

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE BELAS ARTES

Raphaela Benetello Marques

**Pontos de escuta e possibilidades sonoras no uso doméstico
de tecnologias de mixagem orientada a objetos**

Belo Horizonte
2025

Raphaela Benetello Marques

**Pontos de escuta e possibilidades sonoras no uso doméstico
de tecnologias de mixagem orientada a objetos**

Tese apresentada ao Programa de Pós-graduação em Artes da Escola de Belas Artes da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Doutora em Artes.

Área de concentração: Artes.

Linha de pesquisa: Cinema.

Orientador: Prof. Dr. Leonardo Alvares Vidigal.

Belo Horizonte

2025

Ficha catalográfica
(Biblioteca da Escola de Belas Artes da UFMG)

791.43
B465p
2025

Benetello, Raphaela, 1991-
Pontos de escuta e possibilidades sonoras no uso doméstico de
tecnologias de mixagem orientada a objetos [recurso eletrônico] /
Raphaela Benetello Marques. – 2025.
1 recurso online.

Orientador: Leonardo Alvares Vidigal.

Tese (doutorado) – Universidade Federal de Minas Gerais, Escola
de Belas Artes.
Inclui bibliografia.

1. Cinema – Teses. 2. Mixagem (Som) – Teses. 3. Som – Registro
e reprodução – Equipamento e acessórios – Teses. 4. Cinema –
Apreciação – Teses. 5. Som – Técnicas digitais – Teses. I. Vidigal,
Leonardo A., 1967- II. Universidade Federal de Minas Gerais. Escola
de Belas Artes. III. Título.



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Assinatura da Banca Examinadora na Defesa de Tese da aluna **RAPHAELA BENETELLO MARQUES** - Número de Registro **2021699360**.

Título: **“Pontos de escuta e possibilidades sonoras no uso doméstico de tecnologias de mixagem orientada a objetos”**

Prof. Dr. Leonardo Alvares Vidigal – Orientador – EBA/UFMG

Prof. Dr. Pedro Cardoso Aspahan – Titular – EBA/UFMG

Prof. Dr. Rodrigo Octavio D'Azevedo Carreiro – Titular – UFPE

Profa. Dra. Debora Regina Opolski – Titular – IFPR

Profa. Dra. Ana Lúcia Menezes de Andrade – Titular – EBA/UFMG

Belo Horizonte, 31 de outubro de 2025.



Documento assinado eletronicamente por **Leonardo Alvares Vidigal, Professor do Magistério Superior**, em 07/11/2025, às 15:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Pedro Cardoso Aspahan, Professor do Magistério Superior**, em 10/11/2025, às 18:18, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rodrigo Octavio D'Azevedo Carreiro, Usuário Externo**, em 10/11/2025, às 19:46, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ana Lucia Menezes de Andrade, Professora do Magistério Superior**, em 11/11/2025, às 10:13, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Débora Regina Opolski, Usuária Externa**, em 12/11/2025, às 19:28, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rita Lages Rodrigues, Coordenador(a) de curso de pós-graduação**, em 13/11/2025, às 15:44, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4719992** e o código CRC **C831E4E3**.

Referência: Processo nº 23072.256445/2025-61

SEI nº 4719992

AGRADECIMENTOS

Embora a escrita e a pesquisa sejam, em grande medida, tarefas solitárias, não seria possível chegar à conclusão desta tese sem o apoio, a presença e a confiança de tantas pessoas que, de diferentes formas, caminharam comigo. Este trabalho é fruto de gestos, palavras e presenças que me sustentaram ao longo do percurso.

A Deus, que me deu força e serenidade nos momentos de dúvida.

À minha mãe, pelo amor incondicional e pelo exemplo de coragem que me guia.

Ao meu marido, Diego, companheiro de vida e de trajetória acadêmica, que dividiu comigo as dores e as alegrias desta pesquisa, enquanto ele próprio se dedicava à sua dissertação em Artes. Sua escuta, incentivo e partilha cotidiana, além de todo o apoio técnico na criação de slides e ajuda na formatação, transformaram este processo em uma experiência de cumplicidade e afeto.

Ao meu orientador, Prof. Dr. Leonardo Alvares Vidigal, pela leitura atenta, contribuições valiosas e compreensão ao longo desta jornada.

Agradeço também, enquanto aluna, ao Programa de Pós-Graduação em Artes da UFMG e à CAPES, pelo apoio que tornou possível ampliar horizontes em congressos e encontros acadêmicos e aos servidores técnico-administrativos e estagiários(as) da Secretaria da Pós-Graduação, cuja dedicação sustenta o andamento de tantas pesquisas, inclusive a minha.

Enquanto também servidora técnico-administrativa, agradeço novamente à UFMG que me permitiu o afastamento necessário para a dedicação a esta pesquisa e à Faculdade de Direito, pelo apoio e compreensão durante este processo.

Ao amigo e colega de trabalho, Victor Marcelino, pela generosidade em assumir responsabilidades em minha ausência, permitindo que eu me dedicasse com maior tranquilidade a este trabalho.

À criaturinha de quatro patas Toalha, que chegou em meio ao turbilhão da escrita e, com sua presença caótica, ajudou a distrair a mente e acalmar o coração.

E a todos que, de formas diversas, me apoiaram, acreditaram neste caminho e me ofereceram palavras de incentivo. Este trabalho também é resultado do carinho, da confiança e da parceria de cada um de vocês. Se esta tese é também uma escuta, ela carrega em si o som de todas as vozes que me acompanharam, mesmo quando a minha parecia vacilar.

RESUMO

Este trabalho analisa os sistemas e as tecnologias de mixagem atuais que seguem o modelo de áudio orientado a objetos sonoros, como o Dolby Atmos, o DTS:X, o Auro-3D e o MPEG-H, a partir da perspectiva de utilização em ambiente doméstico, principalmente através do uso de fones de ouvido. Partindo de uma análise acerca da representação sonora no cinema desde suas primeiras obras sincronizadas, a pesquisa objetiva abordar a dicotomia entre fidelidade e inteligibilidade e analisar a noção do ponto de escuta a partir da visão de diversos autores, com suas propostas de classificação e categorias possíveis. Como proposta de classificação, foi conceituado o ponto de escuta imersivo para caracterizar momentos narrativos que aproximam o espectador não apenas da escuta do personagem, mas também de suas sensações e sentimentos, através do uso de som imersivo. Além disso, foi traçada uma breve história dos sistemas sonoros no cinema com o objetivo de entender como a evolução técnica e tecnológica aliada aos interesses econômicos desenvolveu narrativamente o produto audiovisual. O desenvolvimento do capítulo segue com a fundamentação sobre a mixagem orientada a objetos sonoros e seus sistemas mais populares atualmente no mercado audiovisual, com as possibilidades no uso em salas de cinema, mas, principalmente, a utilização em ambientes domésticos ou móveis, através da escuta binaural, por meio de fones de ouvido. No quarto capítulo, há entrevistas realizadas com profissionais da área que revelam tanto as potencialidades criativas das tecnologias imersivas quanto os desafios práticos de sua aplicação. Em seguida, há a análise dos filmes *Okja* (2017), *Nada de Novo no Front* (2022) e *A Sociedade da Neve* (2023) que evidenciam como a mixagem orientada a objetos se concretiza na narrativa audiovisual, em uma comparação entre as versões em Dolby Atmos e em estéreo, ressaltando as variações perceptivas na escuta. Por fim, é possível demonstrar através da pesquisa como o áudio imersivo, mesmo em ambientes domésticos, pode transformar a experiência audiovisual contemporânea, principalmente aproximando o espectador da história, inserindo-o sonoramente dentro do universo diegético.

Palavras-chave: cinema; som; ponto de escuta; objetos sonoros; Dolby Atmos; DTS:X; Auro-3D; MPEG-H; ponto de escuta imersivo; *Okja*; *Nada de Novo no Front*; *A Sociedade da Neve*.

ABSTRACT

The work analyzes current mixing systems and technologies that follow the model of object-based audio, such as Dolby Atmos, DTS:X, Auro-3D, and MPEG-H, from the perspective of their use in domestic environments, particularly through headphones. Beginning with an examination of sound representation in cinema since its earliest synchronized works, the research aims to address the dichotomy between fidelity and intelligibility and to analyze the notion of the listening point based on the perspectives of various authors, with their proposed classifications and possible categories. As a proposed classification, the concept of the immersive point of audition was defined to characterize narrative moments that bring the viewer closer not only to the character's listening perspective, but also to their sensations and emotions, through the use of immersive sound. In addition, a brief history of sound systems in cinema is outlined in order to understand how technical and technological evolution, combined with economic interests, has narratively shaped the audiovisual product. The chapter proceeds with a theoretical grounding on object-based mixing and its most popular systems currently available in the audiovisual market, considering their possibilities in cinema auditoriums but especially their use in domestic or mobile contexts, through binaural listening via headphones. The fourth chapter presents interviews with professionals in the field that reveal both the creative potential of immersive technologies and the practical challenges of their application. This is followed by an analysis of the films *Okja* (2017), *All Quiet on the Western Front* (2022), and *Society of the Snow* (2023), which highlight how object-based mixing is realized within audiovisual narrative, in a comparison between Dolby Atmos and stereo versions, emphasizing perceptual variations in listening. Finally, the research demonstrates how immersive audio, even in domestic environments, can transform the contemporary audiovisual experience, particularly by bringing the viewer closer to the story and sonically placing them within the diegetic universe.

Keywords: cinema; sound; point of audition; object-based audio; Dolby Atmos; DTS:X; Auro-3D; MPEG-H; immersive point of audition; *Okja*; *All Quiet on Western Front*; *Society of the Snow*.

LISTA DE ILUSTRAÇÕES

Figura 01 - Localização dos sistemas sonoros na película.....	62
Figura 02 - Produção e consumo do áudio baseado em objetos.....	66
Figura 03 - Microfone binaural.....	70
Figura 04 - Conceito três camadas do Auro-3D.....	73
Figura 05 - Configurações do Auro 11.1 e Auro 13.1.....	74
Figura 06 - Modelo AuroMax 26.1.....	75
Figura 07 - Modelo AuroMax 20.1.....	76
Figura 08 - Visão geral Auro-3D Engine.....	78
Figura 09 - Configuração dos alto-falantes em soundbars.....	78
Figura 10 - Virtualização sonora em soundbar 5.1.2 canais.....	79
Figura 11 - Configuração alto-falante smart.....	79
Figura 12- Configuração Dolby Atmos para salas de cinema.....	83
Figura 13 - Interface Dolby Atmos Renderer.....	86
Figura 14 - Diagrama de fluxo Dolby Renderer.....	87
Figura 15 - Esquema 5.1.2 canais.....	88
Figura 16 - Caixas integradas e módulos complementares habilitados para Atmos.....	89
Figura 17 - Sistema de reflexão das caixas integradas.....	90
Figura 18 - Comparativo entre soundbar estéreo e soundbar com Dolby Atmos.....	91
Figura 19 - Interface mixagem binaural Dolby Atmos e Plug-in Dolby Atmos Binaural Settings.....	93

SUMÁRIO

1 INTRODUÇÃO.....	8
2 REPRESENTAÇÃO SONORA E PONTOS DE ESCUTA.....	13
2.1 Fidelidade e inteligibilidade.....	13
2.2 Tipos de escuta e objetos sonoros.....	23
2.3.1 Michel Chion e os sentidos subjetivo e espacial.....	28
2.3.2 Outras classificações.....	33
2.3.3 Ponto de escuta imersivo.....	41
3 SOM NO CINEMA.....	44
3.1 Breve história dos sistemas sonoros.....	44
3.2 Os sistemas multicanais.....	58
3.3 Mixagem orientada a objetos.....	64
3.3.1 Auro-3D.....	72
3.3.1.1 Auro 11.1.....	72
3.3.1.2 AuroMax.....	73
3.3.1.3 Auro-3D Engine.....	77
3.3.2 DTS:X.....	80
3.3.3 Dolby Atmos.....	81
3.3.4 MPEG-H e a TV 3.0 brasileira.....	95
4 PRÁTICAS E PERSPECTIVAS TECNOLÓGICAS.....	98
4.1 Transição e adaptação.....	100
4.2 Cinema e ambiente doméstico.....	103
4.3 Mixagem binaural.....	109
4.4 MPEG-H e o Fórum SBTVD.....	113
4.5 Perspectivas futuras.....	119
5 EXPERIÊNCIAS SONORAS NO CONTEXTO DOMÉSTICO.....	126
5.1 Okja (2017).....	130
5.1.1 Mikha e Okja na montanha.....	132
5.1.2 O resgate de Okja no centro urbano de Seul.....	134

5.1.3 Laboratórios de testes Mirando.....	136
5.1.4 Mikha invade as indústrias Mirando.....	137
5.1.5 Considerações.....	139
5.2 Nada de Novo no Front (2022).....	140
5.2.1 Abertura.....	142
5.2.2 Chegada dos soldados à trincheira.....	144
5.2.3 Tanque avança sobre a trincheira.....	148
5.2.4 Embate entre Paul Bäumer e soldado francês.....	150
5.2.5 Considerações.....	151
5.3 A Sociedade da Neve (2023).....	152
5.3.1 O acidente nos Andes.....	156
5.3.2 Avalanche.....	159
5.3.3 Travessia final.....	162
5.3.4 Considerações.....	165
6 CONSIDERAÇÕES FINAIS.....	167
REFERÊNCIAS.....	176
GLOSSÁRIO.....	182
ANEXO A - ENTREVISTA ARIEL FERRANTE.....	186
ANEXO B - ENTREVISTA CARLOS KLACHQUIN.....	210

1 INTRODUÇÃO

Esta pesquisa parte de uma curiosidade em relação ao desenvolvimento do som no audiovisual e suas possibilidades de utilização em diversas plataformas. A experiência do cinema, discutida e analisada amplamente por teóricos e estudiosos da área, precisou ser pausada em 2020 com a pandemia de Covid-19 que afligiu todo o mundo e devastou famílias, seja por mortes repentinas ou por conflitos entre a ciência e outras convicções distorcidas. De certo é que não houve um ponto de retorno completo, mesmo após vacinas e uma aparente volta à normalidade. O Informe Anual do Mercado Cinematográfico de 2024, elaborado pelo Observatório Brasileiro do Cinema e do Audiovisual (OCA) e pela Agência Nacional do Cinema (Ancine), revela que o Brasil superou o número de salas de cinema de 2019, contando agora com 3.532, um crescimento ainda tímido se comparado às 3.478 de 2019, ano anterior ao momento crítico pandêmico, ainda que haja uma considerável recuperação se comparado à queda de 2020, com apenas 1.860 salas de exibição. O público frequentador também não se recuperou completamente, com 177 milhões em 2019 e 125 milhões em 2024, ainda que esteja em ascensão, se comparado aos anos anteriores imediatos. O preço dos ingressos e o imediatismo de lançamento em plataformas de *streaming* podem ser indícios para uma investigação mais aprofundada sobre esses dados, o que não é o objetivo deste trabalho.

Ainda de acordo com o relatório, há dados sobre a relação do número de habitantes por sala de exibição, acessibilidade, renda e participação de títulos nacionais, maior público de títulos no geral e títulos nacionais, porcentagem de cinema em shoppings ou complexos e cinemas de rua e diversos outros parâmetros sobre as salas de exibição cinematográfica no Brasil. A ainda insuficiente regulamentação de plataformas de difusão *online* de materiais audiovisuais não dá conta de expor dados sobre *streamings*, como números de assinantes ou maiores exibições. Assim como também não há dados sobre o som das salas, quais formatos possuem, seus sistemas de sonorização, suas atualizações e a evolução ao longo dos anos. Em uma breve busca através de sites de vendas de ingressos sobre os sistemas sonoros disponíveis, é notório no país o reduzido número de salas de exibição com tecnologias de sonorização baseadas em objetos, como Dolby Atmos, DTS:X ou Auro-3D. Em Minas Gerais, por exemplo, há apenas duas salas Atmos, nas cidades de Montes Claros e Uberaba, ou seja, nenhuma na capital Belo Horizonte. Em 2023, havia apenas cerca de 7.800 salas habilitadas com Dolby Atmos em todo o mundo e o alto custo de implementação, podendo chegar a 100 mil dólares, pode explicar em parte a dificuldade de ampla difusão do sistema em salas exibidoras.

Nesta era de hibridização dos meios, o cinema como meio de comunicação é, de fato, chamado a compartilhar as mesmas telas e a se alimentar nas mesmas plataformas que outros meios de comunicação que, há pouco tempo, eram-lhe estranhos (a televisão, por exemplo), ou simplesmente não existiam (internet, por exemplo) (Gaudreault; Marion, 2016, p. 14).

O elevado valor de instalação desses sistemas e o reconhecimento de que as formas de consumo audiovisual têm mudado ao longo das décadas, impulsionada também pela Pandemia, fez com que os sistemas sonoros baseados em objetos fossem implementados também em dispositivos domésticos e móveis. Uma das tentativas de popularização da ideia de “som de cinema em casa”, que sempre atraiu entusiastas do som e consumidores especializados, principalmente com equipamentos como *home theaters*. Além disso, produções audiovisuais lançadas exclusivamente para *streamings* saem com o chamado áudio ‘espacial’, mixado nas tecnologias disponíveis e, com isso, ampliam também as possibilidades de se consumir e também pesquisar a tecnologia fora das salas de cinema. André Gaudreault e Philippe Marion (2016, p. 15) pontuam sobre os momentos de questionamento acerca da sobrevivência do cinema e sua identidade como meio de comunicação:

O que, durante mais de um século, chamamos cinema, sofreu, ao longo de sua história, uma série de mutações tecnológicas sucessivas. Seja no momento da passagem para o cinema falado, seja no momento da introdução dos formatos largos - para citar apenas dois exemplos -, cada tecnologia nova alterou, à sua maneira, de forma progressiva e durável, as práticas de produção e difusão das obras, assim como sua recepção pelos espectadores. Hoje em dia, essa outra mutação tecnológica, chamada comumente de “revolução digital”, provoca, por sua vez, mudanças gigantescas na instituição cinematográfica”.

Kerins (2010) questiona, em sua pesquisa sobre o cinema multicanal e os sistemas Dolby, as diferenças entre a trilha original apresentada em cinema e o produto sonoro incluído em DVDs do filme, explicando que muitas produções são remixadas para o lançamento doméstico. Nestes casos, pode haver diferenças entre a versão para cinema e a assistida em casa, mesmo que, em sua maioria, sejam de pouca importância, visando apenas o balanceamento das diferenças espaciais entre os ambientes exibidores e os domiciliares.

O fato de estúdios e cineastas gastarem tempo, dinheiro e energia para remixar trilhas sonoras de filmes para que correspondam às características espaciais do mercado doméstico é evidência suficiente de sua crença na importância do som *surround* para a experiência cinematográfica moderna¹ (Kerins, 2010, p. 14).

¹ Tradução livre da autora do texto original: “That studios and filmmakers expend the time, money, and energy to remix film soundtracks to match the spatial characteristics of the home market is evidence enough of their belief in surround’s importance to the modern motion picture experience”.

Se, ainda no início dos anos 2000, o mercado doméstico já era de evidente consumo e importância, a disseminação de plataformas de *streaming* e o desenvolvimento de diversos dispositivos capazes de reproduzir som e imagem com qualidade até mesmo para consumo individual fizeram a relevância deste tipo de reprodução se amplificar e as possibilidades tecnológicas se desenvolverem também no mercado fora das salas de cinema. No artigo “*Home Theater(s): Technology, Culture, and Style*”, Kerins (2017) reforça essa afirmação, mencionando que, mesmo com o aumento nos estudos do som e suas tecnologias em salas de cinema, os sistemas domésticos para o som atraíram pouca ou nenhuma atenção dos estudos de cinema e mídia, mesmo que os filmes sejam cada vez mais exibidos em ambientes domiciliares e dispositivos móveis. Portanto, esta pesquisa tenta, de alguma forma, contribuir para o preenchimento da lacuna exposta pelo autor.

Se, antes, ao pensarmos em exibição doméstica, naturalmente considerávamos o meio televisivo e seus alto-falantes estéreo, hoje, podemos inferir uma diversidade de equipamentos e dispositivos capazes de reproduzir os materiais audiovisuais, como *tablets*, celulares, projetores domésticos, óculos de realidade virtual e mesmo televisores que melhoraram sua tecnologia imagética, com resoluções de até 4K e ampliaram suas possibilidades sonoras, com o uso de *home theaters*, *soundbars* e, mais recentemente, até mesmo com alto-falantes sem fio conectados ao aparelho via *wi-fi* ou *bluetooth*.

De fato, a projeção na telona não passa, agora, de uma modalidade de consumo das imagens, entre outras, cuja aura seja talvez ainda a maior, mas uma modalidade mesmo assim. A ‘revolução’ digital já não bateria à nossa porta, teria, sim, invadido nossas casas. Para alguns, até, essa ‘revolução’ seria praticamente coisa do passado (Gaudreault; Marion, 2016, p. 21).

Para dar seguimento à discussão das possibilidades de apreciação do som no contexto doméstico, o segundo capítulo deste trabalho visa expor e entender questões relativas à representação sonora e as escolhas entre fidelidade e inteligibilidade por parte dos técnicos de som, da indústria cinematográfica e do público espectador ainda no início da sonorização sincrônica do cinema, entre o final da década de 1920 e meados da de 1930. Essas discussões e tentativas moldaram a forma de assistir a filmes até hoje e refletem o modo como o audioespectador (Chion, 2008) espera encontrar o som nas obras audiovisuais, tanto em salas de cinema como em ambiente doméstico, pensando também em telenovelas, produtos jornalísticos e esportivos, que absorveram os frutos dessa discussão e aplicaram também em seus produtos. O diálogo, por exemplo, tradicionalmente localizado nos canais frontais, segue com pouca experimentação espacial, mesmo que o problema tecnológico tenha sido superado.

A necessidade de manter o público atento e confortável em um lugar comum mantém sua localização tradicional e abre pouco espaço para uma mudança nos padrões estéticos da mixagem no audiovisual *mainstream*. Além disso, o capítulo também propõe examinar a conceituação do termo “ponto de escuta” ao longo do tempo e através de diversos autores que se debruçaram sobre a temática para elucidar suas características e formas de utilização.

O terceiro capítulo apresenta uma breve história dos sistemas sonoros no cinema e seus reflexos nos dispositivos de uso domiciliar, pontuando esta hibridização midiática do cinema com outros meios de comunicação e arte, seguindo o desenvolvimento da arte cinematográfica e sua característica aglutinadora. O objetivo é a compreensão da evolução técnica e tecnológica ao longo das décadas que, aliada aos interesses econômicos e à aceitação ou rejeição do público, pode se desenvolver e se consolidar como forma de consumo e entretenimento audiovisual, não apenas em salas exibidoras, mas em todo o desenvolvimento audiovisual posterior, com aparelhos televisores, caixas para ambiente doméstico, como *home theaters* e *soundbars*, e dispositivos móveis. Para o entendimento das mais recentes tecnologias e sistemas sonoros no mercado doméstico, abordamos também as diferenças entre a mixagem baseada em canais e a mixagem baseada em objetos sonoros, uma considerável mudança na forma de se produzir, mixar e consumir os sons. O desenvolvimento do capítulo segue com a apresentação dos sistemas de áudio com canais e objetos sonoros mais consumidos atualmente no mercado audiovisual e as possibilidades de utilização em salas de exibição e, principalmente, em ambientes domésticos e em dispositivos móveis, por meio da escuta binaural, através de fones de ouvido.

O quarto capítulo é dedicado a entrevistas realizadas com profissionais envolvidos no desenvolvimento e na aplicação das tecnologias sonoras orientadas a objetos. O *sound designer*, Ariel Ferrante, e o representante Dolby na América Latina, Carlos Klachquin, forneceram um panorama sobre as adaptações e desafios enfrentados no mercado audiovisual acerca das tecnologias sonoras imersivas, tanto no âmbito doméstico, quanto cinematográfico. Os depoimentos revelam tanto as potencialidades criativas dessas ferramentas, capazes de ampliar a imersão e a personalização da experiência sonora, quanto as limitações impostas por questões técnicas, estruturais e de padronização do mercado. As falas destacam também a ausência de convenções estabelecidas para mixagens em formatos como o Dolby Atmos, o que reforça o caráter ainda imaturo e subjetivo da utilização da tecnologia atualmente. Nesse sentido, as entrevistas funcionam como uma base empírica para compreender como a tecnologia se insere no cotidiano de estúdios e no mercado audiovisual, mas também para

problematizar suas possibilidades no ambiente doméstico, marcado por grande diversidade de dispositivos de reprodução.

A análise fílmica, por sua vez, permitiu observar como essas tecnologias se materializam na narrativa audiovisual e de que forma afetam a escuta do espectador. Os estudos de caso de *Okja* (2017), *Nada de Novo no Front* (2022) e *A Sociedade da Neve* (2023) evidenciam diferentes usos da espacialização sonora, ora reforçando a imersão dramática por meio da verticalidade e da densidade do som, ora criando contrastes significativos entre silêncio e som, que tensionam a experiência narrativa e artística de se audiover um filme, mesmo fora das salas de cinema. As análises contemplam uma comparação entre a escuta em Dolby Atmos e em estéreo, objetivando a compreensão das variações na percepção sonora e revelando como a tecnologia imersiva pode aproximar o espectador do ponto de escuta dos personagens, ao mesmo tempo em que o estéreo, mesmo que de forma limitada em relação à espacialização, ainda preserve a inteligibilidade narrativa. A escolha do estéreo como formato base se dá pelo fato de a escuta ser realizada através de fones de ouvido, seja com o modo binaural espacializado, seja entre o lado direito e esquerdo, estereofônico.

Ao final do trabalho, após as referências, há um glossário de termos técnicos utilizados ao longo do texto. Os termos foram prioritariamente descritos em português, entretanto, caso não haja um equivalente usual, foi utilizada a versão em inglês ou no idioma em que são mais reconhecidos. Após o glossário, há dois anexos com as entrevistas transcritas dos profissionais que ajudaram a entender o mercado, a tecnologia e os formatos que compõem o quarto capítulo deste trabalho.

2 REPRESENTAÇÃO SONORA E PONTOS DE ESCUTA

Este capítulo apresenta um referencial teórico para a composição desta tese, pensando primeiramente na representação sonora e suas possibilidades diante da dicotomia entre fidelidade e inteligibilidade. Além disso, foi realizado um levantamento bibliográfico das caracterizações do som do ponto de escuta e seus sentidos, suas propostas de classificação mais recentes e outras categorizações para atribuir significado à localização do som perante os personagens e os espectadores.

2.1 Fidelidade e inteligibilidade

Ainda que atualmente o audiovisual seja reconhecido como linguagem que abrange imagem e som como elementos estruturantes, em seus primeiros anos, o cinema foi permeado por debates acerca da melhor representação sonora que a imagem deveria “transmitir”. Na chamada Era de Transição (1926 - 1934), discussões relativas ao uso do som como mero complemento à imagem e suas possibilidades narrativas permeavam a indústria cinematográfica. Inicialmente, a mudança dos acompanhamentos externos, como pianistas e narradores, para sons pensados e executados em sincronia com a imagem, foi uma transição mais tecnológica do que semântica.

No livro *Sound Technology and the American Cinema*, James Lastra (2000, p. 110) questiona a ideia inicial do uso sonoro no cinema como simples acompanhamento da imagem e tenta abordar o conceito de sincronização de maneira mais abrangente do que apenas a composição de sons relativos às fontes visualizadas na imagem: “Mesmo parecendo cumprir uma função específica (como produzir um diálogo inteligível e narrativamente importante), o som pode simultaneamente desempenhar outras funções, não narrativas ou mesmo não representacionais”². Outra função que Lastra (2000, p. 106) sinaliza como um “capital simbólico” do som é o realismo. A ideia do realismo sonoro conectou, segundo o autor, a performance sonora e a integração narrativa, causando a impressão de que o som emana da tela, ratificando a ideia do realismo “tal como no mundo fora do cinema” (p. 108).

Para além do uso vocálico do som no cinema, os efeitos sonoros, nessa Era de Transição, começaram a ser utilizados a fim de unificar o conjunto de cenas e sequências em uma unidade narrativa. Em *A Woman of Affairs* (1928), filme sincronizado em *Vitaphone*, um

² Tradução livre da autora. Texto original: “Even while appearing to fulfill a particular function (such as providing intelligible, narratively important dialogue), sound may simultaneously be performing other, non narrative or even nonrepresentational ones”.

dos primeiros sistemas de sucesso para reprodução de som e imagem sincrônico³, alguns personagens reagem ao som de uma buzina não visualizada, experienciando o som diegético fora de quadro. Essa forma de utilização já vinha sendo utilizada pelos filmes sem som sincronizado há anos, com músicos se apresentando ao vivo durante a exibição das imagens e, muitas vezes, sendo encarregados também de produzir efeitos sonoros ao vivo em momentos-chave da narrativa, como trovões ou uma batida fatal na porta (Buhler; Neumeyer; Deemer, 2010). Ainda que de forma pouco coordenada entre os exibidores, era sabido que, mesmo como um acompanhamento variável, o som melhorava a representação narrativa, legitimando o que a imagem sugeria.

Uma das discussões que permeavam a produção cinematográfica nos anos 1930 e seguem dividindo teorias é a relação entre original e cópia, ou melhor, entre as diferenças de um som original e o som gravado e reproduzido. Lastra (2000) divide alguns teóricos entre o grupo que vê pouca ou nenhuma diferença entre o som original e o gravado (Béla Balázs, Stanley Cavell, Jean-Louis Baudry, Christian Metz e Gerald Mast) e o “grupo de oposição” (p. 124), composto por Alan Williams, Rick Altman e Thomas Levin, que defendem que as gravações são uma correspondência parcial com o evento original, presumindo que há uma “não-identidade entre o original e a cópia” (p. 125).

Mais do que esmiuçar o que cada teórico defende sobre originais e gravações, é importante destacar alguns pontos: Alan Williams (1980, p. 53), no artigo “*Is Sound Recording Like a Language?*” define o som como um “evento material tridimensional”, isto é, há uma dependência do som com o ambiente em que ele é executado. “A gravação não pode, por definição, reproduzir os próprios sons, uma vez que é obrigado por sua natureza a reproduzir um som (como volume vibratório) gravado a partir de um ponto do espaço através do qual o som existe”⁴ (p. 53). Ainda segundo o autor, os microfones funcionam como ouvidos, ou seja, transmitem o que é possível captar daquele ponto no espaço. O uso de vários microfones e a colocação de diversos alto-falantes durante a reprodução reduz essa especificidade, “mas a ‘amostragem’ permanecerá - assim como na gravação de imagens pode-se aumentar o número de quadros por segundo para 30, 60 ou mais, mas ainda tem apenas uma amostra viável do movimento a ser representado, e não o movimento ‘em si’”⁵ (p.

³ A mecânica de funcionamento do sistema *Vitaphone* será abordada com maior minúcia no próximo capítulo.

⁴ Tradução livre da autora. Texto original: “Recording can not by definition reproduce “the sounds themselves” since it is obligated by its nature to render a sound (as vibrating volume) as recorded from one point of the space in which and through which the sound exists”.

⁵ Tradução livre da autora. Texto original: “The fact of sampling will remain-just as in image recording one may increase the number of frames per second to 30, 60, or more, but still have only a workable sample of the motion to be represented, and not the motion ‘in itself’.”

53). A discussão relativa entre os sons gravados e os sons ‘vividos’ de maneira ao vivo esbarra também no campo da Etnomusicologia, ao pensarmos nas possibilidades sensoriais que uma performance ao vivo pode causar em paralelo a sua mera reprodução sonora que tecnicamente pode ser semelhante, mas sem o sublime estético e cultural. “O ouvinte o apreende apenas como um produto acabado e não como um processo, social e cultural, que se desenvolve, na verdade, como um intertexto, a realidade sonora funcionando apenas como uma abstração analítica a posteriori” (Carvalho, 1999, p. 57).

Outro apontamento interessante é a relação de significado entre os sons identificáveis. Mais do que uma igualdade sonora entre sons originais e suas reproduções, é essencial que os sons tenham identificação com seus significados e, portanto, sejam interpretados da maneira com que foram pensados. O som de uma batida entre dois carros precisa ter uma correspondência sonora com o som original, mesmo que acrescido de efeitos e reverberações. Já o som de um acidente entre um trem e um caminhão precisa apresentar sonoridades diferentes, com verossimilhança com a realidade, mas não necessariamente igual a uma colisão real. No artigo “*Aural Objects*”, Christian Metz (1980, p. 30), defensor da não diferenciação do som original e sua representação, identifica em sua argumentação a tendência de se caracterizar o som mais como atributo do que um objeto próprio cinematográfico, negligenciando “suas próprias características em favor daquelas de sua ‘substância’ correspondente, no caso o objeto visível, que emitiu o som”⁶. Desse modo, o autor privilegia o significado e sua interpretação para além das características da reprodução sonora. Na medida em que dois sons podem ser reconhecidos como provenientes da mesma fonte e desempenhando a mesma função, eles se tornam intercambiáveis. “Parece óbvio insistir em um modelo único de audição apenas se pensarmos simultaneamente a representação como audição e se definirmos a tecnologia como uma orelha substituta”⁷ (Lastra, 2000, p. 132).

As tecnologias empregadas na gravação e reprodução sonoras para cinema preocupam-se menos com o evento ‘original’ daquela sonoridade e mais com sua utilização no contexto filmico; essa equiparação entre a gravação e a audição constrói o que Lastra (2000, p. 133) denomina como “tecnologia em si”⁸, que ocupa o lugar do espectador no evento original e funciona como um substituto com o qual é possível criar uma identificação.

⁶ Tradução livre da autora. Texto original: “Its own characteristics in favor of those of its corresponding ‘substance’, which in this case is the visible object, which has emitted the sound”.

⁷ Tradução livre da autora. Texto original: “It seems obvious to insist on a single model of hearing only if we simultaneously theorize representation as hearing and if we define the technology as a surrogate ear”.

⁸ Tradução livre pela autora do texto original: “technology itself”.

Entretanto, “originais” destinados à reprodução (digamos, os sons em um *set* de filmagem) podem ter apenas uma ligeira semelhança com a representação final. Esta insistência teórica no original autônomo privilegia, portanto, desnecessariamente a presença em detrimento da mediação, em vez de simplesmente ilustrar a transformação⁹.

A utilização de dispositivos tecnológicos como mediadores sonoros entre os sons originais do ambiente e aquilo que é desejável na composição da narrativa fílmica desvendam a ideia de que, pensando cinematograficamente, o processo até então chamado de reprodução sonora deve ser entendido mais como representação sonora, que envolve desde a trilha musical, as ambiências, os silêncios e os sons vocálicos, principalmente diálogos, historicamente privilegiados diante de outras sonoridades. “A representação tecnológica é um processo que envolve desde a encenação pró-fonográfica de um som até a edição final, mixagem, gravação e reprodução”¹⁰ (Lastra, 2000, p. 136). Retornamos dessa forma às ideias de Metz (1980), em que o significado representado pelo som se torna mais importante que sua sonoridade original. Além disso, todo esse processo representativo no audiovisual acontece através da mediação de caixas de som ou alto-falantes, ou seja, através de tecnologias.

Essa questão fundamental entre o original e a reprodução norteia uma outra polêmica que envolve o cinema clássico hollywoodiano, a da hierarquia sonora, fidelidade *versus* inteligibilidade. Inicialmente, os filmes em *Vitaphone* seguiam um modelo subordinado à voz, em que os sons evidenciassem sua origem, identidade e compreensibilidade. “A tarefa essencial era produzir o efeito de sons apropriados e reconhecíveis emanados dos objetos na tela, como se os sons fossem, como a voz, uma propriedade do objeto representado”¹¹ (Lastra, 2000, p. 138). A maior demanda era e continua sendo pela clareza narrativa, principalmente pois quem mantém as receitas gosta e prefere filmes de simples entendimento. Desde as primeiras exhibições silenciosas até às multiplicidades sonoras encontradas atualmente, quem define o que permanece e o que é descartado dentro do audiovisual *mainstream* é o consumidor/espectador. Racionalmente, esses padrões de consumo se impulsionam também

⁹ Tradução livre da autora. Texto original: “However, ‘originals’ destined for reproduction (say, the sounds on a movie set) may have only the slightest resemblance to the finished representation. This theoretical insistence on the autonomous original, therefore, needlessly privileges presence over mediation rather than simply illustrating transformation”.

¹⁰ Tradução livre da autora. Texto original: “Technological representation is a process that involves everything from the pro-phonographic staging of a sound to the final editing, mixing, rerecording, and playback”.

¹¹ Tradução livre da autora. Texto original: “The essential task was to produce the effect of appropriate and recognizable sounds emanating from the objects on the screen, as if the sounds were, like the voice, a property of the depicted object”.

através do mercado como um todo, grandes distribuidoras e o investimento em marketing, cópias e acordos com grandes redes de exibição.

Dois modelos gerais dominaram e, de certa forma, ainda dominam as escolhas e análises das sonoridades, fidelidade e inteligibilidade (Altman, 1992; Lastra, 1992). A fidelidade privilegiaria a originalidade, buscando combinar o som em escala e perspectiva da imagem, isto é, em um *close-up* os sons vocálicos deveriam ser mais evidentes do que em um plano geral. “Todos os aspectos do evento sonoro são significativos, incluindo tempos de reverberação longos ou curtos, proporções de impacto direto e refletido, ou mesmo certas peculiaridades de performance ou espaço”¹² (Lastra, 2000, p. 139). A inteligibilidade, em contrapartida, visaria captar o som direto e não reverberante, criando uma classificação de importância entre os sons captados, podendo gerar uma representação sonora do espaço fílmico, privilegiando os elementos desejados, como os diálogos, “pressupõe que o som possui uma hierarquia intrínseca que torna aspectos essenciais e outros não”¹³ (p. 139).

Lastra (2000) define a fidelidade como uma abordagem sonora fonográfica e a inteligibilidade como telefônica, sendo a primeira a de uma reprodução fiel dentro de um espaço e tempo específico, e a segunda priorizando a legibilidade em detrimento da especificidade material. “Uma gravação ‘fonográfica’ poderia facilmente representar um dos ‘piores’ assentos [de uma orquestra], permanecendo ainda ‘fiel’ ou específica. O telefone, por outro lado, sacrifica a especificidade acústica em favor da renderização da fala sob condições variadas”¹⁴ (p. 139).

A gravação de diálogos, durante a Era de Transição para o sonoro, tendenciou ao modelo de inteligibilidade, reduzindo a reverberação e o ruído de fundo e priorizando a clareza na gravação e nitidez, seguindo a ideia da prioridade narrativa. Mesmo com tentativas de subversão, a inteligibilidade para os sons vocálicos permanece até hoje como o modelo consolidado. Lastra (2000, p. 141) difere esta forma dominante de registro sonoro hierárquico da noção desenvolvida por Rick Altman (1992) do som POA (ponto de escuta), que tende ao modelo de fidelidade sonora:

¹² Tradução livre da autora. Texto original: “All aspects of the sound event are inherently significant, including long or short reverberation times, ratios of direct to reflected sound, or even certain peculiarities of performance or space”.

¹³ Tradução livre da autora. Texto original: “Assumes that sound possesses an intrinsic hierarchy that renders some aspects essential and others not”.

¹⁴ Tradução livre da autora. Texto original: “A ‘phonographic’ recording could just as easily represent one of the ‘worst’ seats, while still remaining ‘faithful’ or specific. A telephone, in contrast, sacrifices acoustic specificity in favor of rendering speech clearly under widely varying conditions”.

O som POA (assim como a tomada POV) tenta representar a experiência de audição dentro da diegese, normalmente a audição de um personagem, às vezes isso é indicado por sons abafados, como se ouvisse através de uma parede, incluindo um efeito Doppler para indicar a passagem rápida de uma fonte sonora (por um aumento ou diminuição do volume indicando a aproximação ou retirada da fonte), ou talvez dando uma noção clara da acústica específica de um local¹⁵.

Lastra (2000) coloca que tanto o som POA, quanto o plano POV (*point of view*) representam espaços diegéticos dentro da narrativa e, por isso, a escuta dessas sonoridades deve ser entendida como se o espectador estivesse dentro deste espaço sonoro o qual a história acontece.

