de espaço e profundidade na mistura de áudio, de modo que o ouvinte experimente o som como se estivesse na mesma sala com a fonte do som. Isso, por sua vez, pode criar um impacto emocional mais poderoso ao ouvinte. Chion também argumenta que o som *surround* pode ser usado para aumentar o realismo da mixagem de áudio, permitindo que o designer de som coloque sons específicos em locais específicos dentro do campo de som *surround*, o que pode criar uma ilusão de realidade mais convincente e ajudar a experiência de imersão de som. No entanto, o autor ressalta que o som *surround* pode ser usado em excesso e que, às vezes, pode

prejudicar o conteúdo da obra. Caso seja colocada muita ênfase nos aspectos espaciais do som, o ouvinte pode se distrair com o campo de som *surround* e perder o foco no conteúdo do som. Portanto, Chion (2011) argumenta que é importante para o designer de som equilibrar o uso do som *surround* com outros elementos da mixagem de áudio, como o timbre e o tom do som.

A principal diferença entre '*surround*' e 'binaural' é que no primeiro as colunas são fisicamente presentes. A reprodução binaural mesmo que via altifalantes possuirá apenas duas colunas, contudo, popularmente, binaural é criado com foco numa escuta por auscultadores (e.g. Roginska, 2018).

3.4 Binaural

O termo 'Binaural' significa relação entre as duas orelhas. Segundo Roginska (2018), a audição binaural consiste na filtragem de frequências realizada pelo córtex auditivo, destacando que isso contribui para a capacidade humana de determinar a direcionalidade dos sons. Essas informações sobre a localização das fontes sonoras incluem ângulo de azimute, elevação e distância. Além disso, o córtex auditivo "desempenha um papel fundamental na nossa percepção consciente do som, incluindo a compreensão da fala, que é inquestionavelmente o estímulo social mais significativo para os seres humanos." (Roginska, 2018, p. 9)

Observa-se um exemplo da escuta de um som que está em (-45°) na Figura 5. Quando um sinal é executado no plano horizontal, o seu ângulo em relação à cabeça é referido como seu azimute, com 0° de azimute diretamente à frente do ouvinte, 90° para a direita e 180° sendo logo atrás. Segundo Roginska (2018), devido ao posicionamento dos ouvidos, identifica-se com maior precisão sons da esquerda (-90°) e da direita (+90°) do que de frente (0°) ou de trás (180°).

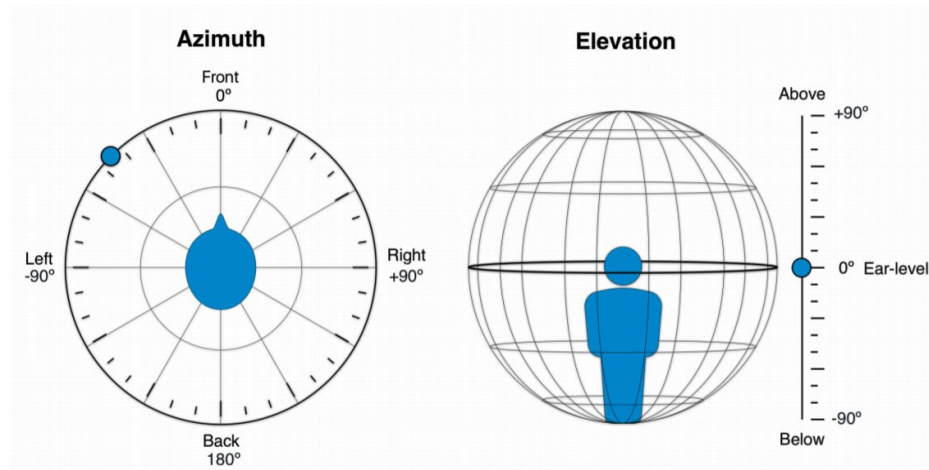


Figura 5 - Ângulo azimute e elevação no espectro de audição em 360°

Fonte: <<https://ageofaudio.com/pt/%C3%A1udio-binaural-3D-e-plug-in-ambeo-orbit-gratuito-de-Sennheiser/>>

A gravação binaural é possível graças a dois microfones omnidirecionais pequenos posicionados no lugar das orelhas em uma cabeça de manequim feita de material com boa absorção de som, como madeira ou fibra de vidro. As orelhas atuais dos modelos são feitas de silicone. A distância entre os ouvidos deve ser semelhante à da fisionomia humana padrão, assim como as outras partes do rosto, como nariz, boca, bochechas, testa, etc. A gravação binaural fornece informações sobre a localização espacial, permitindo reconhecer a direção e a distância dos sons reproduzidos. As reflexões do som nas paredes de um ambiente, seja interno ou externo, também contribuem para a captação binaural. (e.g. Roginska, 2018) Neste método, busca-se imitar as condições reais de audição com a ajuda de pistas binaurais, geralmente reproduzidas com auscultadores. Estas pistas, que incluem as reflexões acústicas do ambiente e do corpo, são conhecidas como HRTF - transferência de função relacionada à cabeça. (ver secção 4.3.1)

Embora a pesquisa sobre tecnologia binaural tenha ganhado destaque no século XXI, suas origens remontam ao século XIX. O primeiro dispositivo estéreo binaural foi criado em 1881 pelo engenheiro francês Clément Ader. Ele criou o Théâtrophone, um equipamento de entretenimento com dois reprodutores de som, ou seja, um par de telefones posicionados nos dois ouvidos do espectador, para dar movimento e espacialidade ao teatro sonoro. De acordo com Mattana (2017), Harvey Fletcher, pesquisador de acústica da Bell Labs, registrou o primeiro microfone de cabeça de manequim. A Figura 6 apresenta o modelo que possuía dois microfones

nas bochechas de um manequim de cera em 1933, dois anos depois da empresa patentear o som estéreo.

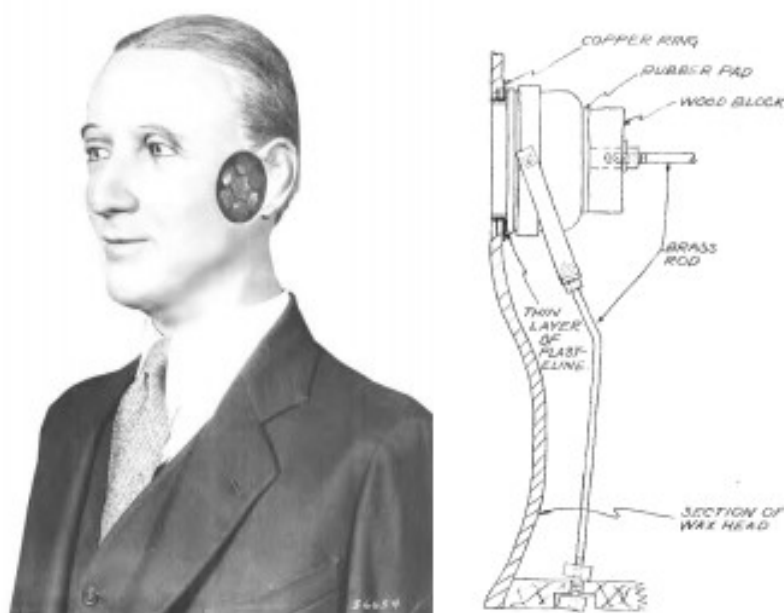


Figura 6 - Oscar, o primeiro *dummy head microphone*

Fonte: <<https://hookeaudio.com/what-is-binaural-audio/>>

A tecnologia de entretenimento imersivo tornou-se mais comum devido ao avanço tecnológico. De acordo com Boren (2018), o ser humano usa a percepção tridimensional do som para se orientar com a ajuda de eco, reverberação, ressonância e reflexões nas paredes. A história da acústica dos teatros remonta aos gregos, barrocos, renascentistas, passando por toda a história cultural, até se tornar moderna no século XIX com o avanço tecnológico que permitiu os estudos sonoros. (e.g. Boren, 2018)

O termo 'som imersivo' é diferente de '*surround*' e '*binaural*'. Para um áudio ser considerado imersivo ou tridimensional, é necessário que tenha canais elevados, e não apenas som vindo da frente e detrás de um altifalante num ambiente físico. Segundo Rumsey (2018), '*Surround*' é usado quando há mais de dois altifalantes que envolvem o ouvinte num plano horizontal. As primeiras técnicas de posicionamento de

mais de um altifalante para reprodução sonora surgiram na década de 1930, com base nas ideias da Bell Labs. A empresa buscava criar efeitos ilusórios ao replicar pares de altifalantes com atrasos e alterações que envolviam mudanças no tempo ou volume da faixa a ser reproduzida.

Rumsey (2001) define que no som estéreo, as saídas de energia no lado esquerdo e direito são ajustadas em volumes separados, às vezes com gravações diferentes, o que cria uma imagem fantasma que pode ser reproduzida por dois altifalantes ou pelos auscultadores. Costuma-se ter dois captadores que são decodificados na DAW para serem distribuídos no lado direito e esquerdo. Além disso, o autor afirma que a configuração de altifalantes mais aceita para reproduzir um som estéreo é chamada *sweet spot*. É baseada num triângulo equilátero, com o ouvinte localizado no meio entre os dois cantos, como mostra a Figura 7.

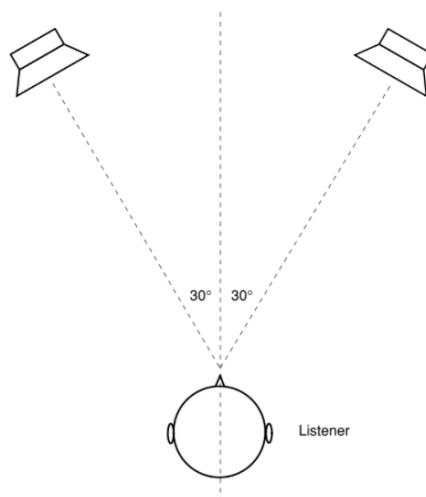


Figura 7 - Configuração do posicionamento estéreo via altifalantes

Fonte: Rumsey, 2001, p.53

Rumsey (2001) afirma que a diferença de tempo entre as fontes de som estéreo pode afetar a sensação de espacialidade. Se, por exemplo, um som estiver adiantado no lado esquerdo, comparado ao direito, a sensação é de que o lado esquerdo é mais importante. Embora isso possa ser considerado som estéreo, geralmente não é o que os engenheiros de som buscam. (e.g. Rumsey, 2001, p.58)

Existem diferenças significativas entre a reprodução de som estéreo através de altifalantes e auscultadores. Ao reproduzir o som através de altifalantes, o som é influenciado pela acústica do ambiente, enquanto com auscultadores o ouvinte ouve o som sem interferência, com as saídas de som do lado esquerdo e direito completamente isoladas. Como resultado, a experiência de um engenheiro de som ao misturar um projeto para altifalantes será diferente da reprodução através de auscultadores. As fontes de som são lateralizadas e há geralmente falta de energia no centro do campo sonoro quando comparado com a reprodução através de altifalantes. (e.g. Roginska, 2018, p.91)

Exemplos de aplicações de som com espacialização são citados por Rumsey (2001) e incluem música, cinema e jogos de computador. Sons são considerados espaciais quando simulam reflexões de uma esfera de 360°, com o objetivo de criar alguma ilusão ou sensação artística ou criativa.

O som binaural é capaz de reproduzir sinais multissensoriais, sendo um som estéreo com características espaciais. Sathian e Ramachandran (2020) explicam:

Nosso corpo externo e interno são especificados por fontes simultâneas de energia eletromagnética, mecânica e quimiossensorial. Quando transduzidas por nossos sistemas sensoriais, essas várias fontes de energia física resultam em sensações visuais, auditivas, somatossensoriais, vestibulares, olfativas e gustativas. Essa mistura de diferentes combinações de entradas multissensoriais nos fornece informações concorrentes e muitas vezes equivalentes que, quando integradas, nos permitem experimentar objetos, eventos e ações autogeradas unitários e psicologicamente significativos. (Sathian, Ramachandran, 2020, tradução nossa)

De acordo com Sathian (2020), é difícil compreender a maioria dos aspectos da funcionalidade humana, incluindo percepção, cognição, emoção, interação social e integração sensório-motora, pois eles dependem criticamente da nossa capacidade de perceber a unidade multissensorial. Na verdade, os indicadores multissensoriais estão relacionados com o corpo e a mente, e aplicados à questão binaural, isso tem relação com a maneira como o ser humano percebe o espaço presente, bem como como ouve sons e identifica sua posição baseada na reverberação.

3.4.1 Gravação e Reprodução Binaural

De acordo com Roginska (2018), os auscultadores são capazes de suprimir as interferências do ambiente físico, proporcionando destaque apenas à reprodução do que foi gravado previamente, sem distrações derivadas do ruído externo. É também possível alcançar a reprodução binaural com altifalantes, mas os dispositivos de reprodução devem ter um sistema de cancelamento de eco e é necessário prestar atenção à acústica da sala. No entanto, em uma condição de reprodução estéreo com auscultadores, torna-se mais fácil, pois os sinais binaurais contêm características espaciais semelhantes às de estarmos com um microfone em cada ouvido. Binaural refere-se ao som em que há dois canais de reprodução que foram filtrados para combinar tempo, intensidade e espacialização.

Pode-se dizer que o objetivo de um sinal binaural é recriar um sinal acústico para ser equivalente a uma audição que simula condições naturais. Um dos exemplos mais comuns atualmente de uma reprodução binaural bem-executada é um áudio disponível no YouTube de uma barbearia virtual¹³, onde um barbeiro corta o cabelo do espectador. A sensação de realidade, como se o profissional estivesse com tesouras ao lado do ouvinte, é reforçada pelo uso de auscultadores, que são a forma mais convincente de reprodução para alcançar o objetivo da reprodução binaural. No entanto, Roginska (2018) explica que um dos problemas na reprodução binaural com auscultadores é a falta de movimentação do ponto de referência. Os auscultadores são reprodutores centrais e os sinais não mudam se o ouvinte muda a posição da cabeça, o que resulta em menos interatividade com o ambiente. No entanto, Kendall (1995) explica que há uma forma de resolver o problema mencionado por Roginska (2018). Para o autor, o Head Tracking consiste na combinação de sensores que captam a direção e orientação da cabeça do ouvinte com a ajuda de uma câmera conectada ao computador. Se a proposta do som binaural é ficar atrás da cabeça e o ouvinte girar a cabeça para a esquerda, o som continuará a ser executado da mesma forma como foi proposto e permanecerá na mesma posição.

Para Grantham (1995), a localização binaural possui três dimensões e inclui a identificação da fala em ambientes com ruído e julgamentos de volume. Ele apresenta o conceito de Lateralização, explicando que uma gravação estéreo escutada com auscultadores pode soar como se viesse de dentro da cabeça, ao invés de altifalantes

¹³ Disponível em: <<https://youtube.com/watch?v=IUDTlvagjJA>>

que criam a sensação de soar externa à cabeça. A lateralização da função cerebral é o nome dado à teoria de que algumas funções neurais e cognitivas são processadas em certos lados do cérebro, seja o esquerdo ou o direito.

De acordo com Kendall (1995), quando um evento sonoro está equidistante dos dois ouvidos, o som pode chegar de maneira semelhante a ambos, mas nunca idêntico devido às assimetrias da cabeça. Observa-se na Figura 8, um exemplo simples sobre a área onde as fontes sonoras estão equidistantes dos dois ouvidos, que é conhecida como Plano Médio. Além disso, o autor menciona o Plano Horizontal, que é o volume captado pelo ouvido, e o Plano Frontal (ou Plano Lateral), que divide a cabeça verticalmente entre a frente e o trás.

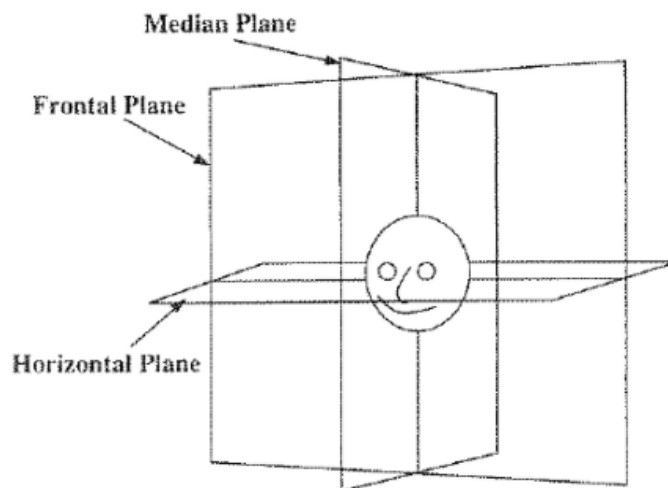


Figura 8 - Plano médio, plano frontal e plano horizontal

Fonte: Kendall (1995)

Segundo Kendall (1995), quando a fonte sonora não está equidistante aos ouvidos, o sinal chega a cada ouvido por uma direção diferente, resultando em HRTFs distintos. O ouvido mais próximo da fonte sonora é chamado de ouvido ipsilateral, enquanto o mais distante é conhecido como ouvido contralateral. A posição da fonte sonora em relação ao centro da cabeça do ouvinte pode ser capturada convenientemente como um vetor expressado em dois ângulos: azimuth e elevação. A

Figura 9 ilustra um ponto de escuta que inclui elevação na audição, enquanto o azimute mede o ângulo entre o vetor no plano horizontal e um vetor que se estende diretamente na frente da cabeça do ouvinte. A elevação é medida pelo ângulo formado entre o vetor e o Plano Horizontal, alinhado a 90 graus acima ou abaixo da mesma angulação.