Rick Altman (1992), no livro *Sound Theory, Sound Practice*, explica que uma das grandes preocupações dos técnicos de som em Hollywood, entre os anos de 1920 e 1930, era em relação à interação entre som e imagem no espaço filmico. Essa questão levava a desdobramentos como o tipo de microfone a ser utilizado, o local de sua instalação, como deveria ser a movimentação em relação às imagens, níveis de intensidade e reverberação, entre diversas outras questões de ordem técnica que visavam construir a relação narrativa entre som e imagem. “Na verdade, como reconheceu o mais perspicaz dos técnicos de som desde o início, a questão da escala sonora vem ao primeiro plano para um inesperado problema de identificação do público, de prazer do espectador e de posicionamento temático”¹⁶ (Altman, 1992, p. 46).

O autor elenca três abordagens principais entre as tentativas de garantir uma localização sonora verossímil em relação à imagem: a instalação de alto-falantes e equipamentos de reprodução nos locais de exibição; a manipulação durante a produção, com escolha de microfones, seus posicionamentos e o controle de níveis e efeitos durante a edição; aperfeiçoamento das tecnologias sonoras, além do progresso dos locais de produção e reprodução. Inicialmente, nas salas de exibição, os alto-falantes se localizavam no fosso apontados para cima onde antes ficava a orquestra, sendo destinado aos sons musicais. Já outras caixas acima da tela, voltadas ao público, recebiam os diálogos. O projecionista era também quem fazia essa troca manual durante a exibição, não havia sons musicais simultâneos aos diálogos e o profissional precisava conhecer minuciosamente a trilha de

¹⁵ Tradução livre da autora. Texto original: “POA sound (like the POV shot) attempts to represent the experience of hearing within the diegesis, normally the hearing of a character. Sometimes this is indicated by muffling sounds, as if hearing them through a wall, by including a Doppler effect to indicate the rapid passage of a sound source (by an increase or decrease in volume indicating the approach or retreat of the source), or perhaps by giving a clear sense of the acoustics specific to a particular location”.

¹⁶ Tradução livre da autora. Texto original: “Indeed, as the most perceptive of sound technicians recognized from the start, the question of sound scale foregrounds to an unexpected extent problems of audience identification, of spectator pleasure, and of subject placement”.

áudio do filme para saber o momento oportuno das trocas entre os alto-falantes. De modo pouco surpreendente, essa mudança manual da saída de som não funcionava bem e precisava ser alterada.

O som multicanal já era discutido entre os técnicos, mas Hollywood e seus distribuidores-exibidores ainda não o consideravam prioridade. “Os expositores estavam bastante dispostos, no entanto, a dispensar a necessidade de mudança manual. Nem instruções melhores nem introduções de curta duração de troca automática através de pistas de controle poderiam conter a maré”¹⁷ (Altman, 1992, p. 48). Ainda no início da década de 1930, com a introdução da troca automática entre alto-falantes, os técnicos puderam se concentrar no desenvolvimento de ideias para a espacialização sonora e as formas como imagem e som poderiam se relacionar.

Altman expõe a ideia de diversos técnicos sobre como o som deveria se comportar em relação à imagem no universo, relações semelhantes às ideias de fidelidade e inteligibilidade abordadas anteriormente. A visão dominante veio do engenheiro de som Joseph P. Maxfield, responsável pela ala de distribuição da *Western Electric, Electric Research Products Incorporated (ERPI)*.

Numa série de artigos que reiteraram inúmeras vezes os mesmos pontos, Maxfield enfatizou a necessidade de limitar a maioria das tomadas a um único microfone (Maxfield 1929, 1930a, 1930b, 1931). Resolvendo assim o problema colocado por Cass, Maxfield mostrou como o uso de um único microfone, colocado próximo à linha de visão da câmera, coordena automaticamente a visão e o som, fornecendo um registro sonoro dos movimentos dos personagens em direção e longe da câmera. Os personagens que se aproximam da câmera também se aproximam automaticamente do microfone, combinando assim uma escala de imagem mais próxima com uma escala de som mais próxima; por outro lado, o personagem que fala ao fundo demonstra características distantes tanto no som quanto na imagem¹⁸ (1992, p. 50).

Assim como os outros engenheiros e técnicos de som da época, a ideia de Maxfield se aproxima ao movimento da fidelidade citado anteriormente por Lastra (2000) e ainda é vigente em dispositivos audiovisuais básicos, isto é, câmeras com microfone direcional

¹⁷ Tradução livre da autora. Texto original: “Exhibitors were quite ready, however, to dispense with the need for manual switching. Neither better instructions nor the short-lived introduction of automatic switching through control tracks could stem the tide”.

¹⁸ Tradução livre da autora. Texto original: “In a series of articles which reiterated numerous times the same points, Maxfield hammered home the need to limit most takes to a single microphone (Maxfield 1929, 1930a, 1930b, 1931). Thus solving the problem posed by Cass, Maxfield showed how the use of a single mike, placed near the camera's line of sight, automatically coordinates sight and sound by providing a sound record of the characters' movements toward and away from the camera. Characters approaching the camera automatically approach the microphone as well, thus matching closer image scale to closer sound scale; conversely, the character who speaks from the background demonstrates distant characteristics in sound and image alike”.

integrado, que captam o som ambiente do local enquanto seguem o enquadramento da imagem, inclusive sendo o padrão para dispositivos móveis, como celulares e *tablets*, pois garante uma correspondência básica entre escala de som e escala de imagem. Pensando em cinema e audiovisual profissional, a ideia de um microfone próximo à câmera, seguindo seus movimentos, apresenta problemas relacionados ao excesso de ruídos ambiente e sons captados que não deveriam integrar a narrativa, considerando a ideia da inteligibilidade.

Maxfield (1931) propõe também outras diretrizes relativas ao uso do som no cinema e suas relações com a imagem, como posições específicas do microfone para cada enquadramento imagético, destacando a importância da reverberação na percepção da escala sonora e o controle em relação ao som direto. Em seus trabalhos teóricos e práticos, o engenheiro também “mostrou que as capacidades de foco da escuta humana binaural, que permite aos humanos cortar seletivamente uma certa porcentagem do som refletido, não estabelece paralelo no sistema de som monaural do cinema”¹⁹ (Altman, 1992, p. 50).

Com o desenvolvimento das tecnologias disponíveis para a captação sonora, a partir dos anos de 1930, os microfones tornaram-se mais leves, compactos e, com isso, ganharam mais mobilidade dentro dos estúdios. Os microfones suspensos e apontados para o chão, como o RCA Type 77-A (ribbon unidirecional, anunciado em 1932) ajudava a rejeitar sons laterais e foi um precursor dos microfones realmente direcionais no estilo boom e padrão polar cardióide ou shotgun, foram convencioneados como os mais adequados para a captação de diálogos das cenas na época, sendo acoplados em hastes móveis e podendo se movimentar de maneira silenciosa dentro do estúdio de forma a captar os sons com maior nitidez. Outra vantagem é que, com a atenção de um bom profissional, o boom móvel conseguia permanecer fora do enquadramento imagético e mantinha uma distância adequada para a gravação sonora.

As possibilidades de movimentação de câmera e variação nos enquadramentos também foram ampliadas com a utilização deste padrão de sonorização, para além do desenvolvimento de trilhos e outros instrumentos de mobilidade. A utilização do *boom* como padrão de captação sonora viabilizou o uso de múltiplas câmeras e de planos em movimento também pudessem ser aperfeiçoados dentro das possibilidades cinematográficas, tornando-se um importante fator econômico (Altman, 1992). Além disso, a padronização no uso do *boom* mudou as características da construção sonora como um todo, desde sua produção até a finalização da trilha de áudio do filme. Antes, com a ideia defendida por Maxfield (1931) do

¹⁹ Tradução livre da autora. Texto original: “Showed that the focusing capabilities of the listening binaural human subject, permitting humans selectively to cut out a certain percentage of reflected sound, have no parallel in the monaural sound collection system of cinema”.

microfone imóvel em relação à câmera, as características espaciais do ambiente já ficavam gravadas na trilha do filme, mantendo uma proporção entre distância, som direto e reverberação, assim como o local de gravação influenciava no resultado final, com variações de volume, frequência e sons refletidos. Com o uso do *boom*, o microfone passa a ser mantido em uma mesma distância da fonte sonora, eliminando grande parte dos fatores ambientais que influenciavam no sistema anterior. A construção de uma banda sonora em vez da gravação direta dos sons já construídos se tornou possível em meados dos anos 1930, com a inclusão de dispositivos de reverberação e equalização e a técnica de *re-recording*, na qual os tipos sonoros - músicas, diálogos, efeitos etc. - eram balanceados e finalizados em uma única trilha, permitindo um maior controle sobre o som nos filmes, principalmente privilegiando a inteligibilidade do diálogo (Altman, 1992, p. 53).

Os modos de representação e a escolha entre fidelidade e inteligibilidade também entram em discussão neste ponto, pois os técnicos de som, vindos em sua maioria das experiências em rádio, estavam acostumados com a reprodução ‘fiel’ das sonoridades, mas verificaram com a experiência cinematográfica que nem sempre manter a escala entre som e imagem é a melhor escolha diante do sentido narrativo. A necessidade de entendimento dos diálogos e dos sons vocálicos em geral foi a grande justificativa para a limitação da reverberação e a gravação próxima da fonte sonora, mesmo com imagens em plano aberto, ou seja, desconsiderando a proporção entre imagem e som. Contudo, a reverberação continuava sendo utilizada em demasia nas reproduções musicais. “Os técnicos de som de Hollywood insistiam regularmente em reproduzir música com um elevado grau de reverberação, correspondendo às grandes salas reverberantes onde estamos habituados a ouvir música ‘séria’ do século XIX”²⁰ (Altman, 1992, p. 55). A escolha das imagens que acompanhavam as sonoridades não dependiam, portanto, das características espaciais do som, e sim de sua capacidade narrativa.

A fidelidade foi a opção teórica mais almejada por técnicos e engenheiros, pois entendiam o cinema sonoro como um reflexo da vida cotidiana, na qual pessoas (fontes) distantes teriam vozes em pequena escala e indivíduos próximos em maior escala, com sonoridades mais evidentes. Na prática, após uma primeira década de experimentações sincrônicas, a opção pela inteligibilidade foi moldando o cinema sonoro, não apenas pela clareza dos diálogos e seu auxílio narrativo, mas também devido ao costume do público que

²⁰ Tradução livre da autora. Texto original: “Hollywood sound technicians regularly insisted on reproducing music with a high degree of reverb, corresponding to the large reverberant halls in which we are accustomed to hearing nineteenth century ‘serious’ music played”.

foi instruído a partir da narrativa visual do teatro, que sempre prezou pelo entendimento do diálogo. “Se o cinema sonoro continuou a praticar a inteligibilidade, apesar dos repetidos apelos à fidelidade acústica, foi porque o cinema continuou a encontrar no teatro um código de realidade há muito consagrado, aplicável às narrativas audiovisuais”²¹ (Altman, 1992, p. 59). Se, no início da década de 1930, a defesa era para a manutenção da escala de imagem e escala de som, começou-se a compreender que gravar preservando a inteligibilidade significava a perda da fidelidade e a existência simultânea não parecia ser uma opção viável.

Enquanto Maxfield da Bell tentava evitar o problema, afirmando casualmente que a correspondência de escala adequada produz inteligibilidade (Maxfield, 1931, p. 71), Dreher da RCA enfrentou abertamente o problema, reconhecendo que existem dois requisitos potencialmente contraditórios para uma boa gravação: “(1) inteligibilidade do diálogo” e “(2) naturalidade, ou fidelidade acústica à interpretação original”, categoria dentro da qual ele incluiu a necessidade de manter as características espaciais do original e, portanto, uma correspondência som/imagem²² (Altman, 1992, p. 58, 59).

A ideia da inteligibilidade do diálogo acabou melhor aceita na indústria como um todo, inclusive entre os técnicos de som, o que Altman (1992) chama de “reação instintiva” dos profissionais, diante das necessidades impostas pela indústria, já no final da década de 1930. O próprio J. P. Maxfield reconhece, em suas publicações, as mudanças nas práticas de captação sonora ao longo da década e que os técnicos deixaram de combinar a escala de som e imagem através do posicionamento ‘imóvel’ do microfone, priorizando a produção de trilhas com volume nivelado. “Ao longo da década de 1930, foi pela clareza das suas bandas sonoras que os técnicos de som foram elogiados e recompensados, não apenas pelo seu realismo espacial. O que antes parecia uma monstruosidade agora se tornou a norma”²³ (Altman, 1992, p. 54). A clareza sonora atraiu mais público, que gostava de entender melhor os filmes através dos diálogos e acabou gradualmente acostumado com uma maior complexidade sonora. Os técnicos foram elogiados e recompensados pelas inovações, já que seus aprimoramentos também vendiam mais ingressos e faziam aumentar o interesse do público pelo cinema.

²¹ Tradução livre da autora. Texto original: “If sound cinema continued to practice intelligibility in spite of repeated appeals for acoustic fidelity, it was because cinema continued to find in the theater a long-consecrated code of reality applicable to audiovisual narratives”.

²² Tradução livre da autora. Texto original: “While Bell's Maxfield tried to dodge the problem, however, by blithely asserting that proper scale-matching produces intelligibility (Maxfield, 1931, p. 71), RCA's Dreher openly faced the problem, recognizing that there are two potentially contradictory requirements of good recording: ‘(1) intelligibility of dialog’ and ‘(2) naturalness, or acoustic fidelity to the original rendition’, within which category he included the need to retain the spatial characteristics of the original, and thus a sound/image match.

²³ Tradução livre da autora. Texto original: “Throughout the thirties it was for the clarity of their soundtracks that sound technicians had been praised and rewarded, rather than for their spatial realism. What once appeared as monstrosity had now become the norm”.

Com os escritos teóricos da época, é possível ainda perceber que as camadas ‘espaciais’ do som, ou seja, ambiências, vozerios e demais camadas de localização, eram deixadas de lado em razão do privilégio do diálogo. A escolha da inteligibilidade sobre a fidelidade, além de perder a correspondência de escala entre som e imagem, negava também os sons ambientes e deixava a história com trilha musical e sons vocálicos. O desenvolvimento do *boom* e de microfones direcionais, como visto anteriormente, ajudou a melhorar as possibilidades de deslocamento do técnico de som pelo estúdio sem comprometer a imagem, aproximando a captação da fonte sonora e deixando os diálogos mais compreensíveis. Com isso, o ideal da inteligibilidade permaneceu como um fator determinante da qualidade sonora nas obras cinematográficas da época, seguindo até os tempos atuais.

A discussão relativa entre original e cópia e entre fidelidade e inteligibilidade apontam o caminho para a ideia das escutas possíveis dentro do cinema, sua relação com os objetos sonoros, os pontos de escuta e suas classificações possíveis, pensando principalmente a partir das tecnologias de mixagem orientada a objetos. É importante refletir que, mesmo em um som original, há a escolha de um ‘ponto de escuta’ e também que toda a acústica do ambiente e o aparato técnico disponível transformam a produção sonora captada e influenciam em sua reprodução, não apenas em um sentido passivo, daquele que ouve, assiste, mas também de quem escolhe o posicionamento da captação sonora, dos técnicos de som direto durante a captação, dos *sound designers* que podem modificar essa escuta durante a mixagem, da direção que seleciona qual sonoridade será aplicada em qual momento e em diversas outras circunstâncias em que o ponto de escuta é selecionado de forma planejada para que o espectador venha a experienciá-lo de modo intencionalmente natural.

2.2 Tipos de escuta e objetos sonoros

Em 1966, o engenheiro eletrotécnico da radiodifusão francesa, Pierre Schaeffer, lança o *Traité des Objets Musicaux*, seu programa de pesquisa interdisciplinar voltado ao estudo dos sons e especificamente da “música concreta”, termo cunhado pelo mesmo autor em 1948, que viria a atualizá-lo como “música eletroacústica”.

Inicialmente, a música concreta poderia ser definida como os sons musicais gravados e exibidos sem a presença de um musicista, fixada sobre um suporte de gravação e exibição. Esse conceito só é possível pelo avanço tecnológico na captação, manipulação e reprodução. “A música concreta pretendia compor obras com sons de todas as proveniências –

notadamente aqueles que chamamos ruídos – judiciosamente escolhidos e colocados em sequência graças às técnicas eletroacústicas da montagem e da mixagem de gravações”²⁴ (Schaeffer, 1966, p. 20).

A palavra “concreta” do termo foi utilizada com base não na origem do ruído, mas nos sons usados em totalidade e as características acústicas que assumem “gravados por um microfone, por exemplo, o som de um violino ou de um trem” (Flôres, 2013, p. 48). Virgínia Flôres (2013, p. 58), em seu trabalho “O Cinema: uma arte sonora”, também explica de maneira sucinta e precisa a proposição de Schaeffer também sobre a existência de quatro tipos de escuta: escutar, ouvir, entender e compreender.

Eu escuto aquilo que me interessa; eu ouço o que se passa de sonoro em torno de mim, independente de minhas atividades, já que não sou surdo; eu entendo, em função daquilo que me interessa e em função do que já sei e do que procuro compreender; e eu compreendo, quando termino de entender, através de uma atividade consciente, o que procurava compreender, porque eu escutava.

Há ainda um quinto tipo de escuta definido pelo autor francês como “escuta reduzida” (termo baseado na noção fenomenológica de Husserl, em 1959), próximo do ouvir, porém mais automático e intuitivo. Seu modo reduzido seria “a escuta que trata das qualidades e das formas específicas do som, independentemente da sua causa e do seu sentido” (Chion, 2008, p. 29). Isto é, aborda o som, de diferentes formas e fontes, como verbal, musical e anedótico, sendo um objeto de observação dele próprio.

Esses conceitos vêm aglutinar também a ideia expressa pelo termo “objeto sonoro”, utilizado por Schaeffer, no *Traité* e em obras anteriores. “O objeto sonoro está no encontro de uma ação acústica e de uma intenção de escuta” (Schaeffer, 1966, p. 271 *apud* Flôres, 2013, p. 72). O autor coloca o objeto sonoro dentro da escuta “entender”, enquanto o objeto musical, aquele inserido em uma linguagem, se enquadra na escuta “compreender”. O objeto sonoro seria ainda diferente do “sinal físico” do som, pois, enquanto o primeiro é aquele mensurável por aparelhos, com corrente elétrica, vibração mecânica e acontece independentemente da figura do ‘ouvinte’, o segundo trata-se de uma percepção qualitativa da escuta e de uma experiência auditiva (Flôres, 2013, p. 60). Além deste sentido, o termo objeto sonoro ganhou, nas últimas décadas, outro significado voltado mais às características técnicas, no qual reflete uma possibilidade de mixagem através de pontos no espaço, com a utilização de metadados²⁵.

²⁴ Tradução livre da autora. Texto original: “La música concreta pretendía componer obras con sonidos de cualquier origen (especialmente los que llaman ruidos) juiciosamente escogidos, y reunidos después mediante técnicas electroacústicas de montaje y mezcla de las grabaciones”.

²⁵ Esta ideia será trabalhada adiante, quando falarmos sobre as atribuições da mixagem orientada a objetos.

Michel Chion (2008) define as escutas em três: causal, semântica e reduzida. A escuta causal é definida como a mais comum, a que usamos para nos informarmos sobre qualquer causa. De modo visível podemos perceber informações adicionais, como ao bater um objeto saber se está cheio ou vazio (p. 27) e se invisível, identificada por um saber ou suposição lógica. Por escuta semântica, o autor refere-se à linguagem ou código para interpretação de uma mensagem, como o código Morse (p. 29). Para a escuta reduzida, o autor se baseia no conceito definido por Schaeffer e o aprofunda no sentido de explicar para que serve essa escuta, que, segundo Chion (2008, p. 31), “tem a imensa vantagem de abrir a escuta e de afinar o ouvido do realizador, do investigador ou do técnico, que assim conhecerão o material de que se servem e dominá-lo-ão melhor”.

A partir dessas definições, cunha-se também o termo “acusmática” para os sons escutados sem a observação de sua origem, sua causa, ou seja, sons desacompanhados de uma visualização, mediados por um sistema de alto-falantes. Michel Chion (2008, p. 61), em *A Audiovisão*, apresenta o conceito, definindo como exemplo de *media* acusmáticos o rádio, o disco ou o telefone e como música acusmática (expressão cunhada por François Bayle) “a música de concerto realizada e escutada em suporte de registro na ausência, voluntária e fundadora, das causas iniciais dos sons e da sua visão”. O autor ainda propõe o conceito inverso da acusmática como escuta visualizada, “acompanhada da visão da causa/fonte” (p. 61). Há também o estabelecimento de uma relação entre a escuta reduzida e a acusmática na medida que a escuta repetida de um mesmo som, como uma gravação reproduzida diversas vezes e, portanto, acusmática, ajuda a se afastar da causa do som e se interessar por suas características específicas, como altura, intensidade e timbre (p. 32).

Essa relação entre acusmática e escuta reduzida é possível de ser percebida no cinema, desde a edição sonora dos filmes, uma vez que os editores se dedicam exclusivamente ao som gravado, seja por técnicos de som direto, bibliotecas sonoras, *foley* e outros possíveis meios, e, ao ouvir repetidamente essas gravações, aguçam suas características e qualidades, estabelecendo a escuta reduzida e muitas vezes podendo ressignificar um som original, como o escorrer de um regador se tornando chuva ou o ecoar de uma placa de metal transformado em trovão. A partir desses conceitos e suas relações, alguns autores tratam da questão do ponto de escuta e suas categorizações, como Rick Altman (1992) e Chion (2008) com sua classificação do ponto de escuta espacial e subjetivo.

2.3 Pontos de escuta e espacialização sonora

Do ponto de vista histórico, os mesmos técnicos de som que discutiam em seus artigos sobre qual a melhor maneira de representar o som em relação à imagem, também iniciaram os debates acerca do ponto de escuta. O engenheiro John L. Cass, um dos primeiros a pensar sobre a inteligibilidade do diálogo em vez da fidelidade em relação ao plano imagético, também começa a dissertar sobre a representação sonora através da perspectiva do espectador, utilizando, possivelmente pela primeira vez, o termo “ponto de escuta” [*point of audition*], em um artigo em 1930, quando comenta que “não se pode dizer que a mistura de som resultante representa qualquer ponto de escuta, mas é o som que seria ouvido por um homem com cinco ou seis orelhas muito longas, estendendo-se em várias direções”²⁶ (Cass, 1930, p. 325 *apud* Altman, 1992, p. 49). A partir dessa argumentação, Cass evidenciava a necessidade de alterar as práticas de edição da época, que não consideravam a experiência do espectador ao alterar um plano distante para um plano próximo, sem considerar a fluidez necessária para manter o público concentrado na narrativa, concluindo que “como é costumeiro entre os humanos tentar manter constante a distância entre o olho e o ouvido, esses órgãos deveriam se mover juntos de um ponto a outro, a fim de manter nossa tão mencionada ilusão [de realidade]”²⁷ (p. 50).

A ideia do ponto de escuta, hoje conhecida e utilizada na caracterização sonora de produções audiovisuais, começa a ser desenvolvida a partir de uma perspectiva auditiva analisada através de mudanças na intensidade e nos níveis de reverberação. Altman (1992) afirma que esse entendimento não deve ser confundido com a correspondência entre escala de som e imagem, nem com a discussão entre fidelidade e inteligibilidade. A denominação como som do “ponto de escuta” tenta, inicialmente, gerar uma comparação com a ideia do plano do “ponto de vista”, proporcionando uma rápida compreensão do conceito. “O som do ponto de escuta é identificado por seu volume, nível de reverberação e outras características que representam o som tal como seria ouvido de um ponto dentro da diegese, normalmente por um personagem ou personagens específicos”²⁸ (Altman, 1992, p. 60). O som do ponto de

²⁶ Tradução livre da autora. Texto original: “The resultant blend of sound may not be said to represent any given point of audition, but is the sound which would be heard by a man with five or six very long ears, said ears extending in various directions”.

²⁷ Tradução livre da autora. Texto original: “Since it is customary among humans to attempt to maintain constant the distance between the eye and the ear, these organs should move together from one point to another in order to maintain our much mentioned illusion [of reality]”.

²⁸ Tradução livre da autora. Texto original: “Point-of-audition sound is identified by its volume, reverb level, and other characteristics as representing sound as it would be heard from a point within the diegesis, normally by a specific character or characters”.

escuta servia, inicialmente, para estabelecer relações espaciais entre locais que não podiam ser apresentados de maneira imagética em uma única cena. Isto é, o som conectava espacialmente os ambientes, dando uma unidade narrativa a partir da utilização de sons contidos na diegese fílmica.

Altman (1992) descreve que, como efeito, o som do ponto de escuta atrai o ouvinte para dentro da narrativa, como se nela pertencesse, funcionando como um ouvinte “interno”, pertencente à história, assim como um personagem. Essa diferenciação é importante, pois foge da discussão anterior, da fidelidade contra a inteligibilidade. Se, no início da década de 1930, os técnicos-teóricos de som estavam voltados à ideia de correspondência de escala ou diálogo privilegiado, em 1938, Hollywood começa a entender a importância dramática do som, inserindo o espectador como um ouvinte interno na narrativa. Enquanto, de maneira basal, a fidelidade era proposta através da manutenção de escala entre imagem e som e a inteligibilidade privilegiava o diálogo perante as demais sonoridades, o som do ponto de escuta chega como uma terceira possibilidade, a de que o espectador escute de maneira semelhante a um personagem diegético. “Somos solicitados não a escutar, mas a nos identificarmos com alguém que escutará por nós. Em vez de nos dar a liberdade de nos movimentarmos à vontade pelo espaço do filme, esta técnica nos localiza num lugar muito específico – o corpo do personagem que ouve por nós”²⁹ (Altman, 1992, p. 60).

Dessa maneira, o som do ponto de escuta passa a constituir uma terceira possibilidade de escolha sonora, funcionando como uma intersecção entre a ideia de escala entre imagem e som e a escolha privilegiada da abordagem inteligível. Além disso, o ponto de escuta evoca uma importante - talvez uma das mais importantes - característica do som no cinema, a espacialidade. Ao conectar espaços que apenas a imagem não seria capaz, o som do ponto de escuta confere a sensação de controle perante esses ambientes conectados pela banda sonora.

Embora o som no ponto de escuta tenha tido algum espaço nas discussões sobre a melhor maneira de sonorizar as imagens para cinema, o formato não chegou a receber o mesmo zelo teórico que a extensa discussão entre fidelidade e inteligibilidade. Mesmo surgindo cedo na história do cinema sonoro, demorou para que fosse integrado no modo de representação sonoro do cinema hollywoodiano. Apenas no final da década de 1930, o som do ponto de escuta começou a despontar como um caminho para uma identidade audiovisual no cinema, da mesma forma que o privilégio do diálogo em relação às demais sonoridades.

²⁹ Tradução livre da autora. Texto original: “We are asked not to hear, but to identify with someone who will hear for us. Instead of giving us the freedom to move about the film's space at will, this technique locates us in a very specific place-the body of the character who hears for us”.

Juntos, o sistema de inteligibilidade (emprestado desde o início de precedentes teatrais e telefônicos) e o sistema de ponto de escuta (elaborado durante a década de 1930 a partir de possibilidades específicas do cinema) constituíram um novo modo de unidade cinematográfica e uma nova posição de sujeito para o cinema hollywoodiano³⁰ (Altman, 1992, p. 64).

2.3.1 Michel Chion e os sentidos subjetivo e espacial

Partindo da discussão sobre perspectiva sonora e espacialidade narrativa, Michel Chion (2008) também concentrou parte de seus estudos a entender e posteriormente classificar o ponto de escuta. Inicialmente, o autor compara o conceito à ideia de ponto de vista, ou seja, um direcionamento fílmico que ajuda a compreender a narrativa e destaca que, no cinema, o ponto de vista pode ser considerado de duas formas: uma estritamente espacial, que considera de onde o espectador vê a cena, e de que ponto de espaço ela é visualizada, seja de cima, por baixo, dentro de algum móvel ou outro lugar “A noção de ponto de vista, neste primeiro sentido espacial, assenta na possibilidade de se deduzir com maior ou menor precisão o lugar de um ‘olho’ a partir da composição da imagem e da sua perspectiva” (Chion, 2008, p. 74). A outra forma é considerada uma maneira subjetiva, pois está ligada a que personagem, durante a ação, também vê aquilo a que o espectador assiste.

A partir desta classificação, Chion (2008), em *A Audiovisão*, compara a divisão de ponto de vista com o ponto de escuta, ao qual atribui também um sentido espacial e outro subjetivo. O subjetivo relaciona-se ao personagem, evidenciando quem ouve e o quê ou ainda “que personagem, num dado momento da ação, está em condições de ouvir aquilo que eu próprio ouço?” (Chion, 2008, p. 74). O espacial busca as fontes sonoras, ou seja, qual o ponto no espaço da tela é a fonte daquilo que pode ser escutado pelo espectador, ou seja, “de onde ouço, de que ponto do espaço representado no ecrã ou no som” (p. 74). Essa distinção básica pode ser relacionada à dicotomia entre fidelidade e inteligibilidade, uma vez que quando um plano geral mostra dois personagens ao longe conversando entre si e os espectadores conseguem ouvir o diálogo como se perto estivessem, o ponto de escuta espacial acaba sendo desconsiderado. Nesse caso, a fim de garantir a inteligibilidade, o público está submetido ao ponto de escuta subjetivo, pois escuta o diálogo como se dele participasse.

³⁰ Tradução livre pela autora. Texto original: “Together, the intelligibility system (borrowed early on from theatrical and telephonic precedents) and the point-of-audition system (elaborated during the thirties out of cinema-specific possibilities) constituted a new mode of cinematic unity and a new subject position for the Hollywood audience”.

Chion (2008) ainda expressa a dificuldade de definir um lugar de escuta específico, devido à natureza omnidirecional do som, propagado em todas as direções, e da escuta, que capta os sons de maneira circular. Por isso, o autor fala que não se deve tratar o ponto de escuta espacial como uma posição exata no espaço, mas como um lugar de escuta, ou ainda, uma área para a escuta, mesmo que haja sons em que é possível atribuir à escuta uma direção, principalmente em sons agudos que tendem a se propagar de maneira mais direcional que os sons graves.

Ainda no sentido espacial, é possível correlacionar este conceito às tecnologias sonoras que permitem a gravação e exibição em multicanais que se utilizam de caixas acústicas espalhadas pelas salas de cinema ou em ambientes domésticos, como *home theaters* e *soundbars*, direcionando os sons de acordo com o espaço que os personagens ocupam na tela. Em 2013, Jeff Smith publicou o artigo “*Atmos, all around: A guest post by Jeff Smith*” no blog de David Bordwell, no qual detalha sua experiência em ‘audiover’ o filme *Valente* (2012), da Disney, estreia do Dolby Atmos, no qual a personagem principal Merida recebe orientações de sua mãe para portar-se como princesa.

Enquanto Merida recita um poema, a Rainha, de pé acima dela, a instrui a projetar sua voz dizendo: “Enuncie! Você deve ser entendida de qualquer lugar da sala ou será tudo em vão.” Corta para Merida. Quando ela responde baixinho, “Isso é tudo em vão”, a Rainha dispara de volta “Eu ouvi isso!” Durante esta breve tomada que se concentra em Merida, a resposta melódica de Emma Thompson como a Rainha sai de um dos alto-falantes *surround* traseiros esquerdos. A localização da voz da Rainha cria um breve “efeito de ponto de escuta”, pois nos coloca realisticamente no meio do espaço diagonal que separa Merida de sua mãe. O momento também demonstra de forma lúdica a instrução da Rainha para Merida ser ouvida de “qualquer lugar na sala”³¹ (Smith, 2013).

No artigo “Ponto de escuta e tridimensionalidade sonora na ficção e no documentário: Para uma estética do envelopamento no cinema contemporâneo”, Leonardo Vidigal e Marina Novais (2020) também se debruçam sobre a temática, trazendo exemplos práticos dessas possibilidades sonoras, como a descrita acima, em que “o sistema sonoro permitiu que o espectador escutasse a fala de Merida a partir do ponto de escuta de sua mãe, para depois ouvir a resposta da mãe do ponto de escuta de Merida” (Vidigal; Novais, 2020, p. 232).

³¹ Tradução livre da autora. Texto original: “As Merida recites a poem, the Queen, standing above her, instructs her to project her voice saying, “Enunciate! You must be understood from anywhere in the room or it’s all for naught.” Cut to Merida. When she replies under her breath, “This is all for naught,” the Queen shoots back “I heard that!”. During this brief shot that holds on Merida, Emma Thompson’s mellifluous response as the Queen issues from one of the left rear surround speakers. The localization of the Queen’s voice creates a brief “point of audition effect” as it realistically places us in the middle of the diagonal space that separates Merida from her mother. The moment also playfully demonstrates the Queen’s instruction to Merida to be heard from “anywhere in the room””.

O desenvolvimento de tecnologias de mixagem orientada a objetos nas últimas décadas pode vir a ocasionar uma mudança em relação a esse pensamento expresso pelo autor. Conforme veremos adiante, os sistemas de som espacializados para salas de cinema e ambientes domésticos permitem incluir objetos sonoros em ângulos específicos e localizações precisas, ocupando um ou mais alto-falantes, de acordo com os metadados criados na pós-produção. Desta forma, o ponto de escuta espacial pode se tornar tão preciso e reconhecível como o ponto de escuta subjetivo.

No entanto, a representação sonora através do ponto de escuta ainda esbarra em outras questões anteriores, como a localização do microfone durante a gravação e a inteligibilidade do diálogo. Chion (2008, p. 76) identifica que quando algum personagem fala virando as costas para o microfone, são percebidos menos harmônicos agudos de sua voz que, conseqüentemente, parece mais ausente que quando falada em frente ao microfone: “Podemos então falar de uma diferença audível entre voz de frente e voz de trás”. Essa diferenciação é mais percebida e valorizada em filmes gravados com som direto, em que o ator possa eventualmente virar de costas para o microfone, e contribui para trazer alguma realidade para a representação, funcionando também como o que Chion (2008, p. 92) chama de “índice materializante”, ou seja, que remete à sensação de materialidade da fonte e dão pistas sobre a matéria que causa o som.

Estas variações sonoras também podem ser construídas pós-sincronização com o deslocamento do ator ou do microfone, mesmo que o usual seja utilizar a voz de frente na maioria das captações sonoras, “com o máximo de agudos possíveis que a técnica lhe permite, e com um objetivo evidente: esses agudos são preciosos para a inteligibilidade” (p. 76). O efeito causado pelo uso da voz de costas nem sempre faz o espectador assimilar o ponto de escuta do plano, muitas vezes por ser um efeito momentâneo e passageiro, sem uma durabilidade suficiente para que seja deduzido como um efeito planejado e também “porque o ponto de escuta, para o som, não está associado à representação de um microfone” (Chion, 2008, p. 76).

Posteriormente, Chion (2009) desenvolve sua teoria sobre os “doze ouvidos” e os pontos de escuta que cada um poderia desenvolver. A ideia parte da noção de que não há um “microfone simbólico” ou um “ouvido simbólico” que represente no espaço fílmico o mesmo que o autor denomina como “olho simbólico”, ou seja, o enquadramento ou a limitação da lente da câmera, que especifica o que o espectador deve visualizar e que raramente se posiciona no ponto de vista de um único personagem (p. 292). A percepção do microfone, segundo o autor, também fica complicada pelo fato de um determinado som ser gravado por

diversos microfones e em locais diferentes, criando “imagens sonoras” que acabam se fundindo em um único som e, por isso, não existiria também alto-falantes simbólicos, uma vez que cada caixa de som pode mesclar diversas sonoridades.

O olho da câmera, localizado direcionalmente (ele vê esse personagem, desse ângulo, dessa altura - daí os termos técnicos como *close-up* médio e plano de ângulo baixo), nos permite fazer distinções entre o ponto de vista da câmera e aquele de um determinado personagem (sugerido pela montagem ou por técnicas “subjetivas”) ou do nosso próprio ponto de vista que a câmera assume para nós. Mas a questão de quem está ouvindo, não podemos responder pelo microfone. Mesmo quando o microfone está visível na tela (como em *One Plus One* [1969] de Godard), ou quando aparece no topo do quadro devido a uma falha no enquadramento da projeção, nada em especial nos faz ouvir através desse microfone, já que pode muito bem sejam outros microfones no *set* ou na pós-produção que misturam os sons que estão capturando³² (Chion, 2009, p. 293).

Essa compactação sonora acontece de maneira ainda mais acentuada ao refletirmos sobre os dispositivos domésticos de projeção audiovisual. Em *home theaters* ainda há a prerrogativa de diversos alto-falantes colocados ao redor do espaço, mas, em televisores comuns ou mesmo com *soundbars* e dispositivos móveis como celulares ou *tablets* que demandam o uso de fones de ouvido para uma experiência sonora estéreo com maior qualidade, a distinção das fontes sonoras e os pontos de escuta possíveis acabam por se entrelaçar, dificultando a percepção detalhada das sonoridades presentes na obra cinematográfica. O uso de mixagem orientada a objetos para aparelhos de uso doméstico e o movimento de captação binaural pode ter auxiliado em uma melhoria dessa percepção sonora, conforme veremos adiante.

Chion (2009) propõe, então, sua teoria sobre os doze ouvidos existentes no cinema e como eles se relacionam. O ouvido técnico não seria localizável, pois é resultante de uma fusão entre múltiplas gravações com microfones no *set* ou durante a mixagem, ou seja, é aqui que todo o som da narrativa está contido e que deveria ser escutado pelo público. O ouvido dos personagens pode ser enquadrado em diversas situações, como quando dois personagens em uma mesma cena ouvem coisas diferentes, por exemplo, um estando ao telefone e o outro não; ou quando os personagens escutam seus próprios sons, seja sua respiração ou batimento

³² Tradução livre pela autora. Texto original: “The camera's eye, localized directionally (it sees as this character, from this angle, at this height—hence the technical terms such as medium closeup and low-angle shot), allows us to make distinctions between the camera's point of view and that of a particular character (suggested by the editing or by “subjective” techniques) or our own point of view that the camera assumes for us. But to the question who is hearing, we cannot answer the microphone. Even when the microphone is visible on screen (as in Godard's *One Plus One* [1969]), or when it appears at the top of the frame because of a glitch in projection framing, nothing especially makes us hear through that microphone, since there may well be other mics on the set or in post production that mix together the sounds they are capturing”.

cardíaco, ou suas falas de maneira consciente, como em uma situação de psicanálise, na qual o “cliente de terapia que toma consciência de sua voz no silêncio do analista”³³ (2009, p. 300).

O ouvido do espectador possui a capacidade de ouvir o que os personagens ouvem e o que não ouvem, como trilhas musicais não diegéticas e narração em *voz over*. Pela descrição inicial, pode-se confundir o ouvido técnico com o ouvido do espectador, pois ambos têm a capacidade de ouvir tudo aquilo que a narrativa apresenta, entretanto o ouvido humano não é capaz de processar toda a complexidade sonora disponível no ambiente e em produções audiovisuais. Há uma seleção natural dos sons mais significativos para o entendimento e, pensando em cinema, a voz e o diálogo sempre são as escolhas privilegiadas, o que Chion (2008, p. 13; 2009, p. 301) chama de vococentrismo.

A escuta linguística envolve excluir tudo o que não é pertinente; é ouvir o sentido e não o conjunto sonoro, através de um processo de seleção que nem sequer tem consciência de si mesmo. Ouvir o significado envolve ser surdo ao resto. Quando ouvimos alguém falando, esquecemos da música e ficamos sujeitos, aproveitados, absorvidos pelas palavras. Isto não nos impede de ouvir passivamente o que não ouvimos, de sermos sensíveis e afetados por isso. Mas não é possível ouvir atentamente todos os sons ao mesmo tempo³⁴ (Chion, 2009, p. 302).