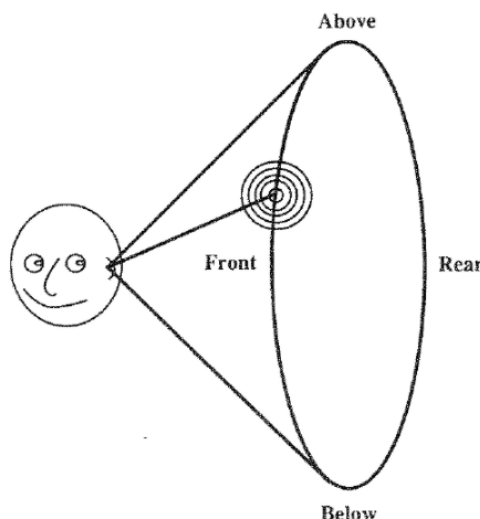


Figura 9 - Exemplo de localização sonora desempenhada pelo ouvido

Fonte: A 3-D Sound Primer: Directional Hearing and Stereo Reproduction (1995)

De acordo com Rayleigh (1907), ele se concentrou na separação dos dois ouvidos propondo experimentos que construíram sua teoria sobre a localização por meio da composição, resultado de muitas experiências realizadas com ondas senoidais. As investigações mostraram que há diferenças entre os sinais acústicos apresentados em cada ouvido, o que afeta diretamente a percepção espacial. Foi constatado que acima de 1500 Hz, há uma obstrução acústica na cabeça, e os julgamentos de localização são determinados pela diferença de intensidade entre os ouvidos. Por outro lado, abaixo de 1500 Hz, a cabeça não é um obstáculo significativo e os julgamentos de localização são determinados pela diferença entre os dois ouvidos.

3.4.2 Técnicas de Gravação Binaural

Desde 1933, a forma mais convincente de registrar um arquivo de áudio binaural é usando uma cabeça de manequim com dois microfones acoplados nos ouvidos. Essa cabeça deve ser uma representação anatômica precisa de um humano adulto, capturando, assim, as interferências relevantes que um ouvido humano sofre no dia a dia, como o tamanho da cabeça, a distância entre os ouvidos e a presença de ombros e tronco, representado pela HRTF (e.g. Roginska, 2018, p. 96). A primeira demonstração de um dispositivo captador binaural foi em 1933 na Exposição Universal Century of Progress¹⁴, e consistia numa cabeça com dois microfones em cada orelha, captando algo essencialmente semelhante à cabeça humana. O registro binaural não será eficaz se o método de reprodução não for correto, sendo necessário que os dois sinais sonoros sejam apresentados independentemente a cada ouvido, de forma equivalente ao registro.



Figura 10 - 3DIO Omni Binaural Microphone / Neumann KU100

Fonte: <<https://3diosound.com> e <https://en-de.neumann.com/ku-100>>

Atualmente, cabeças de manequim com microfones incorporados já estão disponíveis para compra. A Figura 10 e a Figura 11 apresentam alguns produtos que

¹⁴ Foi uma feira mundial que aconteceu em Chicago para celebrar o centenário da cidade. O tema da feira foi de inovações tecnológicas.

captam sinais de maneira binaural, como Neumann, 3DIO, Sennheiser e Roland. As empresas vendem esses modelos para os usuários que desejam usá-los. Nestes modelos, a captação é em estéreo, sendo necessários dois canais amplificados para cada ouvido, que devem ser decodificados na DAW para que o *pan* seja direcionado 100% para cada lado.



Figura 11 - Microfone binaural Roland CS10-EM

Fonte: <<https://roland.com/global/products/cs-10em/>>

Outra forma de capturar um sinal estéreo binaural é usando microfones binaurais que podem ser inseridos nos ouvidos. Embora pareçam com auscultadores, eles contêm microfones e utilizam as características do próprio corpo humano para gravar as pistas binaurais. Embora estas opções sejam mais acessíveis, dependendo da aplicação desejada, a reprodução binaural pode ser diferente em relação a outros equipamentos. Ao compararmos o microfone CS10-EM da Roland (€99) com o dummy head da Neumann KU100 (€7899), é evidente a diferença sonora. Devido ao fato de o CS10-EM estar acoplado ao próprio ouvido do usuário, os sons de movimentos, roupas e respiração serão captados pelos microfones, o que não acontece na cabeça de manequim da Neumann. De acordo com o artigo¹⁵ da própria empresa do microfone da Neumann, a cabeça de quase €8000 observada na Figura 12 possui seu próprio *phantom power* e um padrão polar muito específico, projetado para a tecnologia binaural. A cabeça de manequim obtém a informação espacial devido à

¹⁵ NEUMANN, The Dummy Head - Theory and Practice (1988)

complexa estrutura do ouvido externo (direção dependente da frequência) e ao uso de delay e difração de som ao redor dela.

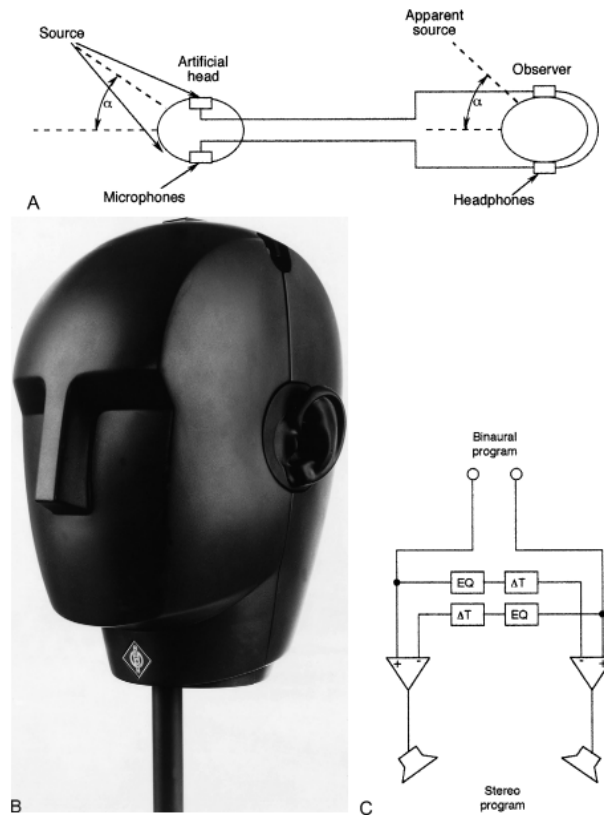


Figura 12 - Fluxo de funcionamento do microfone binaural Neumann KU100

Fonte: Rayburn, 2011 p.242

Rayburn (2011) complementa o processo da gravação binaural. Ele menciona que o campo sonoro envolve a resposta de frequência e ajuste de equalização:

O campo sonoro tridimensional no espaço de gravação original é mapeado com precisão no domínio auditivo do ouvinte por meio de dicas de tempo corretas e modelagem de resposta de frequência ao redor da cabeça, juntamente com equalização para compensar o caminho do sinal secundário entre os auscultadores e os tímpanos. O ouvido externo ou pavilhão da cabeça codifica pistas direcionais no som, incluindo informações verticais. Se o pavilhão da cabeça for semelhante ao do ouvinte, o efeito pode ser muito realista. Por esta razão, as cabeças são muitas vezes fabricadas com pina intercambiável. (Rayburn, 2011, p.241, tradução nossa)¹⁶

¹⁶ Texto original: *the three-dimensional sound field in the original recording space is accurately mapped into the listener's auditory domain via correct time cues and frequency response shaping around the head, along with equalization to compensate for the secondary signal path between headphones and the eardrums. The outer ear or pinna of the head encodes directional*

Após a decodificação de gravações estéreo binaurais na DAW, alguns *plugins* de manipulação de características binaurais podem ser aplicados sobre a gravação original, como é o caso do Neumann KU100, que possui um modelo computacional sugerido pela própria Neumann. Estas ferramentas de pós-produção de som incluem simulações baseadas em resposta impulsiva de sala, como reverb, reguladores de acústica e simuladores de som binaural.

3.4.3 Protótipo de Microfone Binaural

Foi criado um microfone binaural para fins de experimentação. Foi utilizada uma cabeça de manequim de PVC e testadas orelhas de policloreto de vinila (PVC) e de silicone. Os resultados são observados na Figura 13 e na Figura 14, em que há um par de microfones lapela omnidirecionais (Movo LV4-O2) e um par de microfones condensadores (Behringer C2) que foram testados como substitutos dos ouvidos. Dentro da cabeça foram colocadas espumas acústicas de poliuretano para isolar o som e evitar distorções.



Figura 13 - Construção do protótipo do microfone binaural

Fonte: Acervo do autor

clues into the sound including vertical information. If the pinna of the head is similar to that of the listener, the effect can be very realistic. For this reason heads are often manufactured with interchangeable pinna.

Ao comparar com outros microfones já existentes no mercado, como 3DIO, Binaural Enthusiast e Neumann KU100, verificou-se que a combinação de orelhas de silicone com os microfones lapela omnidirecionais oferecem resultados satisfatórios para simulação de escuta binaural. No entanto, para uma captação ainda mais fiel à escuta humana, recomenda-se o uso de uma cabeça mais firme, feita de material com menor vibração, como madeira.



Figura 14 - Protótipo do microfone binaural (Orelhas silicone + Microfones lapelas)

Fonte: Acervo do autor

Consideramos que a captação binaural é afetada por diversos fatores, sendo importante também controlar o ambiente de gravação. Embora construir uma cabeça com microfones lapela possa ser mais acessível do que adquirir um *plugin* de simulação binaural profissional, o ambiente deve ser acusticamente isolado para evitar reverberações e sons externos. Como tal, o tratamento acústico pode ser dispendioso. Portanto, concluímos que a maneira mais acessível de controlar o ambiente virtualmente é através de *plugins* de simulação binaural, que permitem ajustar parâmetros como reflexões, reverberação, equalização e posicionamento da fonte sonora.

3.4.4 Simulação Binaural via *Plugins*

De acordo com Blauert (2013), a simulação de sala binaural é uma técnica que permite escutar em ambientes inexistentes através de um modelo computacional que considera todos os parâmetros acústicos relevantes do ambiente. As percepções precisam ser artificialmente criadas para configurar o ambiente virtual desejado. Blauert destaca que Jeffress foi o primeiro a calcular as pistas ITD (diferença de tempo de chegada) e ILD (diferença de nível) em 1948. O modelo neurocomputacional de Jeffress explica como os sistemas auditivos registram e analisam pequenas diferenças no tempo de chegada dos sons nas duas orelhas para estimar a direção das fontes sonoras no plano azimutal. Esse avanço permitiu o melhor controle dessas pistas através de plugins. O autor conclui:

Desde que Jeffress propôs seu modelo de linha de atraso em 1948, os modelos binaurais cresceram em complexidade e agora são capazes de abordar um grande conjunto de fenômenos perceptivos para realizar tarefas de localização, detecção e reconhecimento de recursos. Os modelos atuais são capazes de simular a periferia auditiva em grande detalhe e dominar a extração de ITD, ILD, bem como pistas monoaurais. (Blauert, 2013, p.27)

No caso de um ouvinte humano, o som que é recebido pelo ouvido externo, cóclea, cabeça e tronco é transformado em sinais de pressão sonora nos dois tímpanos. Durante esse processo de transformação, as pistas espaciais do som são convertidas em pistas temporais e espectrais dos sinais do tímpano. O sistema auditivo pode decodificar essas pistas e utilizar as informações para formar percepções auditivas de acordo com a configuração escolhida. A influência do ambiente acústico é refletida principalmente nos atributos espaciais das percepções auditivas, como direção, distância, impressão espacial e extensão. (e.g. Blauert e Lehner, 1992)

De acordo com Quinot e Katz (2018), os simuladores binaurais funcionam baseados no conhecimento adquirido na última década sobre espacialização sonora. Um exemplo é o plugin binaural Anaglyph, que é composto por um modelo baseado na diferença de tempo interaural e na diferença de nível interaural, além de levar em consideração as características HRTF. A Figura 15 representa a arquitetura do

programa. Os autores explicam que o módulo binaural é composto por um mecanismo de convolução HRIR e um modificador de ILD e ITD. (e.g. Quinot e Katz, 2018, p.2)

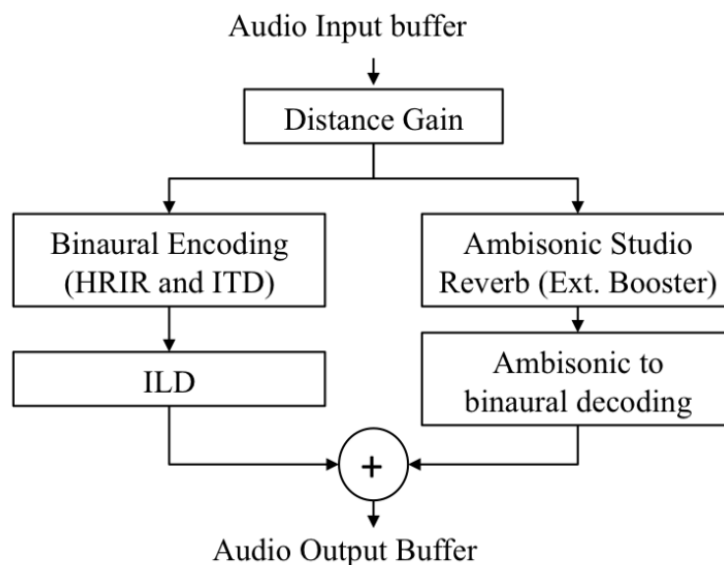


Figura 15 - The Anaglyph binaural audio engine

Fonte: (Quinot, Katz, 2018. p.2)

O processo de codificação dos parâmetros do ambiente acústico em características do campo sonoro pode ser modelado no computador, utilizando uma metodologia de simulação de sala. Da mesma forma, a transformação do som de sinais de campo em sinais de entrada para os tímpanos pode ser modelada por uma metodologia conhecida como tecnologia binaural.

Os ambientes acústicos são especificados por propriedades físicas que, por sua vez, influenciam o campo sonoro nos ambientes. Tais propriedades são os dados geométricos e acústicos das superfícies refletoras de som e as posições, orientações e características direcionais tanto da fonte sonora quanto do receptor. (Blauert, Lehner, 1992)¹⁷

¹⁷ Texto original: *Acoustic environments are specified by physical properties which, in turn, have an influence on the sound field in the environments. Such properties are the geometric and acoustic data of sound reflecting surfaces, and the positions, orientations and directional characteristics of both the sound source and the receiver.*

A simulação de sala binaural pode começar com a modelagem do campo sonoro, na qual se cria um modelo computacional do campo sonoro baseado em um conjunto de dados físicos representativos do ambiente. Nesta etapa, a distribuição espacial das fontes refletidas é calculada a partir dos parâmetros físicos do ambiente. Em seguida, vem a auralização, que é quando o som se torna audível. Nesta etapa, os impulsos de resposta binaural da sala são computados a partir de sinais anecoicos pré-gravados e posteriormente reproduzidos em auscultadores. Em um ambiente composto por superfícies refletoras, o ouvinte recebe sinais sonoros diretamente da fonte sonora. No entanto, o modelo computacional permite a criação de superfícies refletoras e uma variedade de sinais sonoros que podem incidir por diferentes direções, com diferentes tempos de chegada e formas espectrais.

Atualmente, existem alguns plugins para as DAWs (Digital Audio Workstations) que simulam sinais binaurais no mercado de áudio digital. Esses são conhecidos como ‘Binaural Panning’. Alguns exemplos incluem o Binauralizer da NoiseMakers, o Ambeo da Senheiser e o DearVR PRO da Dear Reality que é observado na Figura 16.



Figura 16 - DearVR Pro Binaural Plugin

Fonte: <<https://dear-reality.com>>

Estes simuladores convertem qualquer sinal sonoro em estéreo binaural, com a opção de adicionar reverb. Para criar a sensação de espaço, eles buscam simular os

indicadores binaurais ITD e ILD. Se um sinal for reproduzido na esquerda do azimuth, o mesmo som monofônico é replicado e reproduzido no outro ouvido alguns milissegundos depois, resultando em reflexos sonoros em toda a estrutura do azimuth, que corresponde à posição dos reflexos do som captados pelo ouvido na frente, atrás, acima, abaixo e nas laterais da cabeça. Assim, é possível escolher a direção horizontal e vertical da imagem sonora.

3.4.5 Reprodução Binaural via auscultadores

Segundo Roginska (2018), a melhor maneira de reproduzir um áudio binaural é através de auscultadores. Ao contrário de altifalantes, os auscultadores não sofrem distorção devido a espaço físico como paredes, pois o som do lado esquerdo é direcionado 100% ao ouvido esquerdo e o som do lado direito é direcionado 100% ao ouvido direito. Binaural se refere aos dois canais de som que são reproduzidos entre o ouvido direito e o esquerdo do ouvinte, o que pode levar à confusão sobre a diferença entre os termos 'binaural' e 'estéreo'. A autora explica:

Embora muitos possam argumentar que todo o som estéreo é binaural, o termo binaural é reservado para o som em que o som de dois canais que entram nos ouvidos de um ouvinte foram filtrados por uma combinação de tempo, intensidade e pistas espectrais destinadas a imitar pistas de localização humana. Essas pistas podem ser sobrepostas naturalmente (por exemplo, gravações binaurais) ou por meio de processamento de sinal." (Roginska, 2018, p.88, tradução nossa)

A compreensão da estereofonia vai além de um simples sistema de som. Ela é considerada como um conjunto de relações entre tecnologias de áudio, espaços acústicos (tanto físicos quanto virtuais), técnicas de escuta, discursos científicos e comerciais, condições econômicas e contextos de recepção. Não se pode afirmar que a estereofonia tenha uma história de desenvolvimento linear ou contínuo. A autora também destaca as vantagens da reprodução binaural através de auscultadores.