A descrição do restante dos ouvidos propostos por Chion apresenta situações de escuta específicas para cenas de filmes citados pelo autor, como *Festim Diabólico* (1948), de Hitchcock, em que exemplifica a perspectiva sonora da escuta de uma pessoa morta, e em *Mãe Pálida* (1980), de Helma Sanders-Brahms, com o ouvido de alguém prestes a nascer, assim como também observado mais recentemente em *Beau tem medo* (2023), de Ari Aster. O nono ouvido proposto pelo autor, “O ouvido de um lugar” (2009, p. 305), relaciona-se ao espaço contido na narrativa, o lugar físico da cena que pode ouvir o que os personagens dizem ou ressoar sons que ali já foram amplificados. Essa classificação aproxima-se da ideia do ponto de escuta espacial, com o espectador podendo ser aquele que escuta, assim como as paredes parecem nos ouvir. Chion (2009, p. 307) finaliza sua categorização dos doze ouvidos com “O Grande Ouvido”, aquele que tudo escuta ou o lugar onde tudo pode ser escutado, “o que caracteriza este lugar, designado e suposto pelo filme, é que não há ninguém ali”³⁵. E, por

³³ Tradução livre pela autora. Texto original: “Therapy client who becomes aware of his or her voice in the silence of the analyst”.

³⁴ Tradução livre pela autora. Texto original: “Linguistic listening involves excluding all that is not pertinent; it is listening for meaning and not for the ensemble of sound, through a selection process that is not even aware of itself. Listening for meaning involves being deaf to the rest. When we listen to someone speak, we forget about the music and are subject to, harnessed by, absorbed by the words. This does not prevent us from passively hearing what we're not listening to, from being sensitive to it and affected by it. But it is not possible to listen attentively to all the sounds at once”.

³⁵ Tradução livre pela autora. Texto original: “What characterizes this place, designated and supposed by the film, is that no one is there”.

esse motivo, iguala essa classificação ao décimo ouvido, proposto por ele mesmo, “O Ouvido de Ninguém”.

A classificação dos doze ouvidos proposta por Chion (2009) apresenta nuances específicas para diversas situações comuns em cenas cinematográficas, o que pode representar um ganho na ideia de categorização do ponto de escuta de personagens em perspectivas variadas. Seguir essa ideia pode gerar uma lista que julgamos infinita de “ouvidos” ou pontos de escuta, uma vez que outras cenas e sequências serão desenvolvidas com suas especificidades e necessidades de uma nova denominação, principalmente, pois o ouvido do espectador e o ouvido do personagem abrangem boa parte das demais categorias propostas pelo autor. No entanto, a listagem é interessante nas caracterizações das cenas e nas possibilidades sonoras múltiplas. Assim, pensando em relação às novas realidades de escuta a partir da tridimensionalidade com objetos sonoros, torna-se possível identificar novos ‘ouvidos’ e desenvolver suas características a partir do modelo de mixagem e da escuta.

A partir de críticas quanto à nomenclatura e à dependência do termo com a ideia do ponto de vista, outras classificações do ponto de escuta e do som no audiovisual surgiram nas últimas décadas, tentando suprimir algumas lacunas e acrescentar o som extradiegético na análise sonora.

2.3.2 Outras classificações

O som do ponto de escuta é descrito, pelo *Dictionary of Media and Communication* (2011), no portal *Oxford Reference*, como: “som diegético percebido por um determinado personagem. O equivalente auditivo de uma tomada de ponto de vista: por ex. se uma pessoa está escondida sob cobertores, os sons ouvidos por ela e pelo público são abafados”³⁶, tanto para filmes, quanto para produções televisivas. Todavia, o termo ponto de escuta recebe críticas há décadas principalmente pela definição estar atrelada à imagem e ao plano do ponto de vista.

Birger Langkjær (2000), no artigo “*Den lyttende tilskuer: Perception af lyd og musik i film*” (O Espectador Ouvinte: Percepção de som e música no cinema, tradução da autora), julga o termo ponto de escuta também por sua proximidade com o aspecto imagético. O autor critica as definições de Altman e Chion e descreve o uso da nomenclatura muitas vezes como

³⁶ Tradução livre pela autora. Texto original: “In films and television, diegetic sound is perceived by a particular character. The aural equivalent of a point-of-view shot: e.g. if a person is hiding under blankets the sounds heard by them and the audience are muffled”.

carente de lógica e relevância (Langkjær, 2000, p. 135), por focar excessivamente nos aspectos audiovisuais e pouco no campo da escuta. Langkjær, através de suas críticas, parece almejar uma necessidade de refinamento do termo do ponto de escuta, principalmente para abranger as novas tecnologias sonoras, como o som *surround*.

Svein Høier (2013), no ensaio “*The Relevance of Point of Audition in Television Sound: Rethinking a Problematic Term*”, descreve o termo som do ponto de escuta como problemático por sua dependência com o visual e por ser um fenômeno complexo e de difícil descrição tanto para o som cinematográfico, quanto para televisão. “Por um lado, muitas vezes é difícil falar sobre um ‘ponto’ específico de escuta porque, na maioria dos casos, muitos elementos sonoros diferentes e paralelos estão sendo experimentados simultaneamente”³⁷ (Høier, 2013). Mesmo que o autor concentre seus estudos no som televisivo, sua revisão traz questões relevantes a esta pesquisa, principalmente no enfoque que daremos ao ambiente doméstico e a produções cinematográficas contidas em plataformas de *streaming*, podendo ser assistidas em televisões e dispositivos móveis, como *tablets* e celulares.

O ensaio de Høier considera também que as novas tecnologias de mixagem e representação sonora exigem um aprimoramento do termo do ponto de escuta e, a partir dos estados narrativos de Branigan (1992), propõe uma nova subcategorização do som do ponto de escuta: o espacial se divide também em observacional e ativo; e o subjetivo, em individual e pessoal. Os dois primeiros se direcionam mais a documentários e materiais jornalísticos, enquanto o individual e o pessoal conseguem se adequar a materiais de ficção audiovisual para televisão e cinema.

O ponto de escuta observacional, descreve o autor, acontece quando o microfone é posicionado distante da fonte sonora, muitas vezes com gravações fora de um ambiente controlado, podendo utilizar imagens e sons de câmeras de vigilância ou celulares e em documentários e *reality shows* que tentam disfarçar a presença da câmera do ambiente de captação, como se não houvesse aquele agente modificador do ambiente observacional. Por ser algo mais objetivo, com tendência ao realismo e à manutenção da perspectiva sonora entre enquadramento imagético e sonoro, aproxima-se da definição de ponto de escuta espacial de Chion (2008), ocorrendo com frequência em transmissão de eventos esportivos que combinam narração e comentários em estúdio com microfones espalhados pelo ambiente, a uma certa distância da origem do som. Conforme propõe Høier (2013), o ponto de escuta

³⁷ Tradução livre da autora. Texto original: For one, it is often difficult to talk about one specific “point” of audition because in most cases many different and parallel sound elements are being experienced simultaneously.

observacional funciona, então, como um espectador que vivencia um evento, com o microfone, e, muitas vezes, com a câmera, localizados como testemunhas do acontecimento. Dentro dessa categoria, Høyer (2013) usa o conceito de “adoçar” o áudio³⁸ em situações observacionais em que o som captado e transmitido ao vivo é mixado com gravações prévias: “O POA [observacional] pode ser transformado em POA ativo usando um *sampler* e teclado para acionar diferentes sons, sincronizados com os movimentos visuais, que são produzidos ao vivo e no local”³⁹. A ideia do adoçamento é algo habitual ao imaginarmos uma mixagem sonora em eventos gravados, sejam documentários, esportes ou mesmo materiais ficcionais. Ao misturar som direto, diálogos captados em sincronia com a imagem a outras sonoridades, como *foley* ou sons de banco de dados *online*, uma nova paisagem sonora é construída, atribuindo novos significados e construindo narrativamente a história a ser contada. Com a mixagem orientada a objetos, a ideia de “adoçar” o áudio pode ser amplificada, pensando no direcionamento angular do som de acordo com a necessidade narrativa, inclusive do diálogo, mesmo que ainda pouco usual.

Outra subcategoria do ponto de escuta espacial proposta por Høyer é o POA ativo, vinculado ao registro de situações controladas e a utilização de microfones visíveis, podendo ser direcionais nas mãos de repórteres ou deixados em estandes para participação popular, lapelas em produções jornalísticas ou ainda mais recentemente na criação de conteúdo para redes sociais. Todas essas situações deixam o som em *close-up*, independentemente do enquadramento imagético.

Já o ponto de escuta individual se assemelha ao sentido subjetivo de Chion (2008) e, segundo Høier, acontece em situações subjetivas, na qual o personagem vivencia uma situação física ou psicológica singular. Esse tipo de ponto de escuta geralmente está atrelado a cenas com intensa carga emocional e podem “dar ao público não apenas acesso participativo a uma situação, mas também acesso direto às experiências, pensamentos, sentimentos de um personagem específico e assim por diante”⁴⁰ (Høier, 2013). Outra possibilidade de POA individual acontece em cenas em que o espectador ouve uma voz telefônica diegética na mesma perspectiva sonora do personagem ao telefone.

Por último, em sua proposta de categorização, Høier (2013) cita o ponto de escuta pessoal, o qual também se conecta à subjetividade e às emoções, mas de maneira ainda mais

³⁸ Tradução livre pela autora. Texto original: ““Sweetening” the sound”.

³⁹ Tradução livre pela autora. Texto original: “The POA can be shifted into active POA by using a sampler and keyboard to trigger different sounds, synchronized with the visual movements, which is produced live and on the spot”.

⁴⁰ Tradução livre pela autora. Texto original: “Give the audience not only participatory access to a situation, but also direct access to a specific character’s experiences, thoughts, feelings and so on”.

particular, pois “investigam mais profundamente as mentes dos personagens e fornecem acesso a experiências mais pessoais que não são compartilhadas com outros personagens”⁴¹, como em escritas faladas, memórias sonoras ou conectando dois tempos históricos diferentes, em *flashbacks*. Além disso, outros exemplos possíveis de POA pessoal acontecem a partir de características sobre-humanas dos personagens, como a capacidade de ouvir pensamentos alheios ou conectar-se com os mortos.

A proposta de classificação de Høier surge como uma crítica ao ponto de escuta e sua necessidade de interação com a imagem, mas acaba por subcategorizar o termo, atribuindo a terminologia nomenclaturas mais extensas e pouco autoexplicativas, deixando o conceito mais complexo e sujeito a interpretações errôneas. Além disso, a grande crítica envolvendo sua vinculação ao campo imagético não é suprimida, sendo difícil que em algum momento seja, pois mantemos o que afirma Chion (2008) sobre o contrato audiovisual, com som e imagem sendo modificados e complementados um pelo outro. Mesmo que surjam novas nomenclaturas e outras propostas de classificação, o processo audiovisual será mantido e o som do ponto de escuta não deverá se desvincular da imagem e seu enquadramento.

Se há algum consenso entre os teóricos é que o som do ponto de escuta recebe críticas por sua nomenclatura ao mesmo tempo em que é utilizado para caracterizar as sonoridades. Altman (1992, p. 60) o considerava “um termo desajeitado cujo único mérito é lembrar infalivelmente o plano do ‘ponto de vista’”⁴², Chion (2008, p. 73) definia que “a noção de ponto de escuta é muito complicada e ambígua”, já Høyer (2013) avalia o o POA como “um fenômeno complexo e difícil de descrever”⁴³. As categorizações propostas pelos autores citados levam em conta que o ponto de escuta é sempre diegético, ou seja, um personagem empresta seus ouvidos ao espectador para que ele ouça como alguém que faz parte da história. Os sons considerados como não diegéticos ou extradiegéticos, isto é, sons que não estão contidos no universo filmico, como trilhas musicais e voz *over*, ficam de fora das classificações relativas ao conceito.

Mesmo que rodeado por insatisfações, nenhuma outra expressão conseguiu ganhar força suficiente para deixar de lado o uso do ponto de escuta para a caracterização sonora. Alguns autores tentaram modificar a nomenclatura, mas mantendo seus ‘defeitos’, como a relação com o plano do ponto de vista. É o caso de James Buhler, David Neumeyer e Robert

⁴¹ Tradução livre pela autora. Texto original: “Probe further into the minds of characters and provide access to more personal experiences that are not shared with other characters”.

⁴² Tradução livre pela autora. Texto original: “A clumsy term whose only merit is to recall unfailingly the “point-of-view” shot”.

⁴³ Tradução livre pela autora. Texto original: “POA is a complex phenomenon that is hard to describe”.

Deemer (2010, p. 78) que, no livro “*Hearing the Movies: Music and Sound in Film History*”, tentam disfarçar o ponto de escuta utilizando a expressão “som do ponto de vista”⁴⁴, para caracterizar algo próximo do ponto de escuta subjetivo de Chion: “O efeito geralmente é aumentar nossa identificação com aquele personagem específico cuja audição é imitada pela trilha sonora”⁴⁵. Os autores concentram a análise na proposta de uma nova categorização sonora, que inclui também os sons extradiegéticos e diferencia ainda o som que é visualizado pelo enquadramento e pelo que não é contido na imagem.

Os sons, então, podem ser classificados como diegéticos e extradiegéticos⁴⁶, mas também como dentro ou fora de quadro. “No seu nível mais básico, o som diegético se refere a tudo o que pode ser ouvido por personagens do filme. O som extradiegético, por outro lado, não pode ser ouvido pelos personagens”⁴⁷ (Buhler; Neumeyer; Deemer, 2016, p. 66). Ainda segundo os autores, dentro de quadro se refere a tudo que é contido no enquadramento (*onscreen*) e fora de quadro tudo aquilo que não está enquadrado (*offscreen*), mas que se mostra como uma extensão do espaço-tempo da cena (2010, p. 72), um som acusmático, de uma fonte apresentada ou não, que pode ser “desmascarado” a qualquer momento. A combinação entre essas características cria algumas relações entre som, imagem e espaço narrativo: diegético dentro de quadro, diegético fora de quadro, extradiegético dentro de quadro e extradiegético fora de quadro.

O primeiro deles, diegético-em quadro, é certamente o caso padrão: vemos dentro do quadro o que esperamos ver no mundo do filme [...] Diegético-fora de quadro também é comum: uma sala é exibida na tela, mas ouvimos alguém falando ou tocando música com o volume correto e outras qualidades de som que correspondem a outro quarto conectado ao que vemos. Extradiegético-em quadro, por outro lado, é muito menos comum, mas provavelmente será usado quando estiver claro que um personagem em quadro que imagina ou lembra de uma fala ou música e a performance dessa música é visualizada. Extradiegético-fora de quadro é o caso padrão para narração de voz [*voice-over*] e música incidental ou canção não ouvida pelas personagens⁴⁸ (Buhler; Neumeyer; Deemer, 2010, p. 72).

⁴⁴ Texto original: “Point of view sound”.

⁴⁵ Tradução livre pela autora. Texto original: “The effect is generally to increase our identification with that particular character whose hearing the sound track mimics”.

⁴⁶ O termo “extradiegético” foi adotado no texto como tradução de *Non-diegetic* no livro citado pela justificativa apresentada em Benetello, Vidigal (2022), em que entendemos que a mera negação do som diegético não parece dar conta da complexidade desse elemento sonoro.

⁴⁷ Tradução livre da autora. Texto original: “At its most basic level, diegetic sound refers to everything that can be heard by characters in the film. Non diegetic sound, by contrast, cannot be heard by the characters”.

⁴⁸ Tradução livre da autora. Texto original: “The first of these, diegetic-on screen, is certainly the default case: we see within the frame what we expect to see in the film world [...] Diegetic-offscreen is also common: a room is shown on screen but we hear someone speaking or music playing with the correct volume and other sound qualities that would match another room connected to the one we see. Non diegetic-on screen, on the other hand, is much less common but is likely to be invoked when it is clear that an onscreen character imagines or remembers speech or music and the performance of that music is visualized. Non diegetic-offscreen is the default case for voiceover narration and underscoring”.

A proposta de Neumeyer, Buhler e Deemer (2010, p. 78) inclui os sons extradiegéticos em sua classificação, dá enfoque ao visual para além do áudio, uma vez que parte da categorização engloba o dentro e fora da tela, mas não inclui o ponto de escuta propriamente dito, mesmo que faça uma breve descrição sobre o “som do ponto de vista”, o termo não aparece novamente na publicação. Em 2015, outra publicação de Neumeyer, “*Meaning and Interpretation of Music in Cinema*”, reforça a classificação anterior e detalha suas características.

Pensando em uma escala imaginária de entendimento, o som diegético em quadro seria o mais simples, pois apresenta sincronização entre imagem e som e uma narrativa direta, na qual o espectador conhece o personagem de uma maneira realista, através de suas falas e atitudes. Já o extradiegético fora de quadro, não gera problemas de entendimento do público, pois desde o início do cinema sonoro há o costume de combinar as imagens com fontes sonoras não específicas, como trilhas musicais ou voz *over*. O diegético fora de quadro é compreendido facilmente quando se pensa em sons musicais ou ambiências sonoras que se encaixam na narrativa do filme, mas não possuem uma fonte visível, como o apito de um trem em uma cena dentro de uma estação ou uma música em voz e violão dentro do ambiente de um bar. O extradiegético em quadro é possivelmente o mais complexo em relação ao entendimento, pois “prejudica a sincronização ao visualizar simultaneamente o espaço diegético e extradiegético”⁴⁹ (Neumeyer, 2015, p. 81). O autor separa em interior e exterior, no qual o primeiro aconteceria em um sonho ou em uma visualização direta de sentimentos, pensamentos ou memórias do personagem; já o segundo pode ocorrer através de comentários ou como uma inserção *picture-in-picture*, “geralmente é representado como uma breve intrusão ou comentário narrativo – sendo típico uma aparição em um canto da tela”⁵⁰ (p. 80).

A classificação proposta pelos autores é simples e de fácil entendimento, mas não deixa de apresentar lacunas, por não relativizar com o conceito do ponto de escuta e também não conseguir abranger em totalidade as possibilidades sonoras, como aponta Vidigal (2023, p. 527), ao analisar o filme *Marte Um* (2022), em que o extradiegético em quadro não é suficiente para caracterizar as cenas de ‘destempero’ sensorial da personagem Tércia (Rejane Farias), após o trauma com o falso atentado explosivo. Marie Welde Tranås (2017) é outra

⁴⁹ Tradução livre da autora. Texto original: “Which undermines synchronization by visualizing both diegetic and nondiegetic space simultaneously”.

⁵⁰ Tradução livre da autora. Texto original: “It is usually represented as a brief narrative intrusion or comment—the cameo in a corner of the screen being typical”.

teórica que tenta expor a problemática da terminologia do ponto de escuta e sugerir outras direções. Em sua dissertação em Musicologia *“Unravelling Point of Audition: A discussion of Point of Audition in the sound theory of modern film and television”*, a autora considera o termo ambíguo e expõe a necessidade de compreensão auditiva mais ampla de acordo com as características de cada sonoridade. Ao analisar a terminologia proposta por Høyer (2013), e compará-la com a proposição de Neumeyer (2015), Tranås (2017) aponta algumas diferenças entre os trabalhos dos autores e reforça que não há menção direta entre eles em suas publicações.

Enquanto Høyer concentra-se apenas nos sons contidos na narrativa, diegéticos e foca sua análise apenas nos aspectos sonoros do ponto de escuta, Neumeyer classifica também os sons extradiegéticos, como trilhas musicais e narrações e abrange sua investigação para o espectro audiovisual como um todo, pois sua proposta de classificação atribui à tela parte importante da descrição, com fontes sonoras podendo estar contidas ou não no enquadramento. Além disso, apesar da proximidade temporal e temática entre os autores, suas propostas também diferem em relação à nomenclatura: Neumeyer segue não utilizando o termo som do ponto de escuta em sua classificação, mesmo que sua teoria seja baseada em conceitos próximos ao POA, como objetividade, subjetividade, exterior, interior e diegese (Tranås, 2017, p. 25).

As propostas de análises sonoras vistas até aqui objetivam criar solução para um problema que nem sabemos se realmente existe. A classificação inicial de Chion (2008) entre ponto de escuta subjetivo e espacial parece contemplar a maior parte das necessidades categóricas que a análise sonora demanda. O próprio autor posteriormente tenta subcategorizar com seus doze tipos de ouvido e percebe, ao final, que uma categoria já se confunde com a outra, decaindo um pouco em relevância.

Høyer, conforme dito anteriormente, esmiúça o ponto de escuta de Chion e, em uma tentativa de tornar mais simples seu entendimento, cria subcategorias dentro do subjetivo e espacial, conseguindo trabalhar com exemplos concretos e boas perspectivas para a análise. Entretanto, a subdivisão parece provocar o efeito inverso do esperado, complicando ainda mais o entendimento da questão, ao atribuir nomenclaturas pouco intuitivas e exemplos muito específicos. Além disso, a questão apontada pelo autor como problemática em relação à proximidade do termo com o campo visual não é superada e a vinculação com a imagem se mantém na proposta de Høyer.

Ao deixar de lado o termo do ponto de escuta, Neumeyer gera uma classificação mecanizada do som na obra audiovisual, apenas com o entendimento de perspectiva entre o

universo fílmico, diegético, e sua oposição, a extradiegese, além da inclusão da perspectiva imagética, ao procurar a fonte sonora dentro do chamado contentor visual das imagens, o quadro (Chion, 2008, p. 58). Neumeyer, ao abordar sobre a dimensão realista e psicológica dos sons aproxima-se ainda mais da discussão do ponto de escuta, mesmo que siga recusando em utilizar sua nomenclatura.

A proximidade do ponto de escuta com o plano do ponto de vista não me parece problemática como diversos autores apontam, principalmente por tratarmos todo o tempo de narrativa audiovisual. Seria possível enumerar mais questionamentos sobre como a pesquisa imagética é realizada sem levar em consideração o aspecto sonoro e como as análises em relação à fotografia, direção de arte e demais propostas ‘esquecem’ a relação de Contrato Audiovisual⁵¹, posta por Chion (2008, p. 7), do que estabelecer um problema em aproximar a análise sonora, principalmente a do ponto de escuta, com seu enquadramento visual. A noção de ponto de escuta consegue se desvencilhar da imagem quando necessária, principalmente em seu sentido espacial, mas ainda mantém proximidade com o campo imagético, assim como deve ser em uma análise em que som e imagem se apresentam de forma simultânea.

No Brasil, a produção acadêmica sobre som no audiovisual tem se consolidado nas últimas décadas como um campo plural, articulando reflexões teóricas, estudos históricos e investigações empíricas, incluindo discussões acerca do ponto de escuta. Em espaços coletivos como os Seminários Temáticos relacionados aos estudos do som da Sociedade Brasileira de Estudos de Cinema e Audiovisual (Socine), a pesquisa em sonoridades ganhou visibilidade institucional e redes de interlocução que fomentam diálogos entre pesquisadores e profissionais. Autores como Rodrigo Carreiro (2018; 2019; 2023) têm contribuído para a formação desse campo por meio de organização de coletâneas e com artigos que, entre outras coisas, investigam as práticas de pós-produção no contexto brasileiro. Débora Opolski (2023), ao lado de Carreiro e outros organizadores, também tem marcado o debate com a introdução da metodologia de análise sonora do espectograma, a qual serve como referência para análises sobre função expressiva, montagem sonora e pontos de escuta. A produção de pesquisadores como Leonardo Vidigal (2020; 2023), que também assina trabalhos sobre os pontos de escuta e arranjos audiovisuais, demonstra a relevância de abordagens textuais e contextualizadas, integrando a pesquisa de gêneros, música e práticas de *sound design*, buscando a compreensão de o ponto de escuta opera nas formas contemporâneas de narração sonora. Em articulação

⁵¹ Noção apresentada por Chion (2008, p. 7) de que som e imagem são regidos por acordo de modificação mútua, um influencia e transforma o outro e ambos devem ser levados em conta em uma análise audiovisual: “Uma percepção influencia a outra e a transforma: não vemos a mesma coisa quando ouvimos; não ouvimos a mesma coisa quando vemos”.

com estudos teóricos mais amplos (investigando tridimensionalidade, imersão e estetização e envelopamento sonoro), a produção brasileira vem qualificando tanto o diagnóstico das transformações tecnológicas quanto a proposição de categorias analíticas específicas, como as que fundamentam a presente pesquisa.

Propor uma atualização dessas propostas sem uma mudança real da escuta e seus sentidos parece inicialmente pouco efetiva diante das tentativas antes realizadas, mesmo que as caracterizações sugeridas pelos autores acima citados sejam utilizadas em análises audiovisuais diversas, inclusive a desta pesquisa. Ainda assim, a princípio, acreditamos que a divisão entre subjetivo e espacial já atende boa parte das necessidades relativas à análise sonora. Ao submeter esta pesquisa, concentrando os estudos nos pontos de escuta possíveis através do uso de fones de ouvido e analisando obras audiovisuais com tecnologia de mixagem orientada a objetos, pretendemos investigar e verificar como o(s) conceito(s) do ponto de escuta pode(m) ser observado(s) em sistemas sonoros mais simples, como os domésticos, e entender de que forma a tecnologia imersiva pode influenciar no ponto de escuta do espectador diante da obra audiovisual.

2.3.3 Ponto de escuta imersivo

A partir da revisão teórica e da análise das transformações tecnológicas recentes no campo do som audiovisual, torna-se relevante propor um desdobramento conceitual capaz de abarcar as novas modalidades de recepção e percepção propiciadas pelas tecnologias de áudio imersivo. Entre as categorias já estabelecidas, tratadas anteriormente, observa-se uma lacuna no vocabulário teórico quando se trata de descrever as experiências de escuta mediadas por sistemas de espacialização sonora orientados a objetos, como o Dolby Atmos e o MPEG-H.

Essas tecnologias, ao possibilitarem a movimentação tridimensional dos sons no espaço e a criação de ambientes acústicos que circundam o espectador, promovem uma forma de envolvimento que ultrapassa os limites da representação narrativa tradicional. Nelas, a escuta do espectador não se restringe à imitação da escuta de um personagem, como ocorre no ponto de escuta subjetivo, mas envolve uma vivência sensorial ampliada, em que o som atua diretamente sobre o corpo e sobre as emoções de quem assiste.

Diante dessa constatação, esta pesquisa propõe o conceito de ponto de escuta imersivo, entendido como uma modalidade de experiência auditiva em que a espacialidade sonora constrói uma partilha perceptiva, gerando uma sensação de incorporação e copresença. Trata-se de uma escuta em que o espectador é envolvido pelo campo acústico de modo

tridimensional, participando da diegese como se estivesse inserido fisicamente em seu espaço. No ponto de escuta imersivo, o som passa a operar também como vetor de sensação, fazendo com que o público não apenas ouça como o personagem, mas sinta como ele. O ponto de escuta imersivo caracteriza-se ainda por uma aproximação sensorial e emocional mais intensa, mediada pela espacialização tridimensional e pela densidade acústica.

Nesta conceituação proposta, o som não apenas representa a escuta do personagem, mas envolve o corpo do espectador, circundando-o e inserindo-o dentro da ação fílmica. A espacialidade sonora deixa de ser um elemento de representação e torna-se uma experiência de presença, em que o espectador partilha fisicamente as emoções, angústias e estados de tensão de seu personagem-ouvido.

A formulação desse conceito tem caráter propositivo, buscando contribuir para o campo dos estudos de som no audiovisual ao compreender as possibilidades de transformação do lugar do ouvinte na narrativa a partir das tecnologias de mixagem de áudio imersivo. A partir do momento em que o som pode ser distribuído e movimentado no espaço, criando campos acústicos que ultrapassam os eixos fixos e independem de canais, a escuta passa a operar também como imersão perceptiva e não apenas como mediação narrativa. Enquanto o ponto de escuta subjetivo ancora-se na representação da perspectiva de um personagem, o ponto de escuta imersivo amplia essa relação, deslocando a experiência para um regime de corporeidade e envolvimento afetivo, no qual o espaço fílmico é vivenciado como uma experiência sensorial direta. O ponto de escuta imersivo descreve, portanto, essa nova condição estética, em que o espaço sonoro deixa de se organizar apenas em função da representação da cena e passa a atuar diretamente sobre a sensibilidade corporal do espectador.

Ao introduzir esta categoria, o trabalho visa também oferecer um termo que descreve de forma mais precisa as transformações estéticas e perceptivas a partir da mudança de paradigma que acompanha a evolução tecnológica, de um modelo de escuta mediado pela narrativa para um modelo que se orienta também pela imersividade estética. Essas possibilidades permitem compreender como as tecnologias contemporâneas imersivas deslocam o som para uma dimensão experiencial, em que o espectador passa a integrar o ambiente sonoro como parte ativa da cena.

O ponto de escuta imersivo, portanto, configura-se como uma contribuição teórica desta tese, oferecendo um conceito aplicável às análises fílmicas e aos estudos sobre tecnologias sonoras imersivas. Ao nomear e definir esse novo modo de escuta, a nomenclatura busca delimitar um campo semântico e perceptivo que permita compreender como as práticas

sonoras, tanto em cinema quanto no ambiente doméstico, vêm reformulando as experiências de recepção, o papel da espacialidade e a própria noção de presença sonora na narrativa audiovisual.

A caracterização do ponto de escuta imersivo será aprofundada ao longo das análises fílmicas desenvolvidas nos capítulos seguintes. Nessas análises, o conceito é mobilizado de forma prática, a partir da observação de estratégias de espacialização sonora presentes em obras audiovisuais. Por meio desses estudos de caso, será possível evidenciar como o conceito se manifesta de maneira concreta na construção da experiência audiovisual, revelando de que modo as tecnologias orientadas a objetos intensificam o envolvimento emocional entre o espectador e a narrativa.

3 SOM NO CINEMA

Neste capítulo, apresentamos um breve percurso histórico do som no cinema, com ênfase em sua vertente clássica, em paralelo ao desenvolvimento do som no ambiente doméstico. O objetivo é acompanhar as transformações tanto nos modos de captação quanto nos sistemas de exibição sonora, até chegar às tecnologias contemporâneas de mixagem orientadas a objetos. A história da Dolby também é enumerada com o objetivo de evidenciar os motivos que levaram a empresa a se tornar o padrão sonoro cinematográfico desde a década de 1970. Além disso, abordamos os conceitos relacionados à mixagem orientada a objetos, os principais sistemas que utilizam esta tecnologia e as possibilidades de utilização e integração em ambiente doméstico e *mobile*, através do uso de fones de ouvido.

3.1 Breve história dos sistemas sonoros

A maioria das pesquisas relacionadas ao som no cinema acabam por iniciar reclamando da marginalização recebida pela temática quando comparada aos estudos relativos à imagem cinematográfica. Kerins (2010) critica essa prática, mesmo que ele próprio a tenha utilizado na citação de abertura de seu livro *Beyond Dolby (Stereo)*. O autor julga que comparar os estudos e lamentar sua discrepância “já não parece necessário, não só porque esta queixa tem sido repetida tantas vezes, mas também porque os estudos atuais estão cada vez mais a trabalhar para a resolver”⁵² (p. 3). Entretanto, mesmo tentando fugir desses termos, é preciso ainda reconhecer a disparidade não apenas no número de estudos e análises acadêmicas, como também em relação à elaboração técnica na produção cinematográfica, com profissionais ligados mais à captação e mixagem e pouco sobre criação sonora (Flôres, 2013, p. 17). Nesse sentido, ainda é possível observar que o som busca o reconhecimento de sua fundamental importância à experiência do cinema, mesmo que aos poucos estejamos caminhando rumo à equidade.

Inicialmente, o som no cinema teve sua trajetória marcada pelas dificuldades tecnológicas de sincronia entre as captações de imagem e som, a reprodução em cópias e a exibição dos dois formatos de forma simultânea, criando a impressão de uma única transmissão e desenvolvendo a narrativa audiovisual. Anteriormente à ideia de exibição

⁵² Tradução livre da autora. Texto original: This no longer seems necessary, not only because this complaint has been repeated so often but also because current scholarship is increasingly working to address it.

cinematográfica como é conhecida hoje, houve o chamado Kinetoscópio, de Thomas Edison, que consistia em um aparelho de projeção semelhante a um armário, no qual cada pessoa poderia assistir a pequenos filmes através de uma lente de aumento. O próprio Edison já havia desenvolvido, ainda em 1877, o fonógrafo, aparelho capaz de gravar e reproduzir sons registrados de maneira mecânica, mesmo que de modo ainda limitado. Muitos Kinetoscópios eram acompanhados por fonógrafos, apresentando som e imagem de forma conjunta. No Brasil, algumas exibições do Kinetoscópio com fonógrafos aconteceram em 1894, trazidas por Frederico Figner, empresário tcheco (Cf. Costa, 2006, p. 21-22).

Ainda em 1887, surge o Gramophone, desenvolvido pelo alemão Emile Berliner, que utilizava discos como suporte para a gravação. A reprodução acontecia através da agulha em contato com as ranhuras do disco, emitindo vibrações afuniladas em uma corneta e interligada a uma membrana que sustentava a agulha. A movimentação do ar gerada pela emissão sonora vibra a membrana e faz a agulha riscar a superfície do disco que permanece girando, a partir do esforço mecânico do girar de uma manivela que dava corda no aparelho, uma vez que a energia elétrica ainda não era difundida massivamente na época .

As exibições cinematográficas nos modelos que conhecemos hoje, com salas específicas e apresentações coletivas, iniciaram em 1895 já com acompanhamento musical realizado por pianistas em performance ao vivo dentro das salas, em que acompanhavam o decorrer das imagens e buscavam, através da música, intensificar emoções e aumentar a carga dramática – além de disfarçar o barulho do projetor. Esse acompanhamento sonoro não acontecia de forma padronizada, variando conforme a localidade e até mesmo o horário da sessão.

Em outra parte do livro *Hearing the Movies*, os autores James Buhler, David Neumeyer e Robert Deemer (2010, p. 247) dividem a prática sonora na era dos filmes com som não sincronizado em três fases, todas direcionadas ao cinema clássico: o primeiro cinema, de 1895 a 1905, com sonorização totalmente despadronizada, sendo determinada pelo local e horário da exibição; o *nickelodeon*⁵³, no período entre 1905 e 1915, uma fase de transição, quando o som começou a se tornar uma prática comum em locais com exibição cinematográfica; e o chamado *Picture Palace*, compreendido no período entre 1915 e 1929, no qual já havia uma prática de exibição cinematográfica com filmes médias e longas

⁵³ Primeiras salas de cinema dos Estados Unidos, localizadas próximas a centros comerciais e de transporte, com capacidade de público entre duzentas e trezentas pessoas. Semelhantes aos teatros populares, exibiam, entre as trocas de rolos dos filmes, atrações ao vivo, como palestras ou canções ilustradas.

metragens e uma sonorização voltada para música. As salas de cinema começaram, a partir desse período, a se diferenciar por nível de qualidade e categorias de apresentações musicais.

Por exemplo, em um teatro de *vaudeville*⁵⁴, o local principal para exibição de filmes antes da ascensão do *nickelodeon*, seria mais provável haver acompanhamento musical do que em um show de tenda, que poderia ter apenas um tambor automático; mas mesmo nos cinemas dedicados à exibição cinematográfica, o primeiro show do dia e a sessão no horário do jantar eram executados sem acompanhamento, pelo menos até a década de 1920. Outro exemplo de prática sonora ao vivo era a inclusão de efeitos sonoros marcantes, em momentos-chave (trovões, a fatídica batida na porta etc.). Por fim, era comum alguns filmes, como os populares documentários de viagem, um acompanhamento de narração ao vivo, um modo de apresentação que não desapareceu até os anos 1960; e, embora geralmente pensemos em 1927 como o momento inicial do filme sonoro, tentativas bem-sucedidas, mesmo que comercialmente inviáveis, para sincronizar imagem e som gravado remontam aos primórdios do cinema (Buhler; Neumeyer; Deemer, 2016, p. 247).⁵⁵

Diversas tentativas, criadores e empresas se envolvem na tentativa de consolidação dessa inovação ao processo produtivo, e, em 1919 os pesquisadores Hans Vogt, Joseph Benedict Engl e Joseph Massole registram a patente do Tri-ergon, que utilizava uma tecnologia de densidade variável para a modulação da luz na película.

O sistema traduzia vibrações mecânicas de som em impulsos elétricos que eram convertidos em sinais ópticos, usando a célula fotoelétrica de Jakob Kunz. Esses sinais óticos poderiam então ser gravados diretamente no canto da película por um processo fotográfico, inventado por Eugene Lauste, mas aprimorado por eles. Outra célula fotoelétrica transformava estes sinais em impulsos elétricos, que, por sua vez, eram amplificados e transformados em vibrações sonoras (Eikmeier, 2014).

Na reprodução sonora em ambientes domiciliares, os discos de goma de laca, conhecidos pela reprodução em 78 rotações por minuto, conseguiram alcançar certa popularização e comercialização doméstica nos anos 1910 e a hegemonia da produção do som em disco ampliava também as possibilidades da capacidade de produção em série para o mercado cinematográfico. O suporte era reproduzido em vitrolas e esbarravam na questão da

⁵⁴ Sala de entretenimento popular que reunia diversas atrações, como filmes, apresentações teatrais, espetáculos circenses etc.

⁵⁵ Tradução livre da autora. Texto original: For example, a vaudeville theater, the dominant venue for exhibiting motion pictures before the rise of the nickelodeon, would be more likely to have musical accompaniment than would a tent show, which might have only an automated drum for ballyhoo— if that; but even in theaters dedicated to motion picture exhibition, the first show of the day and the dinner show would occasionally be run without accompaniment even as late as the 1920s. As another example, early sound practice was more likely to include substantial sound effects than was later practice, which tended to restrict effects to key narrative moments (thunderclaps, the fateful knock on the door, etc.). Finally, it was common for certain films, such as the popular travel documentaries, to be accompanied by live narration, a mode of presentation that did not disappear until as late as the 1960s; and, although we usually think of 1927 as the initiating moment of the sound film, successful, if commercially unviable, attempts to synchronize film and recorded sound go back to the earliest days of film.

amplificação, pois ainda não havia um sistema elétrico de qualidade para as salas de exibição. Os discos eram maiores que os LPs⁵⁶ (*long-play*) de 33 rotações, desenvolvidos anos mais tarde, e armazenavam menos tempo de gravação, em média uma ou duas músicas a cada lado.

A Moviola, tão conhecida por sua contribuição em relação à montagem cinematográfica, foi desenvolvida inicialmente como um projetor de filmes domésticos, fechado em um gabinete de madeira, como a vitrola. Por volta de 1919, o dispositivo foi patenteado e o primeiro gabinete utilizava um projetor *Motiongraph*, patenteado em 1909, que projetava imagens em uma tela de 12 por 18 polegadas. A tela, quando não estava em uso, dobrava-se para dentro do gabinete⁵⁷ (Theisen, 1935). O custo elevado - cerca de U\$600, em 1920 - e o contexto da Primeira Guerra, acabaram por minar os planos de uso doméstico do aparelho, que posteriormente passou por adaptações, como a retirada do gabinete de madeira e da lente de projeção, baixando seu custo, conquistando os estúdios e deixando sua marca na história do cinema mundial.

Na década de 1920, as famílias se reuniam em frente ao receptor de rádio para ouvir notícias e programas de entretenimento. A evolução técnica na transmissão sonora no período da Primeira Guerra Mundial possibilitou a popularização do meio de informação e o surgimento de emissoras de rádio com conteúdos jornalísticos, apresentações musicais, esportivas e humorísticas, via AM⁵⁸ e FM⁵⁹. Antes, vinda apenas em meio impresso, o rádio revolucionou a forma de consumo de informação e ampliou as possibilidades de entretenimento sonoro doméstico, antes restrito às vitrolas presentes em casas mais abastadas.

Na mesma década, o sistema *Vitaphone*, desenvolvido pela *Western Electric* e comprado pela *Warner Brothers*, começa a despontar como alternativa. O primeiro filme lançado com a tecnologia foi *Don Juan* (1926), dirigido por Alan Crosland. No *Vitaphone*, projetor e disco eram acionados simultaneamente através de um único motor, consistindo em um sincronismo mecânico em que o disco sonoro rodava de dentro para fora, ao contrário do tradicional de vitrola. O sistema funcionava de maneira satisfatória (Buhler; Neumeyer; Deemer, 2010, p. 285), no entanto, havia dificuldades para se manter o sincronismo entre som e imagem, pois a agulha poderia ‘pular’, o disco poderia estar riscado, a película ser cortada e não se saber exatamente onde a agulha deveria ser reposicionada, entre outras intercorrências.

⁵⁶ Popularizado na década de 1940.