Existem muitas vantagens na reprodução de sinais binaurais pelos auscultadores. A vantagem mais significativa é que os auscultadores oferecem um ambiente de audição controlado. A reprodução do fone de ouvido não é afetada por fatores como localização ou orientação do ouvinte. A menos que a cabeça seja rastreada e o processamento adicional seja usado para compensar o posicionamento do ouvinte, o sinal apresentado será independente da localização ou orientação do ouvinte e resultará no ouvinte sempre estando no ponto ideal. Outro benefício que vem com a entrega direta dos sinais aos ouvidos é a prevenção de quaisquer sinais de diafonia alcançarem o ouvido indesejado. A diafonia é um fenômeno comum na reprodução de altifalantes e pode levar ao colapso completo de uma imagem auditiva espacial. (Roginska, 2018, p.90, tradução nossa)

A história da estereofonia não é linear, mas é composta por interrupções, avanços e retrocessos, movimentos, confusão e competição entre sistemas binaurais e estéreo. É uma história multicanais que envolve múltiplas interseções entre som, música, pesquisa científica, investimento financeiro, inovação e contextos e práticas de escuta que raramente resultam em coesão ou sucesso unificados.

Alguns estudos sobre tecnologia binaural se baseiam nos princípios ASA propostos por Bregman (1994). Nix e Hohmann (2007) realizaram uma pesquisa que aplicou métodos recentes de filtragem estatística multidimensional, integrando o rastreamento da direção da fonte sonora e a dinâmica espectro-temporal de duas vozes misturadas pelos princípios ASA. Os resultados mostraram que o algoritmo rastreou com precisão a direção da fonte sonora, separou envelopes de voz com tempo de convergência algorítmica de até 50 ms e melhorou a relação sinal-ruído em condições adversas, mas exige alta demanda computacional.

De acordo com Wendt, Van de Par e Ewert (2014), a simulação binaural é inflexível, pois a posição e orientação do ouvinte devem ser fixas. Eles mencionam a simulação de ambientes acústicos por meio de programas de computador que usam variáveis de realismo, como reverb simples ou complexo, e estudaram a aplicação de características HRTF em auscultadores.

4. Percepção Auditiva

O som é um fenômeno complexo que pode ser compreendido tanto a partir de sua natureza física, relacionada às ondas longitudinais que propagam pressão nas moléculas de ar, quanto de sua influência fisiológica, que estimula o sentido da audição e permite a percepção de estímulos sonoros. De acordo com estudos realizados por Yost (2007), a evolução dos animais os conduziu a viverem com o propósito de se alimentar, reproduzir e explorar o ambiente, o que resultou em uma percepção do som mais apurada do que a dos seres humanos. Esses animais são capazes de identificar a fonte sonora a distâncias consideravelmente maiores do que nós. Essa capacidade aprimorada está intrinsecamente ligada à sua adaptação para sobreviver e se comunicar eficientemente dentro de suas espécies.

Ainda de acordo com Yost (2007), os receptores auditivos periféricos, no entanto, não evoluíram para processar fontes sonoras, mas para fornecer o código neural relevante para os atributos físicos do som. A falta de receptores periféricos dificulta a leitura das fontes sonoras, prejudicando o sistema auditivo. A percepção auditiva é complexa e é determinada tanto pela fisiologia do sistema auditivo quanto por processos cognitivos que integram informações dos diversos sentidos para produzir uma percepção unificada do mundo. O autor observa:

A sensação auditiva pode ser vista como o processamento dos atributos físicos do som: frequência, nível e tempo. A percepção auditiva pode ser o processamento adicional daqueles atributos que permitem a um organismo lidar com as fontes que produziram o som. Assim, a determinação das fontes sonoras pode ser vista como um estudo da percepção auditiva. A percepção de “fontes especiais”, como fala e música, podem ser subtópicos do tópico geral da percepção de fontes sonoras. (Yost, 2007, p.1/2)

Observa-se na Figura 17 que o aparelho auditivo é composto pelo ouvido externo, ouvido médio e ouvido interno. De acordo com Ballou (2005), o ouvido externo tem a função de capturar e conduzir o som do meato acústico externo até a membrana timpânica, para que seja transmitido ao tímpano. Já o ouvido médio amplifica e ajusta a amplitude das vibrações do ar, usando músculos adjacentes, antes de conduzir as vibrações para o ouvido interno. No ouvido interno ou cóclea, essas

vibrações são convertidas em sinais eletroquímicos, que são transmitidos pelo nervo auditivo para o cérebro, permitindo a audição e o reconhecimento do som.

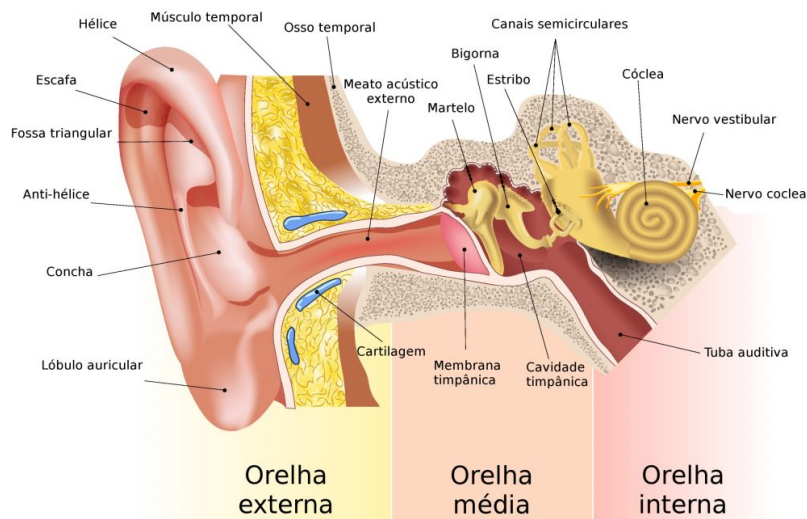


Figura 17 - Anatomia do canal auditivo

Fonte: <<https://infoescola.com/audicao/ouvido/>>

Kendall (1995) estudou como o sistema auditivo integra as informações de ambos os ouvidos em uma percepção única do evento acústico no espaço, permitindo a extração de informações direcionais e a reconstrução da fonte original. Bregman (1994) estudou princípios psicológicos que governam a agregação e segregação dos estímulos sonoros, enquanto Begault e Cooper (2018) analisaram como o sistema auditivo transforma aspectos independentes dos estímulos sonoros em padrões de atividade neural. Esses padrões fornecem informações sobre o conteúdo espectral, amplitude sonora, propriedades temporais e localização no espaço, e são integrados com informações de outros sistemas sensoriais para formar uma percepção unificada e fornecer orientações comportamentais, incluindo a orientação para estímulos acústicos.

Em seguida, vamos explorar a percepção auditiva de pessoas com deficiência visual. Devemos destacar que a percepção auditiva é complementada com outros sentidos, como o tato.

4.1 A Escuta em Pessoas com Deficiência Visual

Esta seção apresenta investigações acadêmicas e testes práticos relacionados com a audição de pessoas invisuais. Segundo Mota (2013), o sistema háptico permite a uma pessoa identificar objetos e temperaturas pelo toque. No campo da fisiologia, a cinestesia é a capacidade de sentir os movimentos musculares e das articulações, o que permite à pessoa se localizar em relação à orientação do corpo. A memória muscular é uma das funções da cinestesia e se desenvolve através da repetição de movimentos. Por exemplo, um caminho que uma pessoa com deficiência visual faz ao sair de casa e ir ao mercado. O sentido vestibular informa sobre a posição vertical do corpo, enquanto o olfato, assim como a ecolocalização, fornece informações sobre lugares, como a diferença entre um açougue e um jardim ou uma livraria.

De acordo com o livro de orientação e mobilidade para pessoas com deficiência visual do Governo Federal Brasileiro (2003), o ouvido é o meio pelo qual as pessoas com deficiência visual percebem a distância e a profundidade. Isso acontece porque as características do som refletido podem fazer com que a pessoa com baixa ou nenhuma visão perceba qual é a verdadeira distância de um objeto. Ao longo dos anos, a pessoa com deficiência visual desenvolve a capacidade de distinguir os diferentes lugares de um espaço através das características acústicas de cada ambiente. (Mota, 2013) A compensação da falta de visão não é automática, é desenvolvida ao longo do tempo e envolve vários recursos neurais.

Além de permitir identificar fontes sonoras, a percepção auditiva também ajuda a identificar obstáculos como árvores e prédios. Para Mota (2013), o termo 'sombra sonora' é usado para descrever como uma pessoa sem visão percebe uma interferência à sua frente, como uma mudança no campo auditivo. Segundo um artigo publicado na PLOS Biology por Gougoux et al. (2005), indivíduos sem visão tendem a ter um desempenho superior em tarefas auditivas em comparação com pessoas com visão. Este estudo comportamental realizado por neurocientistas canadenses mostrou que pessoas que têm deficiência visual de nascença são melhores na distinção de sons. Estudos posteriores usando técnicas de neuroimagem descobriram que áreas do cérebro responsáveis pelo processamento visual podem ser ativadas durante tarefas não visuais, como a leitura em Braille e a recuperação de memória auditiva. A investigação foi feita com pessoas que nasceram invisuais e aquelas que perderam a

visão ao longo da vida, permitindo uma comparação entre audição binaural e não-binaural. Os resultados mostraram que há diferenças perceptíveis entre os dois grupos: quando testados com som em apenas um ouvido, os que nasceram invisuais conseguiram localizar com maior precisão a origem do som. De acordo com Gougoux et al. (2005), tanto pessoas quanto animais sem visão compensam a falta de visão ao desenvolver sua audição, olfato, tato e paladar. O córtex de gatos e ratos sem visão apresenta um grande número de respostas aos estímulos sensoriais, indicando um aumento da sensibilidade das células cerebrais receptoras.

O Jornal oficial da Sociedade de Neurociência Norte-Americana, o JNeurosci, apresentou dois estudos sobre a maior capacidade auditiva de pessoas com deficiência visual em 2019. O primeiro estudo examinou a sensibilidade do cérebro às diferenças sutis entre frequências auditivas e procurou determinar a precisão da informação sobre audição que é transmitida pelos neurônios. Foi constatado que há um pequeno ajuste neuronal no córtex auditivo de pessoas invisuais, que parece ser mais estreito do que em pessoas com visão. O segundo estudo mostrou que a área cerebral hMT+ responsável por rastrear objetos visuais em movimento também produz respostas neurais que refletem outros tipos de movimento, desempenhando assim uma função de suporte para invisuais ao analisarem sons em movimento.

A Universidade Federal de São Paulo realizou um teste de localização sonora e outro de lateralização sonora para compreender se o cérebro dá prioridade à informação vindo de um dos ouvidos (e.g. Dias, Pereira, 2008). O teste, que contou com a participação de 24 pessoas com deficiência visual e com visão, foi realizado com auscultadores estéreo binaural e os participantes foram solicitados a apontar na direção do som que escutaram. Os resultados mostraram diferenças significativas entre pessoas com e sem visão. Foi verificado que a percepção auditiva de pessoas invisuais é mais apurada do lado direito. Embora o estudo tenha incluído um número relativamente pequeno de participantes, outros estudos similares obtiveram resultados semelhantes. Outro estudo recente sobre surdez unilateral e inteligibilidade da fala mostrou que os ouvintes do ouvido direito percebem diálogos melhor do que os ouvintes do ouvido esquerdo (e.g. Wettstein, 2018). Isto poderá ser particularmente interessante quando se escolhe uma posição lateral (portanto não-tradicional) para a narrativa audiodescritiva, e se pretende maximizar a sua inteligibilidade sem cancelar o som original do filme.

4.2 Análise de Cena Auditiva

Análise de Cena Auditiva (*Auditory Scene Analysis*) é um termo proposto pelo psicólogo Albert Bregman (1994). Refere-se ao estudo de como o cérebro é capaz de distinguir os vários componentes de uma fonte sonora. Trata do processo pelo qual todas as evidências auditivas que chegam ao receptor a partir de uma única fonte ambiental, são colocadas juntas em uma única unidade perceptiva. É um modelo que aplica a psicologia Gestalt na análise da percepção auditiva em humanos. A Gestalt estuda os princípios que governam a agregação e segregação dos estímulos, tanto no domínio visual como no auditivo. Esses mecanismos permitem a simplificação do panorama sensorial, reduzindo a necessidade de processamento mental - o que por sua vez facilita a reação aos objectos e eventos do mundo (e.g. Sá, 2013).

A análise de cena auditiva (ASA) tem sido usada há muito tempo no campo da engenharia de áudio para analisar a paisagem sonora e as características psicoacústicas das gravações. Tais recursos são, por exemplo, sonoridade, panorâmica, equalização, direcionalidade ou presença de reverberação. No entanto, o ASA nem sempre é usado em todo o seu potencial: localização de fontes virtuais, espaço auditivo virtual e técnicas binaurais são alguns exemplos em que os resultados dos algoritmos ASA podem fornecer mais informações sobre um determinado conjunto de dados e, portanto, levar a definição de ferramentas de análise mais sofisticadas.

Complementando, Yost (2018) diz que o cérebro usa uma combinação de informações para realizar a análise de cena auditiva. A diferença de tempo de chegada dos sons em cada ouvido é usada para determinar a direção de uma fonte sonora. A diferença de intensidade sonora e o eco também são usados para identificar a localização espacial e a distância da fonte sonora. A análise de cena auditiva está envolvida na identificação de padrões sonoros complexos, como a fala humana.

Os fenômenos relacionados com a agregação de sons e a formação de fluxos auditivos incluem a percepção de fala, de ordem, e de outras propriedades temporais das sequências sonoras. Incluem também a relação entre a informação transmitida pelas duas orelhas, a percepção de numerosidade, a detecção de padrões embutidos em outros sons, e a percepção de camadas sonoras (e.g. Bregman, 1994). Bregman (1994) explica:

(...) que existe um processo responsável pela formação de fluxos auditivos, e pela sensação de continuidade. Este processo é fundamental na resolução de problemas como a necessidade quotidiana de agrupar componentes espectrais corretamente para ouvir que um carro se aproxima quando atravessamos uma rua. Do mesmo modo, é fundamental para distinguir uma voz ou um instrumento de uma performance musical. (Bregman, 1994, p.44)

A investigação de Bregman é extremamente útil para entender como a percepção lida com a complexidade do panorama auditivo. Os estudos de cena auditiva são realizados desde os anos 1960 e podem ajudar o reconhecimento de padrões para que não sejam misturadas informações de diferentes fontes. O autor demonstrou como tendemos a agregar sons de conteúdo espectral semelhante, assim como sons simultâneos ou temporalmente próximos. Inversamente, frequências sonoras distantes favorecem a diferenciação auditiva. O autor distingue dois tipos de segregação auditiva: a segregação primitiva e a segregação orientada por esquemas conceptuais. A segregação primitiva é impulsionada pelos dados auditivos recebidos, e condicionada somente por conhecimentos inatos, inerentes a todos os seres humanos independentemente da sua cultura. Diversamente, a segregação orientada por esquemas conceptuais baseia-se em padrões e conceitos adquiridos através da experiência. O autor define-a como “uma representação mental de alguma regularidade em nossa experiência” (Bregman, 1994, p.43). Essa regularidade torna-se uma referência por meio de repetidas exposições ao ambiente.

Na segregação primitiva, Bregman inclui a integração simultânea e a integração sequencial. Na integração simultânea, o cérebro agrupa estímulos sonoros simultâneos de acordo com o respectivo conteúdo espectral. Além disso, os sons que se sucedem para formar padrões, como as palavras, precisam ser identificados - isto é chamado de integração sequencial. Nesse caso, o cérebro agrupa sons com algumas características semelhantes que ocorrem ao longo de um período de tempo e separa os padrões sequenciais, para que a informação seja transmitida por canais distintos .

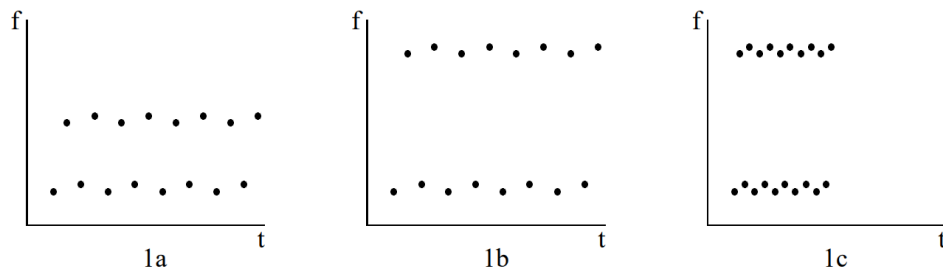


Figura 18 - Comparativo de segregação primitiva

Fonte: Arons, B. A Review of The Cocktail Party Effect. MIT Media Lab. 1992

Na Figura 18, considera-se f a frequência 20Hz a 20Khz e t como tempo. Na figura 1b, a segregação é maior do que na figura 1a, devido a separação das frequências baixas e altas. Na figura 1c há um aumento na velocidade de projeção de som. Normalmente, a análise de cena primitiva agrupa sons que tiveram a mesma origem, e segrega os sons de origens diferentes.

Também é importante considerar a análise da fala numa cena auditiva. De acordo com Arons (1992):

(...) existem extensões e ideias que são específicas para a análise de sinais de fala. Por vezes, é difícil separar os processos primitivos do esquema, e os esquemas de fala tendem a obscurecer as contribuições dos processos primitivos. Considerando as regras primitivas de segregação, é um tanto surpreendente que as vozes se mantenham juntas. A fala consiste em sequências de tons complexos de baixa frequência (vogais) misturados com ruído de alta frequência (fricativas). Com uma taxa de produção de aproximadamente 10 fonemas/s, a fala deve se dividir em dois fluxos de tons altos e baixos alternados. Os ouvintes são capazes de entender e repetir uma sequência rápida de fala, mas não são capazes de relatar a ordem de sons curtos não relacionados (por exemplo, um assobio, zumbido, etc.) (Arons, B. p.10, 1992, tradução nossa)¹⁸

Algumas pessoas conseguem identificar mais frequências em uma mesma fonte sonora do que outras. (Arons, 1992) E quando mais de dois sons ocorrem

¹⁸ Texto original: *In addition to the grouping processes already mentioned, there are additional extensions and ideas that are specific to the analysis of speech signals. Note that it is often difficult to separate primitive processes from schema, and that speech schemas tend to obscure the contributions of primitive processes. Considering the primitive segregation rules, it is somewhat surprising that voices hold together at all. Speech consists of sequences of low frequency complex tones (vowels) intermixed with high frequency noise (fricatives). With a production rate of roughly 10 phonemes/sec, speech should break up into two streams of alternating high and low tones. Listeners are able to understand and repeat a rapid sequence of speech, but are not able to report the order of short unrelated sounds (e.g., a hiss, buzz, etc.)*

simultaneamente, o ouvinte pode perceber sons adicionais que não tenham sido emitidos mas emergem na representação perceptual, a partir da combinação de frequências.

Mais recentemente, Yost (2018) relaciona a análise de cena auditiva com a percepção da origem de som, definindo a segregação do fluxo auditivo¹⁹, o que implica extensão no tempo, ou uma fonte contínua, e sua natureza contínua parece ajudar a definir as possíveis fontes. Os fluxos auditivos são auto-evidentes na experiência humana e podem ser demonstrados também com sequências de tons simples e sons biologicamente mais significativos, como a fala humana e a música.

Rumsey (2018) propõe quatro elementos de foco, que são considerados típicos da escuta: a fonte, o conjunto, o ambiente e a cena. Cada elemento está descrito em termos espaciais que vão do microscópico ao macroscópico, apresentados em um contexto musical. O microscópico refere-se a atributos de elementos individuais dentro de uma cena, e o macroscópico lida com os atributos descritivos da cena como um todo, ou agrupamentos de elementos de dentro dele.

¹⁹ Termo usado para descrever a percepção de som em nosso ambiente. É a combinação de todos os sons que chegam aos nossos ouvidos ao mesmo tempo e é composto por uma mistura de fontes sonoras, incluindo fala, música, ruído ambiente, etc.

4.3 Percepção da Origem Espacial do Som

No domínio da nossa própria investigação, é crucial considerar como a distribuição dos sons no espaço pode influenciar sua diferenciação. A percepção pode agrupar sons quando sua localização no panorama auditivo é semelhante e separá-los quando é diferente. Em uma seção anterior (4.1), mencionamos que pessoas com deficiência visual apresentam uma preferência pelo lado direito no campo auditivo (Dias, 2008; Wettstein, 2018). Na verdade, várias investigações sugerem que isso também ocorre com pessoas com visão e que essa tendência é inata. O fenômeno conhecido como ‘Right-Ear Advantage’ é reconhecido em testes realizados pela Universidade da Califórnia (UCLA) em 2004, assim como por estudos de Roth (2016) e da Universidade de Tel Aviv. Os resultados mostram que a *Right-Ear Advantage* é perceptível em bebês e crianças, indicando que é inata. Outra questão importante diz respeito à compreensão da fala em ambientes reverberantes e ruidosos. O jornal CCADET realizou um teste em que os participantes ouviram palavras em diferentes posições de azimute: em um ambiente sem ruído, a compreensão da fala foi excelente para todos os ângulos (quase 100% das palavras corretas); em ambientes ruidosos ou com eco, a posição relativa do altifalante no lado direito mostrou uma influência significativa na compreensão da fala. Além disso, a compreensão da fala é ligeiramente maior quando o ouvinte não está de frente para o falante e as palavras estão a cerca de $\pm 30^\circ$ de azimute (e.g. Ortiz, Bustamante, 2014).

De acordo com Bamiou (2004), a vantagem do ouvido direito é amplamente documentada e consiste na melhor percepção de estímulos auditivos apresentados ao ouvido direito em comparação com o ouvido esquerdo. Bamiou (2004) apresenta evidências de que essa vantagem é observada tanto em tarefas de identificação quanto em tarefas de discriminação de fala. A investigação de Bamiou (2004) avaliou o papel da atenção na vantagem do ouvido direito usando gravações de magnetoencefalografia²⁰ enquanto os participantes escutavam palavras de duas sílabas. Os resultados confirmaram a vantagem do ouvido direito e mostraram que esta estava relacionada à alocação de atenção e aumento da amplitude da resposta auditiva estável. Isso estava correlacionado com o índice de lateralidade

²⁰ É uma técnica neurofisiológica que mede as variações do campo magnético geradas por atividade elétrica no cérebro. É considerada uma das técnicas de neuroimagem mais precisas para medir a atividade elétrica no cérebro em tempo real.

comportamental. As conclusões sugerem que a vantagem do ouvido direito está relacionada à alocação de atenção e atividade neural no córtex auditivo esquerdo.

Darwin (2007) observa que cada fonte sonora é percebida em uma posição espacial diferente e é possível localizar sons relativamente ao plano horizontal (azimute) com maior precisão do que relativamente ao plano vertical. No entanto, a maior parte das investigações sobre processamento de informação sonora no cérebro humano considera apenas sons individuais apresentados de forma isolada, enquanto na vida real é comum escutarmos múltiplas fontes de som em diferentes posições espaciais.

De acordo com a teoria da Duplex, descrita por Rayleigh em 1907 (e.g. Yost, 2017), nossos dois ouvidos captam som separadamente. Esta teoria afirma que as ondas sonoras consistem em duas partes: a onda de pressão e a onda sonora. A onda de pressão é a perturbação que se propaga pelo ar e outros meios, empurrando objetos. A teoria afirma que a localização de sons é feita através da combinação de diferenças de tempo interaural (ITD) e diferenças de nível/volume interaural (ILD). O ITD é mais importante nas baixas frequências, enquanto o ILD é mais importante nas altas. O ITD para um tom modulado de alta frequência também pode ser uma pista para a localização de sons. A combinação de ITD e ILD permite a localização de sons com maior resolução na frente do que atrás.

De acordo com Roginska (2018) e outros autores no campo da investigação binaural, como Blauert (1992), Ballou (2005) e Darwin (2007), a ITD e a ILD fornecem informações espectrais para inferir a posição da fonte sonora, considerando a função de transferência relacionada à cabeça (HRTF). A HRTF opera com base em frequências dominantes e fornece informações que permitem a localização de diferentes fontes sonoras.

Tanto Blauert (1992) quanto Moller (1992) afirmam que a audição usa uma série de pistas para determinar a direção e distância da fonte sonora, incluindo cor, diferenças de tempo interaural, diferenças de fase interaural e diferenças de nível interaural. Estas pistas são consideradas responsáveis pela audição direcional em diferentes domínios. Nas frequências baixas, a avaliação da direção é feita através de diferenças de fase interaural, nas frequências médias, através de diferenças de tempo interaural, e nas altas, através de diferenças de nível interaural. A coloração é considerada responsável quando não há diferenças interaurais no plano mediano.

4.3.1 Função de Transferência Relacionada com a Cabeça (HRTF)

O termo ‘HRTF’ geralmente aparece juntamente com ‘HRIR’. Ambos são respostas de impulsos de ambientes geradas por cada ouvido (esquerdo e direito) representadas por ângulos de azimute e elevação. O HRIR é uma medida da função de transferência global e contém informações sobre magnitude e fase. Ele é usado para simular as propriedades espaciais e direcionais do som em aplicativos de áudio 3D. Em resumo, o HRTF representa a resposta em frequência da função de transferência, enquanto o HRIR representa a resposta no domínio do tempo da função de transferência, incluindo informações sobre magnitude e fase.

Antes de um som chegar aos ouvidos, ele é filtrado por difração, reflexão da cabeça, torso, ouvidos e ambiente, resultando em atenuação de energia e amplitude. Kendall (1995) explica que a interação da onda sonora original com o corpo, a cabeça, o ouvido interno e os canais auditivos depende de propriedades que podem ser medidas e compreendidas. Essas propriedades são refletidas em uma função matemática chamada ‘Função de Transferência Relacionada à Cabeça’ (Head Related Transfer Function). Quando aplicada ao processamento de amplitude e espectro de frequência, a HRTF produz sensações que nos levam a localizar a posição relativa de várias fontes sonoras no campo auditivo (e.g. Roginska, 2018). Na tecnologia binaural, a HRTF é calculada com base nas frequências dominantes e, embora um algoritmo possa ser incapaz de incluir todas as variáveis da nossa percepção natural, ele fornece pistas de localização extremamente detalhadas e realistas ao comparar o som captado por cada ouvido. Basicamente, ele altera o espectro sonoro original de acordo com a posição e a anatomia do corpo. Ballou (2005) especifica uma implicação importante:

Duas pessoas não compartilham a mesma forma de orelha e cabeça, e apreendemos as nossas próprias orelhas e tamanho/forma da cabeça ao longo de anos de experiência. Se alguém ouvir sons complexos com os HRTFs de outra pessoa, embora a localização esquerda-direita seja boa, haverá muita confusão frente-trás. (Ballou, 2005, p.59)

Para decodificar os parâmetros acústicos do ambiente em características de resposta a impulso, Blauert (2013) afirma que a tecnologia binaural pode ser uma metodologia eficaz na transformação de sinais binaurais em sinais enviados aos

ouvidos. O processo de simulação de um ambiente utilizando tecnologia de simulação de ambiente binaural permite recriar as reflexões e refrações das ondas sonoras. Ele consiste em duas partes: modelagem de campo sonoro e auralização. Enquanto a modelagem de campo é feita através de um modelo computacional baseado na representação física do ambiente, a auralização se refere aos resultados audíveis (e.g. Blauert e Lehnert, 1992, p. 260).

Blauert (2013) menciona que os modelos de localização acústica buscam imitar o sistema auditivo humano, utilizando sinais binaurais divididos em sinais intra-aurais. Embora sinais monoaurais sejam necessários para determinar a posição lateral do ouvinte, eles são menos sensíveis à direção frontal, traseira e elevação.

O método de medição de HRTF implica na captura da resposta ao impulso de uma fonte sonora em várias posições em relação à cabeça do ouvinte. Normalmente, é feita a medição de HRTF em uma câmara anecóica ou em um ambiente controlado, onde a reflexão do som é minimizada. O método de síntese envolve a geração de um modelo HRTF com base em parâmetros físicos, geométricos e fisiológicos, ao invés da medição real. A abordagem mais comum é usar um simulador de cabeça e tronco (HATS) ou um modelo virtual de pavilhão auricular para gerar HRTFs. Em ambos os métodos, os dados de HRTF resultantes podem ser usados para aplicar processamento binaural a uma fonte sonora, criando uma experiência de audição mais imersiva e espacialmente precisa com auscultadores.

4.3.2 Diferença de Tempo Interaural (ITD)

Para um som ser definido como binaural, ele deve possuir primeiramente um sinal estereofônico e fornecer indicações de localização espacial relacionadas ao funcionamento do ouvido humano. O ouvido mais próximo da fonte sonora recebe o som antes do ouvido mais distante, o que é a pista auditiva mais importante para a localização de sons. Na Figura 19 podemos observar a Diferença Temporal Interaural (ITD) e como ela é uma concepção física que expressa a diferença temporal entre duas ondas idênticas. De acordo com os estudos dedicados à teoria Duplex, um som grave emitido perto do ouvinte produz uma diferença de fase sonora inequívoca entre

os ouvidos, mas quando o mesmo som é emitido de uma maior distância, a diferença de fase²¹ não é relevante. No entanto, isso não se aplica às frequências agudas, que produzem grandes diferenças de fase (e.g. Roginska, 2018).

Darwin (2007) observou que as informações derivadas da ITD são degradadas em condições reverberantes e não oferecem dados fiáveis sobre o azimuth da fonte sonora. O autor também debruçou-se sobre outro tópico importante para essa investigação: a interferência da lateralização de sons simultâneos.

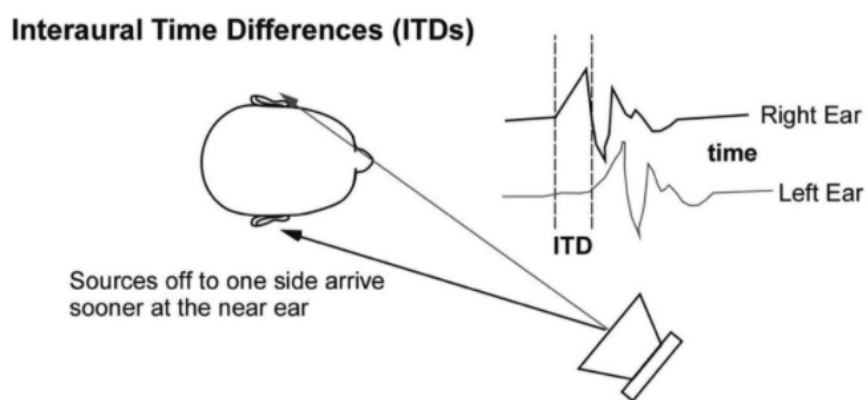


Figura 19 - Exemplo da diferença de tempo interaural

Fonte: Roginska, 2018, p.12

Quando duas regiões de frequências diferentes são emitidas por fontes sonoras independentes e simultâneas, a percepção interpreta-as como sendo separadas. Experiências realizadas por McFadden e Pasan (1976) e posteriormente por Henning (1980) pediram aos participantes para julgar a posição lateral do som. Frequências de 300 Hz e 900 Hz com polaridade oposta foram reproduzidas com ITD e uma situação de escuta diótica²² foi criada. Os resultados sugerem que as diferenças de tempo interaural não fornecem uma pista primária para a agregação auditiva.

²¹ É a diferença, expressa em ângulo ou tempo, entre duas ondas que tenham mesma frequência e em referência ao mesmo ponto no tempo. (Ballou, 2005)

²² Cujas percepção é idêntica a ambos os ouvidos.

4.3.3 Diferença de Nível Interaural (ILD)

A diferença de nível interaural (ILD ou IID) é um conceito importante na teoria Duplex e é uma das principais pistas utilizadas pelo sistema auditivo humano para a localização de fontes sonoras. ILD refere-se à diferença de amplitude entre o som chegando a cada ouvido. É importante destacar que a diferença de nível interaural não depende somente da distância entre a fonte sonora e o ouvido, mas também do posicionamento da cabeça, da fonte sonora e de obstáculos na área.

Um estudo realizado por Blauert (1997) mostrou que o uso de informações ILD é mais efetivo em condições sem eco, ou seja, em ambientes secos. Em condições reverberantes, a informação ILD é prejudicada devido à reflexão de ondas sonoras nas paredes e no teto. Além disso, ILD também pode ser influenciado por diferenças na sensibilidade auditiva dos dois ouvidos. A diferença de sensibilidade auditiva pode afetar a percepção da direção e distância de uma fonte sonora. Portanto, é necessário levar em consideração essas variações individuais ao utilizar informações ILD para avaliar a localização de fontes sonoras. A Figura 20 exemplifica a escuta em conjunto à ILD.

Interaural Intensity Differences (IIDs)

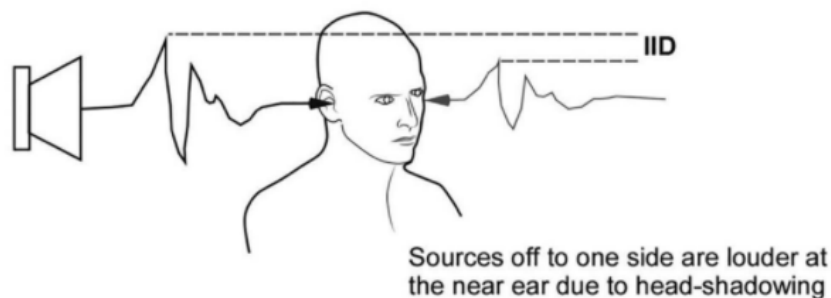


Figura 20 - Fontes sonoras de um lado são mais altas no ouvido próximo devido ao mascaramento da cabeça.

Fonte: Roginska, 2018, p.12

A sensibilidade à ILD varia com a frequência da fonte sonora. De acordo com Roginska (2018), a sensibilidade à ILD é maior para frequências mais baixas,

enquanto que para frequências mais elevadas, a sensibilidade à ILD é menor. Este fato explica porque somos mais capazes de localizar sons graves do que sons agudos. A ILD também é afetada pela absorção acústica dos materiais circundantes, a reverberação do ambiente e a presença de outras fontes sonoras (e.g. Blauert, 1997). Isso significa que em condições acústicas complexas, a percepção da ILD pode ser comprometida.

4.3.4 Mecanismos Perceptivos que Influenciam a Inteligibilidade da Fala

De acordo com Darwin (2007), quando um diálogo é reproduzido, geralmente compete com algum ruído de fundo, no entanto a percepção auditiva implementa mecanismos de mascaramento que favorecem a atenção seletiva. O cérebro ignora componentes mais fracos do som ruidoso, focando-se na parte da informação detectada que corresponde à voz:

O problema de separar dois locutores é, portanto, muito diferente do problema de tentar ouvir um locutor contra um fundo de ruído constante. Esse ruído fornece um nível de mascaramento constante ao longo do tempo (além de flutuações estatísticas) em cada banda de frequência. Consequentemente, apenas as partes mais intensas do sinal serão detectáveis contra esse pano de fundo, e o problema perceptual (ou computacional) é de detecção. (Darwin. 2007 p.5, tradução nossa)²³

O discurso verbalizado é mais inteligível quando está espacialmente separado dos sons concorrentes do que quando ambos provêm da mesma fonte sonora (Cherry, 1953 & Darwin, 2007, p.224). Darwin (2007) afirma que os mecanismos perceptivos possibilitam uma melhor compreensão do que o processamento computacional. Estes mecanismos podem variar dependendo das características do ruído ambiental. O mascaramento é facilitado quando as frequências do ruído são diferentes das frequências de fala, enquanto que a presença de outro diálogo simultâneo pode criar problemas na alocação de informações sensoriais para a outra fonte sonora. As representações intermediárias do som - entre a codificação sensorial no nervo auditivo

²³ Texto original: *The problem of separating two talkers is thus very different from the problem of trying to listen to one talker against a background of steady noise. Such noise provides a level of masking that is constant over time (apart from statistical fluctuations) in each frequency band. Consequently, only the most intense parts of the signal will be detectable against this background, and the perceptual (or computational) problem is one of detection.*

e a representação fonética no cérebro - podem estar relacionadas ao modo como o cérebro diferencia diferentes sons (e.g. Darwin, 2007, p.9).