⁵⁷ Tradução livre da autora. Texto original: The first cabinet utilized a Motiograph projector head of the vintage patented in 1909, and threw pictures to a 12 by 18 inch screen. The screen when not in use folded into the cabinet.

⁵⁸ Amplitude modulada. Alcança maiores distâncias, mas com amplitude variável enquanto se propagam, podendo sofrer com interferências técnicas.

⁵⁹ Frequência modulada. Possui uma melhor qualidade sonora, pois se propaga no espaço alterando seu valor de frequência; entretanto, não alcança longas distâncias como o formato AM.

Dessa maneira, a necessidade de se integrar o som à película diminuiria o custo e eliminaria os problemas da perda de sincronicidade do sistema, tecnologia já viável em outros modelos do mesmo período.

Sistemas concorrentes já mostravam formas diferentes de sincronização entre som e imagem desde 1927, como no caso do *Movietone*, criado pela Fox, lançado no filme *Aurora* de 1927, dirigido por F. W. Murnau, e o *Photophone*, desenvolvido pela RCA. Os sistemas consistiam em uma faixa de áudio mono gravada na película 35 mm. “A diferença entre os dois sistemas era técnica: estava no modo como a energia sonora, convertida em matéria fotossensível, era gravada na película” (Carreiro, 2018, p. 46). Em meados da década de 1920, a forma de sonorização sincrônica era selecionada mais através de interesses comerciais entre exibidores, distribuidoras e executivos do que uma escolha técnica ou artística.

Mesmo sendo descontinuado em 1929, com a Warner adotando o *Photophone*, o filme *O cantor de jazz*⁶⁰, exibido em 1927, é considerado uma das grandes viradas para a exibição sonora no cinema ainda com a tecnologia do *Vitaphone*. O filme revolucionou o modo de assistir cinema, pois, após sua estreia, os espectadores almejavam mais filmes com diálogos e o cinema sonoro tornou-se uma obrigação para manter o público interessado no entretenimento. As salas exibidoras também tiveram que se adaptar para reproduzir o som gravado com qualidade e amplificação esperadas, muitas precisaram de empréstimos para comprar os equipamentos com os bancos exigindo como garantia que os filmes exibidos fossem de empresas para as quais eles também tivessem emprestado recursos, o que significava exibir apenas as produções de grandes estúdios e para grandes públicos. “Esta jogada afetou de forma drástica os pequenos produtores, que ficaram de fora dos circuitos de exibição, fortalecendo ainda mais o poder dos grandes estúdios” (Eikmeier, 2014). A animação *Finding his voice* (1929)⁶¹ explica de maneira educativa como funcionavam os primeiros sistemas de gravação sonora nos filmes da *Western Electric*, mesmo com o *Vitaphone* ainda detendo os direitos de exclusividade, outras mecânicas já vinham sendo testadas paulatinamente em experimentações. Auxiliadas pela compra de um tubo de amplificação sonora e o desenvolvimento dos primeiros microfones condensadores, a sonorização fílmica acontecia através da gravação de voz, a mixagem quase em tempo real com o profissional regulando as sonoridades presentes em cena e a amplificação através de tubos eletrônicos à vácuo *Audion*, que potencializava a voz para tornar possível sua ‘impressão’ no filme, na mesma velocidade que as imagens posteriormente relevadas.

⁶⁰ Hoje considerado um filme problemático e racista, devido ao uso de *blackface* pelo ator principal, Al Jolson.

⁶¹ Disponível em: <https://www.youtube.com/watch?v=MGbTE2kHbOA&t=351s>. Acesso em 12 jul. 2024.

O final da década de 1920 e início da de 1930, conforme mostrado no capítulo anterior, é um período de muita experimentação técnica, criativa e narrativa. Buhler, Neumeyer e Deemer (2010) categorizam o início do filme sonoro de três formas: o ‘sincronizado’, com acompanhamento orquestral gravado e efeitos sonoros; o ‘*part-talkie*’, com acompanhamento musical gravado intercalado com sequências de fala; e o ‘*100% talkie*’, considerado uma obra completamente sonora. “Esta terminologia pode ser confusa porque um filme “sincronizado” nesta tipologia é o que hoje chamaríamos de filme silencioso; “sincronização” aqui se refere ao fato de que a música foi adequada à ação”⁶² (p. 295). Filmes como o já mencionado *Aurora* e *Luzes da Cidade* (1931), de Charles Chaplin, são exemplos dos filmes descritos como “sincronizados” pelos autores. Ainda em 1929, quando a MGM ganhou o Oscar de Melhor Filme com *The Broadway Melody*, de Harry Beaumont, os filmes sonoros viraram sinônimo para filmes falados e ganharam o protagonismo e “no ano de 1930, em Hollywood, a produção de filmes silenciosos, com algumas exceções notáveis como Charlie Chaplin, essencialmente cessaram. [...] Mesmo fora dos Estados Unidos, entretanto, ir ao cinema estava rapidamente começando a significar ir ao cinema falado”⁶³ (p. 299).

A popularização dos filmes falados trouxe questões relativas ao melhor uso do som na narrativa. Questões como as anteriormente vistas relativas à fidelidade e inteligibilidade do diálogo surgem nessa época em que os três componentes básicos da trilha de áudio - diálogo, música e efeitos (além do silêncio) - precisavam encontrar sua melhor forma enquanto representação sonora. Para além disso, era preciso entender como as falas e os efeitos sonoros diegéticos precisavam ser sincronizados e parecerem pertencentes à tela e de que forma e lugar os sons não sincrônicos seriam relacionados à imagem. “Em particular, não estava claro se esse som deveria ser considerado como vindo de um espaço fora da tela ou de algum outro lugar, e se viesse de outro lugar, onde esse lugar estaria localizado”⁶⁴ (Buhler; Neumeyer; Deemer, 2010, p. 301-302) e, principalmente, como o público receberia essas experimentações.

Com o crescimento do interesse do público pelos filmes falados, houve um aumento exponencial das salas de cinema nos Estados Unidos e, por necessidade de padronizar os

⁶² Tradução livre da autora. Texto original: “This terminology can be confusing because a “synchronized” film in this typology is what we would now call a silent film; “synchronization” here refers to the fact that the music has been fit to the action”.

⁶³ Tradução livre da autora. Texto original: “By 1930 in Hollywood, production of silent film had, with a few notable exceptions such as Charlie Chaplin, essentially ceased. [...] Even outside the United States, however, going to the movies was quickly coming to mean going to the talkies”.

⁶⁴ Tradução livre da autora. Texto original: “In particular, it was not clear whether such sound should be considered as coming from offscreen space or from somewhere else, and if it was coming from somewhere else, where that place was located”.

modos de exibição, em 1938, a Academia de Artes e Ciências Cinematográficas organizou e formalizou o registro sonoro, a equalização, especificações das caixas de som, entre outras normatizações. Ficou estabelecido o que hoje se conhece como som monaural – ou seja, todo o som reproduzido por meio de um único canal. “Era um som bastante distorcido, a resposta de frequência penosamente atingia 5 ou 6 Khz, equivalente ao som de rádio AM, e bastante ruidoso também” (Klachquin, 2002).

Ainda que a história do cinema nos conte que a ideia de múltiplos canais sonoros foi discutida e consolidada para cinema apenas após várias décadas de som mono, a técnica e a tecnologia já eram testadas desde antes da própria projeção de imagens, com registros de dois canais sonoros transmitidos por dois telefones em 1879, em experimentos de Alexander Graham Bell (Kerins, 2010). Já no final da década de 1920 e início da de 1930, o *Bell Labs* seguiu com experiências de gravações em dois canais distintos, com um disco em 1932 sendo o primeiro exemplo de registro sonoro em estéreo. Um ano mais tarde, a empresa realizou a transmissão de uma mixagem orquestral em três canais - frontal direito, frontal central e frontal esquerdo - da Filadélfia para Washington.

A escolha de três canais de áudio exclusivos aqui foi ditada por preocupações estéticas: enquanto os engenheiros da *Bell* concordaram que um número infinito de canais de alto-falante frontal era desejável, o uso dos canais esquerdo, central e direito parecia produzir adequadamente o efeito de um campo estereofônico “completo” sem o efeito de “buraco no meio” comum às reproduções usando apenas dois canais (frontal esquerdo e frontal direito)⁶⁵ (Kerins, 2010, p. 21).

Embora o experimento tenha sido considerado um sucesso, ficou ainda a questão de como seria a implantação desse sistema nas salas de cinema e qual seria a melhor maneira de utilizar esse arranjo multicanal em um filme. Kerins (2010) cita Jay Beck (2003), em *Quiet Revolution*, para evidenciar alguns fatores considerados na justificativa da utilização de múltiplos canais na época: relacionar um som com a posição de sua fonte visualizada pela imagem; acentuar a sensação de espetáculo filmico; e criar um espaço narrativo. Justificativas que são consideradas válidas e usuais até o presente, mesmo com os modelos mais atuais de mixagem e utilização do som em audiovisual. Um filme de demonstração da *Bell Labs*, em 1937, acabou se tornando a primeira apresentação documentada de áudio multicanal

⁶⁵ Tradução livre da autora. Texto original: “The choice of three unique audio channels here was dictated by aesthetic concerns: while Bell engineers agreed that an infinite number of front loudspeaker channels was desirable, the use of left, center, and right channels seemed to adequately produce the effect of a “full” stereophonic field without the “hole in the middle” effect common to reproductions using only two (front left and front right) channels”.

sincronizado com imagens, combinando três canais frontais mixados com recursos visuais para três cenas, cada uma com trilha sonora multicanal de uma maneira diferente.

Este filme de demonstração mostrou que três fatores básicos foram considerados no desenvolvimento do som multitrilha. Primeiro, um “modelo binaural” musical onde a fidelidade estava relacionada à posição fixa do ouvinte e a uma fonte sonora estacionária. Em segundo lugar, um “modelo estereofônico” espetacular onde a intenção era imitar o movimento mesmo que exagerasse a escala do evento sonoro. E terceiro, um “modelo construído” mais complexo, onde objetos estacionários e em movimento eram representados em acordo espacial com a imagem⁶⁶ (Kerins, 2010, p. 22 *apud* Beck, 2003, p. 68-69).

Na mesma época, a indústria de exibição cinematográfica experimentava enviar diferentes sonoridades para diferentes caixas nas salas de cinema, mesmo que a motivação fosse mais financeira que inovadora. A ideia era manter a trilha musical saindo do fosso a frente da tela - como no cinema silencioso, em que a música era tocada ao vivo - e deixar os diálogos em caixas atrás da tela, numa tentativa de manter uma ideia teatral para o cinema. Até 1933, havia a dificuldade de misturar múltiplas faixas de áudio que tivessem sido gravadas separadamente, por isso, o projecionista era quem ficava responsável pela alternância entre os dois conjuntos de alto-falantes durante a exibição e precisava deter o conhecimento total da trilha de som do filme para saber o momento exato da troca. “Não é de surpreender que esse arranjo complicado não tenha durado muito, pois os técnicos de som rapidamente reconheceram que ter música vindo de trás da tela com o diálogo era perfeitamente aceitável para o público”⁶⁷ (Kerins, 2010, p. 22). Além disso, nessas primeiras décadas do cinema sincronizado, o chamado ruído nominal era inerente aos sistemas de gravação e reprodução de áudio da época, e acabava sendo considerado aceitável ou “dentro da norma” para o padrão tecnológico daquele período. As limitações técnicas resultavam em uma qualidade de áudio que hoje seria considerada baixa, com a presença constante de chiados, estática ou zumbidos. Esse “ piso de ruído ” era o nível padrão de interferência com o qual os engenheiros e o público tinham que lidar. Com isso, mesmo que de maneira breve, é possível perceber que, desde o princípio, o cinema sonoro tenta incluir as configurações de múltiplas sonoridades em suas características basais.

⁶⁶ Tradução livre da autora. Texto original: This demonstration film showed that three basic factors were considered in the development of multi-track sound. First, a musical “binaural model” where the fidelity was related to the fixed position of the auditor and a stationary sound source. Second, a spectacular “stereophonic model” where the intent was to imitate motion even if it exaggerated the scale of the sound event. And third, a more complex “constructed model” where stationary and moving objects were depicted in spatial accord with the image.

⁶⁷ Tradução livre da autora. Texto original: Not surprisingly, this convoluted arrangement did not last long, as sound technicians quickly recognized that having music coming from behind the screen with the dialogue was perfectly acceptable to audiences.

Enquanto isso, o processo de captação avançava com a chegada com gravadores ópticos e a separação entre montagem da imagem e edição de som, a mixagem com música e bandas bilaterais. As salas de exibição se mantinham com caixas atrás da tela e alguns problemas com amplificação. Na década de 1940, algumas melhorias começaram a despontar na tentativa de ampliar o número de canais de áudio na captação e exibição cinematográfica. O Vitasound da Warner Brothers surgiu como um dos primeiros sistemas “multi-alto-falante padronizado”⁶⁸ (Kerins, 2010, p. 23). Mesmo não sendo exatamente um sistema multicanal, o Vitasound representa um avanço tecnológico por colocar alto-falantes em torno do público, que seriam ativados em momentos nos quais a trilha musical e as ambiências ficam em destaque, enquanto os diálogos seguiram nas caixas atrás da tela. Outra inovação do sistema é que a alternância entre as caixas acontecia através de uma trilha de controle “entre os orifícios da roda dentada da película para ligar e desligar esses alto-falantes surround; um método semelhante seria usado por sistemas multicanais da década de 1950”⁶⁹ (p. 23).

Em 1941, o filme *Fantasia*, da Disney, inaugurou o sistema Fantasound, considerado o primeiro filme sonoro multicanal (Kerins, 2010) e o primórdio do som surround, cuja premissa era elaborar uma experiência de orquestra ao espectador. Conforme mostra Buhler, Neumeyer e Deemer (2010, p. 338), a captação da Orquestra da Filadélfia ocorreu através de oito canais com um profissional em mixagem necessário para monitorar cada canal. O som foi exibido em três canais diferentes e alimentava três alto-falantes atrás da tela de forma horizontal, possibilitando que o som viesse de vários lugares ao longo da largura da imagem, balanceando a potência relativa entre as caixas, criando a ideia de que o som se move entre elas. “É como o 'balance' do som estéreo, mas aplicado a um elemento pontual dentro da mixagem. Nasce assim o ‘potenciômetro panorâmico’ ou pan-pot” (Klachquin, 2002). Além disso, na estreia em Los Angeles, havia uma ideia de canal surround, com noventa e seis alto-falantes menores ao redor da sala, inclusive na parte de trás, utilizados em efeitos sonoros específicos e para o musical do final do filme. “Curiosamente, uma das configurações tentadas [...] durante o processo de três anos de desenvolvimento do Fantasound, foi uma configuração frontal de três canais e surround de dois canais semelhante ao arranjo de 5.1 canais que os sistemas DSS adotariam meio século depois”⁷⁰ (Kerins, 2010, p. 24). Devido ao alto custo de implementação e a complexidade técnica da exibição, a experiência do

⁶⁸ Tradução livre da autora. Texto original: “Multi-speaker system”.

⁶⁹ Tradução livre da autora. Texto original: “Between the sprocket holes of the filmstrip to turn these surround speakers on and off; a similar method would be used by multi-channel systems of the 1950s”.

⁷⁰ Tradução livre da autora. Texto original: “Intriguingly, one of the configurations tried but ultimately abandoned during the three-year process of developing Fantasound was a 3-channel front and 2-channel surround setup similar to the 5.1-channel arrangement DSS systems would adopt half a century later”.

Fantasound chegou a apenas seis salas nos Estados Unidos; nas outras, o filme foi reproduzido em som monofônico padrão da época. O sistema exigia, além dos alto-falantes adicionais atrás da tela e ao redor do espaço, um projetor extra que continha toda a trilha de áudio, enquanto outro projetor ficava apenas com a imagem. Além disso, cada canal exigia o controle manual de uma pessoa diferente,

A experimentação do Fantasound resultou no desenvolvimento de alguns conceitos, como o de atribuir ganho dos amplificadores a canais de áudio, mesmo que ainda incipientes pela dificuldade tecnológica de executar todo esse sistema com gravação óptica do som. A gravação magnética desponta apenas após a Segunda Guerra e se consolida na década de 1950, separando a gravação de imagem e som e produzindo conteúdo sonoro com maior qualidade, menor distorção e ruído, faixas mais amplas de frequências e um custo menor em comparação às gravações ópticas de 35 mm que estavam em uso.

As melhorias vindas da gravação magnética pouco influenciavam o resultados nas salas de cinema, pois ainda se utilizava o som monofônico óptico como padrão de exibição. A partir da década de 1950, o sistema evoluiu com o uso de fitas magnéticas e a partir da possibilidade de se colocar caixas ao redor da sala, ampliando a ideia de som com espacialidade. Com isso, surge a ideia do som estéreo ainda de maneira experimental, com a utilização de “um canal traseiro (então chamado de “canal de efeitos”) e três ou mais canais dispostos horizontalmente atrás da tela”⁷¹ (Kerins, 2010, p. 25).

Nos ambientes domiciliares, os sistemas *hi-fi*, ou de alta fidelidade, substituíram os antigos rádios, oferecendo em um mesmo módulo o toca-discos e o receptor de rádio, combinado a alto-falantes e amplificadores elétricos. Alguns modelos contavam, inclusive, com aparelhos televisores acoplados, a grande inovação da época. Com a chegada da televisão aos lares dos espectadores a partir de 1945, a indústria cinematográfica viu seu público diminuir de forma numerosa e precisou modificar seus parâmetros de exibição para permanecer atrativa ao público em geral. Em relação à imagem, a proporção e o tamanho da tela foram modificados, deixando as telas mais largas e muitas vezes ampliando a gama de cores dos filmes, chegando a um formato que a TV não conseguiria imitar na época. Com a imagem mais larga, o som mono não era mais viável, pois não conseguia cobrir a extensão da tela e dar veracidade às imagens. Elsaesser e Hagener (2018, p. 167) destacam que, durante os

⁷¹ Tradução livre da autora. Texto original: one rear channel (then called the “effects channel”) and three or more channels arranged in a horizontal spread behind the screen.

anos 1970, Hollywood respondeu ao problema de como atrair novos públicos ao cinema investindo em novas tecnologias sonoras:

Filmes como *Nashville* (Robert Altman, Estados Unidos, 1975), *Guerra nas estrelas* (George Lucas, Estados Unidos, 1977) e *Apocalypse now* (Francis Ford Coppola, Estados Unidos, 1979) representam marcos na história do cinema, notadamente porque questionaram radicalmente a experiência cinematográfica e seus modos de gravar e reproduzir o som. Outros filmes da mesma época, como *A conversação* (Francis Ford Coppola, Estados Unidos, 1974) e *Um tiro na noite* (Brian de Palma, Estados Unidos 1981), dramatizam e destacam quase obsessivamente as possibilidades oferecidas pelas novas tecnologias sonoras como microfones ocultos, microfones direcionais, sistemas de escuta e de pós-produção.

O som *surround* é consolidado a partir dessa necessidade do cinema se diferenciar da televisão, tanto em imagem, com projeções maiores e a tecnologia Technicolor com três cores, quanto em som, que firma a ideia do multicanal, já que o padrão mono não trazia veracidade à telas tão extensas. “Ficou absolutamente evidente a necessidade de colocar mais canais por trás da tela, fazer uma ‘imagem de som’ tão larga quanto a imagem. Da mesma maneira, a nossa percepção auditiva não é ‘mono’” (Klachquin, 2002).

A qualidade na captação estava demasiadamente melhor, mas as salas ainda enfrentavam um limite técnico, como a potência dos amplificadores e a qualidade das caixas de som, uma vez que o som mono óptico chegava a no máximo 5 Khz; já o magnético atingia 15 Khz ou mais. Uma solução seria distribuir o som pela sala de cinema, utilizando os canais disponíveis e os alto-falantes, criando para a plateia uma maior sensação de envolvimento e realidade. No processo descobriram que nem todo som deveria ser espacializado de acordo com o plano da imagem, principalmente diálogos, pois tornava-se incômodo assistir a um personagem à direita do plano com som saindo nas laterais correspondentes, logo muda o plano e o personagem está no centro com seu diálogo sendo transmitido nas caixas centrais.

A sensação deste “pulo” na fonte dessa voz é desconfortável e não funciona. Então esta experiência mostrou que era melhor deixar o diálogo no centro. Neste caso, se o personagem que está falando está à direita na tela, e a sua voz sai pelo canal central, interpretamos como se o som tivesse vindo da direita, é um efeito psicoacústico. Alguns filmes dessa época têm diálogos endereçados para esquerda ou direita, mas só em longos planos onde não há mudança na posição dos atores ou da câmera. Esta prática de diálogo ancorado no centro é utilizada hoje como padrão, embora em alguns filmes se utilizem planos rápidos de voz em algumas circunstâncias (Klachquin, 2002).

As películas de 70 mm se tornam uma realidade em nível industrial, permitindo um total de seis canais de áudio gravados de forma magnética, cinco canais por trás da tela e um canal *surround*. Na película, quatro canais se localizam em uma pista por fora da perfuração e

um canal por dentro, ao lado da imagem, em cada lado do negativo. Diversas tecnologias e formatos surgiram a partir dessa ambição do mercado em reconquistar o público cativado pela televisão, dentre elas, destacam-se: o *Cinerama* (1950), com som *Warnerphonic*, que utilizava projeção dupla, com banda de áudio magnética de 35 mm, separado da película da imagem. O som contava com sete canais independentes, cinco atrás da tela e dois *surrounds*, além de uma tela grande e curva. De custo altíssimo, o formato não ganhou a indústria e, em 1965, a empresa fechou. O *Super Panavision 70* e o *Cinemascope* (1953), da *Twentieth Century Fox*, conseguiram unir imagem e som no mesmo filme por meio da técnica chamada de “*striping*” (Kerins, 2010, p. 38), uma espécie de ‘colagem’ da fita magnética ao longo das bordas de uma impressão de filme, possuindo ainda quatro canais distribuídos em esquerdo, centro, direito e *surround* mono. O sistema tinha ainda outra vantagem em relação aos outros, pois funcionava utilizando os projetores padrão 35 mm que os expositores já possuíam e conseguia oferecer imagem *widescreen* e som multicanal, apenas com a troca da lente e uma nova cabeça magnética para o som.

No entanto, o sucesso do sistema da Fox resultou tanto das práticas comerciais do estúdio quanto de quaisquer vantagens técnicas: a Fox simplesmente exigiu que os cinemas que desejassem exibir qualquer um de seus filmes *Cinemascope*, incluindo muitos dos lançamentos de alto perfil do estúdio, instalassem tanto no sistema auditivo quanto no visual componentes do sistema *Cinemascope*⁷² (Kerins, 2010, p. 25).

A experimentação do chamado ‘canal de efeitos’ dos sistemas multicanais vigentes acabou identificando problemas em seu uso na época, como a qualidade inferior do som em relação a outros canais e principalmente em como utilizar o canal que dava relativa espacialidade e conferia uma profundidade ao visualizado pela tela achatada. Algumas tentativas foram realizadas, até mesmo com diálogo, mas nenhuma se destacou. Aliado a isso, a política de exclusividade da Fox acabou por minar aos poucos o sucesso do *Cinemascope* e, devido ao alto custo de produção, finalização e exibição, fazendo com que a indústria cinematográfica mantivesse o padrão de gravação em 35 mm e o som óptico e mono. Nem mesmo o padrão magnético, que parecia ter agradado a todos, conseguiu seu lugar em larga escala nesse momento, pois trocar os equipamentos já instalados para óptico geraria um custo extra que os exibidores não estavam dispostos a pagar, principalmente com a receita ainda em queda desde a chegada da televisão.

⁷² Tradução livre da autora. Texto original: Yet the success of Fox’s system stemmed as much from the studio’s business practices as from any technical advantages: Fox simply required theaters wanting to show any of its CinemaScope films, including many of the studio’s high-profile releases, to install both the aural and visual components of the CinemaScope system.

O estéreo não deixou de existir totalmente e seguiu acontecendo em grandes filmes, voltados ao ‘espetáculo’, como *Oklahoma!* (1955), *West Side Story* (1961) e *Ben-Hur* (1959), o que acabou por criar um paradoxo em relação à aceitação do público, que de certa forma permanece até a atualidade. A tecnologia inicialmente monofônica do cinema fez com que os espectadores acostumassem-se a que os sons deveriam ser centralizados atrás da tela, principalmente os diálogos. O público aceitava bem o estéreo em filmes de grandes produções, ‘espetaculares’ e em cinemas especiais, com telas mais largas, como *Cinerama* ou 70 mm, mas não percebiam a necessidade para filmes a que assistiam em cinemas mais populares e de maneira cotidiana. Ao mesmo tempo, a motivação econômica deixou o som estéreo reservado a locais reduzidos e sublimes, deixando o som mono como padrão comum. Logo, estabeleceu-se um paradoxo, uma vez que “o estéreo foi afastado do *mainstream* porque era visto como espetacular demais, e visto como espetacular porque foi usado em apresentações de ‘espetáculos de grande sucesso’ em telas grandes, mas não para exibição *mainstream*”⁷³ (Kerins, 2010, p. 28).

Nos ambientes domésticos, a reprodução em estéreo era uma das importantes evoluções dos dispositivos, pois, na década de 1960, os sistemas integrados, popularmente conhecidos como *hi-fi*, foram aos poucos sendo substituídos por aparelhos modulares que poderiam incluir amplificador, rádio, toca-discos, toca-fitas e duas caixas acústicas separadas, gerando uma reprodução em dois canais (estéreo). Os consumidores mais interessados começaram a montar seus próprios sistemas domésticos, com módulos de marcas diferentes, visando uma melhor qualidade na reprodução sonora. A melhoria nos sistemas sonoros domésticos precede, então, a evolução da exibição sonora em salas de cinema e, de algum modo, acaba por incentivá-la, uma vez que o público já consumia um som de qualidade dentro de casa e entendia que melhores equipamentos intensificavam a experiência de escuta.

Na sonorização cinematográfica, os canais *surround* melhoram em qualidade e diminuição de ruído nas décadas de 1950 e 1960, mas permanecem ‘acessórios’, sendo utilizados mais para ambientação e efeitos ocasionais, devendo se tornar quase imperceptível ao público, a menos que houvesse uma falha de execução. Na tentativa de diminuir gastos e aumentar a produtividade, outras tecnologias surgem e uma das principais é a apresentada pela empresa de Ray Dolby que reduz os ruídos do magnético, que dependia do contato físico entre a fita e a cabeça de gravação, e o coloca novamente como a principal forma de captação

⁷³ Tradução livre da autora. Texto original: stereo was pushed out of the mainstream because it was viewed as too spectacular, and it was viewed as spectacular because it was used in large-screen presentations of “blockbuster spectacles” but not for mainstream exhibition.

sonora em cinema, uma vez que seus custos eram bem menores que os do óptico. Inicialmente, a utilização do sistema Dolby estava voltada à indústria fonográfica, melhorando o som das gravações sonoras com o objetivo principal de agradar o público doméstico. A indústria cinematográfica e sua característica intermediária entende que essa melhoria através do uso dos dispositivos e produtos Dolby poderia aprimorar o som cinematográfico e começa a utilizar essas inovações em suas produções. Com isso, a empresa, segundo Mark Kerins (2010, p. 15), torna-se sinônimo de qualidade sonora cinematográfica e sua história começa a se confundir com a própria história do som no cinema:

No limitado trabalho acadêmico existente sobre o som do cinema no contemporâneo ou sobre as tecnologias sonoras do cinema, os estudiosos muitas vezes até descrevem o som do cinema moderno como “som Dolby” e assumem – aparentemente sem reconhecer que a tecnologia sonora mudou sob os seus pés – que as bandas sonoras de hoje contam com as mesmas regras de seus precursores baseados em Dolby Stereo. Em particular, “som Dolby” é frequentemente usado de forma intercambiável com “som *surround*”, sugerindo que Dolby Stereo é um substituto adequado para todos os sistemas de som *surround*. A verdade é que, em vários aspectos, o Dolby Stereo tem mais em comum com os sistemas monofônicos que substituiu do que com os sistemas *surround* digitais que o substituíram⁷⁴.

Embora o Dolby Stereo seja frequentemente celebrado como o início da era do som *surround*, sua arquitetura técnica ainda mantinha uma estrutura essencialmente frontal e canalizada, muito próxima do modelo monofônico. Conforme veremos adiante, o sistema operava com quatro canais codificados (esquerdo, central, direito e *surround*) comprimidos em uma única trilha óptica estéreo, o que reduzia significativamente a separação espacial e a independência dos sons. Assim, o chamado *surround* era, na verdade, um canal único de ambiência difusa, sem direcionamento preciso nem espacialização envolvente. Desse modo, o Dolby Stereo funcionou mais como uma etapa de transição do que como uma transformação radical da escuta cinematográfica, introduzindo certas possibilidades espaciais para as expansões multicanais e orientadas a objetos que viriam com os sistemas digitais das décadas seguintes.

⁷⁴ Tradução livre da autora. Texto original: “In what limited academic work exists on contemporary cinema sound or on film sound technologies, scholars often even describe modern film sound as “Dolby sound” and assume—apparently not recognizing that sound technology has shifted under their feet—that today’s soundtracks rely on the same rules as their Dolby Stereo-based forerunners. In particular, “Dolby sound” is often used interchangeably with “surround sound,” suggesting that Dolby Stereo is a suitable stand-in for all surround sound systems. The truth is that in key ways Dolby Stereo actually has more in common with the monophonic systems it replaced than it does with the digital surround systems that replaced it”.

3.2 Os sistemas multicanais

Em 1965, é fundada a Dolby, empresa que leva o sobrenome de seu criador, responsável por desenvolver técnicas de redução de ruídos e melhorias em todas as áreas de gravação e reprodução sonora para cinema (Kerins, 2010, p. 1). Inicialmente, a empresa coloca novamente a gravação magnética para cinema no protagonismo tecnológico da época, com redução de ruídos e custo mais baixo. Elsaesser e Hagener (2018, p. 167-168) destacam que a experiência cinematográfica foi modificada ao longo dos anos principalmente pelo som e a forma de captação do ouvido humano:

O sistema de atenuação de ruído Dolby possibilitou a separação de várias trilhas e acelerou a composição de trilhas sonoras sobrepostas. Ele entrou nos cinemas combinado com diferentes sistemas de som *surround* e, juntos, alteraram a paisagem sonora dos filmes. Graças a um conjunto de alto-falantes distribuídos pelo auditório (em vez de escondidos atrás da tela, como nos sistemas de som tradicionais dos cinemas da época clássica), os novos sistemas de som multiplicaram a possibilidade de sobrepor os sons, bem como permitir que o filme "ultrapassasse" os limites da tela e "entrasse" no espaço do espectador. O filme já não existia apenas na tela, estendia-se para o auditório também.

Entre 1965 e 1967, a empresa apresenta o *Dolby A*, um codificador que conquista relativa importância no cenário musical em Londres, adotado primeiramente pela gravadora Decca, despertando a curiosidade sobre sua aplicação no cinema e evidenciando mais uma vez a ideia intermediária do cinema e sua capacidade de absorver tecnologias de diversas áreas em prol de seu desenvolvimento.

Logo foi descoberto que esse sistema permitia que a edição de som no cinema fosse realizada com muito mais "camadas", pois era possível, a partir de então, sincronizar fitas de áudio diferentes e regrava-las em uma fita única, para um formato próprio de reprodução para a montagem e, depois, para a inserção ótica na película a ser exibida nos cinemas, sem aumentar muito o nível de ruído, como acontecia antes. O Dolby também era o mais barato para ser licenciado pelos estúdios, tornando-se, ao longo dos anos, o principal sistema sonoro trabalhado no cinema mundial (Vidigal; Novais, 2020).

O grande problema do custo das cópias e tecnologias de exibição ainda persistia. Seria possível trazer a redução de ruído e os multicanais para o sistema óptico de 35 mm? Uma parceria da Dolby com a Kodak e a RCA propõe alguns experimentos de utilização do sistema Dolby aplicado ao mono de 35mm. "Em 1974 se introduz a matriz MP (*motion pictures*), que codifica quatro canais em dois [...] e nasce assim o sistema *Dolby Stereo* no cinema" (Klachquin, 2002).

A ideia principal do sistema *Dolby Stereo* era desenvolver um som com a qualidade do magnético, mas que poderia ser copiado na forma óptica e ao mesmo tempo da imagem, sem gerar um custo extra. O formato 35mm era o mais popular do cinema até então e inserir os multicanais neste modelo seria fundamental para a consolidação dos sistemas Dolby. A matriz MP codificava quatro canais de áudio em dois para serem gravados na película 35mm e decodificava a trilha de áudio novamente em quatro canais durante a projeção. “O cinema passa a ter som em quatro canais com alta qualidade, equivalente ao magnético, porque não tem um alto nível de chiado, nem tem a distorção do óptico mono e a sua resposta em frequência pode atingir 14 KHz” (Klachquin, 2002). O sistema tinha também algumas desvantagens, como a ausência de um canal *surround* e, posteriormente, quando adicionado, recebia apenas metade da faixa dinâmica dos canais frontais. Além disso, para garantir a compatibilidade, nenhum som determinante para o filme poderia ser adicionado ao canal *surround*, com o risco de não ser tocado nos cinemas mono.

A consolidação do *Dolby Stereo* como um novo padrão a ser conquistado veio com a estreia de *Star Wars*, em 1977, especificamente em sua cena de abertura, quando o som da nave espacial chegando acompanha seu deslocamento imagético, da parte de trás do cinema para a frente, evidenciando o potencial sonoro do sistema. O sucesso iminente agradou o público e os expositores e, menos de três anos após a estreia da saga de George Lucas, o formato já havia sido usado em 85 longas-metragens e os equipamentos instalados em pelo menos 1.200 salas de cinema. Entretanto, muitos desses filmes lançados como Dolby Stereo recebiam configurações de áudio especiais fora das especificações padrão do formato, como em *Contatos Imediatos do Terceiro Grau* (1977), com uma faixa de *subwoofer* dedicada ou *surround* dividido em esquerdo e direito, em *Superman* (1978) (Kerins, 2010, p. 32).

Nos ambientes domésticos, o sistema estéreo ainda era o padrão, com tocadores *Stereo 8* na década de 1960, populares em sistemas automotivos, que incluíam quatro programas estereofônicos por fita; daí, a ideia de oito faixas de áudio gravadas na fita magnética, que apresentava qualidade e estabilidade, mesmo com o veículo em alta velocidade. Além disso, no início da década de 1970, houve a volta dos sistemas “tudo em um” nas casas, com grandes módulos estéreo com duas caixas amplificadas integrados com rádio, toca-fitas e, por vezes, tocadores de discos de vinil. As fitas de áudio desses *players*, conhecidas como *compact-cassettes*, ou apenas fitas-cassete, apresentavam chiados e tinham um nível de fidelidade bem abaixo dos vinis, mas a praticidade, o baixo custo e a possibilidade de o público gravar suas músicas a partir de outras fitas ou de emissoras de rádio fizeram o formato se popularizar e conquistar o mercado. O êxito também se deve à criação do *Walkman*

pela Sony, em 1979, primeiro dispositivo de música portátil, que consistia em um leitor de fitas-cassete que funcionava através de duas pilhas comuns e um par de fones de ouvido estéreo.

Em paralelo ao desenvolvimento fonográfico, em 1971 há o registro da invenção dos dispositivos de armazenamento de dados, chamados popularmente de disquetes ou, em inglês, *floppy-disk*. Esses equipamentos foram amplamente utilizados até a década de 1990 no armazenamento de dados variados, como *softwares*, áudios e imagens, no qual era possível copiar também programas originais para os dispositivos e reproduzi-los em outros computadores. Em 1982, o surgimento do *Compact Disc* (CD) e seus tocadores revolucionou a reprodução de áudio digital, sendo utilizado em residências, escritórios, automóveis e também de maneira portátil, com o lançamento do *Discman*, também pela Sony, em 1984. A grande capacidade de armazenamento do formato, fez com que o CD fosse também o padrão utilizado pela informática na época, em *softwares* originais. A interação entre informática e som era ampliada, na medida em que era possível codificar *softwares* em áudio e gravá-los em fitas-cassete, podendo transmitir os dados via telefone diretamente a outras pessoas e também via rádio, havendo, inclusive, programação dedicada, na época, a transmissão de programas de computador codificados em formato de áudio, como no caso do *Clip Informática*⁷⁵, programa da *Rádio USP* da década de 1980, que trazia reportagens e entrevistas sobre informática e transmitia *softwares* em áudio para que seus ouvintes pudessem gravar em fitas cassetes e posteriormente decodificar o material em computadores compatíveis.

Os aparelhos televisores já recebiam transmissão a cores desde a década de 1970, mas a reprodução sonora em estéreo chega nos dispositivos apenas a partir da década de 1980, com dois alto-falantes acoplados ao aparelho, uma aurora para os *home theaters* que, aliado à popularização do aparelho de videocassete e de amplificadores receptores, começou a difundir a ideia do “cinema em casa”. Em 1985, já havia 10 padrões televisivos, combinados com três sistemas de cores, resultando em 14 sistemas de transmissão de TV diferentes em todo o mundo. Mesmo com aparelhos saindo com configuração em estéreo, as estações de TV comprimiam (e ainda comprimem) o som antes da transmissão, um processo que reduz consideravelmente a faixa dinâmica, ampliando a distorção e reduzindo o conteúdo de alta frequência. Formatos como o VHS, *LaserDisc* e posteriormente o DVD popularizaram o consumo de conteúdo cinematográfico fora das salas de cinema e sem depender das exibições televisivas. O mercado audiovisual doméstico desenvolve a partir dessas inovações

⁷⁵ Um exemplar do programa pode ser escutado no link disponível no *YouTube*: <https://youtu.be/TdHhqd7Uus?t=1361>. Acesso em 31/07/2024.

técnicas e conquista o público que, com a popularização de videolocadoras, começa a ter acesso a filmes antes inalcançáveis por diversos motivos, como a falta de salas de cinema nas cidades, o custo do ingresso e também a curadoria das salas, que acabam de certa forma sempre privilegiando lançamentos comerciais.

Na reprodução cinematográfica, a Dolby apresenta, em 1987, uma atualização através do sistema SR (*Spectral Recording*), que funcionava com a mesma premissa do *Fantasound*, mas utilizando agora a gravação magnética, conseguindo reduzir ainda mais o ruído dos filmes 35 mm. Em 1992, a Dolby lança sua versão digital no filme *Batman Returns*, com redução de ruídos ainda maior que a versão anterior, potência máxima dez vezes maior que o sistema SR e canal *surround* dividido entre direita e esquerda. O novo formato seguia o padrão de seis canais, cinco frontais e um *surround* (5.1), com a trilha localizada entre as perfurações da película e lida através dos processadores Dolby. O sistema também tinha a vantagem de incluir na trilha das laterais da imagem o formato análogo ao SR, criando o formato SRD, que era possível ser reproduzido em qualquer sala de cinema da época, mesmo as que continham apenas a tecnologia mono.

Como a trilha análoga Dolby SR ocupa o mesmo lugar na película que a trilha mono de 1937, significa que quando projetamos um filme SRD num projetor mono, ele vai reproduzir o som sem inconvenientes porque as pistas Left Total e Right Total são somadas oticamente e compõe um som mono compatível (Klachquin, 2002).

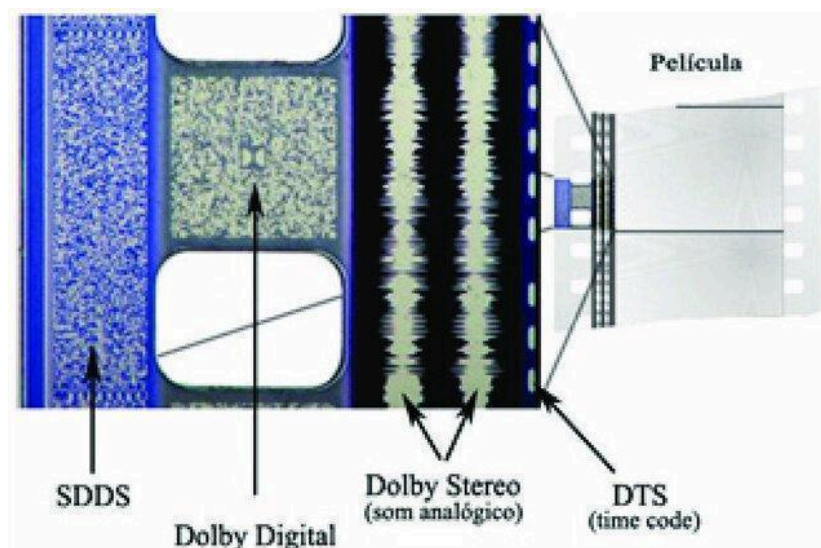
A modernização das salas para o formato digital ainda era um empecilho por custar em torno de US\$ 20 mil na época. Por isso, na estreia apenas dez cinemas nos Estados Unidos e no Canadá tinham a tecnologia e apenas uma sala na Grã-bretanha conseguiu exibir o *Batman* (1989) dirigido por Tim Burton em SRD. Esse alto custo de implantação acabou abrindo espaço para a concorrência, que lançou o sistema DTS, em 1993, com a estreia de *Jurassic Park*. Desenvolvido pela *Digital Theatre Systems*, o sistema mantinha os 5.1 canais e custava quatro mil dólares, um quinto do valor de instalação do equivalente da Dolby. Uma das possíveis justificativas para o menor custo é que a trilha sonora do filme chegava em um CD-ROM - e, posteriormente, em DVD - separado, e uma trilha simples de *timecode* no filme mantinha a sincronização entre disco e projetor. “Este sistema de mídia dupla permitiu que a DTS usasse uma taxa de bits mais alta e, portanto, menos compactação dos dados de áudio originais do que seus concorrentes”⁷⁶ (Kerins, 2010, p. 42) e acabou se espalhando de maneira

⁷⁶ Tradução livre da autora. Texto original: “This dual-media system allowed DTS to use a higher bitrate, and hence less compression of the original audio data, than its sound-on-film competitors”.

mais efetiva pelas salas, atingindo a marca de mais de mil cinemas equipados para a estreia do filme de Steven Spielberg. Logo após o DTS, a Sony também lançou sua versão de som digital através do sistema *Sony Dynamic Digital Sound* (SDDS), com o filme *O último grande herói* (1993). O SDDS adicionava mais dois canais atrás da tela, além do 5.1 já oferecido pelas outras empresas, resultando em uma configuração 7.1, com cinco canais frontais e dois *surround*, a área utilizada para imprimir o som eram as bordas externas da película fora dos furos da roda dentada, uma área menos protegida do que entre os furos, como o sistema Dolby oferecia.

Essa diferença de apresentação foi crucial para que os estúdios, após alguns anos de negociações, concordassem em incluir os três tipos de sons digitais disponíveis no mercado em um mesmo rolo de filme, acrescentando a mídia sonora do DTS e a faixa de *Dolby Stereo* para as salas ainda desatualizadas, facilitando para o mercado exibidor, que antes precisava instalar três tecnologias diferentes ou se contentar em exibir o filme em uma qualidade sonora inferior.

Figura 01 - Localização dos sistemas sonoros na película.



Fonte: Revista Tecnologia e Tendências, 2019.

Três anos após os lançamentos, os sistemas de som *surround* digital já eram considerados o padrão para os grandes estúdios de Hollywood. “O som cinematográfico é, daí em diante, multicamada, multidirecional e consiste, entre outras coisas, de ruídos e sons que não são nem naturais na origem nem produzidos por instrumentos musicais (ruídos eletrônicos, amostras sonoras, som digital)” (Elsaesser; Hagener, 2018, p. 168).

Entre as décadas de 1980 e 1990, outras tentativas de inserção de formatos foram realizadas no âmbito do mercado doméstico, como o *Digital Audio Tape*, que consistia em um tipo mais avançado de fita magnética, capaz de registrar bits e apresentar maior fidelidade e capacidade de memória e o MiniDisc, mais compacto e barato que o modelo anterior e com boa fidelidade sonora. Apesar das vantagens dos lançamentos, o público permaneceu fiel aos CDs, principalmente pelo surgimento dos modelos regraváveis, o custo baixo e a boa compatibilidade em diversos dispositivos, como domiciliares, empresariais, automotivos e portáteis. No final dos anos 1990 e início dos anos 2000, surge o formato MPEG-3, que viria a se popularizar como MP3, um compactador que permite comprimir os sinais de áudio e armazenar em dispositivos de memória. Com o surgimento da internet, o formato começou a ser amplamente utilizado para armazenar arquivos sonoros que poderiam ser reproduzidos em qualquer computador. Logo, surgiram também tocadores portáteis do formato, sendo o mais popular o iPod da Apple, lançado em 2001, que armazenava cerca de 1000 músicas simultâneas em formato em MP3 e as reproduzia em estéreo através de fones de ouvido.

No mercado cinematográfico, os sistemas Dolby e DTS permanecem ativos, inclusive com lançamentos também para ambiente domiciliar, enquanto o SDDS recebeu sua última atualização em 2007 e foi descontinuado pela Sony. Com a digitalização das projeções, o som pode se descompactar e ganhar mais qualidade com menor compressão. O lançamento de atualizações nos formatos, como *Dolby EX* (1999), com um canal *surround* no fundo da sala, em formato 6:1 e o Dolby 7:1 em 2010, apontam o caminho das duas empresas continuarem coexistindo. Além disso, a utilização de *codecs*⁷⁷, como *Dolby TrueHD* e *DTS-HD Master Audio*, oferecem redução no tamanho do armazenamento de áudio, deixando mais espaço direcionado para imagens ou outras informações sem perda de qualidade. Kerins (2010) aponta alguns caminhos que o som cinematográfico e doméstico deve tomar, pensando principalmente no aumento do número de canais e expansão do *surround*, mas alerta que o desenvolvimento da tecnologia não é suficiente para uma nova padronização de formato, pois a economia, a logística e as preferências do espectador também têm papel importante nesse imbróglcio audiovisual.

Em 2012, a Dolby lançou, com a estreia de *Valente*, da Pixar-Disney, o sistema *Dolby Atmos*, que oferece uma ampla possibilidade de inclusão de objetos sonoros pela sala, sendo compatível com, no mínimo, 9:1 canais (Vidigal; Novais, 2020). A partir do Atmos para cinema se desenvolve o *Home Atmos*, voltado a ambientes domésticos e dispositivos móveis.

⁷⁷ Compactador e descompactador que permite que o áudio master original possa ser reconstruído sem perdas a partir de dados comprimidos.

Na mesma época, a empresa Auro lança outra tecnologia de mixagem orientada a objetos, o Auro 11.1 e Auro 13.1 e, em parceria com a empresa de *hardwares* Barco, começa a instalar seu sistema em algumas salas de cinema e faz sua estreia com o filme *Red Tails* (2012), da *Lucasfilm*. Já a DTS atualizou seu sistema apenas em 2015, quando apresentou sua versão orientada a objetos, o DTS:X. As três empresas possuem atualmente versões para uso doméstico e possibilidades de escuta através de fones de ouvido, que abordaremos adiante.

3.3 Mixagem orientada a objetos

Se o desenvolvimento do som para cinema e uso doméstico foram pautados pelo número de canais sonoros disponíveis e como eram divididos e utilizados, a ideia da mixagem orientada a objetos sonoros vem desconstruir um pouco o modelo de som direcionado a canais e passa a integrar a possibilidade dos sons serem atribuídos a ângulos de posicionamento, mesmo que ocupem mais de um canal simultaneamente.

A diferença basal entre os sistemas anteriores e os especializados é a distinção entre a mixagem baseada em canais e a mixagem baseada em objetos. A tecnologia baseada em canais ainda é a forma mais reconhecida pelo público de se criar e utilizar sons para filmes e audiovisual em geral. Os formatos sempre citados como 5.1, 7.1 e outros se referem a exemplos de organizações de alto-falantes em sistemas bidimensionais, com o eixo X sendo representado pelas caixas frontais e o eixo Y pelo surround. O padrão Dolby 5.1, por exemplo, é constituído por cinco canais de escala completa conectados aos alto-falantes distribuídos em cinemas geralmente na forma de esquerda, centro, direita, surround esquerdo e surround direito. Já o “.1” canal se refere ao subwoofer, um alto-falante dedicado à reprodução de sons de baixa frequência, ou seja, mais graves, localizado usualmente na frente do público, duas vezes mais forte que os outros canais, uma vez que para os canais graves necessitam de mais energia (potência) para serem percebidos pelo ouvido humano com o mesmo volume subjetivo. Concorrentes da Dolby, como DTS e Sony, também utilizam sistemas com layouts semelhantes. As atualizações como o 7.1 acrescentam mais dois canais surround, separando os sons laterais dos sons vindos do fundo da sala.

A adição de mais alto-falantes ao redor e também acima do espectador e, conseqüentemente, a inclusão de canais de altura, reproduzidos por alto-falantes localizados acima da área de audição (geralmente embutidos no teto ou voltados para cima), ampliaram a capacidade do som surround, chegando ao que hoje conhecemos como som imersivo ou envolvente. O desenvolvimento das plataformas de som baseadas em objetos e independentes

de canal surgiu a partir da necessidade da tecnologia se basear em qualquer configuração de alto-falante e em atendimento às formas de consumo de entretenimento atuais, nas quais não são consideradas apenas as salas de cinema e aparelhos de televisão ou de som, mas a utilização dos alto-falantes dos celulares, dos soundbars e toda a possibilidade de escuta inerente ao uso de fones de ouvido em dispositivos domésticos e móveis.

No áudio baseado em objeto, a posição do som pode ser definida em relação a coordenadas em três eixos independentes de alto-falantes, ou seja, uma determinada sonoridade pode ser posicionada a partir da escolha de um ângulo, não necessariamente em um canal ou direcionado a alguma caixa específica. A distribuição dos sinais de áudio acontece nos eixos XYZ e cria os objetos sonoros que podem se mover dentro da sala. A inovação não é apenas a mixagem em si, mas principalmente os metadados⁷⁸ complementares, ou seja, indicações em texto que informam ao sistema qual caixa ou em qual ângulo determinada sonoridade deve ser tocada.

Um som pode estar vinculado a um ou mais canais, em intensidades diferentes, desde que descrito nos metadados que são transferidos junto com o filme, podendo também ser adaptável a diversas configurações sonoras. Dessa forma, os objetos sonoros adicionados funcionam como uma camada complementar de informação ao som transmitido junto com o conteúdo assistido. “Livre de restrições de canal, o som pode ser colocado e movido para onde você ouviria no mundo real – desde aviões acima de você, até multidões ao seu redor, até tudo no meio – proporcionando excelente som em praticamente qualquer *layout* de alto-falantes” (DTS *White paper*, 2015).

Os sistemas ou plataformas lançados com a ideia de mixagem orientada a objetos funcionam a partir da utilização de dois tipos de conteúdo de áudio: *beds* e objetos. As *beds* (ou camas, em português)⁷⁹ são estruturas semelhantes aos sistemas anteriores já conhecidos e integram o conteúdo baseado em canais, funcionando melhor com sons mais estáticos, como ambiências e reverberação. O desenvolvimento desse tipo de áudio pode se dar de maneira análoga aos tradicionais 5.1 e 7.1 ou incluir os canais de altura, como no formato 7.1.2, em que há dez canais: 7 tradicionais (esquerdo, central, direito, *surround* esquerdo, *surround* direito, *surround* traseiro esquerdo e *surround* traseiro direito); .1 *subwoofer* amplificado; .2 canais suspensos (no teto).

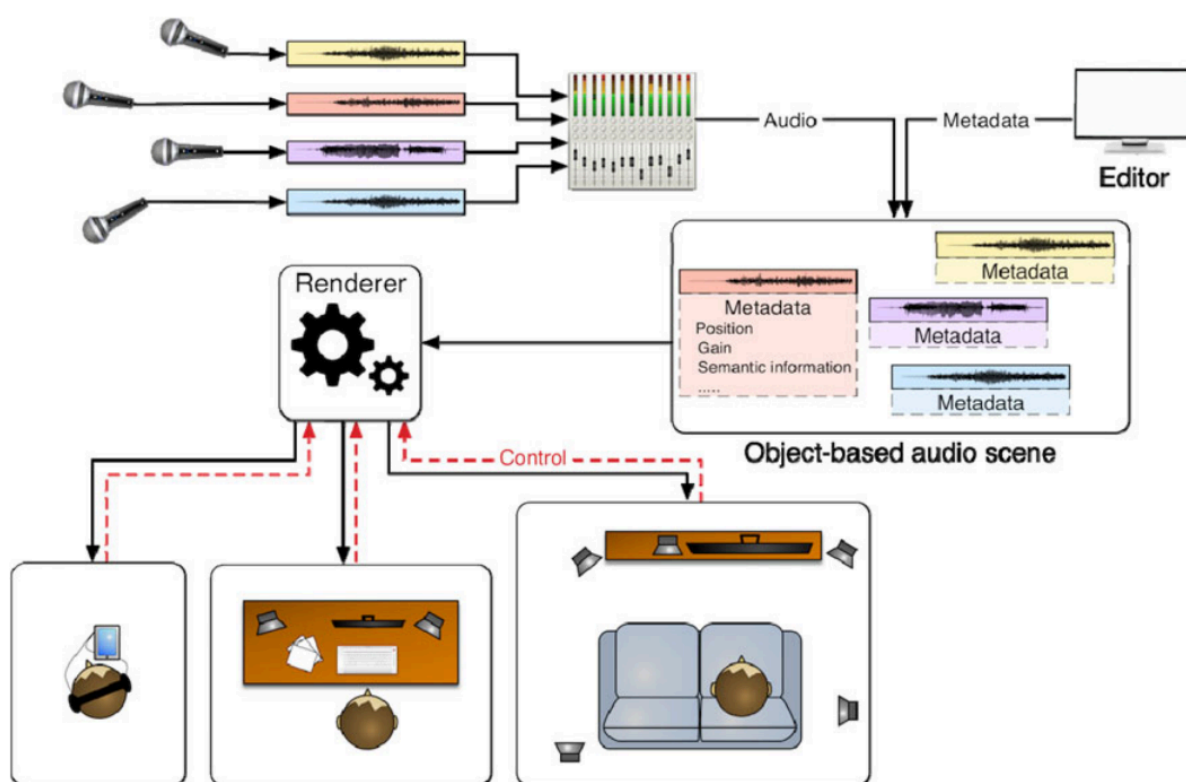
⁷⁸ Entende-se por metadados um conjunto de informações, como pontos de referência, formato e direitos de uso sobre outros dados, permitindo um detalhamento do material sob todas as formas e conteúdos de uma fonte.

⁷⁹ A autora optou por utilizar a tradução *cama* ou *cama sonora* para definir os canais do tipo *beds* dentro da possibilidades de mixagem baseada em objetos.

Já os objetos sonoros são os componentes que podem ser localizados de maneira angular com precisão. A renderização acontece de maneira individual e com a inclusão dos metadados que indicam a posição e o momento de reprodução. Além disso, os objetos podem se mover pelos alto-falantes e são mais apropriados para elementos de áudio dinâmico. Sun (2021, p. 2) aponta que essa descrição por metadados mudou fundamentalmente o paradigma do fluxo de trabalho de criação de conteúdo, pois os profissionais de mixagem não ficam mais limitados a um arranjo fixo de alto-falantes ao descrever uma cena sonora:

Quer se trate de alto-falantes 5.1, fones de ouvido estéreo ou outros dispositivos, a mixagem é feita por algoritmos de *software* no mecanismo de renderização. O conceito de “mixar uma vez, tocar em qualquer lugar” é apontado como um dos pontos de venda mais fortes dos novos sistemas de áudio envolvente⁸⁰.

Figura 02 - Produção e consumo do áudio baseado em objetos.



Fonte: Xuejing Sun, 2021.

Algumas padronizações foram estabelecidas desde 2018, em publicações da *Society of Motion Picture and Television Engineers* (SMPTE), como o ST 2098-1 para metadados de

⁸⁰ Tradução livre da autora. Texto original: “Whether it’s 5.1 speakers, stereo headphones, or other devices, mixing is done by software algorithms in the rendering engine. The concept of “mix once, play everywhere” is flagged as one of the strongest selling point of the new immersive audio systems”.

áudio imersivo, o ST 2098-2 sobre o mapeamentos de dados auxiliares verticais de metadados, além da especificação *Immersive Audio Bitstream* (IAB) e o ST 2098-5 que trata dos canais de áudio imersivo e grupos de campo sonoro. Essas orientações objetivam a entrega de um fluxo de dados coerente e padronizado, simplificando a distribuição e permitindo a entrega de uma mixagem que se adapte aos diferentes sistemas ou plataformas disponíveis. Segundo a publicação *Immersive audio, capture, transport, and rendering: a review* (2021), de Xuejing Sun, o padrão SMPTE é amplamente baseado no formato Dolby Atmos, mas foi projetado para ser extensível e compatível com versões anteriores. “Testes de interoperabilidade⁸¹ entre os principais sistemas de áudio imersivo (Dolby Atmos, Barco Auromax e DTS:X) já foram realizados com sucesso”⁸² (p. 14).

Para a realização da mixagem tridimensional, Tsingos (2018) classifica os referenciais sonoros como alocêntrico, ou ambiental e egocêntrico, focado no observador. “Um referencial egocêntrico codifica a localização do objeto em relação à posição (localização e orientação) do observador ou ‘eu’. Um referencial alocêntrico codifica a localização do objeto usando localizações e direções de referência em relação a outros objetos no ambiente” (p. 245). As referências egocêntricas definem a localização de um som a partir da posição e orientação do ouvinte, ou seja, a experiência sonora é descrita em relação ao corpo do sujeito. Por exemplo, quando se diz que um som é percebido “à direita do espectador”, adota-se um ponto de vista (ou de escuta) egocêntrico, comumente utilizado nos estudos perceptivos e nas representações cognitivas do espaço sonoro. Já o referencial alocêntrico descreve a posição dos sons com base em um sistema externo ao ouvinte, geralmente atrelado à cena ou ao ambiente. Nesse caso, a posição do som é codificada em relação a outros objetos ou coordenadas fixas do espaço, como ao afirmar que um som “ocorre atrás da árvore no canto da sala”, independentemente de onde o ouvinte esteja localizado. Essa abordagem é predominante em aplicações técnicas, como nas interfaces de programação de áudio em videogames ou ambientes imersivos, que geralmente expressam a posição dos objetos sonoros em coordenadas cartesianas do espaço tridimensional. A distinção entre esses referenciais também evidencia a complexidade dos processos de codificação e renderização do som em sistemas avançados de áudio imersivo, nos quais a cena sonora precisa ser ajustada dinamicamente em resposta aos movimentos do ouvinte, convertendo posições alocêntricas em percepções egocêntricas coerentes.

⁸¹ Interoperabilidade são protocolos e tecnologias que permitem que os dados fluam entre diversos sistemas com o mínimo de intervenção humana, permitindo que os sistemas conversem entre si e compartilhem informações.

⁸² Tradução livre da autora. Texto original: Interoperability tests among major immersive audio systems (Dolby ATMOS, Barco Auromax, and DTS:X) have already been conducted with success.

O som do cinema moderno (da era do som *surround*) utiliza um referencial aloccêntrico. Os pontos de referência são as localizações nominais dos alto-falantes (por exemplo, L, C, R) ou zonas de alto-falantes (por exemplo, matriz *surround* lateral esquerda, matriz *surround* superior direita). Essas localizações possuem uma correspondência conhecida e consistente com os elementos importantes do ambiente cinematográfico: a tela, o público e a sala. Elas também têm uma correspondência conhecida e consistente com as ferramentas de autoria: controle deslizante esquerda/direita (*fader*), posição do *joystick* e a interface gráfica do usuário (GUI). Dessa forma, quando um controle de mixagem está totalmente à esquerda e à frente, entende-se que o som será reproduzido por um alto-falante localizado nominalmente na borda esquerda da tela. Todos os metadados de localização são gerados e decodificados usando esse referencial. Isso se aplica tanto a objetos quanto a canais. Somente ao utilizar o mesmo referencial para objetos e canais é possível garantir que a relação espacial entre eles seja preservada (Tsingos, 2018)⁸³.

Diversos sistemas surgiram com a ideia da utilização de camas e objetos sonoros, mas alguns se destacaram, como o AuroMax, o DTS:X, o Dolby Atmos e o MPEG-H. As possibilidades deste tipo de mixagem vão muito além das salas exibidoras, incluindo o consumo doméstico em televisores, *home theaters*, *soundbars* e dispositivos móveis, como celulares e *tablets*, o uso através de fones de ouvido, com o sistema binaural e até mesmo em automóveis com os alto-falantes dos carros.

O chamado binaural é um termo que descreve a audição humana cuja intenção é reproduzir a maneira como usamos os ouvidos para criar um campo sonoro. Ainda que o termo binaural esteja frequentemente associado a tecnologias contemporâneas de áudio imersivo, sua origem remonta a experimentos realizados no final do século XIX, quando as primeiras tentativas de capturar e reproduzir a espacialidade do som começaram a surgir. Em 1881, o engenheiro Clément Ader demonstrou o *Théâtrophone*, na Exposição Internacional de Paris. O sistema usava dois microfones posicionados à esquerda e à direita do palco da Ópera, transmitindo os sinais sonoros por linhas telefônicas para dois fones de ouvido. A experiência visava simular a escuta em estéreo natural, e já antecipava princípios fundamentais da escuta binaural moderna.

A noção de escuta binaural parte da própria fisiologia da audição humana que por meio da diferença de tempo (ITD – *Interaural Time Difference*) e intensidade (ILD –

⁸³ Tradução livre da autora. Texto original: “Modern (surround-sound era) cinema sound uses an allocentric frame of reference. The reference points are the nominal location of loudspeakers (e.g., L, C, R) or loudspeaker zones (e.g., left-side surround array, right-top surround array). These locations have a known and consistent mapping to the important features of the cinema environment: the screen, the audience, the room. These locations also have a known and consistent mapping to authoring tools: left/right fader, joystick position, and the GUI. In this way, when a mix control is full-left and full-front it is understood that sound will be reproduced by a loudspeaker that is nominally located at the left edge of the screen. All location metadata are generated and decoded using this reference. This applies to both objects and channels. Only by using the same frame of reference for both objects and channels can we ensure that the spatial relationship between objects and channels is preserved”.

Interaural Level Difference) entre os sinais que chegam a cada um dos ouvidos é possível localizar sons no espaço. Baseado nesse princípio, a gravação e reprodução de conteúdo de áudio e audiovisual visam simular a forma de escuta humana. Experimentos de audição binaural exigem que dois sinais distintos em amplitude ou fase cheguem aos ouvidos esquerdo e direito. Cada ouvido recebe ondas sonoras de uma maneira particular, dependendo do formato e das características físicas da cabeça, orelhas, tronco, além das cavidades nasais e orais, que são interpretadas pelo cérebro e nos permitem localizar o som no espaço. A acústica denomina esses dados anatômicos e sua influência da percepção sonora no ambiente como *Head-Related Transfer Functions*⁸⁴ (HRTF), ou Funções de transferência relacionadas à cabeça, que são medições que representam de forma matemática as pistas espectrais para determinar a melhor localização sonora de acordo com as características médias humanas. O uso de fones de ouvido faz com que os sons sejam escutados através de uma mesma distância, seja perto ou dentro das orelhas, mas, com o rastreamento dos movimentos e distâncias da cabeça e suas interpretações cerebrais, as gravações com microfones binaurais (Figura 3) e a mixagem baseada em objetos conseguem simular as distâncias sonoras. “As cabeças falsas e as pistas binaurais/HRTF simuladas não são uma combinação perfeita para nenhum ouvinte individual, mas para muitas pessoas elas funcionam bem”⁸⁵ (DellaSala, 2014). Isso acontece pois as medições HRTF são uma média relativa a variadas medidas humanas, que se aproximam entre si, mas nunca se repetem entre os indivíduos.

As pistas para sons que chegam de cima estão associadas a ressonâncias direcionais sensíveis no ouvido externo na faixa de frequência de cerca de 7 a 12 kHz. Elas podem ser muito diferentes para pessoas diferentes porque nossos ouvidos são todos fisicamente diferentes. No entanto, tomando uma média de muitos ouvidos, encontra-se uma tendência geral indicando que, à medida que uma fonte sonora é elevada no plano mediano (diretamente para a frente neste caso), há um aumento no nível sonoro que atinge o tímpano em frequências em torno de 7 a 8 kHz. Há outra ressonância direcionalmente sensível no ouvido externo em torno de 12 kHz, mas ela é muito menos previsível porque, à medida que as frequências aumentam, diferenças físicas menores entre as orelhas resultam em maiores variações de resposta entre os indivíduos⁸⁶ (DellaSala, 2014).

⁸⁴ HRTF são dados de engenharia sonora e técnicas de gravação binaural, que capturam sons através de microfones posicionados como os ouvidos humanos com a intenção de criar uma experiência de áudio tridimensional e espacial.

⁸⁵ Tradução livre da autora. Texto original: “The dummy heads and simulated binaural/HRTF clues will not be a perfect match for any individual listener, but for many people they work well”.

⁸⁶ Tradução livre da autora. Texto original: “The cues for sounds arriving from above are associated with directionally sensitive resonances in the external ear in the frequency range of about 7 - 12 kHz. They can be very different for different people because our ears are all physically different. Taking an average of many ears though, one finds a general trend indicating that as a sound source is elevated on the median plane (directly forward in this case) there is an increase in sound level reaching the eardrum at frequencies around 7-8 kHz. There is another directionally sensitive resonance in the external ear around 12 kHz, but it is much less

Figura 03 - Microfone binaural.



Fonte: 3Dio Sound, 2024.

Essa tecnologia permaneceu durante algum tempo restrita a contextos científicos, musicais experimentais ou demonstrações técnicas. Contudo, com o crescimento das tecnologias de consumo baseadas em realidade virtual, jogos imersivos, experiências sonoras em 360 graus e os sistemas de áudio orientado a objetos, o binaural ganhou relevância nas discussões sobre espacialidade no som. Diferentemente da mixagem estéreo convencional, a abordagem binaural contemporânea explora modelagens acústicas precisas, capazes de simular a percepção espacial a partir de qualquer ponto de escuta, levando a uma maior imersão para conteúdos em realidade virtual e também cada vez mais presente no audiovisual consumido através de dispositivos móveis.

A anatomia binaural humana permite, portanto, a localização da origem dos sons no espaço através da análise das diferenças entre os tempos de chegada dos sons nos dois ouvidos e é essa capacidade que consegue transformar a audição com uma ilusão de reprodução multicanal e multidirecional. Da mesma forma, utilizando a mixagem orientada a objetos e fones de ouvido convencionais, é possível gerar um campo sonoro que se aproxima da escuta do mundo real, com aplicações diversas, como cinema, jogos, músicas, *podcasts* e outros formatos. Ao realizar uma análise filmica com foco na construção do som é interessante que se retome o conceito do “ouvido técnico”, proposto por Michel Chion (2009), isto é, aquele que contém todo o som da narrativa e que deveria ser escutado pelo público. Essa noção não deve ser compreendida apenas como uma metáfora, mas como uma dimensão

predictable because as frequencies get higher smaller physical differences between pinna result in greater response variations between individuals”.

a ser explicitada em qualquer análise sonora. Trata-se de reconhecer que toda escuta é atravessada por mediações técnicas, contextuais e culturais, o que implica situar sob quais parâmetros a escuta analítica foi conduzida. Assim, torna-se necessário indicar de maneira clara o sistema de reprodução utilizado, as condições de audição (sala de cinema, estúdio, fones de ouvido, ambiente doméstico, entre outros) e o tipo de “ouvido técnico” mobilizado pelo pesquisador. Essa explicitação não apenas confere rigor metodológico à análise, mas também permite compreender como as condições materiais e tecnológicas moldam a percepção e a interpretação do som no audiovisual.

O sistema auditivo humano é capaz de detectar sonoridades na faixa de frequência de cerca de 20 Hz a 20 kHz. Em experimentos binaurais realizados pelos autores Florian Völk, Jörg Encke, Jasmin Kreh e Werner Hemmert e publicados no artigo *“Relevance of Headphone Characteristics in Binaural Listening Experiments: A Case Study”* (2017), foi possível verificar, ainda que de maneira primária, que fones de ouvido do tipo circumaural, isto é, os que cobrem totalmente as orelhas e “isolam” o som podem induzir uma amplitude audível entre os ouvidos e produzir diferenças de fase, ou seja, tanto em aplicações domésticas, quanto em estúdio, podem haver disparidades sonoras audíveis que podem resultar em sensações auditivas ampliadas ou deslocadas lateralmente. “Em casos extremos e dependendo do material, estas disparidades podem provocar sensações auditivas adicionais, em posições não pretendidas pelo engenheiro de mixagem”⁸⁷ (p. 9).

Conforme veremos no decorrer do capítulo, o uso de fones de ouvido para consumo audiovisual doméstico tem se tornado cada vez mais comum e, com a popularização de tecnologias de mixagem orientada a objetos e a virtualização de canais e sons posicionados, as possibilidades de escuta se ampliaram e tentam oferecer uma experiência sonora tridimensional a partir de aparelhos domésticos e dispositivos móveis. “Oferecer uma experiência de áudio envolvente por meio de fones de ouvido sempre atraiu muitos interesses devido ao seu amplo acesso à base de usuários em massa”⁸⁸ (Sun, 2021, p. 21). A partir desses conhecimentos em mixagem orientada a objetos e camadas sonoras e em gravação e reprodução em sistemas binaurais, o objetivo a seguir é investigar o estado da arte, apresentando as principais empresas e o funcionamento de seus sistemas de mixagem mais modernos.

⁸⁷ Tradução livre da autora. Texto original: In extreme cases and depending on the material, these disparities may elicit additional hearing sensations, at positions not intended by the mixing engineer.

⁸⁸ Tradução livre da autora. Texto original: Delivering immersive audio experience through headphones has always attracted a lot of interests due to its wide access to mass user base.

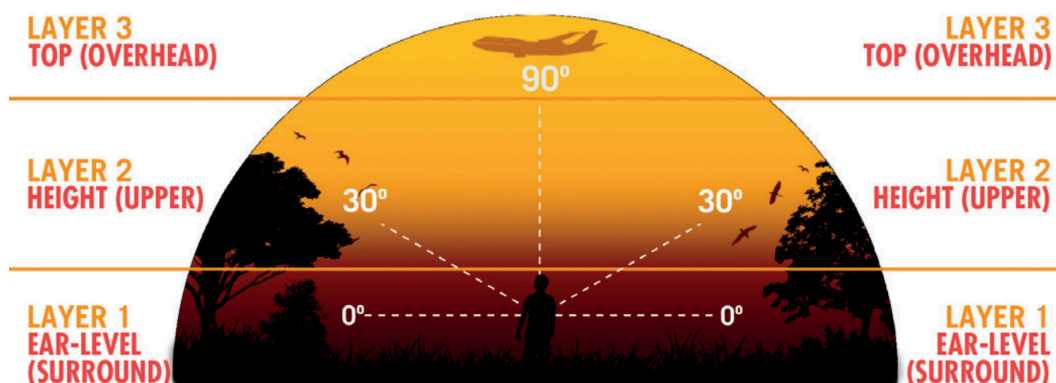
3.3.1 Auro-3D

Auro-3D é uma tecnologia de áudio imersivo desenvolvida pela empresa belga Auro Technologies em parceria com a Barco. Os primeiros conceitos do sistema surgiram em 2005, através do estúdio de criação *Galaxy Studios* e por seu responsável Wilfried Van Baelen, que objetivava o desenvolvimento da tecnologia de áudio tridimensional. Em 2011, a Auro fez seu primeiro lançamento cinematográfico com *Red tails*, da LucasFilm, utilizando o sistema Auro 11.1. Desde então, manteve outros lançamentos e, em 2012, fechou uma parceria com a empresa de animação DreamWorks para mixar 15 de seus próximos lançamentos, entre eles *Os Croods* (2013) e *Turbo* (2013).

3.3.1.1 Auro 11.1

O sistema Auro 11.1 compõe a ideia de som imersivo da Auro-3D adicionando uma terceira dimensão ao som mixado, a partir de um *layout* específico com três camadas de alto-falantes, inclusive uma camada de altura, e seis canais na tela. A primeira camada, ao nível do ouvido, é baseada no padrão *surround* 5.1 canais e é utilizada para ambiências e sons mais realistas. A camada de altura é uma cópia da camada *surround*, mas posicionada mais acima nas paredes, a cerca de 30° de elevação, auxiliando na reprodução de sons elevados e na inclusão de reflexos sonoros que ajudam na criação do chamado campo sonoro envolvente. “A combinação de ambas as camadas cria um ‘campo estéreo vertical’ ao redor do público, que é o elemento chave para a experiência sonora natural e envolvente do Auro-3D®” (*Auro Technologies*, p. 6). Já a camada superior é posicionada a 90° do público, ou seja, no teto e deve ser utilizada, segundo as diretrizes da empresa, para a reprodução de efeitos especiais como sobrevoos e o voz *over*. O sistema 11.1 ainda baseava-se em canais e não em objetos, como suas atualizações. A empresa afirma, em seu material de divulgação, que a camada de altura consegue criar um campo sonoro imersivo ao redor do público ao utilizar os alto-falantes do teto e, como segunda camada é uma duplicação da camada *surround*, haveria uma compatibilidade com as versões anteriores.

Figura 04 - Conceito três camadas do Auro-3D.



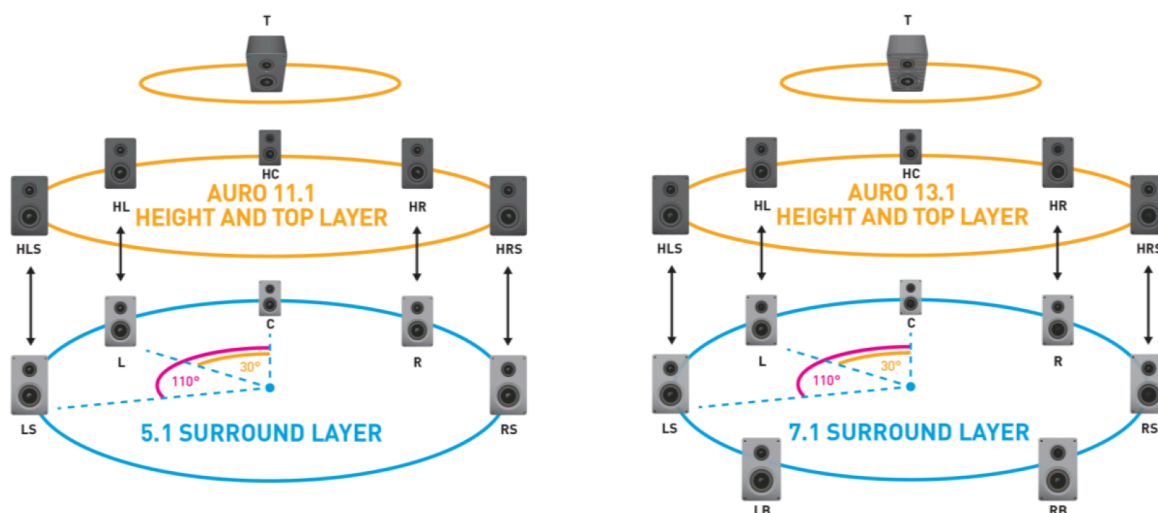
Fonte: AuroMax White Paper (2015).

3.3.1.2 AuroMax

A AuroMax é a versão atualizada do som do Auro 11.1, com compatibilidade da experiência auditiva da versão anterior, adicionada à mixagem orientada a objetos e um fluxo de bits padronizado para cinema. O sistema foi desenvolvido para funcionar de maneira híbrida, combinando canais e objetos sonoros ao mesmo tempo. O fluxo de trabalho idealizado por seus criadores envolve primeiro o desenvolvimento de canais com atmosfera imersiva e depois a adição de objetos sonoros específicos na mixagem. Esses objetos podem ser localizados em posições ao redor ou acima do espectador, dependendo principalmente do número de canais de reprodução, alto-falantes e amplificadores.

O AuroMax utiliza uma arquitetura de canal baseada no *surround* digital 5.1 canais, mantendo a experiência auditiva e a compatibilidade com formatos distintos como 7.1, 11.1 e 13.1, além de múltiplos canais de altura, dedicados à camada superior. Os auto-falantes mantêm as matrizes laterais, traseiras e superiores, posicionados em camadas ao redor do ambiente e permitindo a reprodução do áudio em três dimensões, ao nível do ouvido, a 30° de altura em relação ao espectador e a 90°, no teto. A ideia híbrida do sistema parte da combinação das camadas de quaisquer formato Auro-3D, seja com 12, 14 ou mais canais, conforme ilustra a figura abaixo (5), com os objetos sonoros adicionados. Além disso, em configurações para cinema, há a adição de mais alto-falantes, que não são definidos de maneira separada, mas como zonas de espaço auditivo, isto é, pequenos grupos de caixas distribuídos nas paredes e no teto das salas de cinema.

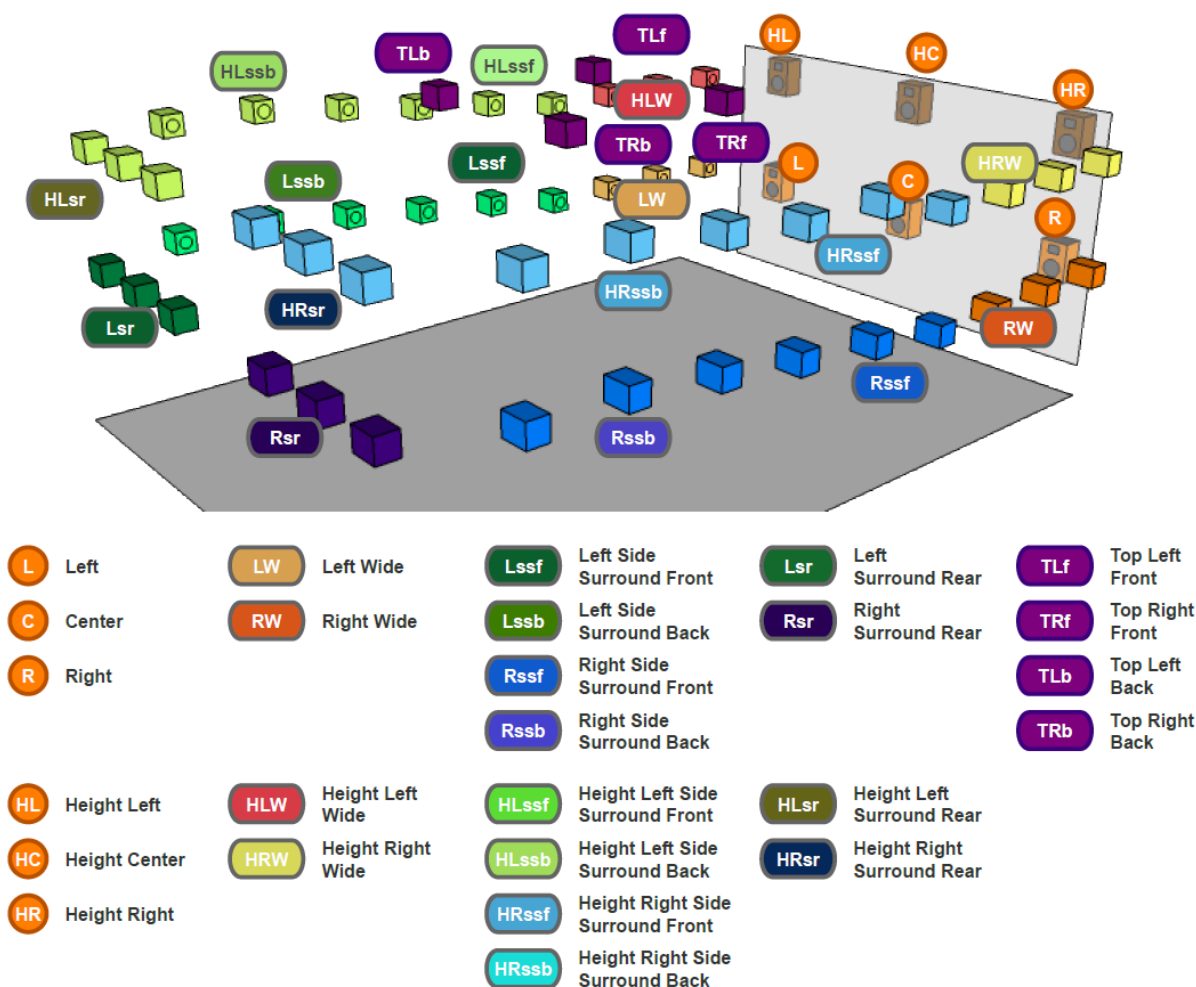
Figura 05 - Configurações do Auro 11.1 e Auro 13.1.