O processamento neural de informações auditivas pode ser influenciado pelo treinamento e experiência auditiva. A presença de eco pode interferir na diferenciação entre fontes sonoras concorrentes, e o ruído branco também pode levar a uma redução significativa na inteligibilidade da fala.

5. Estudo de caso: o posicionamento da audiodescrição no campo auditivo

A percepção tende a agrupar sons temporalmente próximos, especialmente quando seus espectros frequentes são semelhantes ou temporariamente próximos (e.g. Bregman, 1994). Isso significa que a sobreposição de uma narração descritiva na banda sonora de um filme pode prejudicar a compreensão da história. Por esse motivo, na mistura sonora é comum diminuir o volume da banda original para dar destaque à audiodescrição, o que infelizmente também acaba diminuindo o som do filme em si. A questão é como isso pode ser evitado. O problema é complexo e as soluções podem variar de acordo com diferentes parâmetros e critérios. Por isso, decidimos conduzir um estudo de caso experimentando misturar a audiodescrição e a banda sonora de uma curta-metragem de quatro maneiras diferentes.

Referimos anteriormente que a distribuição de múltiplos sons no campo auditivo facilita a respectiva segregação, contrariamente ao que sucede quando esses sons provêm de uma mesma fonte. A audiodescrição com som binaural foi estudada por Mariana Lopez, no âmbito do projeto *Enhancing Audio Description*²⁴. Contudo, esse estudo focou-se na espacialização de efeitos sonoros, e refazer o som de um filme em função da audiodescrição envolve mais recursos do que habitualmente se dispõe na indústria de distribuição. Haverá outras maneiras de incrementar a inteligibilidade, que podem ser implementadas com menos recursos técnicos.

O objetivo do nosso estudo foi experimentar várias maneiras de tratar a audiodescrição, e perceber se o formato binaural pode favorecer a inteligibilidade do todo. O estudo foi exploratório: pretendemos testar algumas possibilidades, e analisar resultados tendo em conta dois gêneros de informação. Por um lado, extrapolamos da investigação apresentada nos capítulos anteriores de modo a discutir efeitos perceptivos. Por outro, realizamos uma entrevista a pessoas invisuais de modo a monitorizar a inteligibilidade do enredo de uma obra.

²⁴ Disponível em <<https://enhancingaudiodescription.com>>

5.1 Pressupostos

Este estudo foca-se na própria mistura de som, e na escuta com auscultadores. Sendo assim, as características do espaço envolvente não afetarão o resultado final percebido.

Sabe-se que:

- Uma mistura estereofônica produz uma imagem auditiva de 180°. A posição relativa de cada som é inferida a partir da comparação entre duas pistas áudio, tendo em conta a respectiva amplitude e espectro frequencial.
- Quando adicionamos uma faixa monofônica a uma faixa estereofônica, a sua mistura resulta numa faixa estereofônica. O sinal da monofônica é igual na pista direita e esquerda, e o sinal da estereofônica é diferente.
- Os *plugins* de simulação binaural, que operam com base na HRTF, podem ser aplicados a uma faixa estéreo de modo a produzir uma imagem sonora de 360°. Podemos experienciar isto com auscultadores, pois bastam duas pistas áudio para manipular a posição dos sons no campo auditivo; tendo em conta a respectiva amplitude e espectro frequencial.
- A inteligibilidade da fala em ambientes ruidosos é ligeiramente maior quando o ouvinte não está de frente para o locutor, e os sons-palavras se localizam em ângulos de azimute em torno de $\pm 30^\circ$. (Ortiz, Bustamante, 2014).
- As diferenças interaurais de tempo e amplitude fornecem mais pistas sobre a posição da fonte sonora no plano horizontal do que no plano vertical. A percepção de posição do som no plano vertical depende do filtro exercido sobre a onda sonora pela orelha, além das assimetrias espectrais criadas pelos reflexos das ondas sobre os ombros e sobre a cabeça do ouvinte. (Roginska, 2018).
- Num contexto de escuta de uma fonte sonora que se encontra a uma distância maior do que cerca de 1m, as pistas binaurais proporcionam poucas sugestões da distância da fonte. (Kapralos, 2003).
- Para permitir uma localização da fonte de forma precisa é necessário que a fonte contenha uma grande variedade de frequências. Algumas investigações

indicam que a precisão de localização diminui à medida que a largura de banda da fonte diminui. (Kapralos, 2003).

A inteligibilidade da audiodescrição e da banda sonora original depende grandemente dos princípios que governam os mecanismos de agregação e segregação perceptual. Bregman (1994) e Toiviainen (2008), notaram como a percepção utiliza a regularidade de estímulos para organizar a recepção dos sons:

- Sons não relacionados raramente começam e terminam exatamente ao mesmo tempo.
- Gradações de mudança são percebidas como regulares quando: a) um som único tende a mudar suas propriedades suave e lentamente; b) uma sequência de sons de uma mesma fonte tende a mudar suas propriedades lentamente.
- As mudanças que ocorrem em um evento acústico afetam todas as componentes do som percebido ao mesmo tempo.

A posição relativa de fontes sonoras simultâneas é um dos fatores que influencia a agregação e segregação perceptual, e a distância a que fonte sonora se encontra do ouvinte é determinante para a atenção seletiva. A percepção opera com base em vários indicadores:

- A amplitude pode ser considerada uma pista primária para a identificação da distância (Yost, 2005). Mas não deve ser considerada isoladamente, uma vez que a percepção também se serve de outras pistas. Toiviainen (2008) mostra-nos como a intensidade relativa de um som também depende das suas qualidades específicas. Por exemplo, um grito a grande distância pode ter um volume percebido mais alto do que um sussurro a curta distância.
- O movimento de translação do ouvinte causa uma mudança de azimuth maior para objetos próximos do que para objetos distantes (Toiviainen 2008). Este fenômeno também tem implicações quando o ouvinte se mantém imóvel: a mudança de posição dos sons será mais perceptível quando os sons se encontram próximos do que quando se encontram distantes.
- A relação de amplitude entre o som direto e o reverberante é igualmente importante. Se a amplitude da reverberação for proporcionalmente muito elevada, a fonte sonora parecerá distante (Toiviainen 2008).

- Um som produzido ao ar livre, quando percorre em frequências nas zonas agudas e médias, pode dar a impressão de uma fonte distante. (Toiviainen, 2008) De fato, as altas frequências atenuam-se mais rapidamente do que as baixas, e o som precisa cobrir uma grande distância através do ar antes que o percebamos.

Numa secção anterior referimos que a percepção auditiva implementa mecanismos de mascaramento que favorecem a atenção seletiva, privilegiando a fala humana (e.g. Darwin, 2007). Toiviainen (2008) fornece detalhes importantes sobre outros tipos de mascaramento, descrevendo-o como um fenómeno pelo qual um sinal de baixa intensidade (mascarado) é tornado inaudível por um sinal mais forte (mascarador). O mascaramento pode ser simultâneo ou temporal.

Relativamente ao mascaramento simultâneo, o autor explica que:

- É causado pela sobreposição de padrões de atividade no ouvido.
- A força do mascaramento depende do conteúdo de frequência dos dois sinais de áudio, bem como de sua amplitude relativa.
- Com tons puros, o efeito de mascaramento é mais forte quanto mais próxima a frequência do mascarado estiver da frequência do mascarador.
- É mais forte para as frequências acima do mascarador do que para as frequências abaixo.
- A faixa de frequências em que ocorre o mascaramento é de cerca de 90 Hz para sons abaixo de 200 Hz.
- Aumenta para cerca de 900 Hz para frequências em torno de 5000 Hz.
- O grau em que sons complexos mascaram ou são mascarados por outros sons depende da sua intensidade e de seu conteúdo espectral. Por exemplo, uma diferença de intensidade maior é necessária para um mascarador senoidal mascarar um mascarado semelhante a ruído do que é necessário para um mascarador semelhante a ruído mascarar um mascarado senoidal.
- Um mascarador de média ou alta frequência é mais propenso a mascarar frequências altas do que frequências baixas.
- Um mascarador de baixa frequência mascara as frequências mais baixas.

Relativamente ao mascaramento temporal, Toiviainen descreve-o como o fenômeno pelo qual um tom suave é mascarado por um tom mais alto que ocorre um pouco antes (pós-mascaramento) ou depois (pré-mascaramento). Observa que:

- O pós-mascaramento ocorre em cerca de 50-200 ms após a remoção do mascaramento.
- O pré-mascaramento tem menor alcance; o intervalo é de cerca de um décimo de segundo.

McAdams (1993) nota que o timbre faz com que o ouvinte identifique a textura sonora, permitindo que sons que vibram na mesma frequência possam ser identificados como distintos. O timbre é formado pela frequência fundamental somada aos harmônicos, que são as demais frequências secundárias que compõem o objeto sonoro. A frequência fundamental relaciona-se diretamente com o tom. Toiviainen (2008) explica que tanto a informação de lugar (vibração máxima no ouvido) quanto a informação de taxa (estrutura periódica de estímulos sonoros, codificados por padrões de disparos neurais) são usadas pelo sistema auditivo para determinar o tom. Além disso,

- A taxa de informação domina para baixas frequências e a informação de lugar domina para altas frequências. Com efeito, frequências baixas fornecem menos pistas sobre localização do que frequências médias e altas.
- Para tons harmônicos complexos, o tom percebido é igual à sua frequência fundamental.
- Para tons não-harmônicos, como um som de sino, o tom percebido pode não ser claro – esses tons geralmente resultam na percepção de vários tons simultâneos.

São muitas as variáveis que determinam a percepção do som. E, como McAdams (1993) observou, muitas podem ser de natureza cultural, independente da física acústica. O nosso estudo de caso assume a complexidade e a ambivalência como motor exploratório, mas ancora-se em fatos sólidos. A partir da análise espectral da banda sonora do filme e das quatro versões de audiodescrição, podem-se prever possíveis efeitos de mascaramento, de modo a tornar possível-se jogar com a posição da AD no campo auditivo para atenuar esses efeitos.

5.2 A Obra Escolhida e a Gravação da Audiodescrição

O estudo foi realizado com a curta-metragem *All Through The House*, do episódio seis da segunda temporada da série premiada "*Love, Death and Robots*" de Tim Miller. A versão estéreo do filme está disponível na plataforma Netflix. A série apresenta curtas-metragens abordando temas de ficção científica, terror, fantasia e comédia. O episódio em questão conta a história de duas crianças, Billie e sua irmã, na noite de Natal. Elas acordam com um barulho vindo da sala de estar e descobrem a presença de um monstro assustador na forma do Papai Noel. A série possui a opção de audiodescrição em inglês na Netflix, e para fins de investigação, foi realizada a tradução para o português.

As análises de frequência foram realizadas usando o Izotope RX7 para comparar a banda sonora original com as diferentes audiodescrições. A Figura 21 apresenta o espectro sonoro da banda sonora original, incluindo efeitos sonoros, diálogos e música, com amplitude (dB) e frequência (Hz) do sinal estéreo.

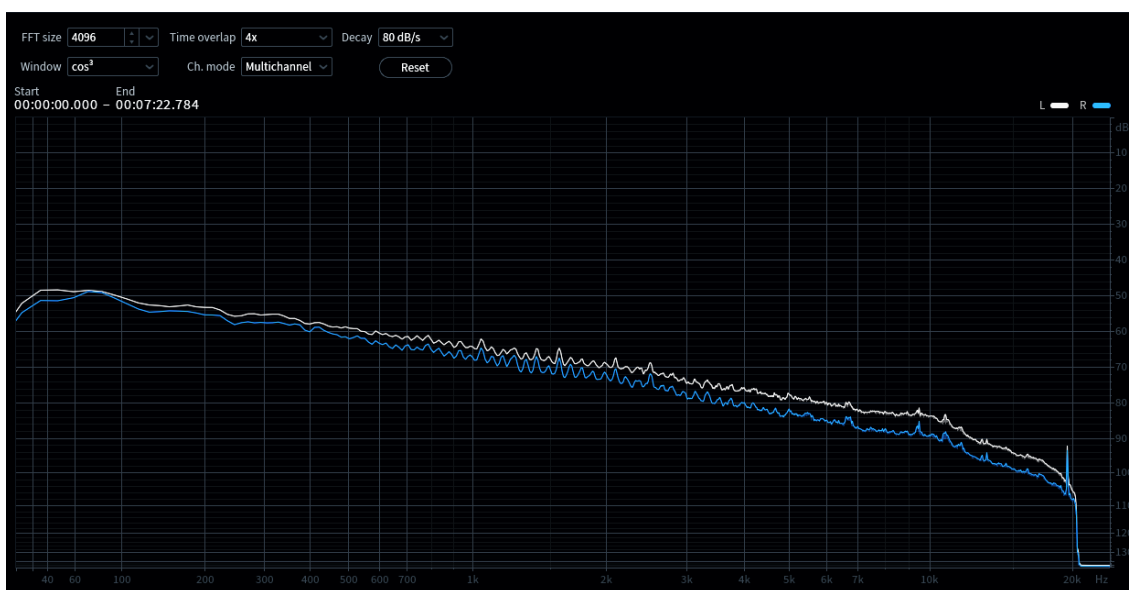


Figura 21 - Análise espectral do som da curta-metragem

Fonte: Análise do autor via Izotope RX7

A análise espectral do filme revela uma predominância de frequências médio-graves, com maiores níveis de amplitude abaixo dos 150 Hz. Para melhorar a inteligibilidade dos sons sobrepostos, é importante elaborar uma AD que considere a

diferença espectral entre eles (e.g. Bregman, 1994). As frequências da voz masculina variam de 80 Hz a 150 Hz e da voz feminina de 150 Hz a 200 Hz (e.g. Behlau, 2001).

Portanto, neste caso, seria mais adequado usar uma AD com uma voz de frequências médias e agudas, ao invés de graves. No entanto, para agradar ao público infantil, raramente se utilizam vozes muito graves. Escolhemos uma voz feminina, como a opção de idioma na audiodescrição oferecida pela Netflix. Era de se esperar, portanto, que a inteligibilidade pudesse ser mais complicada acima de 150 Hz.

A audiodescrição foi traduzida para o português do Brasil por Gabriela Pina e narrada por Anaí Corrêa em uma cabine acústica da Medical Company, com 0,9m x 1,8m. Foi utilizado um microfone condensador de captura cardióide Behringer B2-Pro para a gravação, e não um sistema de captação binaural. Embora se acredite que é possível simular as características binaurais com *plugins* de edição de áudio, considera-se, de acordo com Making Music Pro²⁵ e Producer Hive²⁶, que o melhor microfone para tal é o Neumann KU100²⁷ (€7899). No entanto, este estudo não teve acesso a esse equipamento. Outros microfones mais acessíveis, como o Roland CS-10EM²⁸ e 3DIO Free Space²⁹, são menos sensíveis na direção vertical, apenas na horizontal.

²⁵ <<https://makingmusicpro.com>>

²⁶ <<https://producerhive.com/buyer-guides/studio-gear/neumann-ku100-review-binaural-microphone-head>>

²⁷ <<https://en-de.neumann.com/ku-100>>

²⁸ <<https://roland.com/products/cs-10em>>

²⁹ <<https://3diosound.com/collections/microphones/products/free-space-binaural-microphone>>

5.3 As Versões De Mistura

A gravação da audiodescrição foi misturada com a banda sonora original. Elaboramos quatro misturas³⁰, mantendo intacta a imagem estereofônica do filme. Tratamos a audiodescrição de modo diverso em cada mistura, adotando diversas estratégias para potenciar a sua inteligibilidade.

Mistura #1	Audiodescrição Monofônica
-------------------	----------------------------------

A Mistura #1³¹ implementa a estratégia comumente adotada na indústria. O sinal de áudio da audiodescrição monofônica é adicionado em igual medida nas duas pistas da mistura final, que é estereofônica, como a banda sonora do filme. Isso posiciona a audiodescrição no centro do campo auditivo, à frente do ouvinte. Ao comparar o som captado pelos dois ouvidos, o cérebro extrai informações sobre o azimute.

Adicionalmente, a amplitude sonora do filme foi diminuída em 6 dB para tornar a narração audível e atender às normas de loudness da Netflix (-27 LKFS ($\pm 2,0$ LU)). A amplitude serve como indicador de proximidade entre o ouvinte e a fonte sonora (e.g. Bregman, 1994; Yost, 2005; Toivainen, 2008). Assim, na Mistura #1, a audiodescrição está no primeiro plano, à frente do ouvinte, enquanto o som original do filme passa para o plano de fundo.

³⁰ <https://drive.google.com/drive/folders/1R0lwGYj93k_Mh1Nb7Mq4RP6q88dJEoKd?usp=sharing>

³¹ <https://drive.google.com/drive/folders/1SFECqTtW0pn1n6U5E7Ae8ulrpnITlyi?usp=share_link>

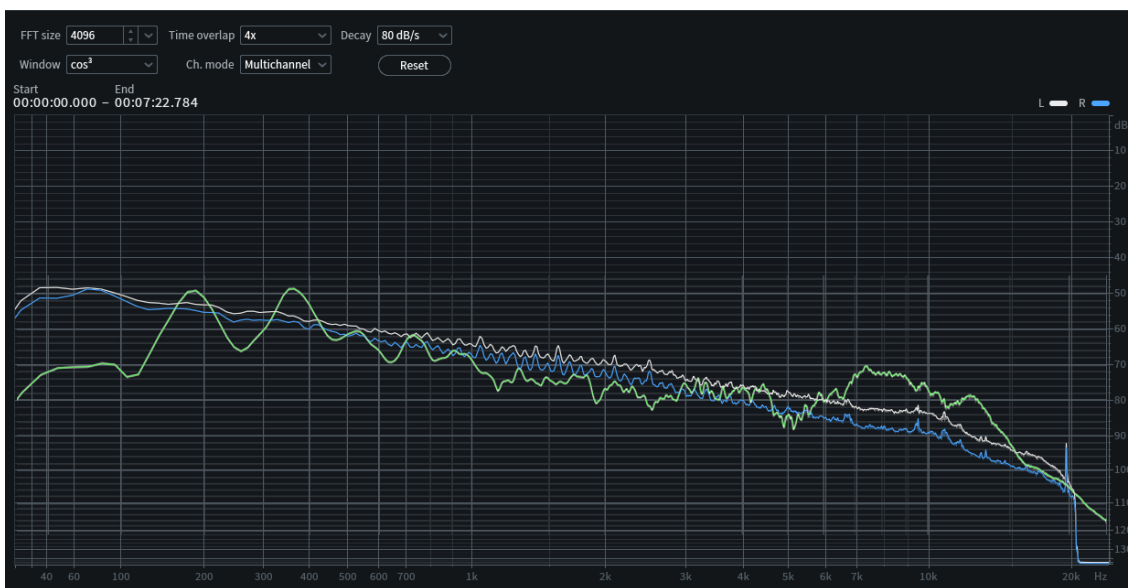


Figura 22 - Mistura #1 - espectro (Hz) e amplitude(dB) da AD e da banda sonora

Fonte: Análise do autor via Izotope RX7

A Figura 22 permite-nos comparar o espectro da AD com o do som original da obra, e prever possíveis zonas de mascaramento. Sabemos que o efeito de mascaramento é mais forte quando a AD e a banda sonora ocupam posições semelhantes no espectro, especialmente na zona de frequências médias, que são dominantes na AD (e.g. Toiviainen, 2008). Além disso, o mascaramento é mais forte para as frequências acima do mascarador do que para as frequências abaixo. O sinal monofônico da AD, representado por verde, mostra que a narração tem maior amplitude na zona acima de 5000 Hz, o que pode mascarar a banda sonora. A banda sonora original é mais audível nas frequências mais baixas, abaixo do espectro da voz feminina.

Mistura #2**Audiodescrição Estéreo →**

Na Mistura #2³², a banda sonora do filme mantém a amplitude original, e a amplitude da audiodescrição é mais alta na pista direita do que na esquerda, diferença de aproximadamente 6 dB. Ao comparar o som captado pelo ouvido direito e esquerdo, o cérebro conclui que a narradora se encontra à direita do ouvinte.

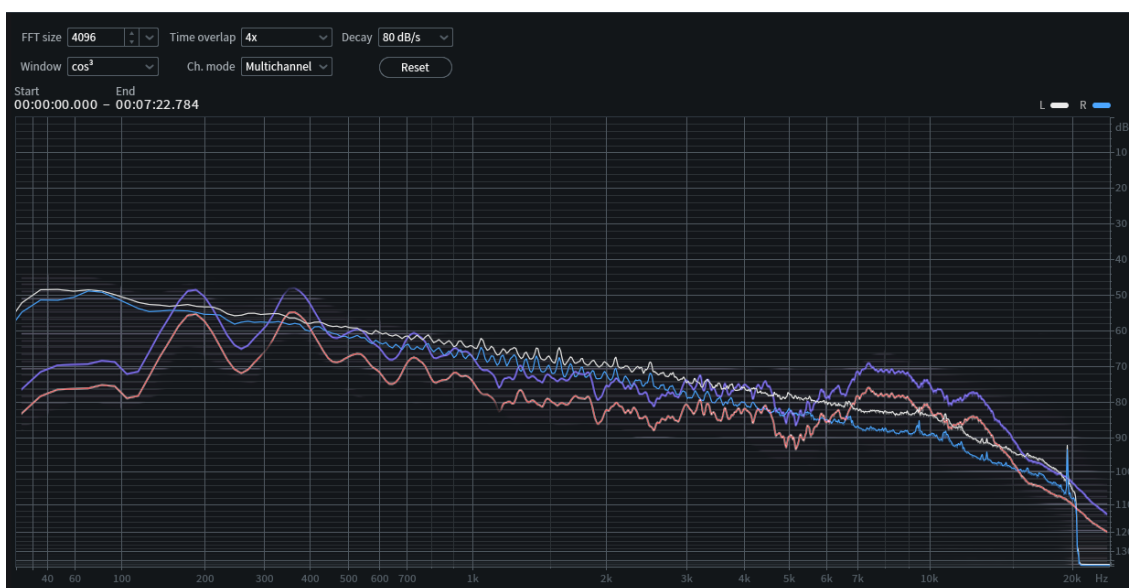


Figura 23 - Mistura #2 - espectro (Hz) e amplitude(dB) da AD e da banda sonora

Fonte: Análise do autor via Izotope RX7

Na Figura 23, a audiodescrição em formato estéreo com posicionamento à direita (13%) é representada pelos sinais rosa e roxo. Em geral, a sua amplitude é menor que a da banda sonora, inclusive na zona de altas frequências. Isso pode evitar que a banda sonora seja relegada para o plano de fundo e as componentes mais agudas sejam mascaradas pela audiodescrição. Além disso, a posição lateral pode ajudar a evitar ou reduzir o mascaramento da audiodescrição, facilitando a segregação perceptual. No entanto, também pode haver desvantagens: a atenção seletiva do ouvinte pode ficar concentrada demais no extremo direito do campo auditivo, onde se encontra a audiodescrição.

³² <https://drive.google.com/drive/folders/1xf7Q3pqP5We0xqs6ZS2SSStNnOJEkW1u3?usp=share_link>

Mistura #3**Audiodescrição Binaural →**

Na Mistura #3³³, a narração da audiodescrição também se encontra à direita do ouvinte. O formato da mistura é Binaural, mas mantivemos duas pistas. É também possível utilizar o mesmo software para trabalhar com *surround* e outros formatos mais complexos. Buscamos uma posição de azimuth que não distraísse totalmente a atenção do ouvinte para o lado direito. Alguns estudos que compararam a inteligibilidade da voz frontal e lateral mostraram que, sob condições de ruído, os ouvintes percebem melhor o diálogo que está no ângulo lateral azimuth $\pm 30^\circ$ do que na posição frontal (e.g. Ortiz & Bustamante, 2014). Levando isso em consideração, posicionamos a audiodescrição em um ângulo de 35° à direita. A Figura 24 mostra a respectiva configuração digital. A escolha pela posição foi baseada em testes próprios que privilegiam a banda sonora original.



Figura 24 - *Plugin* de simulação binaural DearVR

Fonte: Acervo do autor

Simultaneamente, buscamos integrar a audiodescrição no espaço acústico do filme. Aplicamos um pouco de reverb do próprio *plugin* para criar a sensação de que há uma pessoa presente narrando a audiodescrição. Fizemos isso no modo Direct Signal, o que significa que o único som da audiodescrição é a simulação que conta com as características pré-estabelecidas de HRTF. Configuramos o software para

³³ <https://drive.google.com/drive/folders/1rmuYEsYvkuRXT5ENTDATkgX64wgerXf?usp=share_link>

simular uma distância de pouco mais de 1 metro sabendo que a uma distância maior, as pistas binaurais oferecem poucas sugestões sobre a distância da fonte (e.g. Kapralos, 2003).

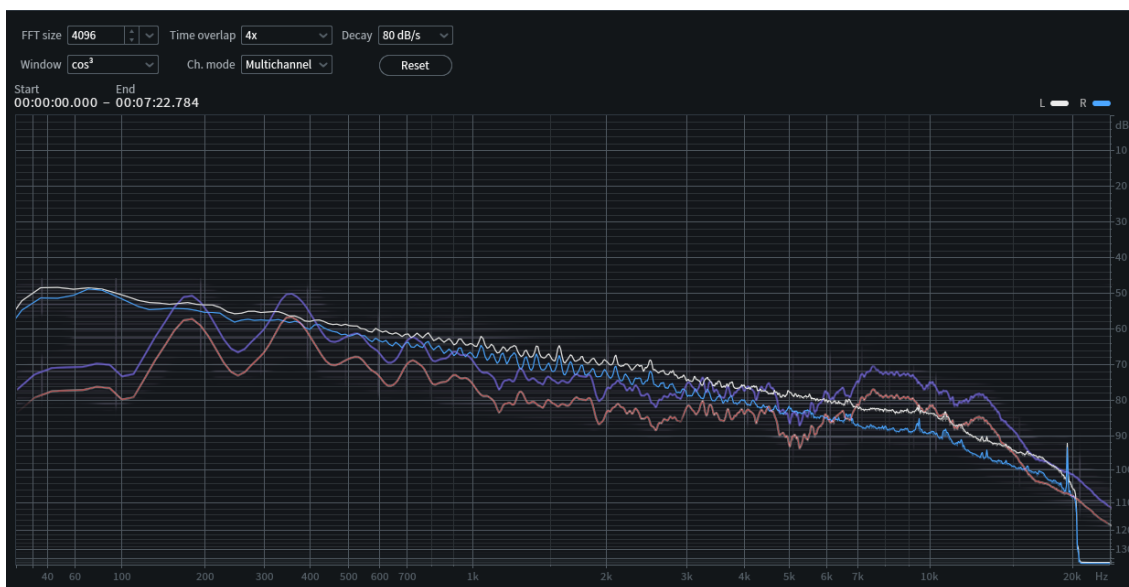


Figura 25 - Análise de espectro de frequências e nível (dB) da AD binaural à direita

Fonte: Análise do autor via Izotope RX7

O software utilizou o efeito Diferença de Tempo Interaural (ITD), onde o lado direito do sinal estéreo foi reproduzido 150 milissegundos antes do lado esquerdo. Isso permite que a percepção integre os sinais em um único evento, mas até certo ponto, esta diferença temporal pode reduzir o efeito de mascaramento. O pós-mascaramento ocorre em torno de 50 ms a 200 ms após a remoção do mascaramento, enquanto que o pré-mascaramento tem um alcance menor, com um intervalo de cerca de um décimo de tempo (e.g. Toiviainen, 2008).

Mistura #4	Audiodescrição Binaural ↑
-------------------	----------------------------------

Na Mistura #4³⁴ a audiodescrição também se encontra em formato binaural. Desta vez, situa-se numa posição elevada, por detrás da cabeça do ouvinte. Utilizou-se o *plugin* observado na Figura 26, o Senheiser Ambeo. Ao comparado com o DearVR, pareceu-nos que processa melhor a sensação de binaural com elevação.



Figura 26: *Plugin* de simulação binaural Ambeo

Fonte: Acervo do autor

Optou-se pelo azimute a 180° e por uma elevação superior a 47°. O próprio *plugin* oferece impulsos de resposta de reverberação, permitindo o controle do tamanho das reflexões e da sala. Os parâmetros para a decisão foram estruturados para fornecer clareza ao perceber a narração da AD por outra direção espacial do azimute, que fosse distante do diálogo da curta-metragem.

³⁴ <https://drive.google.com/drive/folders/11Fk-M9x-7qIHF3mBoAMemxPChYPI5Q_v?usp=share_link>

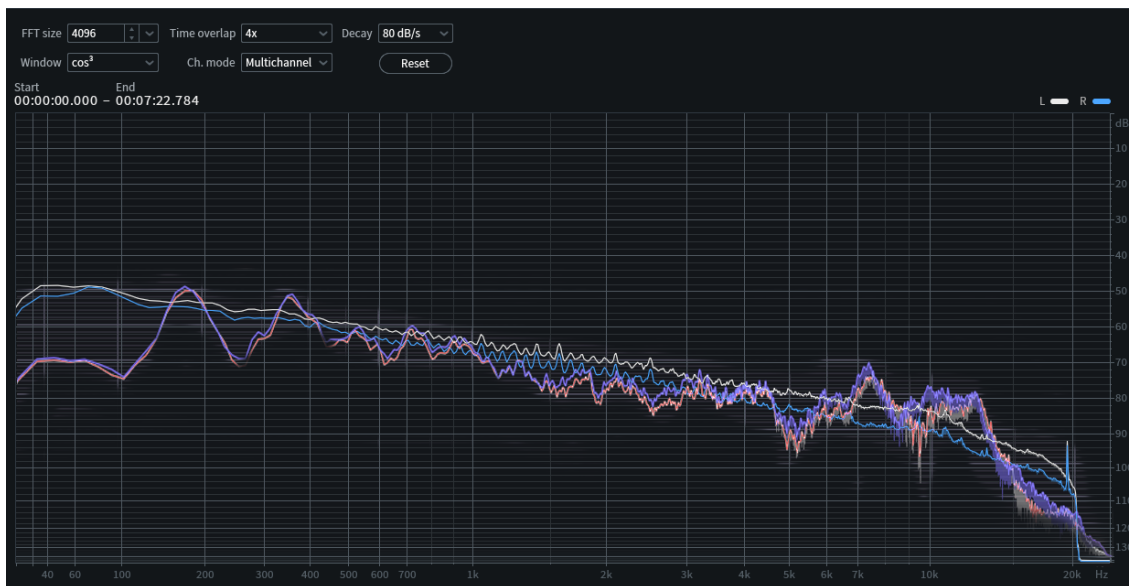


Figura 27 - Análise de espectro de frequências e nível (dB) da AD binaural com elevação

Fonte: Análise do autor via Izotope RX7

A análise espectral da Mistura #4 vista na Figura 27 difere-se das outras por mostrar mais instabilidade de frequências no decorrer da obra. Isso pode influenciar mecanismos de localização, tornando a posição da AD mais ambígua; até porque é mais fácil discernir a posição de fontes sonoras no plano horizontal do que no vertical (e.g. Roginska, 2018). No entanto, esta colocação da AD pode facilitar a segregação perceptiva, sem relegar o som do filme para segundo plano nem dar primazia ao lado direito do campo auditivo.

5.4 Entrevistas

Para averiguar o nível de inteligibilidade das nossas quatro misturas, entrevistamos 23 participantes com perda total e com visão reduzida de origem, congênita ou adquirida. A entrevista incluiu 11 perguntas de múltipla escolha e 2 de resposta livre. (p.f. ver anexo)

As entrevistas foram realizadas à distância. De acordo com Gil (2008), as entrevistas remotas possuem o benefício de permitir uma amostragem mais ampla. As pessoas foram recrutadas a partir de grupos do Facebook destinados a pessoas com algum tipo de deficiência visual. Foi utilizado um formulário de perguntas do Google.

Conforme a Revista Pesquisa Qualitativa³⁵ (2020), os leitores de tela³⁶ utilizados em smartphones por pessoas com deficiência visual são compatíveis com esse tipo de pesquisa online. A entrevista incluiu 13 perguntas de múltipla escolha e duas de resposta livre.

Com o objetivo de compreender a relação dos participantes com a audiodescrição, não foi imposto um limite de tempo. Conforme afirma Gil, este tipo de entrevista é uma forma de diálogo assimétrico, na qual uma das partes busca coletar dados e a outra se apresenta como fonte de informação (e.g. Gil, p.109). O autor observa que:

Este tipo de entrevista é bastante empregado em situações experimentais, com o objetivo de explorar a fundo alguma experiência vivida em condições precisas. Também é bastante utilizada com grupos de pessoas que passaram por uma experiência específica, como assistir a um filme, presenciar um acidente etc. Nestes casos, o entrevistador confere ao entrevistado ampla liberdade para expressar-se sobre o assunto. (Gil, p.113, 2008)

Antes de os voluntários assistirem ao curta-metragem, eles foram solicitados a compartilhar sua idade, o tipo de deficiência visual e sua opinião sobre a audiodescrição na prática em suas rotinas. Depois do filme, eles começaram informando o tipo de fone de ouvido que utilizaram - pequeno e interno (intra-auricular) ou médio/grande e externo (auricular). Em seguida, responderam a diversas perguntas de múltipla escolha sobre aspectos narrativos específicos do curta-metragem. Por fim, foi solicitado que os participantes dessem sua opinião sobre a audiodescrição apresentada de forma livre e dissertativa.

³⁵ <<https://editora.sepq.org.br/rpq/article/view/250>>

³⁶ Aplicação de software de smartphones para implementar acessibilidade por meio da leitura de discurso sintetizado, permitindo o usuário a ouvir o texto.

5.5 Discussão De Resultados

Os participantes do estudo não apresentaram dificuldades quanto ao formulário e às perguntas referentes à obra apresentada. Foi verificado que as pessoas com baixa visão consomem menos audiodescrição do que as que são completamente cegas. Aqueles com baixa visão consomem a audiodescrição apenas raramente ou ao menos uma vez por mês, enquanto as pessoas com deficiência visual a consomem de duas a três vezes por semana. Pessoas com mais de 46 anos consomem menos audiodescrição do que as mais jovens. A maioria dos participantes (90%) utilizou auriculares *in-ear*. Não foi identificada diferença na inteligibilidade em função do tamanho dos auscultadores, o que pode ser explicado pela ampla variedade de dispositivos de som disponíveis com características diferentes.

Houve poucas respostas incorretas às perguntas sobre o enredo da obra, mas estas podem ser significativas. A Tabela 1 apresenta a média de acertos. A mistura monofônica #1 obteve mais acertos, enquanto a mistura binaural com elevação #4 teve menos acertos. A mistura binaural à direita empatou com a estéreo.