Fonte: AuroMax White Paper (2015).

O sistema baseado em grupos detém a vantagem de não exigir uma amplificação separada para cada alto-falante da sala de cinema, o que reduz seus custos de implementação (Auro Technologies, 2015, p. 21). A configuração cinematográfica do AuroMax se divide em zonas ou grupos de alto-falantes, podendo incluir até 26 zonas endereçáveis individualmente, com canais *surround* e de altura divididos em duas zonas por parede e a camada superior em quatro zonas, dispostas como um quadrado (figura 6). Além disso, há o chamado “Proscenium” (p. 21), um grupo de alto-falantes entre os canais da tela e os alto-falantes *surround* frontais, com a pretensão de aperfeiçoar o espaço sonoro existente entre os canais de tela e os canais *surround* em configurações normais, permitindo transições mais suaves entre os sons da tela e os da sala.

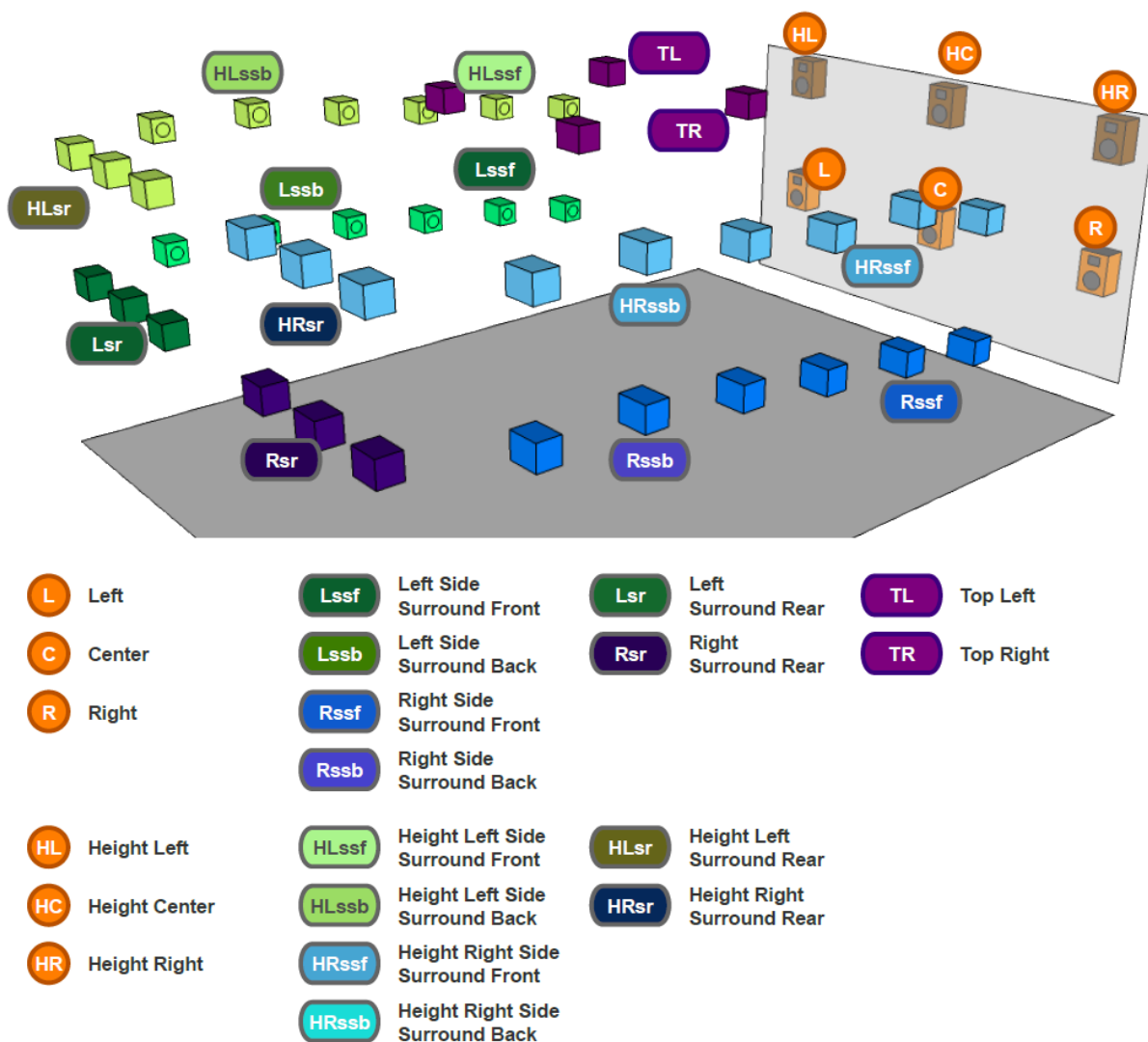
Figura 06 - Modelo AuroMax 26.1



Fonte: AuroMax White Paper (2015).

Outra possibilidade ainda é o formato 20.1 (figura 7), com 20 diferentes zonas de alto-falantes, reduzindo a camada superior a duas zonas e não incluindo a zona entre os canais da tela e o *surround* frontais. E ainda o 22.1, com a camada superior ainda dividida em quatro zonas distintas, baseado no *layout* original 11.1 canais, recomendado para cinemas que desejam converter um sistema Auro 11.1 sem alterar a instalação de alto-falantes.

Figura 07 - Modelo AuroMax 20.1



Fonte: AuroMax White Paper (2015).

O sistema também inclui o Bloco de Mídia Integrado (IMB), que funciona como um *codec* responsável pela descritografia e decodificação da trilha de áudio principal, podendo incluir também as trilhas para deficientes auditivos e visuais e dados de movimento para utilização em aplicações específicas, como D-BOX⁸⁹.

A indicação da empresa é que se utilize um gerenciamento de graves que faça o envio das informações de baixa frequência dos alto-falantes para os *subwoofers*, sugerindo ainda a instalação de um *subwoofer* em cada lado da sala, preferencialmente nas paredes laterais. Em relação às ferramentas de criação de conteúdo, o sistema utiliza de maneira nativa o *Auro-3D Creative Tool Suite*, um conjunto de *plugins* que oferece a mixagem em Auro-3D para

⁸⁹ D-BOX é uma empresa que, desde 2009, produz sistemas de controle de movimento em cadeiras e controles remotos, tanto para salas de cinema, quanto para *home theaters* e consoles de *videogame*.

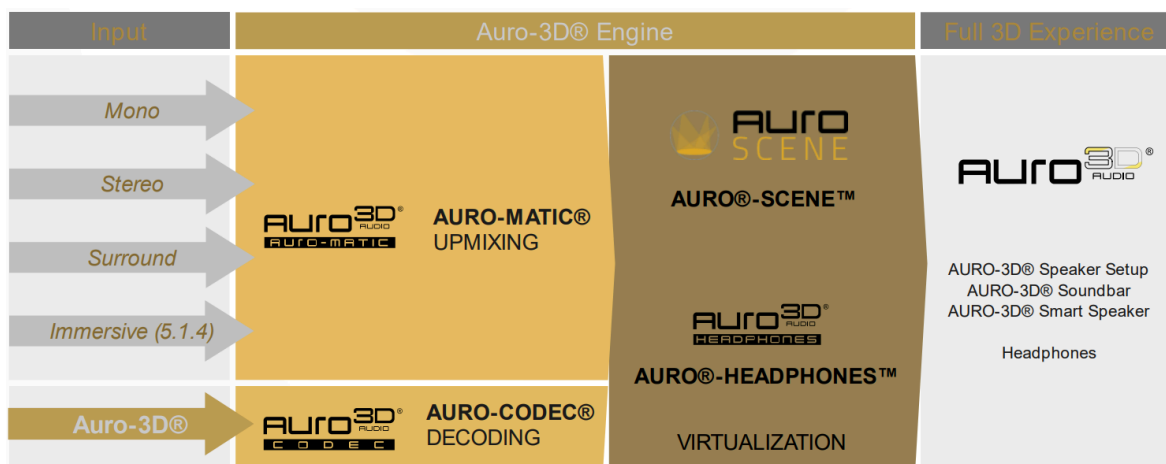
softwares como *Pyramix*, *Nuendo*, *Reaper* e *DaVinci Resolve* para todos os formatos até Auro 13.1 e *ProTools* e *Logic Pro X* apenas para os formatos 5.1.4 e 7.1.4 (*Auro Creative*, 2024).

De acordo com o site da empresa, os arquivos de áudio em Auro-3D podem ser disponibilizados em diversos formatos, como para *download online*, mídia física, em lojas virtuais ou serviços de música e em plataformas de *streaming*. Para isso, dois *codecs* são indicados: *Auro-Codec* que codifica sem perdas os áudios de 6 ou 8 canais em arquivos com formato PCM-WAV (*Pulse Code Modulation - Waveform Audio File Format*). O uso de arquivos FLAC (*Free Lossless Audio Codec*) também é recomendado pelo amplo suporte em aplicativos de *streaming*; já o *Auro-Cx NGA Codec* foi projetado especificamente para o uso em *streaming*, com transmissão de alta resolução sem perdas no conteúdo, utilizando largura de banda menor e, com isso, economizando os custos para armazenamento de dados, mantendo a qualidade. Através deste *codec* também é possível distribuir a mixagem de áudio orientado a objetos. Nos sistemas de reprodução, os decodificadores AuroMax interpretam o conteúdo codificado e distribuem os sinais para os alto-falantes apropriados. Isso inclui o gerenciamento de diferentes configurações de caixas e a adaptação acústica do ambiente.

3.3.1.3 Auro-3D Engine

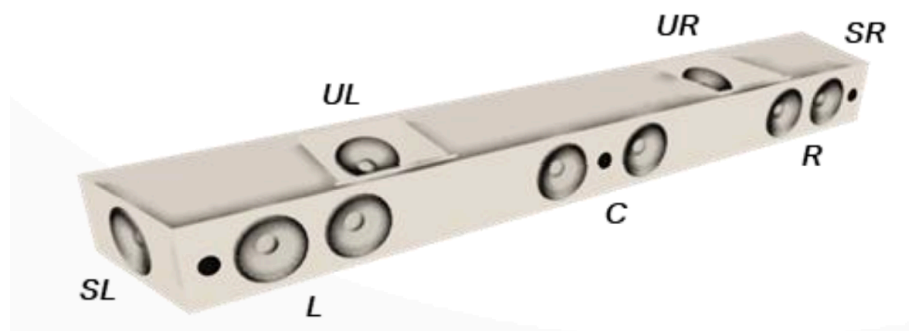
A partir do desenvolvimento de dispositivos capazes de receber um volume de dados considerável e reproduzir sons com certa qualidade, tornou-se necessário a ampliação das possibilidades de reprodução dos sons chamados de envolventes ou imersivos em aparelhos domésticos e dispositivos móveis, como receptores e processadores AV em *home theaters*, *soundbars*, alto-falantes do tipo *smart* e aparelhos diversos com a tecnologia e a possibilidade de escuta através de fones de ouvido. Técnicas como o *upmixing*, em que o sistema converte o conteúdo de áudio em um formato tridimensional, com a criação e adição de canais de conteúdo existente, transformando, por exemplo, um estéreo em Auro 13.1, e também o inverso, o *downmixing*, que combina múltiplos canais de áudio já existentes em um número reduzido, como de Auro 9.1 a 5.1 *surround*, possibilitam essa transição entre dispositivos. Há ainda a possibilidade chamada de *unmixing*, que consiste em separar em múltiplos canais incluídos em um único mix existente, com a utilização de metadados, tal como um mix Auro 9.1 de um sinal codificado 5.1 (*Auro-3D Engine*, 2022, p. 5). O Auro-3D Engine foi a solução apresentada pela empresa para aproveitar a experiência auditiva do AuroMax em outros ambientes diferentes da sala de cinema, adaptando automaticamente o conteúdo de entrada no sistema de som instalado.

Figura 08 - Visão geral Auro-3D Engine



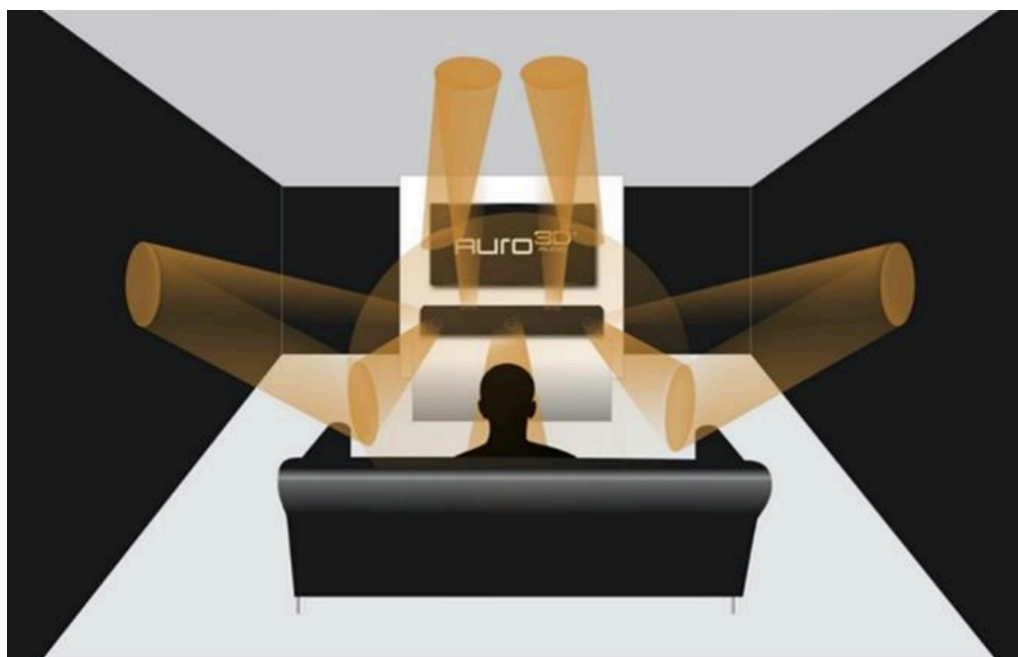
Fonte: Auro-3D Engine White Paper (2022)

A partir da combinação de quatro tecnologias principais - *Auro-Codec*, *Auro-Matic*, *Auro-Scene* e *Auro-Headphones* - o *Auro-3D Engine* apresenta-se como uma solução para utilização em ambientes domésticos, tanto para filmes e séries, como para reprodução musical e utilização em *videogames* e outros consoles. Dentre os dispositivos disponíveis, inclui-se os *soundbars* com configurações que podem ir de estéreo a 5.1.2, incluindo dois alto-falantes superiores para reflexão no teto e dois laterais para reflexão nas paredes do ambiente (figura 09).

Figura 09 - Configuração dos alto-falantes em *soundbars*

Fonte: Auro-3D Engine White Paper (2022)

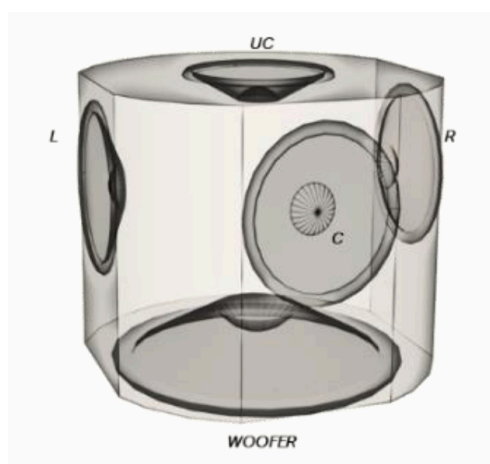
Figura 10 - Virtualização sonora em *soundbar* 5.1.2 canais.



Fonte: Auro-3D Engine White Paper (2022)

Além disso, podem ser adicionados alto-falantes extras, para o *surround* ou para os frontais. No modo alto-falante *smart*, o sistema comporta a configuração 3.1.1 com ou sem *subwoofer*, dependendo da capacidade do dispositivo

Figura 11 - Configuração alto-falante *smart*



Fonte: Auro-3D Engine White Paper (2022)

O modo de utilização através de fones de ouvido, *Auro-Headphones*, consegue processar até 14 fontes de entrada, transformando o som em uma versão virtualizada das configurações dos alto-falantes físicos. O sistema da empresa se baseia nas funções de

transferência relacionadas à cabeça (HRTF) e oferece um controle individualizado - ITD *Offset (Inter-aural Time Difference)*⁹⁰ -, no qual o usuário profissional consegue ajustar de maneira mais precisa o posicionamento das fontes. Além disso, alguns modos pré-definidos sobre o local de reprodução são possíveis de serem selecionados, como: *Home Cinema (default)*, *Lounge*, *Cinema Theater*, *Concert Hall* (*Auro-3D Engine*, 2022, p. 14).

Por último, o sistema *Auro-3D Engine* apresenta quatro modos possíveis de escuta a serem selecionados pelo usuário final de acordo com o material reproduzido. O modo estéreo compacta os canais frontais, *surround* e superiores em dois canais; o modo *surround* utiliza os alto-falantes disponíveis para criar o modo de escuta ou ainda age por meio da virtualização em *soundbars* e alto-falantes *smart*, mas não criará canais de altura, geralmente sistemas 5.1 ou 7.1 ou sua equivalência virtual. A seleção Auro-3D utiliza todos os recursos disponíveis de espacialização, seja real ou virtualizado. Há ainda o modo nativo que apresenta ao usuário a versão mais próxima do conteúdo original, seja estéreo, *surround* ou formatos tridimensionais, reais ou virtualizados.

3.3.2 DTS:X

Lançado em 2012, o sistema precursor dos objetos sonoros DTS - empresa já conhecida concorrente dos sistemas de mixagem e reprodução Dolby - foi inicialmente batizado de DTS Neo:X e despontou oferecendo até 11.1 canais, adicionando em altura frontal e largura e realizando *downmix* de fontes 11.1 para sistemas 5.1 e 7.1. Em 2015, a DTS adaptou seu sistema para a mixagem e reprodução de objetos sonoros e lançou a plataforma DTS:X, que, diferentemente dos outros sistemas, não recomenda um *layout* específico de alto-falantes, pois, segundo a empresa, funciona em todas as configurações e é funcional nos mais diversos *layouts*. O processador de áudio renderiza dinamicamente a saída de som dependendo do número e da posição dos alto-falantes disponíveis, se tornando um formato flexível inclusive com compatibilidade nas configurações e nos receptores de áudio e vídeo com Dolby Atmos e Auro-3D (Palenchar, 2015).

Mesmo instalado em algumas salas de cinema, principalmente nos Estados Unidos, o sistema DTS:X foca sua funcionalidade mais para mercado doméstico e de entretenimento fora das salas de cinema. Para *home theaters*, o sistema oferece até 12 canais com DTS *Play-Fi*, já para *soundbars*, receptores a/v habilitados e conexões móveis, há o DTS Virtual:X, que “cria” alto-falantes *surround* ou de altura virtualizados utilizando processamento

⁹⁰ Tradução da autora: Compensação ITD (diferença de tempo inter-aural).

psicoacústico de trilhas já existentes, inclusive com a criação de mais canais de altura, caso haja necessidade. A partir do DTS *AutoStage*, o sistema consegue também disponibilizar a reprodução sonora tridimensional em automóveis, através dos alto-falantes dos carros e de transmissões em rádios HD ou mídia física.

Para o sistema binaural, o DTS *Headphones:X* consegue espacializar o som através do uso de fones de ouvido estéreo comuns em aparelhos diversos que contenham as configurações DTS:X. Alguns consoles de *videogame*, televisões, computadores, celulares, *tablets* e outros dispositivos diversos já saem com o *software* decodificador nativo nos aparelhos. Os metadados salvos pelo *codec* reproduzem até 12 canais de sons *surround* binaural e funções de transferências relacionadas à cabeça (HRTF) são manipuladas para incluírem compensação para ambiências, como reflexões e atrasos, utilizando as características acústicas do estúdio de mixagem original ou ainda um laboratório de áudio profissional de referência.

Para a criação do conteúdo, há o pacote de *plugins* DTS:X *Studio Creator Suite*, compatível com *ProTools*, que entrega o conteúdo de áudio para cinema e outras demandas, como jogos e música. A plataforma ganhou destaque com o lançamento de *blu-rays* novos na época, como *Ex Machina* (2014), mas principalmente com relançamentos de filmes com remasterização do som, como *ET: O Extraterrestre* (1982) e *Apollo 13* (1995) (DTS, 2024).

3.3.3 Dolby Atmos

Dolby Atmos é outra tecnologia de som *surround* semelhante aos sistemas descritos anteriormente, mas que leva a marca da empresa com maior reconhecimento mundial em relação à temática sonora (Sergi, 2012, p. 3). Assim como seus concorrentes, o Atmos expande os sistemas de som existentes e adiciona canais de altura, permitindo que os sons sejam interpretados como objetos tridimensionais sem limitações horizontais ou verticais. A primeira exibição aconteceu com o filme *Valente* (2012) em salas de cinema e desde então houve uma expansão para dispositivos domésticos e de uso *mobile*.

Conforme observa Sergi (2012), no artigo *Knocking at the door of cinematic artifice: Dolby Atmos, challenges and opportunities*, o Atmos pode não ser o único entre toda uma série de novas tecnologias chamadas imersivas, mas sua entrada no mercado das salas de cinema e principalmente a rápida evolução para o doméstico, aponta a necessidade de análise do consumo audiovisual além dos locais exibidores. O autor explica que, mesmo estando presente em praticamente todos os cinemas mundiais, as receitas da Dolby são geradas cada

vez mais pelos setores de informática, dispositivos domésticos e móveis, e não dentro das salas de cinema.

Nos três anos entre 2010 e 2012 (ano em que o Atmos foi lançado), os números são inequívocos: enquanto o segmento de cinema digital da Dolby viu as suas receitas continuarem a cair (de 20% da receita total da Dolby em 2010 para apenas 11% em 2012) o oposto é verdadeiro para as receitas de licenciamento provenientes de vários outros contextos de entretenimento (incluindo televisões digitais, decodificadores, leitores de DVD e *Blu-ray Disc*, consoles de jogos de vídeo, receptores de áudio/vídeo, dispositivos móveis, computadores pessoais e vários dispositivos móveis) com receitas aumentando de 77% em 2010 para massivos 86% em 2012 (Sergi, 2012, p. 2-3).

Desde então, o desenvolvimento do Atmos permaneceu em salas de cinema, mas conseguiu atingir um número maior em relação a aparelhos domésticos e dispositivos móveis, principalmente pelo custo de implantação. Com diversos sistemas concorrentes já desenvolvidos, o Atmos não chega como uma grande inovação, mas com o objetivo de manter a hegemonia da Dolby no mercado audiovisual. E, com o investimento para ambientes domésticos, sonorização de automóveis, sistemas binaurais e as facilidades de distribuição e difusão - além de uma receita de US\$325,6 milhões no primeiro trimestre fiscal de 2024⁹¹ - consegue atualmente ser o sistema mais utilizado dentre os disponíveis para mixagem de objetos sonoros.

Ainda pensando em salas de projeção, a plataforma Atmos despontou com a ideia de distribuir até 64 alto-falantes endereçados individualmente, acrescentando a dimensão vertical e metadados que permitem a colocação de até 128 objetos sonoros, que podem ser trabalhados de maneira independente e direcionados para qualquer alto-falante, em qualquer ponto da sala. Assim como na Auro e no DTS:X, os objetos podem ser movidos utilizando o chamado *panner*⁹² ao invés de atribuir sons a canais específicos em modelos como o 7.1, por exemplo. [Dolby Atmos] “permite que um mixador pegue qualquer trilha, seja um efeito, voz ou elemento de trilha musical, e manipule-a como um objeto que pode ser colocado ou movido em qualquer lugar dentro do espaço 3D da sala – sem atribuí-lo diretamente a um alto-falante específico”⁹³ (Dolby, 2020, p. 3). Ou seja, ao contrário das configurações baseadas em canais,

⁹¹ O trimestre fiscal da empresa é composto pelo período de outubro a dezembro de 2023.

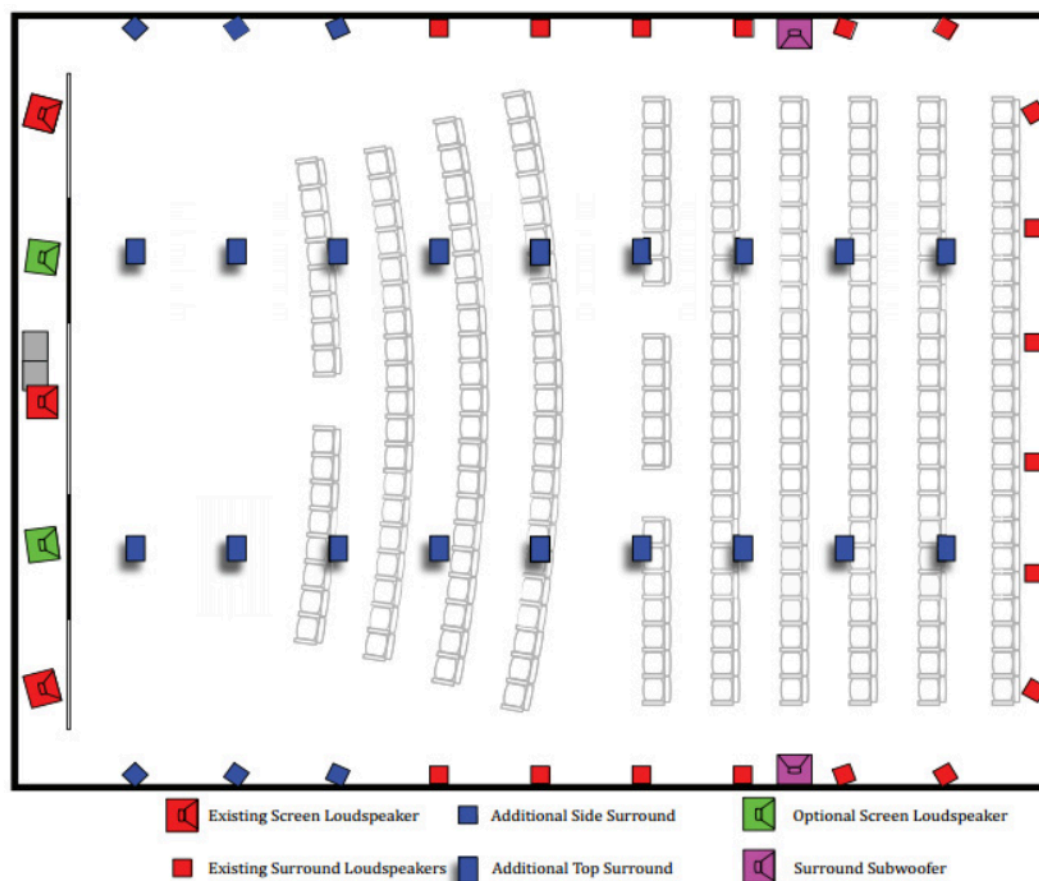
⁹² Plugin ou conjuntos de plugins que auxiliam na mixagem, conseguindo movimentar as faixas de áudio com base na intensidade. Vem da ideia de Panning, que é o ato de colocar ou mover um som para qualquer lugar do campo sonoro. Com o panning, as fontes sonoras podem ser posicionadas de forma que sejam percebidas como provenientes do alto-falante esquerdo, do alto-falante direito ou de qualquer lugar intermediário, como no caso dos objetos sonoros do Atmos e dos outros sistemas vistos anteriormente.

⁹³ Tradução livre da autora. Texto original: “Allows a mixer to take any track, whether an effect, voice, or music track element, and manipulate it as an object that can be placed or moved anywhere within the 3D space of the room — without directly assigning it to a particular speaker”.

na qual vários alto-falantes fazem parte de uma mesma zona, cada alto-falante em Atmos se torna uma zona independente, melhorando também a transição entre as fontes da sala, como da tela para as laterais e vice-versa.

Na imagem abaixo (Figura 12), há a exemplificação do posicionamento dos alto-falantes no sistema Atmos. Em vermelho o sistema já existente antes da atualização: o ícone maior representa os alto-falantes da tela, o menor, os *surround*. Em azul aparece os alto-falantes adicionais do tipo *surround*, o retângulo demonstra os instalados no teto e o quadrado o das laterais. Em verde há a possibilidade opcional de instalação de mais alto-falantes na tela e, em roxo, o posicionamento do *subwoofer surround*, além dos já tradicionais *subwoofers* localizados atrás da tela.

Figura 12- Configuração Dolby Atmos para salas de cinema



Fonte: Dolby White Paper, 2018.

O Dolby Atmos apresenta também a compatibilidade com versões anteriores, através do sistema de entrega CDP (*Content delivery package*), em que, ao invés de múltiplas versões e arquivos, gera apenas um modelo adaptável para qualquer formato que o filme possa ser

reproduzido, seja a mais completa estrutura Atmos, até sistemas 5.1 canais, isto é, “um filme mixado no Atmos poderia então ser '*down-mixado*' para qualquer configuração de canal específica disponível onde está sendo reproduzido”⁹⁴ (Sergi, 2012, p. 6).

Para os exibidores, a vantagem de implementação do Atmos fica por conta da ideia de que o filme seja escutado como o planejado pela mixagem independentemente da arquitetura da sala de exibição, do tamanho da tela ou da configuração de canais.

A Dolby também reconhece especificamente que este é o seu primeiro sistema 'centrado na sala', onde o 'ponto ideal' (um lugar ideal no auditório onde o som pode ser ouvido como 'pretendido') não é mais relevante, pois o som se dirige a públicos de todas as direções com igual potência, alcance e direcionalidade⁹⁵ (*Ibid.*, p.6).

O Dolby Atmos, assim como os sistemas anteriores, funciona a partir da ideia de camas e objetos, ou mixagem baseada em canais e objetos. “Isto, por sua vez, apresenta uma série de questões e escolhas relacionadas: o quanto do som deve ser usado como 'objetos' e o quanto deve ser incluído nas camas sonoras que constituem o ponto de partida de uma trilha sonora do Atmos?”⁹⁶ (Sergi, 2012, p. 11). A ideia dos canais ainda funciona melhor para algumas sonoridades, como reverberações e, segundo orientações da própria plataforma, nem todo som deve ser inserido como objeto (Dolby, 2016), mas definir essa escolha ainda é um ponto de discussão entre profissionais da área.

Nas configurações para sala de cinema, o Atmos consegue processar até 128 objetos sonoros direcionados para até 64 canais individuais, em um sistema (cama) de 7.1.2 canais. As camas são mapeadas para os alto-falantes atribuídos a cada canal e os objetos passam por um processo de renderização em tempo real com o posicionamento baseado na localização física dos alto-falantes. A renderização é um método que utiliza o mapeamento dos canais e objetos para localizar os alto-falantes da sala, ambiente doméstico ou fones de ouvido. Um dos diferenciais do Atmos é a utilização de diferentes tipos de renderizadores que funcionam para cada ambiente específico, promovendo uma compatibilidade com os mais variados tipos de sistemas sonoros. Em um ambiente 7.1, por exemplo, o renderizador precisa “espalhar” o

⁹⁴ Tradução livre da autora. Texto original: “A film mixed in Atmos could then be ‘down-mixed’ to any particular channel configuration available where it is playing”.

⁹⁵ Tradução livre da autora. Texto original: “Dolby also specifically acknowledges that this is their first ‘room-centric’ system, one where the ‘sweet spot’ (an ideal seat in the auditorium where sound can be heard as ‘intended’) is no longer relevant as sound addresses audiences from all directions with equal power, range and directionality”.

⁹⁶ Tradução livre da autora. Texto original: “This in turn presents a number of questions and related choices: how much sound to use as ‘objects’ and how much to leave in the ‘sound beds’ that constitute the starting point of an Atmos soundtrack?”.

áudio baseado na mixagem em objetos para os canais. Já para fones de ouvido, o ideal é que tenha uma entrega consistente com a escuta individualizada (Dolby, 2016).

No artigo “*Immersive Audio: Optimizing Creative Impact without Increasing Production Costs*”, Sexton Connor (2017) coloca que um arquivo mestre renderizado em Atmos é capaz de produzir sub-renderizações adaptáveis aos formatos tradicionais baseados em canais, como estéreo, 5.1 e 7.1, gerando uma compatibilidade entre os formatos de forma automatizada. Os sons mixados como objetos sonoros se ajustam aos canais disponíveis e, com o direcionamento profissional e o uso de metadados, selecionam em qual zona de alto-falantes o som ficará melhor posicionado. Também por essa característica de auto-adaptação, salas de cinema recebem os arquivos mestre em Atmos, mesmo que não tenham a configuração instalada em sua estrutura e, consequentemente, o formato se populariza como um padrão, uma vez que funciona em variados sistemas sem maiores perdas ou problemas de compatibilidade.

Em relação à criação de conteúdo, os *plugins* e o renderizador *Dolby Atmos Renderer* são compatíveis com os principais *softwares* de mixagem de áudio profissional, como *ProTools*, *Nuendo*, *Cubase*, *Studio One*, *Pyramix* e *DaVinci Resolve* (Dolby Professional, 2024). O som 7.1.2, mais os objetos sonoros e os metadados são utilizados pelo renderizador para monitorar, gravar e reproduzir o conteúdo em Atmos, assim como poderá também converter para modelos de mixagem baseados em canais, para monitoramento e entrega. O *Dolby Atmos Renderer* também grava arquivos mestre Atmos usados para codificação de discos *Blu-ray 4K Ultra HD* e pode exportar para ADM BWF⁹⁷, formato comumente usado para codificação por serviços de *streaming*.

⁹⁷ ADM BWF é um formato de áudio padronizado baseado em arquivos de ondas de transmissão. Os metadados são armazenados em um pedaço XML dentro do arquivo de ondas de transmissão.

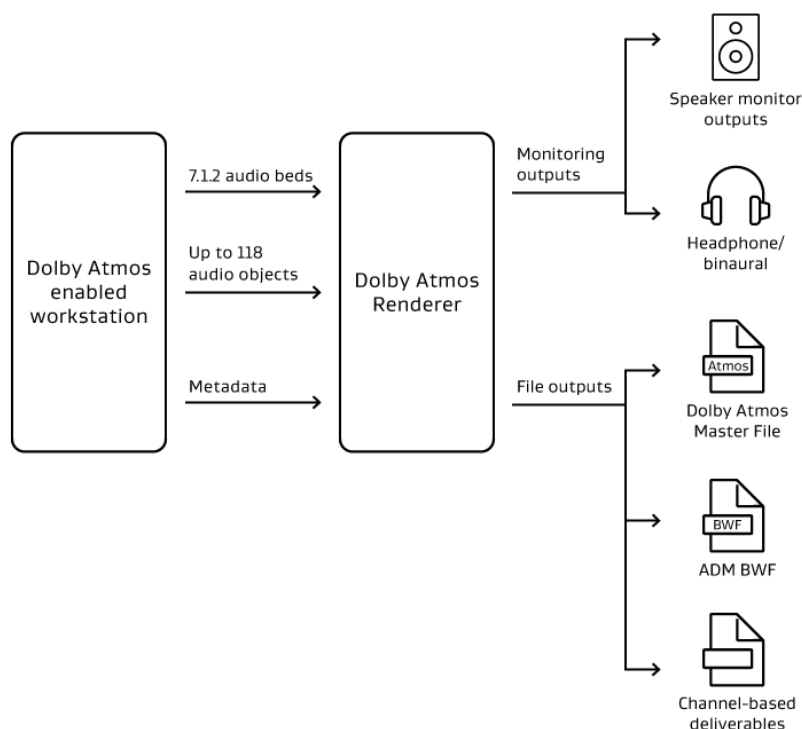
Figura 13 - Interface Dolby Atmos Renderer



Fonte: Sound on Sound, 2022.

Os codecs *Dolby TrueHD* e *Dolby Digital Plus* seguem sendo os utilizados para transmissão dos arquivos da mixagem, incluindo os metadados para os objetos sonoros. O *Dolby TrueHD* é voltado mais ao uso em *Blu-Ray Disc*, pois entrega o som sem compressão e, consequentemente, com maior qualidade possível. Entretanto, os arquivos são extensos e demandam uma largura de banda considerável para a transmissão, o que torna seu uso para *streaming* limitado. Os sinais Dolby Atmos codificados em *Dolby TrueHD* são transmitidos de um reprodutor *Blu-ray* para o sistema de som habilitado com Atmos por meio de uma conexão HDMI, que recebe a trilha de áudio e os objetos sonoros associados a metadados posicionais e, em seguida, decodifica, processa e renderiza o som para a configuração específica do alto-falante no dispositivo (Dolby, 2018). O *Dolby Digital Plus* promove uma alta compressão, demandando uma largura de banda menor e, por isso, é o mais adotado em dispositivos que dependem da transmissão de conteúdo Atmos através de plataformas de *streaming*.

Figura 14 - Diagrama de fluxo Dolby Renderer



Fonte: Dolby Professional, 2024.

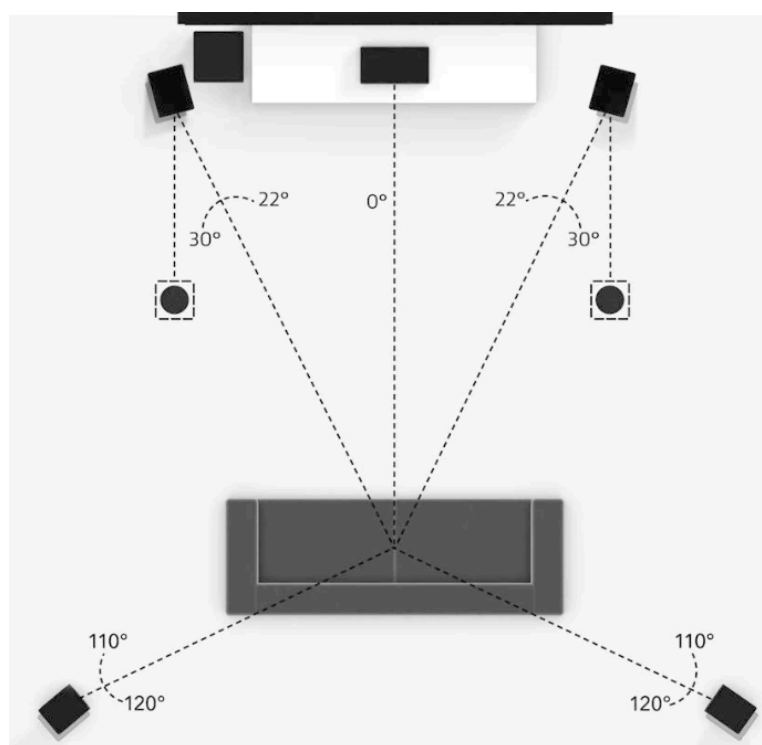
Em 2015, a Dolby começou a integrar o sistema Atmos a dispositivos domésticos e móveis, como *home theaters*, *soundbars*, televisores, computadores, consoles de jogos, *tablets* e celulares, que, na maioria das vezes, saem de fábrica com a tecnologia em suas configurações. Em computadores, o *software Dolby Access* pode ser adquirido para integrar o sistema Atmos aos alto-falantes do aparelho e através de fones de ouvido.

A mixagem em Atmos para uso doméstico também acontece de maneira diferenciada, com recursos mais simples e menor custo. No livro sobre pós-produção sonora no Brasil, Rodrigo Carreiro (2019) entrevista profissionais da área, abordando as perspectivas em relação ao Atmos. Entre opiniões distintas sobre a consolidação da tecnologia para cinema, é possível perceber um consenso em relação ao seu uso doméstico, principalmente no caso de produções nacionais. Como relata Daniel Turini ao autor:

O Atmos doméstico é bem mais simples do que o Atmos para cinema. O que algumas pessoas estão fazendo – a tendência no momento seria essa – é mixar o produto em 5.1 e levar um ou dois dias para uma sala Atmos de mixagem, para espalhar mais, abrir mais o som. (...) Atmos para cinema é muito caro. Na produção audiovisual brasileira não deve ter muito espaço, por causa dos orçamentos pequenos. Para o Atmos doméstico existe mais espaço (Turini, 2019 *In*. Carreiro, 2019, p. 37-38).