	Mono	Estéreo (direita)	Binaural (direita)	Binaural (elevação)
Acertos sobre inteligibilidade	100%	80%	80%	60%

Tabela 1 - Ordem de acertos sobre inteligibilidade da obra das versões de audiodescrição

Os trechos das questões #9 e #10 obtiveram mais erros, portanto foram analisados através da análise de espectro de frequência. Eles seguem o padrão médio de frequências da obra como um todo, o que também nos permite extrapolar conclusões sobre a inteligibilidade geral das nossas quatro versões de mistura.

A questão #9 de múltipla escolha ‘Como era a criatura?’ foi a mais controversa em todas as misturas. A Figura 28 apresenta a análise espectral do trecho da obra que contém a resposta da narradora. É possível notar que na obra original, há uma maior amplitude do que na narrativa audiodescritiva. Na mistura #1, a

audiodescrição não é mascarada, devido à sua posição frontal que monopoliza a atenção e à redução do som original do filme, que passa a ficar no segundo plano. É provável que o mascaramento ocorra nas outras misturas.

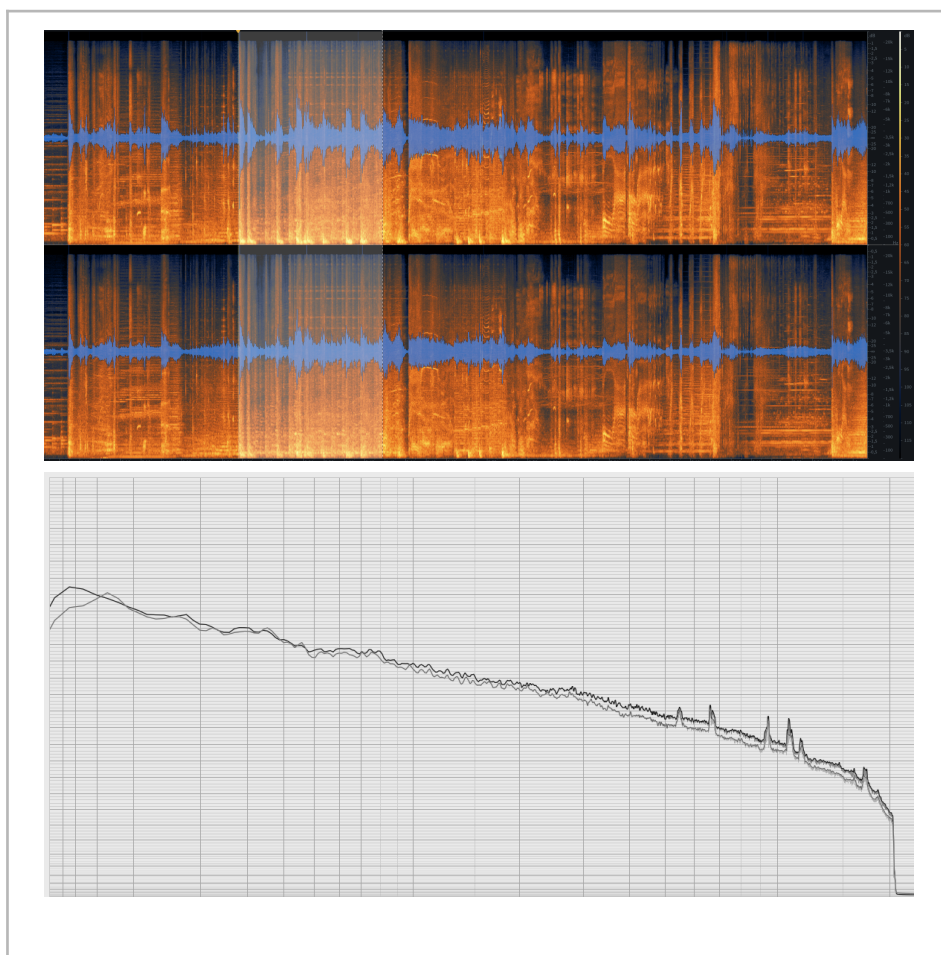


Figura 28 - Espectrograma com onda sonora
Análise média de frequências da banda sonora do trecho da pergunta #9

Fonte: Análise do autor via Izotope RX7

A Tabela 2 apresenta as frequências médias das quatro misturas em comparação com a banda sonora original. É possível verificar como a diferença de tempo (ITD) e nível interaural (ILD) afetam as versões binaurais. O *plugin* binaural utiliza padrões de equalização para tornar os impulsos de resposta mais claros. Para isso, foi feita uma cortina entre 6 kHz e 10 kHz com a narração inserida na curta-metragem.

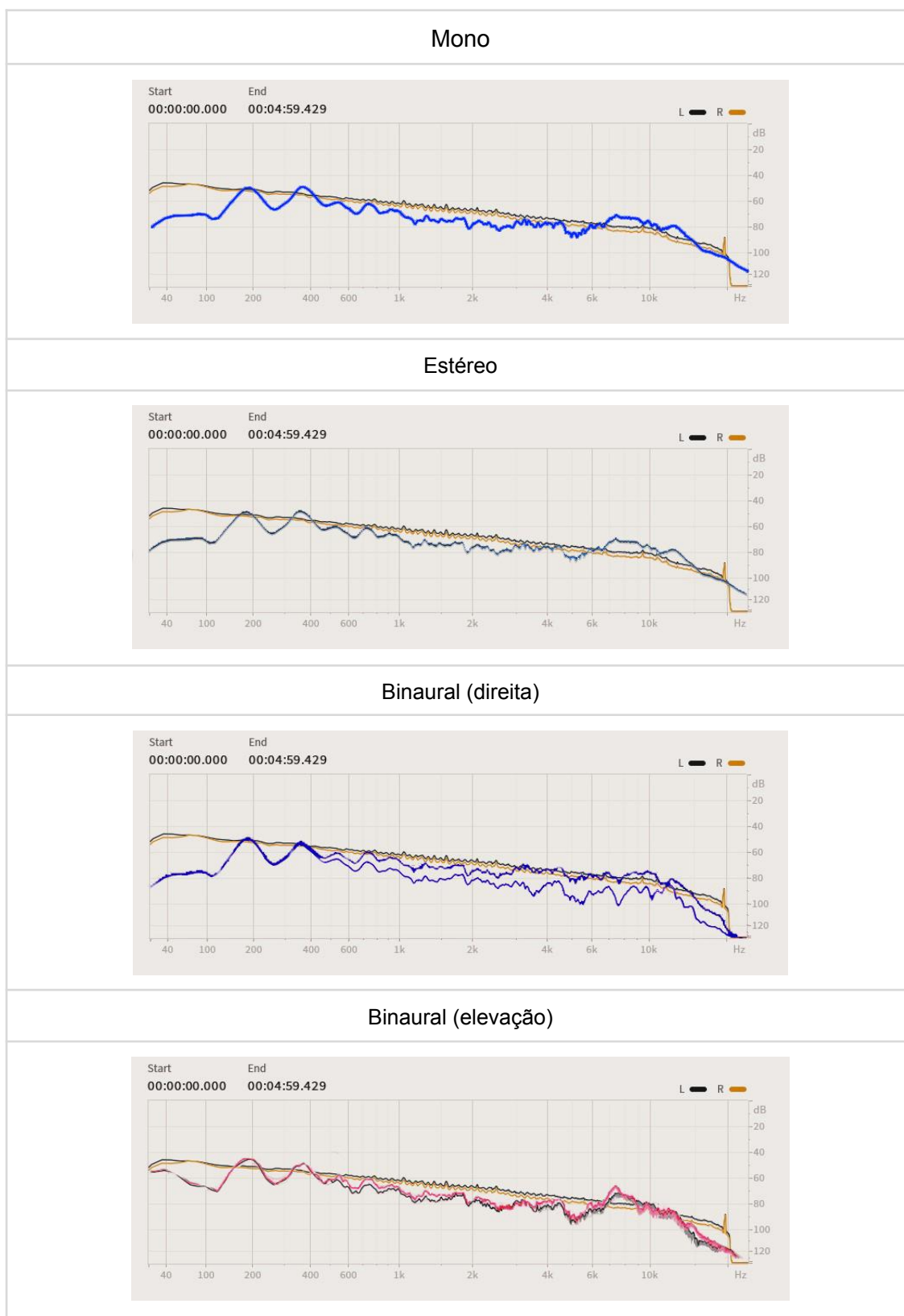


Tabela 2 - Análise de frequências das quatro misturas em relação a banda sonora original

Fonte: Análise do autor via Izotope RX7

O mascaramento sonoro ocorre principalmente quando o conteúdo das frequências dos sinais são parecidos (e.g. Darwin, 2007; Toiviainen, 2008), e isso pode explicar o motivo pelo qual esta questão obteve erros. A voz da narração (150-250 Hz) compete com a música devido a abordar frequência próximas nos instrumentos de cordas como o contra-baixo (41-246 Hz), o violoncelo (65-659 Hz) e o violino (196-2637 Hz).

A falta de inteligibilidade da AD em formato binaural, situada atrás e acima do ouvinte na Mistura #4, parece mais surpreendente, visto que o mascaramento ocorre menos quando as fontes de fala e ruído são separadas espacialmente. (Blauert, 2013, p.428) Pode ser explicada por vários fatores:

1. A propensão natural de um ouvinte para prestar mais atenção à zona frontal do campo auditivo.
2. A cadência de localização auditiva no eixo vertical. (Krapalos, 2013)
3. O fato de um som traseiro ser recebido de modo filtrado, devido à anatomia das orelhas pode contribuir para que a AD pareça mais imprecisa.

As respostas à pergunta #10 'Qual era o nome das crianças?' apresentaram mais equívocos na Mistura #1, onde a AD é monofônica. Há alguns aspectos que o podem explicar:

1. Embora o nome do irmão seja repetido mais de quatro vezes, o nome da irmã é dito apenas duas vezes, tanto na audiodescrição como na banda sonora original.
2. Quando a narradora AD diz 'Uma luz se acende, (pausa) Billie acorda' (00:59) é o primeiro momento que o nome do irmão aparece, seguido de uma pausa, que incrementa a inteligibilidade.
3. O nome da irmã é referido pela primeira vez quando a narradora da audiodescrição diz 'A irmã mais velha Lia senta na cama' (01:02). Neste caso, não faz pausa, e as sílabas '(ve)lha' e 'Lia' possuem fonema parecido, o que pode confundir o ouvinte.

A Tabela 3 relativa a esta secção do filme mostra o espectro do som original. Na obra audiovisual, os intervalos de menor amplitude de som - que não possuem a visualização de onda sonora em azul - possuem ruídos representados pelas notas harmônicas do violino da banda sonora musical. A música representa o suspense e a curiosidade da cena. O violino é representado pelos riscos alaranjados contínuos ao centro, que predominam nos 1300 Hz - 5600 Hz. As frequências dominantes - excluindo o violino - rondam os 200 Hz e 1300 Hz.

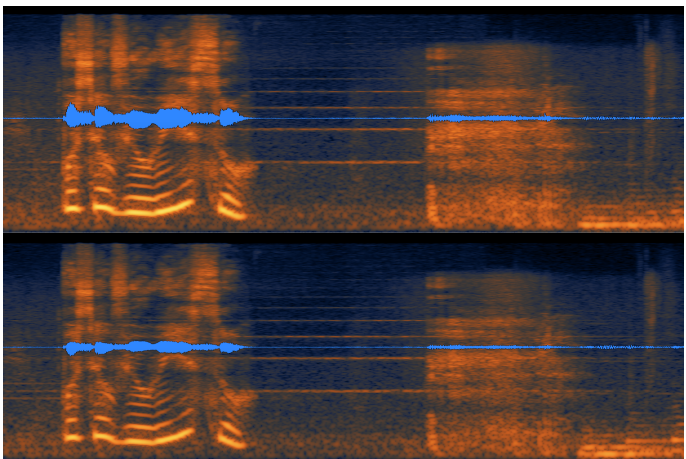
	Frequências dominantes	Espectrograma (frequência e amplitude)
Banda sonora	80 Hz (-55dB) 253 Hz (-51dB) 321 Hz (-48dB) 400 Hz (-48dB) 1200 Hz (-61dB) 2811 Hz(-61dB)	

Tabela 3 - Análise espectral de frequências da banda sonora da obra

Fonte: Análise do autor via Izotope RX7

A análise espectral da AD neste trecho do filme mostra que a Mistura #1, na qual é monofônica, possui menos frequências agudas (máx. 7400Hz) do que nas misturas #2 (máx. 8260Hz) e #3 (máx. 9999Hz). Isto pode ser uma explicação adicional para o mascaramento presente: quanto menor é a gama de frequências, mais difícil se torna a segregação das fontes sonoras individuais (e.g. Bregman, 1994).

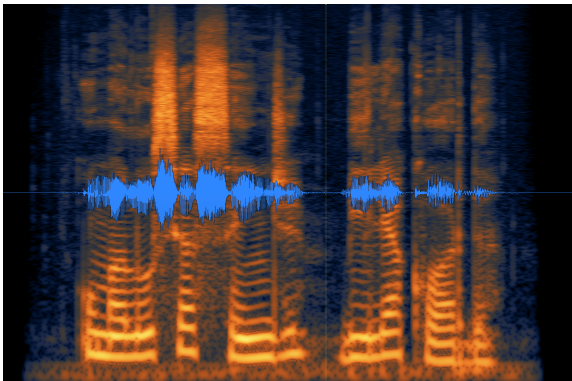
#1	Frequências dominantes	Espectrograma (frequência e amplitude)
Mono	182 Hz (-42dB)	
	349 Hz (-39dB)	
	763 Hz (-55dB)	
	966 Hz (-58dB)	
	7400 Hz (-61dB)	

Tabela 4 - Análise espectral de frequências da Mistura #1 de AD

Fonte: Análise do autor via Izotope RX7

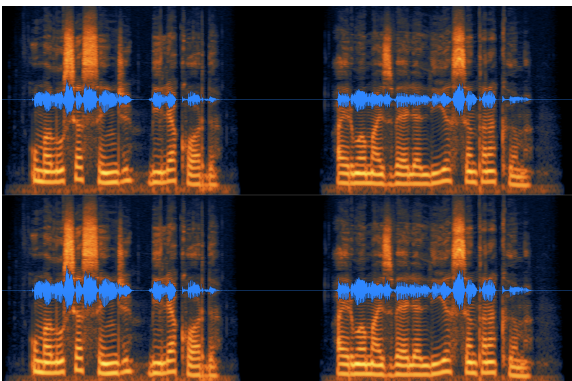
#2	Frequências dominantes	Espectrograma (frequência e amplitude)
Estéreo direita	182 Hz (-41dB)	
	349 Hz (-41dB)	
	760 Hz (-54dB)	
	966 Hz (-58dB)	
	8260 Hz (-61dB)	

Tabela 5 - Análise espectral de frequências da Mistura #2 de AD

Fonte: Análise do autor via Izotope RX7

Na versão binaural à direita (Mistura #3), a maior amplitude dos agudos deve-se ao *plugin* DearVR modular o efeito de reflexões e as características HRTF, o que adiciona mais informação espectral. Embora a maior quantidade de informação espectral possa aumentar o mascaramento, a audiodescrição não está na posição central, tornando a inteligibilidade de ambas as fontes aceitável.

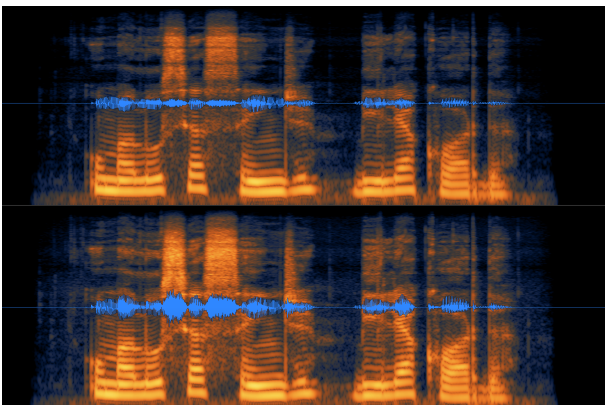
#3	Frequências dominantes	Espectrograma (frequência e amplitude)
Binaural direita	183 Hz (-42dB)	
	347 Hz (-45dB)	
	763 Hz (-53dB)	
	966 Hz (-56dB)	
	3325 Hz (-59dB)	
	9999 Hz (-63dB)	

Tabela 6 - Análise espectral de frequências da Mistura #3 de AD

Fonte: Análise do autor via Izotope RX7

A respeito da Mistura #3 em que a audiodescrição é binaural com foco no ouvido direito, alguns participantes relataram estranhamento referente a posição da narração, embora tenham gostado da experiência. Por exemplo, um participante escreveu:

"Esplêndida, voz boa, cadência boa, e acompanhou bem as cenas no tempo. Particularmente nunca ouvi a audiodescrição dividida, então não poderia me expressar a respeito dessa parte em si, mas de início, posso dizer que prefiro com os áudios normais. Talvez seja só falta de costume."

Outros participantes aplaudiram a experiência, como se vê neste comentário:

"Perfeita a descrição foi bem detalhada e não foi tão rápida e além de ouvir áudio descrição podemos acompanhar as falas do filme em tempo real que é importante também além de entendermos o que está acontecendo por meio do recurso."