Em *Home Theaters*, a implantação do Atmos pode ser realizada com até 34 alto-falantes, sendo 24 frontais e *surround* e 10 aparelhos suspensos. No entanto, as configurações mais comuns ficam limitadas a até 11.1 canais, com opções como 5.1.2, 5.1.4, 7.1.2, 7.1.4 e 9.1.2. Há pelo menos três tipos de caixas acústicas para essa implementação: as embutidas no teto, as habilitadas especificamente para Atmos e as adicionais. Os embutidos no teto, ou ainda chamados de *in-ceiling*, devem estar alinhados com as caixas frontais esquerda e direita e com os respectivos *surrounds*. Além disso, devem possuir um padrão de dispersão amplo - em torno de 45° - para tetos até 4 metros de altura. Acima disso, a fabricante diz ser suficiente padrões normais de dispersão (Dolby, 2016).

Figura 15 - Esquema 5.1.2 canais

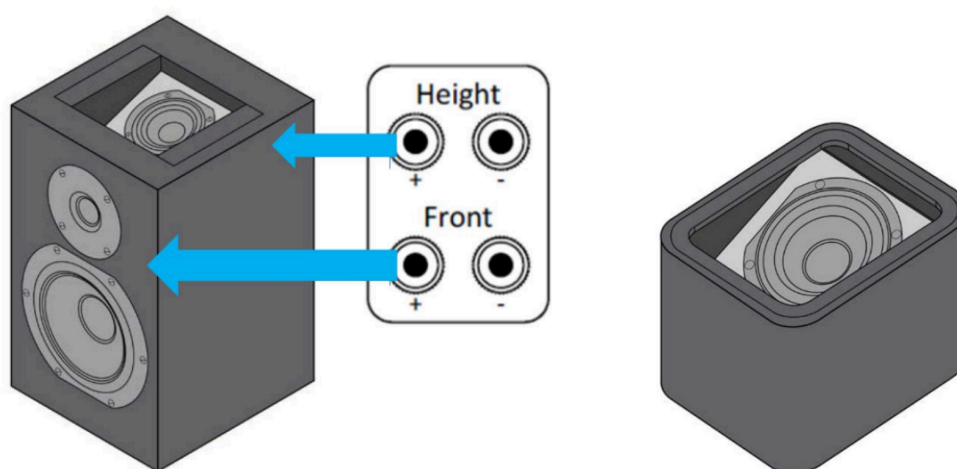


Fonte: Dolby Home Theater Speaker Guide, 2024.

A dificuldade de instalação de caixas de som no teto cria um empecilho para *home theaters* em Atmos, uma vez que seu grande diferencial é o som verticalizado e de alguma forma isso precisa acontecer, mesmo que as caixas não estejam necessariamente ali instaladas. Para produzir esse efeito sem a necessidade de fixar os aparelhos no teto, foram desenvolvidas

algumas caixas com dois alto-falantes e habilitação para Dolby Atmos. Além do tradicional posicionado na horizontal, essas caixas especiais também chamadas de *Atmos-enabled*, adicionam um segundo alto-falante no topo, posicionado de forma a direcionar o som angulado no teto, chegando até o espectador (Figura 17). Além disso, há as caixas complementares ou *add-on* (Figura 16), semelhantes aos módulos superiores das caixas habilitadas, mas que podem ser montadas em caixas acústicas convencionais.

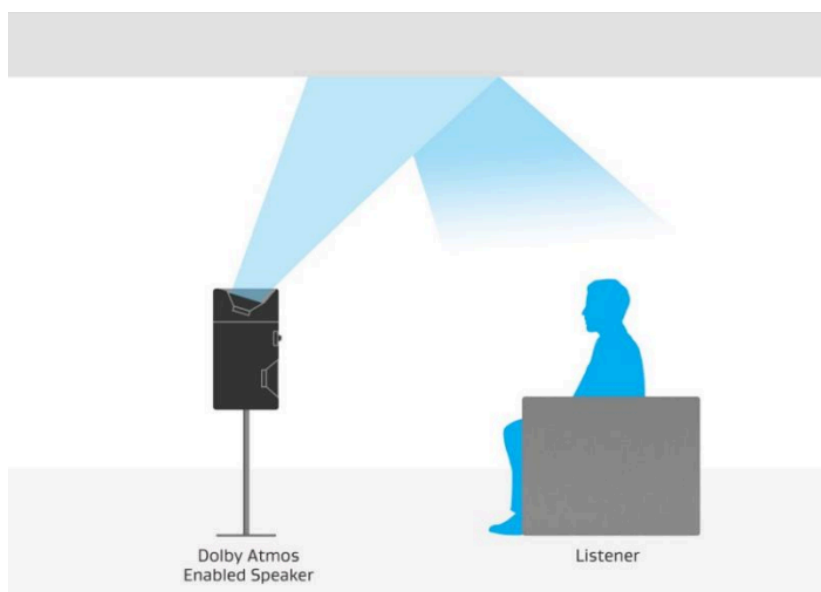
Figura 16 - Caixas integradas e módulos complementares habilitados para Atmos



Fonte: Dolby Atmos Enabled Speaker Technology, 2016.

As caixas integradas e as adicionais possuem alimentação elétrica própria e a percepção do efeito de som vindo do teto acontece pela reflexão da onda sonora, mas alguns materiais absorventes e tetos rebaixados não são indicados para a instalação.

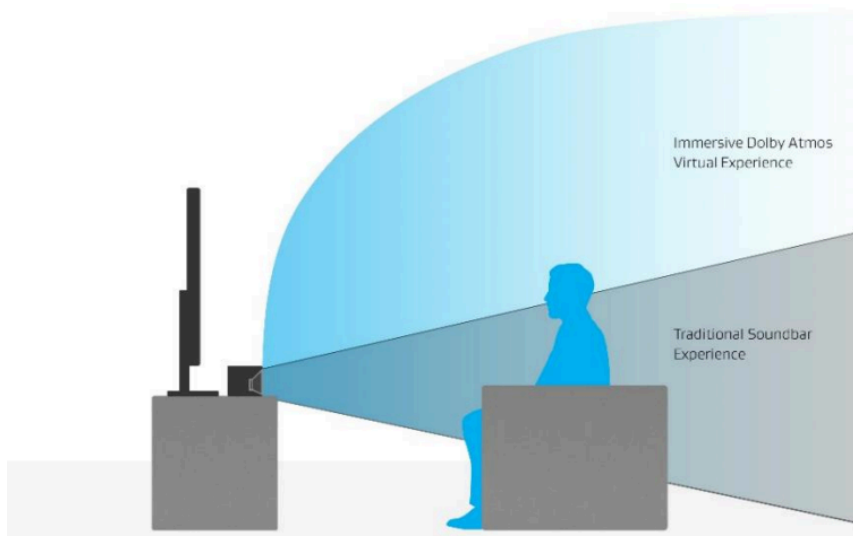
Figura 17 - Sistema de reflexão das caixas integradas.



Fonte: Dolby Atmos Enabled Speaker Technology, 2016.

Mesmo com essas soluções, instalar um *home theater* exige um orçamento considerável e um espaço dedicado a receber tantos equipamentos, o que não é prática comum nos lares medianos. Com isso, surge a necessidade de um aparelho mais compacto que melhore a qualidade sonora do ambiente e não ocupe tanto espaço ou exija grandes habilidades para instalação. Os *soundbars* despontam suprimindo essa necessidade, com duas caixas na parte superior direcionando o som para o teto e rebatendo para o espectador, criando a verticalização esperada com o Atmos, em um esquema semelhante ao das caixas acústicas integradas com a virtualização do canal de altura. O aparelho também possui os alto-falantes frontais e *surround* na mesma barra de som, utilizando da reflexão também para o som das caixas *surround*, fazendo com que o formato do ambiente também influencie no resultado. Tanto os *home theaters* quanto os *soundbars* precisam de um teto uniforme para que o som seja rebatido de maneira adequada a criar o efeito desejado.

Figura 18 - Comparativo entre *soundbar* estéreo e *soundbar* com Dolby Atmos.



Fonte: Dolby Atmos for soundbar applications, 2018.

A virtualização sonora obtida através de *soundbars* e caixas integradas acontece através da aplicação de filtros de sinalização de altura projetado nos componentes de áudio antes da mixagem nos alto-falantes. “Esses filtros simulam as pistas espectrais naturais transmitidas pelo ouvido humano aos sons que chegam de cima”⁹⁸ (Dolby, 2018, p. 6). Para a virtualização dos sons *surround* são utilizados dados de HRTF e cancelamento de interferências para promover uma aproximação sonora entre os ouvidos do espectador e as pistas binaurais dos alto-falantes *surround*, mesmo que haja mais de um espectador e eles estejam fora do chamado ‘ponto ideal’ da escuta. “O virtualizador *surround* aprimora os canais frontal esquerdo, frontal direito, *surround* e vertical do sinal multicanal para criar um efeito *surround* virtual envolvente, compensando o formato retangular da barra de som”⁹⁹ (Dolby, 2018, p. 12).

A utilização de *soundbars* ainda oferece, segundo a Dolby, outras vantagens para os ouvintes, como nivelador de volume e modulador de volume, que mantém o equilíbrio tonal na diminuição e aumento de volume do dispositivo e na alternância entre produtos, como filmes e músicas; há ainda o intensificador de diálogo, que melhora a inteligibilidade da fala, acentuando o conteúdo do diálogo de forma que permita ao espectador o entendimento mais

⁹⁸ Tradução livre da autora. Texto original: These filters simulate the natural spectral cues imparted by the human ear to sounds arriving from overhead.

⁹⁹ Tradução livre da autora. Texto original: The Surround Virtualizer enhances the Front Left, Front Right, Surround, and Overhead channels of the multichannel signal to create an enveloping virtual surround effect, compensating for the rectangular form factor of the sound bar.

claro da fala; e o equalizador inteligente, que mantém o equilíbrio espectral, isto é, de timbre e tom, a partir da comparação com um tom de referência.

O chamado Dolby Atmos virtualizado ou *Home Atmos*, ou seja, uma espécie de emulador do Atmos original, através de *softwares*, funciona como uma adaptação e uma tentativa de popularização do Dolby Atmos real. A partir de duas saídas de som já é possível obter virtualmente as possibilidades sonoras do Atmos, gerando um aumento dos dispositivos possíveis de se inserir a tecnologia e diminuindo os custos de implementação. Com isso, veio o crescimento expressivo de dispositivos domésticos e *mobile* com habilitação para Dolby Atmos: *soundbars*, televisores, carros, celulares e *tablets*, que, também através de fones de ouvido, conseguem promover a experiência binaural.

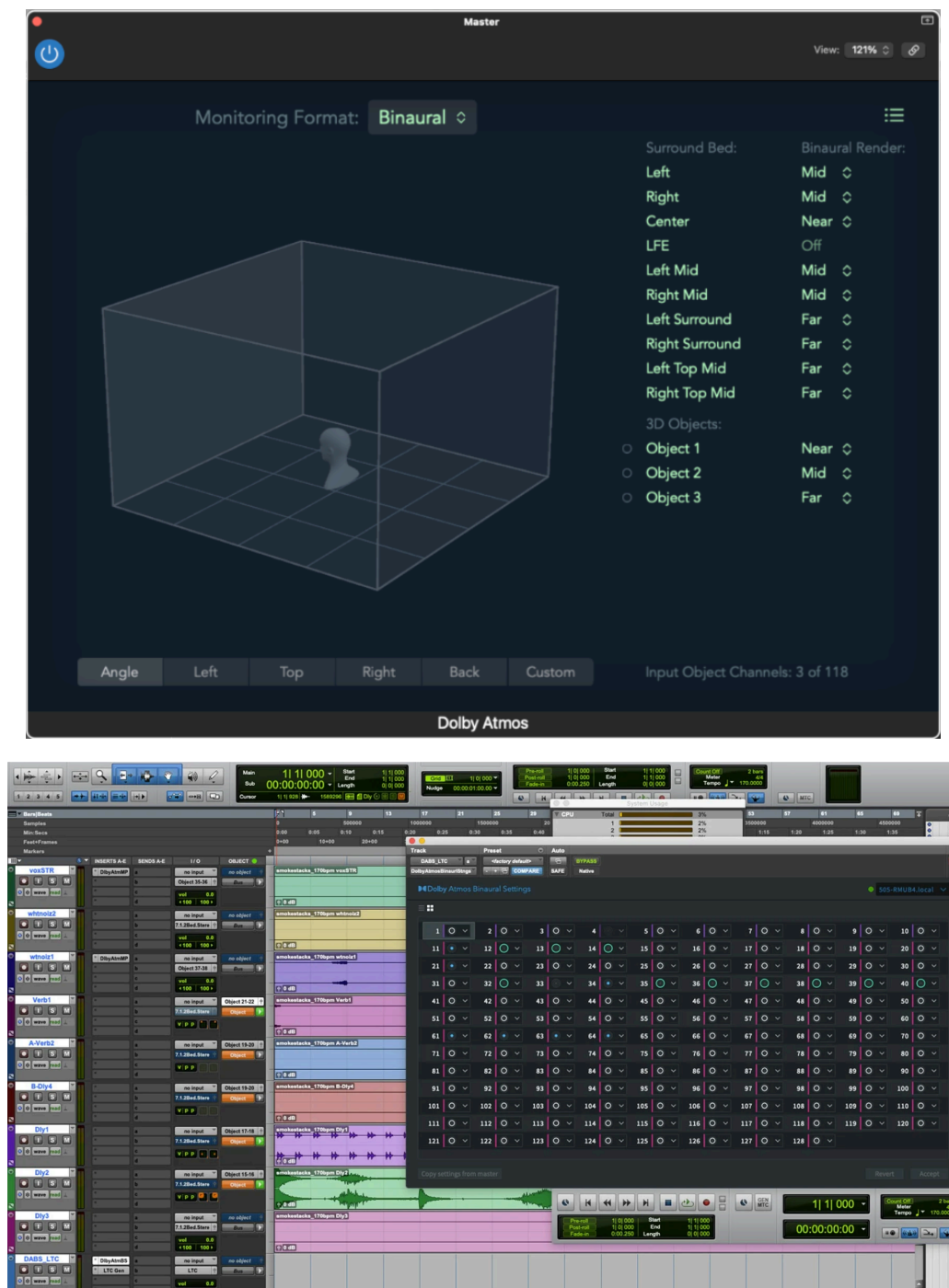
Carlos Klachquin (2023), representante da Dolby na América Latina, explica que o funcionamento do Atmos para fones de ouvido se assemelha a um sistema binaural “chamado de *Synthesis Sound*, na qual é possível criar um som virtual em várias posições, mesmo usando fones comuns, uma vez que o processamento acontece no equipamento”. Esse processamento citado determina frequências de onda compatíveis com o que seria percebido em uma sala de cinema e, através dessa síntese de campo sonoro, cria virtualmente um som espacializado, mesmo que escutado através de fones comuns.

A mixagem acontece de maneira semelhante às realizadas para demais dispositivos domésticos. A recomendação da Dolby é que o som seja mixado tanto em alto-falantes quanto em fones de ouvido, uma vez que o produto final poderá ser escutado em diversos dispositivos. A renderização binaural acontece a partir da aplicação de um modelo de distância por meio de metadados relativos aos objetos e camadas sonoras. “As alterações nas configurações do modo de renderização binaural serão aparentes apenas no feed de monitoramento de fones de ouvido e na renderização binaural da sua mixagem” (Dolby, 2021). Os arquivos são codificados como estéreo imersivo AC-4, específico para audição em fones de ouvido, que permite configurar os sons como “desligado, próximo, médio e distantes”, fornecendo ainda níveis de espacialização para cada objeto ou canal aplicado. Esses modos são como medidas da distância virtualizada entre um objeto ou canal e a cabeça do espectador. “Os metadados binaurais são aplicados à saída do fone de ouvido durante o monitoramento, gravação ou reprodução de um master. Esses metadados não estão incluídos no processamento do alto-falante” (Dolby, 2021).

A renderização é realizada através do *plug-in Dolby Atmos Binaural Settings* que trabalha com o *Dolby Atmos Renderer*, visto anteriormente, e permite o controle de entrada,

grupos, descrições e metadados binaurais diretamente por DAW¹⁰⁰ que o profissional trabalha, podendo ser utilizado o *Ableton Live*, *Apple Logic Pro*; *Avid Pro Tools Ultimate* e *Steinberg Nuendo*.

Figura 19 - Interface mixagem binaural Dolby Atmos e Plug-in Dolby Atmos Binaural Settings



Fonte: Dolby Professional Support, 2021.

¹⁰⁰ Expressão para *Digital Work Station*, ou seja, é o *software* que o profissional utiliza para mixar o som.

O Dolby também faz algumas recomendações de mixagem para reprodução binaural, como evitar colocar objetos sonoros centralizados na “sala”, pois nesta posição as configurações do modo de renderização binaural não terão efeito. Além disso, caso o objetivo seja criar uma sonoridade frontal, o objeto deve ser posicionado no centro da parede frontal e o modo de renderização binaural deve estar como “desligado” ou “próximo”. Na publicação *Frontal sound localization with headphone systems using characteristics of each device*, de 2017, os autores Yuya Ono, Akira Saji e Jie Huang chegam a conclusão semelhante da recomendação dada pela Dolby para simulação de sons de alto-falantes frontais com o uso de fones de ouvido. Os experimentos evidenciam que é possível diminuir a localização interna do som dentro da cabeça cancelando as características de frequência das sonoridades medidas com fones de ouvido e adaptando para características de frequência medidas com alto-falantes frontais. No entanto, após os testes, foi constatado que houve a diminuição da localização dos sons frontais na cabeça, mas a naturalidade foi prejudicada, não sendo viável recomendar a configuração.

Outra tecnologia voltada aos equipamentos domésticos foi lançada em 2024 pela Dolby, o Atmos *FlexConnect* está sendo anunciado como um *software* capaz de identificar o posicionando de alto-falantes sem fio já presentes na casa, como dispositivos Echo Dot, popularmente conhecido como Alexa, ou caixas portáteis *bluetooth*. De acordo com o divulgado, o algoritmo identifica as características e o posicionamento das caixas conectadas através de dos microfones da TV e otimiza o som de forma inteligente para qualquer *layout* de sala e tipo configuração dos alto-falantes. Inicialmente, a atualização chegará em aparelhos de televisão da marca TCL, mas com perspectivas de expansão, segundo a própria desenvolvedora.

Todas essas possibilidades de utilização da tecnologia virtualizada podem aproximar o público da experiência tridimensional sonora, mas possivelmente não conseguem transmitir a mesma experiência de *home theaters* completos e, principalmente, em salas de cinema com Atmos real. Comparar essas possibilidades não é o objetivo desta pesquisa; mesmo assim, é inevitável perceber que o *Home Atmos* é uma tentativa de popularização do Atmos real, com menor custo de implementação dos dispositivos e de investimento profissional em equipamentos para estúdio de mixagem.

3.3.4 MPEG-H e a TV 3.0 brasileira

O MPEG-H é um padrão de áudio de terceira geração, voltado à codificação de áudio 3D, permitindo a reprodução imersiva e a interatividade por parte do espectador. Desenvolvido pela *Fraunhofer Institute for Integrated Circuits* (IIS), da Alemanha - empresa criadora de formatos como MP3 e AAC -, o MPEG-H foi projetado para permitir a mixagem baseada em canais, objetos e cenas, possibilitando a chamada entrega universal, ou seja, saída para uma variada gama de configurações, desde sistemas estéreo tradicionais a instalações com som 3D e *soundbars* com *upmixing* automatizado. “Independentemente do dispositivo, o sistema de áudio MPEG-H proporciona a melhor experiência sonora possível, tanto no *home theater* quanto em *smartphones*, *tablets* e dispositivos de realidade virtual”¹⁰¹ (Fraunhofer IIS, 2023).

Além do desenvolvimento dos algoritmos de configuração, renderização e os recursos de interatividade, a Fraunhofer IIS também foi parte integrante do grupo *Moving Picture Experts Group* (MPEG) que estabeleceu o padrão através do ISO/IEC 23008-3 (MPEG-H 3D Audio), com testes perceptuais para aferição de qualidade, em parceria com a *European Broadcast Union* (EBU), e lançou as primeiras implementações comerciais do formato, como *codecs* para *broadcasters*, *set-up boxers*, televisores, *soundbars* e aplicativos para dispositivos móveis.

Uma das grandes diferenças do MPEG-H e dos demais formatos é a possibilidade de personalização da experiência sonora pelo ouvinte, como o ajuste do volume de diálogos ou a escolha entre diferentes comentaristas em uma transmissão esportiva, por exemplo, ampliando a capacidade interativa das transmissões do tipo *broadcast*. Essa característica se mostra relevante na medida em que o som deixa de ser apenas um componente passivo da narrativa audiovisual televisiva para assumir um papel ativo, ajustável em tempo real, de acordo com os contextos de escuta e as preferências do usuário. Além disso, através do uso de metadados, recursos de acessibilidade como audiodescrição podem ser incluídos como objetos sonoros ativáveis a depender da escolha do espectador, diminuindo o uso de banda da transmissão, podendo aumentar a qualidade sonora mesmo com a inserção de “canais” interativos. Tsingos (2018, p. 259) nomeia essa interatividade como metadados interativos:

Os usuários podem desejar controle total sobre a personalização de um programa de

¹⁰¹ Tradução livre da autora. Texto original: Regardless of the device, the MPEG-H Audio system delivers the best sound experience possible - in the home theatre as well as on smartphones, tablets and Virtual Reality devices.

áudio. Para garantir que cada personalização resulte em uma mixagem com qualidade sonora ideal, o criador de conteúdo ou a emissora pode fornecer metadados interativos que definem um conjunto de regras de renderização a serem usadas exclusivamente para a personalização. Os metadados interativos podem especificar valores mínimos ou máximos para parâmetros de objetos; exclusão mútua entre objetos; regras de posição, volume ou *ducking* entre objetos; e regras gerais de mixagem. Esses metadados interativos são normalmente utilizados por uma interface de usuário no dispositivo do consumidor para evitar a criação de uma renderização (mixagem) não ideal. Por exemplo, se houver objetos de diálogo em inglês e em espanhol presentes no fluxo de áudio, os metadados interativos impediriam que o usuário ativasse ambos os objetos simultaneamente¹⁰².

No cinema, o MPEG-H pode ser utilizado para mixagens compatíveis com sistemas de múltiplos canais e orientação espacial, permitindo também que produções independentes ou distribuídas digitalmente implementem a mixagem orientada a objetos, sem a necessidade de licenciamento exclusivo. Outra vantagem do formato é a possibilidade de renderização em diferentes formas, adaptando-se ao equipamento disponível e a compatibilidade com o *Audio Definition Model* (ADM), padrão utilizado para a descrição formal de mixagens em áudio 3D, facilitando os fluxos de trabalho profissional e o intercâmbio entre estúdios.

Em ambiente doméstico, o formato possui ainda maior destaque por sua característica de promoção da interatividade. A popularização de dispositivos como smart TVs, *soundbars* e sistemas de reprodução multicanal, bem como o aumento do consumo audiovisual através de plataformas de *streaming* - também chamadas de OTT (*over the top*), pois entregam o conteúdo através da internet, sem a necessidade de operadoras de cabo ou satélite tradicionais - reforça o interesse por formatos que entreguem qualidade sonora superior, incluindo também a adaptação ao espaço de escuta e às configurações técnicas disponíveis. O formato ainda possui suporte ao modelo binaural, criando possibilidades de experiências imersivas através do uso de fones de ouvido comuns.

O MPEG-H também foi o formato escolhido para o serviço de transmissão de TV UHD baseado no padrão ATSC 3.0 da Coreia do Sul e está em funcionamento desde maio de 2017, sendo atualmente o único sistema de áudio utilizado na difusão televisiva do país. No Brasil, o MPEG-H foi selecionado em dezembro de 2021 como o formato obrigatório para o futuro padrão de transmissão digital terrestre, a TV 3.0. Desde então, emissoras como *Rede Amazônica*, *TV Globo* e *TV Cultura* iniciaram transmissões regulares com MPEG-H Audio

¹⁰² Tradução livre da autora. Texto original: Users may want complete control over personalizing an audio program. To ensure that every customization results in an optimal sounding mix, the content creator or broadcaster can provide interactive metadata defining a set of rendering rules to be used for personalization only. Interactive metadata can specify object parameter minimum or maximum values; inter-object mutual exclusion; inter-object position, volume or ducking rules; and overall mix rules. The interactive metadata are typically leveraged by a consumer user interface to prevent the creation of a non-ideal rendering (mix). For example, if both English and Spanish dialogue objects are present in the audio stream, the interactive metadata would prevent a user from enabling both objects simultaneously.

em seus canais ISDB-Tb (TV 2.5), preparando-se para a transição para a TV 3.0, com início previsto para 2026 e implementação gradual no território brasileiro.

A escolha se deu através do Fórum do Sistema Brasileiro de TV Digital Terrestre (SBTVD) e do Ministério das Comunicações em discussões conduzidas entre 2020 e 2022, nas quais foram realizados testes de avaliação com docentes de universidades brasileiras e profissionais da área sonora em sistemas de áudio variados que pudessem suprir as necessidades da TV 3.0, com critérios que envolviam compressão eficiente, imersividade, recursos de acessibilidade, interatividade e compatibilidade com estruturas já consolidadas. O MPEG-H foi escolhido como padrão oficial de áudio após obter os melhores resultados nas simulações técnicas, sobretudo por sua capacidade de entrega do áudio espacial de alta qualidade com taxas de bits reduzidas, um aspecto importante ao levarmos em conta a transmissão em contextos de cobertura irregular da transmissão.

A transição para a TV 3.0 será gradual, com a expectativa de que os televisores compatíveis com o novo padrão comecem a chegar no país a partir da implementação do novo formato. Além disso, haverá dispositivos adaptadores que permitirão que os aparelhos atuais recebam o novo sinal, que será transmitido por meio de aplicativos, substituindo os tradicionais canais numéricos.

4 PRÁTICAS E PERSPECTIVAS TECNOLÓGICAS

Conforme observado nos capítulos anteriores, as discussões referentes à representação sonora permeiam já há algum tempo as análises e os estudos relativos a cinema e audiovisual. Além disso, as tecnologias representativas avançaram nas últimas décadas com a possibilidade da mixagem orientada a objetos sonoros e o uso de metadados no posicionamento angular dos sons narrativos. Com o objetivo de entender se o desenvolvimento dessas tecnologias tem promovido mudanças significativas nos modos de produção e recepção audiovisual, especialmente em ambientes domésticos, foram realizadas entrevistas com profissionais da área, buscando também compreender as dinâmicas de implementação dessas tecnologias a partir das perspectivas de especialistas atuantes no setor, de modo a articular os referenciais teóricos discutidos previamente com as práticas concretas da indústria contemporânea.

O livro *A pós-produção de som no audiovisual brasileiro*, de Rodrigo Carreiro (2019), de certa forma contribui para essas entrevistas, pois dá uma perspectiva do trabalho dos profissionais de som no país, as visões acerca da produção sonora nacional e internacional e apresenta, mesmo que de maneira ainda incipiente, uma ideia que parece ter permanecido e até mesmo ampliado, de acordo com os entrevistados: o mercado audiovisual para aparelhos domésticos e dispositivos móveis tem abarcado um maior uso das tecnologias de mixagem orientada a objetos do que o cinema convencional, principalmente pela possibilidade de exploração binaural e seus custos mais enxutos.

As entrevistas foram conduzidas de maneira semi estruturada, seguindo um roteiro principal voltado ao entendimento do uso das novas tecnologias em produções nacionais e internacionais, suas abrangências e possibilidades a partir do uso de dispositivos domésticos. Eventualmente outras questões surgiram ao longo da conversa, sempre dentro da temática principal. As entrevistas foram gravadas com a autorização dos participantes¹⁰³ e a transcrição foi realizada pela autora; primeiramente, através do uso de ferramentas de inteligência artificial e, posteriormente, com a escuta ativa para as correções necessárias, tentando manter a originalidade e o sentido de cada frase.

Para conduzir uma boa entrevista, você precisa conhecer o assunto e o entrevistado. Precisa chegar preparado. Levar uma lista de perguntas, mas não se limitar a fazê-las numa ordem pré-determinada. Se você prestar atenção ao que o entrevistado está

¹⁰³ A proposta para a realização das entrevistas foi aprovada pelo Comitê de Ética em Pesquisa da UFMG, através da assinatura do Termo de consentimento livre esclarecido (TCLE) pelos participantes.

falando, outras perguntas vão surgir naturalmente. A transcrição das conversas também é uma arte em si mesma. É uma atividade monótona, mas você não pode fazê-la mecanicamente, porque traduzir a linguagem oral para o meio escrito dá trabalho. Declarações que soam brilhantes no ouvido resultam incompreensíveis para o olho. Transcrever bem significa traduzir para o papel, com a máxima fidelidade possível, aquilo que o entrevistado falou. Nem sempre você pode manter as palavras exatas que ele usou, mas o sentido precisa ser preservado. O equilíbrio entre ser fiel ao sentido da fala e as palavras exatas é delicado (Carreiro, 2019, p.11-12).

A entrevista com Ariel Ferrante aconteceu em 25 de março de 2025 de forma *online*. Profissional da área de mixagem com ampla experiência no mercado audiovisual brasileiro, Ariel graduou-se inicialmente em Geografia, mas cresceu muito envolvido no meio sonoro, com pai músico e mãe cantora. Ao final da faculdade, fez um curso técnico no Instituto de Áudio e Vídeo (IAV) de São Paulo e se identificou com a parte técnica e operacional da música. Com a finalização do curso, conseguiu um estágio nos estúdios Mega em 2007, também em São Paulo, para trabalhar na parte de gravação musical. Como o estúdio atuava também na parte audiovisual, Ariel acabou migrando da música para o cinema, mesmo que inicialmente a contragosto. “O meu chefe lá, uma figura saudosa, falou ‘esses grandes estúdios de música estão acabando com crise do mercado fonográfico, o futuro é audiovisual’ e eu bem resistente no início, mas comecei a me dividir: era assistente no estúdio de música e no estúdio de cinema” (Ferrante, 2025). Desde 2010, Ariel concentra seu trabalho na área audiovisual, participando de gravações e mixagens no mercado brasileiro, se estabelecendo como técnico de mixagem e supervisor de som, com extensa filmografia entre longas de ficção, documentários e séries. Atualmente, também atua como consultor de áudio para a *Fraunhofer Institute for Integrated Circuits* no Brasil, na qual trabalha dando capacitações sobre o formato MPEG-H no país. Sua trajetória oferece subsídios importantes para a compreensão dos desafios técnicos e criativos enfrentados diante das exigências dos modelos de codificação sonora.

Em 15 de abril de 2025, foi realizada a entrevista com Carlos Klachquin, engenheiro de som e funcionário da Dolby há mais de 30 anos, onde exerce diversas funções, como supervisor na implantação de sistemas de som em salas de cinema na América Latina e na habilitação de estúdios e laboratórios que fazem a cópia e a transcrição de som com tecnologia Dolby. Klachquin participou ativamente na discussão e implementação técnica das tecnologias de áudio multicanais nos fóruns de desenvolvimento da televisão digital na América Latina. Desde 2013, é responsável pela implementação do Dolby Atmos na mesma região, incluindo a supervisão da instalação e a regulação dos sistemas em salas de cinema e estúdios de finalização. A posição mercadológica de Klachquin permite o acesso a uma visão

estratégica mais ampla sobre os processos de introdução do Dolby Atmos na cadeia de produção e distribuição de conteúdos audiovisuais, pensando principalmente no consumo doméstico.

As análises e reflexões apresentadas ao longo deste capítulo buscam elucidar os testemunhos acerca das transformações estéticas e técnicas relativas ao mercado e ao consumo observados na atualidade, almejando a construção de um, ainda que breve, panorama crítico sobre o estado da arte do som imersivo em ambientes domésticos. As transcrições completas estão anexadas ao fim do trabalho.

4.1 Transição e adaptação

A transformação tecnológica dos formatos sonoros reconhecidos para as tecnologias de som imersivo exigiu um período de adaptação pelos profissionais da área, sobretudo porque, de certa forma, a tecnologia passou a estar disponível antes mesmo de sua demanda se consolidar no mercado. Inicialmente, houve um certo temor relativo a como esse novo modo de mixar conteúdo sonoro seria absorvido pelo mercado, conforme detalha o mixador e supervisor de som Ariel Ferrante (2025), em sua entrevista para a pesquisadora:

Em 2018, 2019, um pouco antes da pandemia, teve um certo momento assim de todo mundo falar: “Nossa, agora só vamos mixar assim. Quem não mixar dessa forma não vai ter mais trabalho.”, ou ainda: “Ah, se você quer produzir conteúdo para Netflix, você só vai fazer assim porque eles não vão aceitar de outro jeito.” Teve um certo exagero do tipo “o 5.1 morreu, o 5.1 acabou”, mas que não se concretizou nem de longe e tem todo um *hype* em torno. Legal, o negócio [Atmos] está andando mesmo, está evoluindo, a gente está cada vez mais em contato com, mas nem de longe chegou nisso. Noventa por cento das coisas que eu faço é finalização em 5.1, seja para cinema, seja para plataforma.

Segundo Ariel, em suas atuações nas produções para cinema e plataformas de *streaming*, nos primeiros anos, a entrega de projetos com Atmos era pouco frequente, mas foi uma época interessante para que os profissionais se familiarizassem com os novos recursos e possibilidades criativas da tecnologia, mesmo sem que houvesse ainda uma expectativa real de exibição no formato. Atualmente, no mercado sonoro brasileiro ainda não é consenso o uso do Atmos ou de outra tecnologia semelhante na finalização de um projeto, mesmo que venha crescendo com consistência, tudo ainda depende do orçamento disponível. “Tem muitos projetos que eu faço orçamentos das duas coisas, [Atmos e 5.1] e a maioria ainda fala para ir no 5.1 mesmo. Isso vem do cliente final da plataforma, Disney, Netflix etc. Está ganhando

espaço, é uma coisa que vai dominar, mas não com aquele desespero de inicialmente” (Ferrante, 2025).

Parte desta ‘agonia’ profissional sentida por Ariel é também uma estratégia comercial das empresas para lançar seus produtos e atrair novos consumidores com promessas de maior público, foi assim na transição do cinema silencioso para o falado, as telas maiores e com o cinema multicanal, logo não seria diferente com os objetos sonoros. “O entretenimento evoluiu muito assim, inventava uma coisa para se ir ao cinema e logo o *home video*, o consumo em casa chegava junto, esse jogo comercial que permanece com a ideia de que as plataformas tão dominando o mundo, o cinema vai morrer? E assim continua” (Ferrante, 2025). O modelo de mixagem com objetos sonoros se consolidou no mercado brasileiro principalmente através de plataformas de *streaming*, que, a partir da chegada de dispositivos móveis e aparelhos domésticos como televisores e *soundbars* com tecnologia imersiva ou áudio espacializado, começou a se tornar parte efetiva do fluxo de trabalho, estabelecendo padrões de entrega e normas específicas dos conteúdos com mixagem orientada a objetos sonoros. Nesse contexto, o Atmos e outras tecnologias semelhantes passaram a ser não apenas uma ferramenta de mixagem criativa, mas também uma exigência de distribuição, podendo, de certa maneira, moldar a forma como os técnicos e mixadores pensam o som em diversas etapas dos projetos.

Do ponto de vista empresarial, a transição tecnológica ocorreu em paralelo ao uso de maquinários e dispositivos mais acessíveis para a produção sonora, isto é, construir um pequeno estúdio em casa ou realizar atividades com profissionais em diferentes localidades geográficas é uma realidade da área de *sound design*. Carlos Klachquin (2025) comenta que o acesso à tecnologia atualmente é muito mais facilitado do que já foi em épocas passadas, o que acaba por gerar, na visão do entrevistado, uma problemática envolvendo o controle técnico e o padrão de qualidade das obras audiovisuais:

A produção foi democratizada, porque o mundo digital permitiu que você já não precisasse trabalhar mais com 35 mm, dava para trabalhar em um ambiente simples, com investimento menor, não só a produção se democratizou, é possível você poder trabalhar em casa para mixar. Mas isso, por outro lado, tem um calcanhar de Aquiles, acabou o filtro, acabou o controle de qualidade e também quem faz isso não tem a experiência que tinha o pessoal de som nos estúdios, que sabiam como trabalhar, lidar com esse tipo de conteúdo para cinema. Isso criou um ambiente que temos hoje na qual é necessário reeducar, treinar e conscientizar pessoas sobre esses aspectos técnicos que se deixam de lado. E não só isso na produção, mas também na exibição, especialmente em festivais, é fundamental também cuidar dos padrões, existe um padrão para cinema, é o mesmo padrão. Mudou a tecnologia mas a forma que ouvimos, monitoramos, com a exceção de Atmos, que introduz uma nova tecnologia, o resto está padronizado e nem sempre se segue esse padrão. Então, qual é a consequência disto, a intenção do criador do conteúdo não chega ao público da

mesma maneira. É como um pintor de quadros que expõe um quadro num local mal iluminado. Então a essência do que você criou, não chega ao destinatário. Então eu diria assim que é importantíssimo, principalmente numa Universidade, todo o âmbito educar nesse sentido, de tentar civilizar de novo isso que tínhamos conseguido, porque as salas inclusive melhoraram muito. Festivais meio que se improvisam, mas aí mesmo improvisando, é necessário colocar dentro da pauta para respeitar a obra do criador.

Desde a chegada das tecnologias de mixagem orientadas a objetos e a difusão comercial de *plugins* e *renders*, o Dolby Atmos foi o formato que dominou e tem dominado o mercado em relação à produção em cinema e *streaming*. “O que dominou o mercado de pós-produção de filmes e séries foi o Dolby Atmos e ponto. Não tem nem sombra de nenhum outro que chegue perto. Existem outros padrões, você deve saber melhor que eu, mas o Dolby Atmos dominou” (Ferrante, 2025). Inicialmente para desenvolver a mixagem em Atmos havia, além da estação de trabalho comum, ou seja, *softwares* como *ProTools* ou *Nuendo*, a necessidade de um processamento externo com o *Dolby Atmos Renderer*. “Nos últimos anos as próprias fabricantes integraram esse renderizador dentro das suas plataformas. *Nuendo* foi o primeiro a integrar o Atmos e, no ano passado ou retrasado, a *Avid* também incluiu o render dentro do *ProTools*” (Ferrante, 2025).

Para a Dolby, o renderizador age como um conjunto de algoritmos que codificam a informação dada pelo profissional de mixagem, ou seja, o técnico seleciona ali as intenções de posicionamento sonoro e o *software* codifica e reconstrói de acordo com o ambiente de escuta posterior. E dentro desses meios de escuta, existem dois caminhos principais, o cinematográfico e o doméstico. Ariel exemplifica inclusive que as denominações dos *codecs* se diferenciam, para cinema AC-4 é o nome do *codec* popularmente conhecido como Dolby Atmos, mas “o Atmos é um produto da Dolby que em cada plataforma é uma tecnologia diferente, para cinema é de um jeito, o que eles estão usando agora no *streaming* ainda não é a AC-4, é o AC-3, que é o Dolby Digital antigo melhorado, um AC-3 com os objetos adicionais” (Ferrante, 2025). Klachquin (2025) explica que tanto para cinema como para consumo doméstico a filosofia de aplicação é a mesma, mas existem diferenças técnicas cruciais em cada modelo:

Existem, na verdade, duas avenidas principais que são a de *Home Theater*, de entretenimento doméstico e a de cinemas, as duas são de Dolby Atmos, mas são, eu diria assim, animais de diferentes pesos. A filosofia subjacente é a mesma, é a primeira definição que eu dei, o sistema que transmite informações espaciais, mas por diferenças técnicas, muito grandes no ambiente porque cinema é um espaço necessariamente grande e doméstico é um espaço reduzido, pequeno, é a sala, é o quarto. Então o número de caixas que você utiliza é diferente, a largura de banda você dispõe para trabalhar em cinema é muito diferente da que se dispõe para

streaming, para doméstico e isso obriga a implementar de forma diferente. Ou seja, a filosofia é a mesma, mas os dados como são enviados são diferentes e têm nomes diferentes e a forma de monitorar tanto na produção, quanto como ouvimos são muito diferentes entre os cinemas e o ambiente doméstico. Embora seja muito flexível no Atmos, estas diferenças são necessárias, tem a ver como a mídia trafega. Isso é importantíssimo levar em conta, porque afeta o *codec*, que passa a ser outro, e afeta a forma da distribuição de caixas e como se processa essa informação para poder reproduzir.