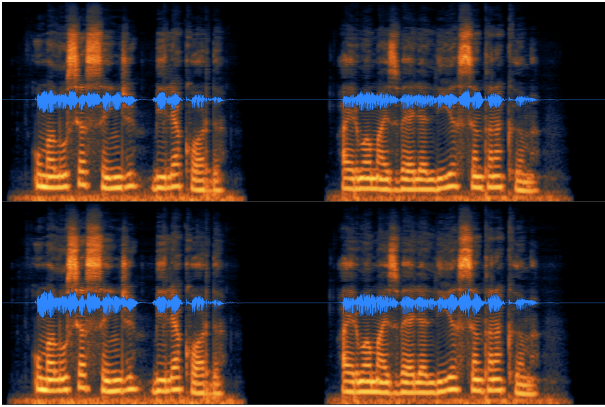
#4	Frequências dominantes	Espectrograma (frequência e amplitude)
Binaural elevação	190 Hz (-40dB)	
	340 Hz (-46dB)	
	480 Hz (-54dB)	
	760 Hz (-61dB)	
	7400 Hz (-61dB)	

Tabela 7 - Análise espectral de frequências da Mistura #4 de AD

Fonte: Análise do autor via Izotope RX7

Pode parecer surpreendente que a Mistura #4, também binaural, apresente um limite de agudos inferior, semelhante ao da Mistura #1. Isso é devido ao modo como a tecnologia simula uma fonte sonora atrás da cabeça, elevada em relação ao plano horizontal. A localização perceptiva de um som nesta posição depende muito do filtro exercido pelas orelhas, e o software reproduz isso (e.g. Roginska, 2018).

Embora os resultados do estudo de caso indiquem a versão mono como a mais inteligível, esta se sobressai apenas por uma questão na amostra que consta 20 pessoas. Acreditamos que deve ser realizado um estudo futuro com mais participantes, de modo a comparar e testar mais posicionamentos de audiodescrição no campo auditivo, para testar simulação binaural e gravação binaural.

A utilização do som binaural permite aprimorar a experiência de ouvir a audiodescrição ao mesmo tempo que a banda sonora de um filme. Com uma audiodescrição binaural situada lateralmente - esta voz torna-se tão compreensível como no modelo tradicional - com a vantagem de não ser necessário reduzir o nível (dB) da banda sonora original.

6. Conclusão

Esta investigação sobre narração audiodescritiva binaural pretende servir de auxílio ao debate sobre novos métodos de acessibilidade. Acreditamos que a pessoa com deficiência visual deva ter mais opções para consumir materiais acessíveis. O nosso problema de investigação não pode ter uma resolução universal: a combinação puramente auditiva de uma experiência audiovisual é impossível por natureza. Em primeiro lugar, porque a junção entre som e imagem produz significados e efeitos que não estão contidos na própria imagem e no próprio som. Em segundo lugar, porque a memória a curto prazo e a dinâmica da atenção tornam a combinação de uma audiodescrição com o som original do filme num problema altamente complexo que excede as preocupações criativas consideradas na concepção das obras audiovisuais.

Nossa investigação iniciou-se com a compreensão da audiodescrição. Atualmente, a maioria das obras cinematográficas conta com a opção de audiodescrição, que é uma faixa adicional de áudio que serve como uma narrativa complementar às cenas visuais relevantes para a compreensão da história. Embora tenha havido um aumento na atenção dos distribuidores de entretenimento em relação ao setor de acessibilidade devido à implementação de novas políticas públicas, as emissoras e cinemas ainda têm dificuldade em atender ao número mínimo de horas de exibição de obras acessíveis. A audiodescrição é inserida como uma faixa adicional entre os diálogos dos personagens e pode prejudicar a percepção da música e dos efeitos sonoros, que são elementos importantes na narrativa cinematográfica.

Estudamos os princípios que governam a percepção auditiva, e compreendemos algumas vantagens auditivas que pessoas com deficiência visual possuem: sobressaem em tarefas que incluem apenas escutar, quando comparadas às pessoas com visão apurada. Além disso, apresentamos investigações que indicam que as pessoas escutam mais com o ouvido direito, o fenômeno conhecido como Right Ear Advantage, e levamos isso em consideração ao desenvolver o nosso estudo de caso. Para o estudo, consideramos investigações prévias sobre o campo auditivo. Estudamos que fatores influenciam a inteligibilidade do som, e que fatores influenciam especificamente a inteligibilidade de discurso falado. No caso, quando a escuta possui

um foco, normalmente existem ruídos que podem competir com a percepção auditiva, de modo a causar mascaramento sonoro.

A partir de nossa investigação, compreendemos que a simulação binaural pode ser convincente para o ouvinte se for reproduzida através de auscultadores e realizada com as configurações ideais de posicionamento, reflexão, reverberação e controle de resposta de impulsos. Desenvolvemos um protótipo de microfone binaural e constatamos que, embora a simulação binaural não seja equivalente à gravação binaural realizada com microfones profissionais, os resultados podem ser satisfatórios no que se refere à identificação precisa de sons e à simulação de ambientes virtuais. Além disso, a simulação permite a edição de parâmetros de arquivos sonoros, uma vez que o microfone depende da captação e de um ambiente tratado acusticamente. Os resultados das investigações sobre a localização sonora permitiram-nos afirmar que os *plugins* binaurais para software de edição de áudio são inferiores às gravações de microfones topo de gama no que se refere ao parâmetro de elevação vertical ($+90^\circ$).

Realizamos um estudo de caso sobre diferentes formas de tratar o posicionamento de uma narração audiodescritiva gravada em mono. O estudo se concentrou na simulação binaural através de auscultadores para permitir o controle dos parâmetros de um ambiente virtual, independentemente das condições físicas do ambiente em que o ouvinte se encontra. Com o uso de um *plugin* binaural para *softwares* de edição de áudio - que considera características HRTF e que aplica a diferença de tempo interaural (ITD) e a diferença de nível interaural (ILD) ao processamento de sinal sonoro - tornou-se possível posicionar a AD em qualquer ponto do campo auditivo de azimuth e elevação. No entanto, é fundamental ter em conta a inteligibilidade da AD, para que os sons não se misturem e não se mascarem uns aos outros. Contrariamente ao que se poderia prever, o nosso estudo sugere que uma colocação lateral é preferível a uma colocação traseira com elevação. É possível que isso se deva a hábitos instaurados, que podem ser modificados. Em todo o caso, é provável que isso também varie com as características específicas de cada segmento sonoro.

Ao utilizarmos um modelo computacional que converte uma AD mono ou estéreo em binaural, podemos manipular diferentes parâmetros de modo a dar destaque às frequências desejadas. Isso permite-nos evitar o mascaramento sonoro;

deixa de ser necessário diminuir a amplitude do som original do filme para garantir a inteligibilidade da AD.

Nossa pesquisa reforça a ideia de que são necessárias mais opções de mistura que sejam focadas nas pessoas com deficiência visual. Sugerimos que as próprias produtoras de filmes considerem previamente o público com deficiência visual, o que não é comum. Recomendamos que seja feita uma mistura específica de audiodescrição para cada banda sonora original, considerando uma análise das frequências dominantes da obra. Visto isso, uma possibilidade para futuros estudos é a criação de um modelo computacional focado na mistura de audiodescrição, que simule áudios mono e estéreo binaural, considerando equalização, posicionamento azimutal e as frequências dominantes para evitar o mascaramento que pode ser causado por sons sobrepostos.

Referências Bibliográficas

- Arons, B. (1992). *A Review of The Cocktail Party Effect*. MIT Media Lab.
- Ballou, G. (2005). *Handbook for sound engineers*. 3rd ed. [S.l.]: Focal Press, Gulf Professional Publishing.
- Behlau, M. (2001). *Voz: o livro do especialista*. Revinter.
- Blauert, J. (2013). *The Technology of Binaural Listening: Modern Acoustics and Signal Processing*. Springer.
- Blauert, J. & Lehnert, H. (1992). *Principles of Binaural Room Simulation*. Isevier Ltd.
- Bourne, J. (2008). *The Didactics of audiovisual translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bregman, A. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.
- Callaghan, C. (2009). *Auditory Perception*. <<https://plato.stanford.edu/entries/perception-auditory>> Acesso em 29 de abril de 2021.
- Chion, M. (2011). *A Audiovisão*. Columbia University Press New York Chichester, Texto & Grafia, ed.
- Chion, M. (1994). *The Audiovisual Contract: Projections of Sound on Image*. In *Audio-Vision: Sound on Screen*. New York: Columbia UP, annotation by Deborah Wolfson (Theories of Media, Winter 2003).
- Dias, T. & Pereira, L. (2008). *Habilidade de localização e lateralização sonora em deficientes visuais*. UNIFESP. <<https://scielo.br/j/rsbf/a/NbGwM9XSGQHvMqW8bsVLZ4x/?format=pdf&lang=pt>> Acesso em 18 de abril de 2021.
- Darwin, C.J. (2008). *Listening to speech in the presence of other sounds*. Phil. Trans. R. Soc. B.
- Franco, E.P.C. & Silva, M.C.C. (2010). *Audiodescrição: Breve Passeio Histórico*. In: MOTTA, L. M. V.; ROMEU FILHO, P. (orgs.). *Audiodescrição: Transformando Imagens*

em Palavras. Secretaria dos Direitos da Pessoa com Deficiência do Estado de São Paulo.

Gougoux, F., Zatorre, R., Lassonde, M., Vos, P & Lepore, F. (2005). *A Functional Neuroimaging Study of Sound Localization: Visual Cortex Activity Predicts Performance in Early-Blind Individuals*. PLOS Biology. <<https://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.0030027&type=printable>> Acesso em 25 de abril de 2020.

Gil, A.C. (2008). *Métodos e Técnicas de Pesquisa Social*. Editora Atlas, 6ª edição.

Jnerosci, J. (2020). *Early Blindness Shapes Cortical Representations of Auditory Frequency within Auditory Cortex*. Journal of Neuroscience, 39(26), 5143. Retrieved from <https://jneurosci.org/content/39/26/5143> on April 28, 2020.

Kapralos, B., Jenkin, M. R. M., & Milios, E. (2003). *Auditory Perception and Spatial (3D) Auditory Systems*. York University, Ontario.

Kendall, G. (1995). *A 3D Sound Primer: Directional Hearing and Stereo Reproduction*. Massachusetts Institute of Technology.

Kleiner, M., Dalenbäck, B., & Svensson, P. (1993). *Auralization-An Overview*. Audio Engineering Society. Retrieved from <http://aes.org/e-lib/browse.cfm?elib=6976> on September 20, 2022.

Lopez, Kearney, & Hofstadter. (2018). *Audio Description in the UK: What Works, What Doesn't, and Understanding the Need for Personalising Access*. British Journal of Visual Impairment.

Lopez, Kearney, Hofstadter, & Balla. (2020). *Enhancing audio description: accessible filmmaking, sound design and the importance of educating filmmakers*. Media Practice and Education Volume 21.

Mattana, A. (2017). *What is binaural audio*. <<https://hookeaudio.com/what-is-binaural-audio>> on November 24, 2021.

McAdams, S., & Bigand, E. (1993). *Auditory Scene Analysis: Hearing in Complex Environments*. In: Thinking in Sound: the Cognitive Psychology of Human Audition, Oxford University Press.

Miller, T., & Fincher, D. (2021). *Love, Death & Robots* (2nd season): All Through The House (episode 6). Blur Studio, Netflix Studios.

MøLLER, H. (1992). *Fundamentals of Binaural Technology*. *Applied Acoustics*, 36(3/4), 171-218.

Mota, M.G. (2003). *Orientação e Mobilidade: Conhecimentos Básicos para a Inclusão da Pessoa com Deficiência Visual*. <http://portal.mec.gov.br/seesp/arquivos/pdf/ori_mobi.pdf> Acesso em 20 de setembro de 2022

Moylan, W. (2007). *The Art of Recording*. Focal Press.

Naves, M., Alves. & Araújo. (2016). Guia para Produções Audiovisuais Acessíveis. Ministério da Cultura.

Neumann. *The Dummy Head - Theory and Practice*. <<https://lesonbinaural.fr/EDIT/DOCS/Neumann%20KU%20100.PDF>> Acesso em 25 de agosto de 2022.

Neves, CC. (2012). *Televisão e Deficiência Visual: O Sonoro na Produção de Imagens Mentais*. <<https://tede2.pucsp.br/bitstream/handle/4487/1/Camila%20Portella%20Neves.pdf>> Acesso 29 de abril de 2020.

Ortiz, A.L., Padilla, Bustamente, & F. Orduña. (2014). *Improving speech intelligibility for binaural voice transmission under disturbing noise and reverberation using virtual speaker lateralization*. *Journal of Applied Research and Technology*.

Quinot, D. & Katz, B. (2018). *The Anaglyph binaural audio engine*. Audio Engineering Society, Maio.

Roginska, A. & Geluso, P. (2018). *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Routledge; 1 edition, 2018.

Sá, A. (2013) *How An Audio-Visual Instrument Can Foster The Sonic Experience*. <<https://journals.gold.ac.uk/index.php/lea/article/view/84/65>> Acesso em 1 de junho de 2021.

Silva, M. O. & Mól, S. (2020). *O Uso Da Ferramenta Formulários Do Google Para Pesquisas Com Pessoas Com Deficiência Visual*. *Revista Pesquisa Qualitativa*, 8(17), 221-235.

Snyder, J. (2014). *The Visual Made Verbal: A Comprehensive Training Manual and Guide to the History and Applications of Audio Description*. Dog Ear Publishing, LLC.

Toiviainen, P. (2008). *The psychology of electronic music in The Cambridge Companion to Electronic Music*. In Nick Collins and Julio Escriván (Eds.), Cambridge University Press.

Ucla. (2004). *Left And Right Ears Not Created Equal As Newborns Process Sound, Finds UCLA/UA Research*. ScienceDaily. [Online]. Available at: <http://sciencedaily.com/releases/2004/09/040910082553.htm>.

Wendt, Van,, Ewert. (2014). *A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation*. J. Audio Eng. Soc., 62.

Wettstein V. & Probst, R. (2018). *Right Ear Advantage of Speech Audiometry in Single-sided Deafness*. Wolters Kluwer Health, Inc.

Yost, W. & Gourevitch, G. (1985). *Directional Hearing*. Springer-Verlag New York Inc.

Yost, W. (2017). *History of sound source localization: 1850-1950*. Acoustical Society of America.

Glossário

Azimuth	Medida de abertura angular do sistema de coordenadas horizontal.
Binaural	Sinal sonoro estéreo com pistas binaurais que dão a sensação realista ao ouvinte. É um áudio que simula a audição humana natural, usando dois microfones posicionados como orelhas humanas para capturar o som. Isso permite ao ouvinte ter uma experiência de áudio mais imersiva e realista, com informações sobre a direção e distância dos sons. Também é possível converter sons mono em binaural via modelo computacional.
Estéreo	Método de gravação e reprodução sonora que utiliza dois sinais independentes de modo a produzir uma perspectiva auditiva de 180°.
Mono	Sinal sonoro com um canal de captação ou reprodução, onde todos os sons são combinados em um único canal. Isso resulta em uma única fonte sonora centralizada, sem qualquer informação espacial.
Pistas Binaurais	Método de gravação sonora que utiliza dois microfones posicionados de modo a emular as condições auditivas do ser humano, somado com a simulação de reflexões de ambiente natural e produzir uma perspectiva auditiva de 360° com canais elevados.
Plugin	Ferramentas complementares a DAW.
Reverb	Reverberação ou Reverb é o som criado quando ondas sonoras são refletidas em superfícies, o que cria uma sensação de espaço e continuidade.
Surround	Sistema de reprodução sonora que utiliza mais de um altifalante. O som é executado de várias direções e transmite ao ouvinte um efeito de envolvimento e espacialidade.

Anexos

QUESTIONÁRIO

1. Qual é a sua idade?
 - ☐ menos de 25
 - ☐ entre 25 e 29
 - ☐ entre 30 e 45
 - ☐ entre 46 e 64
 - ☐ mais de 65

2. Qual é a sua idade e a sua imparidade visual?
 - ☐ Baixa visão
 - ☐ Cegueira

3. Está familiarizado com audiodescrição de filmes e séries?
 - a. Se sim, costuma assistir obras com audiodescrição? Com que frequência?
 - ☐ Uma vez por semana
 - ☐ De duas a três vezes por semana
 - ☐ Uma vez por mês
 - ☐ Quase nunca
 - ☐ Outro _____

4. Qual é a sua opinião sobre audiodescrição? Acha que funciona bem ou gostaria que fosse diferente? Explique.

PERGUNTAS SOBRE INTELIGIBILIDADE

5. Qual era o principal objetivo da criatura na casa?
 - ☐ Destruir a casa
 - ☐ Presentear as crianças
 - ☐ Assustar as crianças
 - ☐ Roubar a casa

6. Por onde a criatura foi embora? (04:08)
 - ☐ Pela porta de entrada
 - ☐ Pela porta dos fundos
 - ☐ Pela lareira
 - ☐ Desapareceu magicamente

7. Como foi descrito o horário e o clima? (0:39)
 - ☐ Inverno e dia
 - ☐ Inverno e noite
 - ☐ Outono e dia
 - ☐ Outono e fim de tarde

8. Como era a criatura? (02:19)
 - ☐ Grande, peluda, com patas de cavalo
 - ☐ Grande, com dentes sujos, grandes garras e pés cor-de-rosa
 - ☐ Grande e alta, forte, peluda, com mãos e pés humanos
 - ☐ Grande, húmida, sem pelos, de pele azul

9. Qual é o nome das crianças? (01:02 - 01:05)
 - ☐ Lia e Vincent
 - ☐ Lara e Vini
 - ☐ Lara e Billie

- ☐ Lia e Billie

PERGUNTAS SOBRE A AUDIODESCRIÇÃO

10. Sobre a obra exibida. Acredita que compreendeu todos os diálogos e a narração da audiodescrição?
11. No curta-metragem escutado, o que você entendeu da história?
12. O que achou da AD binaural?
13. Qual foi o tipo de fone de ouvido utilizado?
 - ☐ Pequeno (in ear)
 - ☐ Grande (over-ear)
 - ☐ Não sei