As diferenças entre Atmos para cinema e ambiente doméstico não se concentram apenas nos aspectos técnicos relativos a formatos e entrega, mas também nos modos de monitoramento e reprodução, destacando-se a variedade de dispositivos possíveis que incluem a tecnologia e o controle do ambiente de escuta, que fora da sala de cinema, compete com toda a paisagem sonora ao redor.

4.2 Cinema e ambiente doméstico

As diferenças entre os ambientes de exibição apontam para a necessidade de regulação estrutural para a aplicação das tecnologias de mixagem orientada a objetos. Enquanto o cinema tradicionalmente opera com um grau de controle quase absoluto sobre o espaço acústico, o ambiente doméstico impõe uma série de variáveis que desestabilizam essa lógica. Carlos Klachquin (2025) elenca dois aspectos principais que diferem a mixagem para salas de cinema (*Atmos Theatrical*) e a direcionada a ambientes domésticos, também chamada de Atmos HE (*Home Entertainment*), primeiro em relação à monitorização e reprodução em estúdio daquilo que será o resultado final em uma grande sala cinematográfica:

A primeira é a questão da monitoração, embora seja uma questão física muito básica, é o seguinte: em um ambiente pequeno, doméstico, você tem o tempo de reverberação relativamente mais curto, já que o som tem uma velocidade definida e a acústica tem a ver com as reflexões do som no ambiente que ele limita, a condição de contorno que são os muros onde se reflete o som. Então, você imagina o seguinte, entrando um pouquinho em questões técnicas, basicamente o que é a acústica, você tem que considerar que cada elemento da superfície, teto, muro e objetos agem como um refletor, por um lado reflete o som, por outro absorve, e não reflete igual nem absorve igual em toda as frequências, normalmente absorve mais as frequências mais altas, mas às vezes acontece o contrário, depende da substância que você está encontrando nesse muro. [...] Agora pensa que isto é para cada elemento da superfície, imagina que cada elemento poderia se tornar como se fossem pequenos azulejos, que vai ter essa capacidade de filtragem dependendo da matéria. Outro aspecto é que entre o som direto que você recebeu de um alto-falante e o som que se é refletido nesse mesmo alto-falante, nessa superfície e que depois atinge o teu ouvido, existe uma diferença de tempos porque os percursos são diferentes. Existem inclusive reflexões secundárias, um som bateu nesse muro e depois bateu no outro e retorna... Isso tudo cria um ambiente extremamente complexo de sons que chegam no teu ouvido com tempos diferentes e filtrado de formas diferentes na frequência, isto tem um efeito que se acumula, o que você percebe é a soma de cada um de todos

elementos de grande quantidade de elementos com diferentes retardos. Então dá para entender, porque a ciência da acústica é tão complexa, porque cada ambiente tem sua função de transferência, ou seja, como esses sons originais que saiu de um alto-falante é modificado e chega ao teu ouvido.

Essas questões de ordem acústicas afetam tanto as salas de cinema quanto o ambiente doméstico, mesmo que em menor escala. Em cinema, o fato de haver uma plateia numerosa gera a necessidade de se gerenciar a forma como cada espectador receberá o som da produção. “Quando você projeta um cinema, especialmente o *surround*, você precisa que, na medida do possível, cada alto-falante atinja toda a plateia. E isso não é possível, fisicamente é muito difícil” (Klachquin, 2025).

A concepção comum de que a posição central da sala oferece a melhor experiência para assistir e, sobretudo, ouvir um filme está ancorada nesse princípio da acústica espacial, na qual os espectadores posicionados próximos às paredes não desfrutam da mesma qualidade sonora percebida no centro, especialmente no que se refere à reprodução dos canais *surround*. A configuração ideal de audição em sistemas *surround*, como o 5.1, pressupõe que o ouvinte esteja equidistante das caixas acústicas, que são responsáveis também por criar a espacialidade e a imersão sonora. Quando o espectador se posiciona próximo às paredes laterais ou traseiras da sala, a assimetria na distribuição das fontes sonoras compromete a coerência do campo sonoro, resultando em perda de definição espacial e alteração na percepção dos efeitos direcionais planejados na mixagem. Uma forma de amenizar essa situação é inserindo mais caixas em canais *surround* para uniformizar a distribuição. “Em Atmos existem outros compromissos além desses porque exige uma densidade de caixa maior para poder criar esse ambiente virtual, então temos fileiras de caixa no teto, além das caixas que ficam na parte de trás e na lateral” (Klachquin, 2025).

Em ambientes domésticos o problema da acústica espacial também acontece, mas em menor escala, exatamente por se tratar de ambientes menores, com menos lugares para se ocupar, como sala ou quarto. Entretanto, o som vindo dos aparelhos reprodutores compete com o ambiente, janelas abertas, ruídos de outros habitantes da casa etc., enquanto no cinema espera-se um tratamento acústico e uma sala silenciosa.

Em ambiente doméstico as necessidades são diferentes e o número de caixa é bem menor, tão menor que o espaço por cima de você, o que você sente para assistir é tão reduzido que permite implementar ou que conhecido como *soundbars*, que é muito prático, porque você não tem que fazer muitas integrações e mais ou menos consegue recriar o ambiente multicanal com um elemento só, que é uma barra de caixas que trabalham, seja por reflexão, ou por criar uma imagem virtual e mais ou menos conseguindo ter um resultado espacial (Klachquin, 2025).

O segundo aspecto apontado por Klachquin relaciona-se aos *codecs* e a forma de entrega do material mixado. Em cinema, a distribuição do conteúdo acontece através de discos rígidos ou por *download* via satélite. “Um filme pode conter um tamanho de 150 GB, 250 GB, depende se é 2D, 3D, quanto tempo dura o filme, mas está nessa ordem de centenas de gigas” (Klachquin, 2025). O arquivo baixado é lido pelos sistemas da *Integrated Media Service* e do aparelho *Integrated Media Block*, um dispositivo que decodifica os arquivos e garante que os filmes sejam reproduzidos de forma confiável, com a qualidade de imagem e áudio correta, incluindo *trailers* e anúncios.

E que aproveito para adicionar um erro bastante comum que as pessoas confundem DCP com DCI. DCI é o padrão para cinema comercial, DCP é o *Digital Cinema Package* que é o arquivão grande, o arquivo enorme, um conjunto de arquivos com o qual se entrega o filme para o cinema, ele contém tudo, legenda, áudio e imagem, *Watermark*, marca d'água, enfim, uma série de elementos de segurança. E quando você passa o arquivo que contém um som dentro do DCP, é um formato chamado MXF [*Material Exchange Format*](Klachquin, 2025).

Para manter a ideia de entrega universal em apenas um formato de arquivo, é obrigatório a exportação no padrão DCI também em áudio 5.1 ou 7.1 de maneira não comprimida, através do arquivo PCM (*Pulse Code Modulation*) do áudio, para que qualquer sala consiga exibir o material enviado. “Um cinema que não tem Atmos não vai decodificar, mas vai buscar essa informação aceita, isso é obrigatório. Então, quando se cria o conteúdo, coloca no áudio, no pacote do DCP, o MXF e também o PCM do áudio 5.1” (Klachquin, 2025). Portanto, durante a exportação da mixagem, há dois formatos principais de áudio, dependendo da plataforma que exibirá o conteúdo, o VWD, ligado à *Cubig Digital Recorder Sound* e o atualmente mais utilizado IAB (*Interactive Audio Bitstream*), que permite que conteúdos de som imersivo sejam transmitidos de forma adaptável para diversos sistemas de som, isto é, há a garantia que uma única mixagem em áudio IAB irá ser reproduzida em variados sistemas ou dispositivos sonoros.

No som doméstico, há também esses mesmos *codecs* como forma de entrega, mas existe a possibilidade do uso de um formato mais comum, o MP4 (MPEG-4 Part 14), que reúne áudio e vídeo e pode ser exportado diretamente pelo renderizador Atmos de maneira conjunta à imagem, ou seja, o arquivo *embed*. Outra forma comum de exportação é o voltado ao binaural, que será abordado com maior profundidade adiante. Klachquin enfatiza que uma mixagem sonora realizada para ambiente doméstico, utilizando ambientes de monitoração próximos ou salas de mixagem pequenas, não funcionará bem se utilizadas em salas de cinema, pois as salas de mixagem precisam ter um volume mínimo para reproduzir

aproximadamente as condições de um cinema médio. “Espaços, volumes diferentes, relações de largura, altura e comprimentos diferentes. Atmos resolve essa história da tecnicidade, mas na parte acústica você precisa ter um ambiente de monitoração próximo ao que você vai ter depois no cinema” (Klachquin, 2025). Para produções voltadas ao ambiente doméstico, como conteúdos originais de plataformas, há o chamado *Near Field Audio*, no qual as caixas de som são posicionadas de forma muito próxima e, portanto, o som direto é dominante sobre o campo de reverberação.

Isso cria uma diferença vital, se você mixar monitorando em *Near Field*, doméstico, não tente levar isso para um cinema porque vai funcionar mal e muito mal. Mixagem para cinema se faz em ambiente com um volume mínimo e com um sistema de monitorização adequado, tendo a necessidade de caixa bem maior que para o doméstico. [...] É muito importante insistir sobre isso, porque eu vejo muito realizador que desconhece esta questão. Esse material, ainda não codificado, mixado para o cinema pode ser utilizado para doméstico depois. Aí funciona, funciona bem e isso tem uma longa experiência da época de Dolby Digital de cinema que depois se colocava esse conteúdo em DVD. Então essa mesma mixagem também funciona bem em casa, mas ao contrário não, por essas questões acústicas que eu mencionei (Klachquin, 2025).

Na visão de Ariel, a produção nacional de Atmos para cinema ainda é incipiente, principalmente pelo fato de o parque exibidor do país ainda ser, em sua maioria, em 5.1. Ariel aponta que há poucas diferenças no modo de trabalhar objetos sonoros para cinema ou *streaming*, o que realmente é modificado é a amplitude de variação de volume do sinal de áudio, ou seja a diferença entre um som mais silencioso e o som mais alto que o sistema pode reproduzir sem apresentar distorção, o também chamado de *range* dinâmico.

No Brasil, a gente tem pouca gente fazendo Atmos para cinema, a maioria da nossa produção e do nosso parque de exibição é 5.1, mas independente disso mesmo se tivesse, eu sinto que a grande diferença não seria na abordagem de como eu iria trabalhar os objetos ou como eu iria espacializar as coisas. “Ah, se for cinema, eu vou fazer assim. Se for *streaming*, eu vou fazer assado”. Tem alguma diferença. Mas, eu acho que a grande diferença entre os dois não é nem de espacialidade, acho que é muito de *range* dinâmico, de possibilidade dinâmica, entendendo dinâmica como a diferença entre os altos e os baixos. [...] a maneira como vai ser consumido, é no celular, no *tablet*, com o fonezinho ou em casa numa televisão. As TVs estão dessa espessura [pequena], os alto-falantes não conseguem reproduzir muito grave, muito médio, você vai concorrer com o mundo real, o cara vai ver a TV com o celular tocando, com o Instagram aberto, com o cachorro latindo. Então eu acho que a grande diferença para mim é dinâmica, de como colocar as coisas do ponto de vista de dinâmica de alto e baixo, até mais do que espacialidade (Ferrante, 2025).

Outro ponto levantado por Ariel é uma mudança estrutural nas dinâmicas de produção sonora, motivada menos pelos avanços tecnológicos de espacialidade e mais pelas transformações no ecossistema de distribuição audiovisual, isto é, a velocidade que o produto

sai do cinema e chega em plataformas de *streaming*. Esse encurtamento entre as janelas de exibição e a consequente sobreposição entre formatos originalmente distintos acaba impondo uma exigência de versatilidade e agilidade por parte dos profissionais de som, que precisam adaptar suas entregas a múltiplos contextos de exibição quase simultaneamente, muitas vezes sem a compensação de recursos ou tempo adicional, o que pode acabar ocasionando o que Klachquin comenta em sua entrevista, nos problemas de se levar uma mixagem em campo próximo (*near field*) para salas grandes, de cinema. Ariel Ferrante (2025) aponta para as tensões entre estética, técnica e o mercado, que acabam por caracterizar o atual cenário da pós-produção sonora:

Há 10 anos, eu mixava um filme para cinema, ele era lançado no cinema e passava meses ou quiçá um ano para o produtor falar “poxa, a gente vai passar na Globo. Vamos fazer uma masterização aqui para televisão, porque vai ter que tocar na televisão”. E mil coisas mudam, não é? Eu não vou entrar nisso, mas a principal coisa é o *range* dinâmico. A TV tem muito menos range dinâmico que o cinema. Então, eu tinha muito tempo entre uma janela e outra, era muito diferente. Agora, está muito misturado, por exemplo, tem acontecido o filme original de plataforma, que foi concebido o projeto desde sempre dentro da plataforma, um produto que vai para a plataforma, mas quando a gente está finalizando o produto, o produtor vai e fala “olha, a gente quer um DCP”, que é um pacote para cinema. “A gente quer uma cópia para fazer uma estreia em cinema”. Então eu já tenho que entregar um produto só praticamente, a gente dá uma mexida, lógico, mas já não é assim, é muito misturado, se eu fiquei ali quatro semanas mixando um filme para uma plataforma, imagina eu não posso cobrar mais uma semana para ter a versão cinema, se não eles não vão pagar, vão achar que eu estou querendo extorquir, é intolerável isso para eles, do ponto de vista de recurso, de dinheiro, de tempo também. Então o que a gente está passando é muito isso, é muito emendado uma coisa na outra. Eu entrego a master da plataforma e já entrego junto a do cinema ou mesmo que seja o contrário, o filme para cinema, eu finalizo ele para cinema e na semana seguinte - aconteceu isso recentemente - um filme grande que eu fiz, na semana seguinte, eles queriam uma saída para o Telecine, que era um dos patrocinadores e uma saída para a Netflix, porque dali alguns meses ia entrar na Netflix e tudo ali de imediato. Então a gente não está tendo muito esse tempo de fazer muito diferente. Está tendo que dar uma homogeneizada nas coisas assim, o que não é bom, mas a gente é empurrado para isso.

Essa realidade apontada por Ariel vai ao encontro das críticas feitas por Klachquin em sua entrevista. Do ponto de vista de quem analisa e recebe o produto pronto, misturar a mixagem de cinema e *streaming* pode parecer um descuido ou uma falta de preparo profissional. Entretanto, diante do exposto pelo mixador, percebe-se que a situação mercadológica acaba impondo certos padrões que geram as inconveniências técnicas. Ao evidenciar a problemática é possível entender que essa questão não se relaciona apenas a um desconhecimento técnico ou uma falta de zelo na produção para cinema, mas se dá principalmente pelo encurtamento de prazos entre o lançamento para cinema e plataformas e a

limitação de recurso para pós-produção sonora que acaba por reduzir o prazo disponível para adequar perfeitamente a mixagem ao ambiente de destino.

Uma saída vislumbrada por Ariel é o uso da chamada entrega universal que as tecnologias de mixagem orientada a objetos promovem. Mesmo que o som da produção seja finalizado em 7.1 ou semelhantes, o técnico já realiza a mixagem em Atmos, por exemplo, pois o renderizador e seus algoritmos entregam o *downmix* com qualidade, ou seja, mesmo que o produto acabe sendo escutado em fones de ouvido, estéreo normal, há vantagens em mixar em Dolby Atmos, com espacialidades, pois a entrega do material final parece vir com maior qualidade, mesmo que o dispositivo final não tenha suporte para a leitura de objetos sonoros ou binaural.

Eu acho que essas ferramentas novas, como o Atmos, é muito interessante a maneira como o *codec* calcula ali os as espacialidades. E na hora de reduzir isso, do Atmos que pode estar lá multicanal, multipista, multi objeto e isso é reduzido, isso é empacotado no estéreo normal ou no binaural. Mas a maneira como isso se dá realmente é muito interessante. Então, eu acho, e muitos colegas meus estão fazendo isso, que às vezes nem são para ser Dolby Atmos, que vão ter sua vida em 5.1 ou mesmo em estéreo, a gente já mixa em Dolby Atmos, já usa o renderizador para fazer as espacialidades, porque quando ele faz esse *re-render*, essa re-renderização para binaural, realmente o resultado é muito legal. Você tem um estéreo rico ali, mais espacialidade para as coisas mais definidas espacialmente (Ferrante, 2025).

Os pontos levantados sobre as diferenças entre a mixagem para cinema e a mixagem para ambiente doméstico, principalmente plataformas de *streaming* apontam para o desafio de construir fluxos de trabalho compatíveis com essas múltiplas saídas e, como Klachquin salienta, há pelo menos três: a de cinema, a doméstica padrão e a otimizada para objetos sonoros, o que demanda tempo e organização. Com isso, a questão deixa de ser apenas técnica para se tornar também logística e financeira, refletindo as condições de produção em diversos contextos. Os relatos mostram que a migração da mixagem tradicional para a baseada em objetos não é apenas uma atualização técnica, mas uma transformação que acaba por evidenciar o próprio papel da escuta no audiovisual. No cinema, as condições de recepção são mais previsíveis e a tecnologia acaba servindo sobretudo para potencializar a imersão. Já no ambiente doméstico, há também a necessidade de atuar como uma mediação entre o som da produção e a paisagem sonora do ambiente, como trânsito, outras pessoas e demais sonoridades, oferecendo possibilidades que nem sempre se traduzem em experiências superiores.

4.3 Mixagem binaural

Entre as possibilidades oferecidas pelas tecnologias de áudio orientadas a objetos, a mixagem binaural tem ocupado um lugar de destaque, especialmente no contexto das plataformas de *streaming* e da escuta em dispositivos móveis. A simulação de espacialidade tridimensional por meio de dois canais, buscando preservar, ainda que de forma adaptada, a proposta imersiva do som multicanal, abre outro campo de desafios técnicos e estéticos. No ambiente doméstico, onde o som muitas vezes é consumido via fones de ouvido, o binaural passa a ser uma solução não apenas viável, mas estratégica para garantir uma experiência de envolvimento sensorial com a narrativa audiovisual. De acordo com as entrevistas realizadas, a mixagem binaural parece ser um campo em transformação contínua, no qual a fidelidade sonora, a adaptabilidade técnica e a recepção do público estão em constante negociação.

Carlos Klachquin afirma que o setor de pesquisa e aplicação binaural da Dolby é onde há mais pessoas trabalhando atualmente e onde a empresa mais emprega esforços de constante atualização e melhorias. Além disso, segundo Klachquin, a escuta binaural tornou-se um fenômeno global, acompanhando a crescente ubiquidade do uso de fones de ouvido no cotidiano urbano, seja em transportes públicos, ruas ou espaços domiciliares. Embora os dados quantitativos sobre o número exato de ouvintes não sejam disponibilizados, Carlos destaca que a popularização desse formato é notável e segue em expansão contínua. Ele observa que, independentemente da região, o cenário parece similar, isto é, há um uso amplamente disseminado dos fones para consumir conteúdo audiovisual. Nesse contexto, a Dolby tem direcionado grande parte de seus esforços à difusão da escuta binaural, inclusive no campo musical. A praticidade do formato também é um diferencial, pois a compatibilidade com dispositivos como *notebooks* e fones convencionais, somada ao fato de não exigir sistemas complexos para acompanhamento, uma vez que quando o material mixado em binaural pode ser monitorado no mesmo modo, favorecendo a adoção tanto por usuários quanto por profissionais na etapa de criação e finalização de conteúdos.

O binaural parece ser o arroz com feijão atual, o renderizador Dolby Atmos, você tem como exportar em binaural. Ou seja, há todo um suporte de Dolby, inclusive a Dolby incentiva a produção de conteúdo, porque interessa a Dolby que os dispositivos que têm Dolby Atmos tenham conteúdo, essa é a questão básica. Então, o peso atual sobre ensino, sobre licenciamento tem muito a ver com essas tecnologias domésticas e, dentro das tecnologias domésticas, binaural tem um peso enorme. E tem uma outra coisa linda no binaural: você pode utilizar basicamente ou praticamente qualquer tipo de fone. O fone não é parte especializada, o truque está no algoritmo que decodifica, no renderizador inserido que decodifica segundo o Atmos. E tem um aspecto que não é muito evidente, mas a forma que está construído

esse *streaming* de dados, como está codificado em Atmos permite que o conteúdo praticamente possa ser reproduzido para qualquer dispositivo, mesmo que não tenha Atmos, você não vai perder informação, vai perder informação espacial, claro, mas você não vai perder o conteúdo e isso é importantíssimo, sempre tem que cuidar da compatibilidade e da retrocompatibilidade, isso é uma filosofia permanente (Klachquin, 2025).

A premissa da retrocompatibilidade citada por Klachquin é utilizada pela Dolby desde seus primeiros sistemas, ainda analógicos e com fixação nas laterais da película. Isso não parece ter mudado com a digitalização audiovisual e o uso do áudio imersivo e da mixagem orientada a objetos. Klachquin relembra que na transição tecnológica para o som digital 5.1, mesmo após a incorporação do formato, as cópias dos filmes continuaram a incluir a faixa analógica, uma estratégia que evitava problemas logísticos na distribuição. Essa duplicidade permitia que uma mesma cópia fosse reproduzida tanto em salas equipadas com Dolby Digital quanto em cinemas com sistemas analógicos, inclusive mono. Além da praticidade, esse modelo também oferecia segurança, pois em caso de falha na reprodução digital, o sistema automaticamente poderia recorrer à faixa analógica. A multiplicidade de formatos visava garantir uma ampla compatibilidade e evitar interrupções de exibição, adaptando aos recursos disponíveis em cada sala. Trazendo aos tempos atuais, Klachquin destaca que essa mesma filosofia de compatibilidade segue presente na distribuição digital via DCP. Os arquivos entregues aos cinemas possuem tanto a mixagem em Dolby Atmos quanto versões convencionais em PCM 5.1 ou 7.1. Assim, se a sala não estiver equipada com sistema Atmos, o player automaticamente ignora os arquivos imersivos e reproduz a versão compatível. Por outro lado, salas com suporte a Atmos também se adaptam, uma vez que se o DCP não inclui conteúdo específico para Atmos, o processador recorre ao áudio convencional. O mesmo princípio permanece para o *streaming* com áudio binaural. Mesmo quando o conteúdo é mixado especificamente para esse formato, ele pode ser distribuído de forma que se mantenha reproduzível em qualquer dispositivo, independentemente de decodificadores avançados.

Outra questão central relacionada aos sistemas binaurais, abordada brevemente no capítulo anterior, é a chamada funções de transferência relacionadas à cabeça (HRTF *Head-Related Transfer Function*), responsável por modelar como o som interage com as características físicas de cada indivíduo, como o formato da orelha, da cabeça e do tronco antes de atingir o canal auditivo. Klachquin aponta que, embora os sistemas binaurais utilizem uma HRTF média baseada em dados estatísticos populacionais, essa generalização nem sempre produz os mesmos efeitos sonoros para todos, pois há uma limitação estrutural dessa forma de escuta que não se relaciona com os equipamentos ou os padrões de distribuição, mas

sim com a própria fisiologia humana. Alguns ouvintes, cujas características anatômicas coincidem mais com esse modelo médio, terão uma experiência mais convincente de espacialidade; outros, mais distantes da média, podem perceber os efeitos com menor precisão ou realismo.

Acontece que cada um de nós, cada indivíduo, tem diferenças de forma, digamos assim, assim como o nosso rosto é diferente para cada pessoa, a nossa constituição do ouvido, embora tenha os mesmos princípios, não é exatamente igual de uma pessoa para outra. Então, tem essa função de transferência, que você disse muito bem, mas quem tem ouvido mágico que tem essa função de transferência, porque cada indivíduo vai ter uma função de transferência diferente. Afortunadamente, parece ser que a dispersão, digamos assim, o sigma funciona, a diferença não é tão grande assim, mas realmente cada um tem sua própria função de transferência, então essa função de transferência se utiliza de uma transferência média, de um estudo estatístico (Klachquin, 2025).

Ariel Ferrante corrobora com a visão de Klachquin sobre o binaural e os modos de escuta na atualidade e no futuro. O mixador observa que, na prática, o consumo de conteúdo com som imersivo tende a se concentrar em dispositivos portáteis e acessíveis, como fones de ouvido, *tablets*, celulares e televisores. Embora existam tecnologias como o *Home Theater*, a adesão a esses sistemas por parte do público geral sempre foi limitada. O uso equivocado desses equipamentos, como a instalação incorreta das caixas, acabou por comprometer a experiência prometida e contribuiu para a queda de seu apelo no mercado. “Ninguém monta aquilo direito, um monte de fio pendurado, o pessoal comprava *Home Theater* 9.1 e botava as nove caixas em frente a televisão” (Ferrante, 2025). Nesse contexto, os *soundbars* surgem como uma alternativa prática e mais eficaz, oferecendo uma solução integrada e simplificada, que exige menos conhecimento profissional e esforço do consumidor. Do ponto de vista técnico, Ariel distingue que o binaural, especialmente quando viabilizado por *codecs* recentes como o Dolby Atmos ou MPEG-H, tem potencial para oferecer uma experiência auditiva mais rica e com maior sensação de espacialidade, mesmo quando a reprodução é reduzida ao estéreo, reforçando a ideia de que o futuro do áudio imersivo pode estar mais vinculado a soluções compactas e portáteis do que a montagem de sistemas dedicados.

Essa aposta no binaural se articula também com os movimentos da própria indústria em facilitar o acesso às ferramentas de criação. Um exemplo citado pelos dois entrevistados é o caso do renderizador Dolby Atmos, que hoje se encontra integrado a *softwares* como o *ProTools* ou disponível para aquisição a preços relativamente acessíveis, especialmente para instituições de ensino. Ainda que a operação do Atmos envolva certos requisitos de *hardware*, como uma estação de trabalho potente e uma interface multicanal com suporte a múltiplas

saídas, Klachquin destaca que essas configurações não são inalcançáveis. Em termos comparativos, o custo do renderizador não ultrapassa o de muitos *plugins* utilizados na pós-produção, sendo uma ferramenta viável dentro de orçamentos médios de trabalho. Segundo Klachquin, o avanço do mercado tem tornado cada vez mais democráticas as possibilidades de montar um estúdio funcional, com opções de interfaces em faixas de preço entre 2.700 a 5.000 dólares, e monitores de áudio que vão de modelos sofisticados a alternativas mais econômicas e satisfatórias.

Embora esse discurso de acessibilidade financeira possa ser levado em conta, os valores mencionados ainda estão distantes da realidade de muitos profissionais no Brasil, mesmo que os preços de sistemas e equipamentos anteriores fossem muito mais elevados devido a uma complexidade superada através de DAWs computacionais. Há uma certa estratégia industrial que busca ampliar a adesão à tecnologia por meio da redução das barreiras de entrada, como, por exemplo, a inexistência da necessidade de uma mesa de som física, apontada por Klachquin como uma simplificação importante no processo de mixagem, mesmo que diversos profissionais prefiram seguir utilizando o equipamento na garantia de controle, fluidez e precisão no trabalho sonoro.

Se, conforme Klachquin afirma, a Dolby oferece incentivos para a utilização de seus materiais, esse esforço parece se dar apenas na disponibilização técnica de ferramentas e no desenho de um ecossistema que estimula o uso do Atmos em escalas variadas de produção, das mais industriais às independentes. Com exceção do desconto para instituições de ensino, parece haver pouca perspectiva de ampliação de uma política de incentivo ao uso dos produtos Dolby por produções menores e com pouco orçamento, talvez causada pela hegemonia de mercado da empresa nas mais diversas etapas audiovisuais, de pós-produção à exibição. Além dos produtos da própria Dolby, todo o ecossistema técnico ainda é significativo, pois para operar o Atmos de forma eficiente, é necessário um computador potente, uma interface de áudio multicanal capaz de manejar pelo menos 16 canais e, em caso de monitoramento com caixas, um sistema de alto-falantes compatível com configurações como 7.1.4 ou 7.1.6¹⁰⁴. Klachquin afirma que o renderizador da Dolby é flexível o bastante para se adaptar a diferentes configurações e fluxos de trabalho e destaca que existem no mercado diversas interfaces que oferecem recursos suficientes por preços que, embora expressivos, não são considerados tão elevados diante do panorama geral da indústria.

¹⁰⁴ Configuração de sistema de som com sete canais *surround*, um canal de *subwoofer* (LFE – *Low Frequency Effects*) e quatro ou seis canais superiores (*overhead*), comumente usada em mixagens Dolby Atmos para cinema ou *home theater*.

Hoje em dia é muito popular a *Matrix Studio*, da *Avid*, porque eles reúnem uma quantidade enorme de possibilidades e de formas de entrar e sair, mas por outro lado, ela é enxuta no sentido que tem tudo o que você precisa. Tem outra opção, possivelmente mais barato que é a *Audient Oria*, também é uma interface multicanal e deve ter mais por aí, tem um mercado e cada vez que você entra para ver o que que apareceu, muda muito, mas eu diria que hoje está bastante democrático (Klachquin, 2025).

Embora haja uma uma fala enfática sobre a “democratização” da tecnologia, principalmente se comparado ao elevado custo de implementação de antigos mega estúdios de mixagem e dublagem, os valores envolvidos ainda delimitam um certo perfil de usuário, normalmente vinculado a estruturas já profissionalizadas. A ideia de que é possível montar um estúdio Atmos com investimento reduzido precisa ser contextualizada à realidade econômica local e à disparidade de acesso a recursos técnicos, mesmo com opções variadas de acesso à tecnologia.

4.4 MPEG-H e o Fórum SBTVD

O campo da radiodifusão aberta, com suas particularidades regulatórias, históricas e culturais, também passou a integrar o debate sobre formatos sonoros avançados, sobretudo no contexto do desenvolvimento da chamada TV 3.0. Essa nova geração de televisão digital mobilizou uma ampla articulação entre entidades públicas, fabricantes de tecnologia, emissoras e especialistas em engenharia, reunidos no Fórum do Sistema Brasileiro de TV Digital Terrestre (Fórum SBTVD). No centro dessas discussões esteve a definição do padrão tecnológico que norteará o futuro da televisão no país e a escolha do sistema de áudio capaz de atender às demandas de qualidade, acessibilidade e interoperabilidade. Entre as tecnologias avaliadas, o sistema MPEG-H destacou-se por suas características de mixagem orientada a objetos, pela capacidade de personalização da escuta e por seu alinhamento com modelos internacionais de inovação. No entanto, sua adoção implicou uma série de negociações e adaptações, tanto do ponto de vista técnico quanto político.

Um dos profissionais envolvidos com o Fórum SBTVD e na escolha do formato do áudio a ser adotado pela nova TV brasileira foi o mixador e supervisor de som Ariel Ferrante, que aponta desde o princípio que não enxerga o MPEG-H como um concorrente do Dolby Atmos, pois acredita ter aplicações diferentes.

O MPEG-H, é desenvolvido pela Fraunhofer, que é um instituto de pesquisa na Alemanha com vários braços lá dentro e um deles é o Instituto de Circuitos Integrados Fraunhofer IIS, a sigla em alemão. E eles pesquisam tecnologias de codificação de mídia em geral, mas acho que eles são célebres porque há 25 anos ou um pouco mais eles criaram o MP3, então que é um *codec* que mudou o mundo, né? Basicamente mudou a indústria fonográfica, indústria musical, então o MP3 foi criado pela Fraunhofer e depois disso, várias outras maneiras de codificar áudio e vídeo também. O h264 é deles, o h265, hoje em dia o AAC, o *enhanced AAC*, que é o AAC melhorado, um monte de dispositivos ao redor do mundo usam essas tecnologias. As emissoras transmitem em AAC, emissora de TV digital, os nossos celulares tocam na grande maioria também em AAC. [...] O que muita gente vê como concorrente do Dolby Atmos, mas eu não acho. Acho que realmente são universos, são objetivos diferentes. O MPEG-H é um *codec* de áudio desenvolvido pela Fraunhofer com o olhar para a emissora de TV, para *broadcasting*, tanto para conteúdo ao vivo, quanto para conteúdo preparado, pós-produzido, que vai ao ar, mas dentro do universo de TV. E qual é o grande lance do MPEG-H? Ele permite áudio imersivo, é um *codec* de áudio imersivo, ou seja, você também tem vários formatos, tipo o Atmos que têm 5.1.4, 7.1.4, 9.1.6¹⁰⁵, tem várias configurações possíveis de áudio imersivo dentro do MPEG-H (Ferrante, 2025).

A escolha do MPEG-H como sistema de áudio para a TV 3.0 no Brasil foi influenciada pelas características técnicas e também pela adequação às particularidades operacionais da radiodifusão. Ariel destaca como o formato se diferencia em termos de aplicabilidade prática quando comparado a outras soluções de áudio imersivo, como o Dolby Atmos, especialmente no contexto televisivo. O MPEG-H também utiliza a lógica de objetos sonoros, mas com uma abordagem distinta daquela proposta pela Dolby, especialmente no que diz respeito à interatividade. Enquanto o Atmos é frequentemente associado a produções de alto padrão técnico e longos processos de pós-produção, como séries e filmes em plataformas de *streaming*, o MPEG-H foi concebido com foco na eficiência e na simplificação dos fluxos de trabalho típicos da televisão aberta, sendo um diferencial decisivo, uma vez que a produção televisiva é caracterizada por prazos apertados, conteúdos ao vivo e um volume elevado de entregas diárias. Nessa realidade, recursos complexos de mixagem se tornariam impraticáveis.

O MPEG-H usa objetos de áudio e tem uma conceituação um pouquinho diferente da Dolby para prover interatividade e oferecer possibilidades do usuário final interagir e fazer escolhas no conteúdo de áudio. E de uma forma muito prática, muito simplificada, que eu acho que é o grande trunfo deles para emissoras de TV, porque quando a gente está falando de uma série da Netflix que vai ser mixada em Atmos, a gente está falando de um processo de pós-produção que você tem muito tempo, é então uma série de ficção complexa eu tenho 2, 3, 4 semanas para trabalhar em um episódio de 40 minutos, eu com uma equipe de 5, 6, 8 pessoas. Então, o Dolby Atmos exige uma complexidade que você consegue fazer ali. Conteúdo para a

¹⁰⁵ As nomenclaturas indicam configurações de sistemas de som multicanal. O primeiro número representa a quantidade de canais ao nível do ouvinte, como caixas frontais e surround; o segundo número indica o canal dedicado aos graves (subwoofer); e o terceiro número refere-se aos canais posicionados acima do ouvinte, ou seja, canais de altura ou teto, usados em formatos de áudio imersivo como o Dolby Atmos.

televisão não tem esse tempo, a televisão está no ar todo dia, então o cara tem quantos jornais no ar, eventos esportivos, reportagens televisivas que são gravadas antes, mas os fluxos de produção são muito intensos. E aí que a Fraunhofer tem essa proximidade com as emissoras de TV, porque esse *codec* que ele é muito prático e muito objetivo para esse universo da emissora de televisão, pois ele permite fazer áudio imersivo igual o Dolby Atmos, então você tem não somente o *surround*, tem a possibilidade da camada superior, mas o MPEG-H usa os objetos para oferecer interatividade, para o usuário poder interagir, por exemplo, num evento esportivo, um futebol no Brasil, se produtor do conteúdo quiser oferecer isso para o espectador, o espectador pode desligar o narrador, “só quero ouvir a atmosfera do estádio e desligar a narração”. Exemplo: um clássico lá, um São Paulo e Palmeiras, é possível ouvir apenas a torcida do Palmeiras, ou a do São Paulo (Ferrante, 2025).

Outro exemplo citado é o da audiodescrição, uma vez que, atualmente, para garantir acessibilidade, muitas emissoras precisam transmitir duas versões do áudio, uma padrão e outra com a descrição embutida, o que implica em duplicação de arquivos, maior uso da banda e aumento da complexidade operacional. “Então, isso é ineficiente, dá mais trabalho e gasta banda, gasta espaço de transmissão da emissora, ou seja, precisa diminuir a qualidade, o quanto de áudio está sendo ocupado, para poder caber naquela banda duas vezes a mesma coisa” (Ferrante, 2025). O MPEG-H, por outro lado, possibilita que essas camadas adicionais de áudio sejam transmitidas em um único pacote, utilizando metadados que permitem ao usuário final ativar ou desativar recursos conforme sua necessidade, promovendo uma personalização da experiência de escuta no sistema televisivo.

Esse termo de mixagem orientada a objetos, a Fraunhofer faz questão de falar *Real Objects*, objetos reais, porque é uma concepção diferente do que foi usado no Dolby Atmos, mais nesse sentido de que em uma cena de MPEG-H, em uma mixagem, por exemplo, se você está transmitindo um documentário jornalístico você tem a cama, a *bed*, que eles falam em inglês, a base de canais que está com tudo o que você quis mixar dentro do seu universo e se você tem um objeto que é a voz, então tem toda a mixagem de vozes, se ele for falado em português, por exemplo, você tem um objeto lá que é a voz em português e esse objeto vai dentro do mesmo pacote, mas separado da cama, com metadados que explicam para o receptor “ó esse canal aqui é um objeto chamado diálogo em português” e você pode mandar um também em espanhol e um em inglês. E o usuário chaveia, troca o que ele está querendo de forma muito fácil e é isso, sem você repetir tudo. Igual por exemplo, um conteúdo que era muito comum até um tempo atrás, acho que ainda existe, a tecla SAP que mudava para a língua original, eram duas vezes o mesmo negócio sendo transmitido, ocupando espaço. E quando pega o MPEG-H, você tem uma transmissão única e você vai chaviando, *switching*, que eles falam, alternando esses objetos. Esse seria meio que um funcionamento básico do MPEG-H (Ferrante, 2025).

Pela descrição de Ariel, é possível compreender que a grande vantagem do formato em comparação a concorrentes como o Dolby Atmos parece consistir na adaptabilidade operacional e em sua proposta de interatividade simplificada, tratando-se de uma solução pensada menos para o requinte da mixagem artística e mais para a eficiência da cadeia de produção e distribuição em larga escala. “Existem vários tipos de interatividade possíveis e

tudo já está sendo transmitido no mesmo pacote descrito com metadados, no qual o receptor lê esses metadados e o usuário tem esse controle. Então, o MPEG-H é muito eficiente, muito prático para uma emissora de TV” (Ferrante, 2025).

O envolvimento de Ariel com o MPEG-H decorre diretamente do contexto da implantação da TV 3.0 no Brasil, um processo iniciado em 2020 e ainda em fase de implementação. Assim como a TV digital, também chamada de TV 2.0, marcou a substituição do sinal analógico pela transmissão digital, agora novamente o país passará por uma nova reestruturação, motivada pelos avanços nas tecnologias de áudio e vídeo nas últimas décadas. Essa mudança não é exclusiva do Brasil, trata-se de uma tendência internacional que responde à evolução acelerada dos padrões de produção, transmissão e recepção de conteúdo audiovisual.

A condução dessa proposta de transformação foi realizada pelo Fórum SBTVD e o processo aconteceu por meio de um chamamento internacional redigido em inglês para atrair propostas de empresas e organizações com atuação global. Esse documento especificava, com base em critérios técnicos detalhados, os requisitos que deveriam ser atendidos pelas tecnologias candidatas nas diferentes camadas do sistema de transmissão: vídeo, áudio, legendas, interatividade, entre outras.

No caso específico do áudio, o documento estabelecia exigências como eficiência de compressão, qualidade sonora, preservação do sincronismo com o vídeo, suporte a áudio imersivo e possibilidade de interatividade. Embora houvesse também itens classificados como desejáveis, não obrigatórios, o conjunto das demandas traçava um perfil técnico específico. Apenas dois candidatos apresentaram soluções concretas: o *codec* AC-4, da Dolby (associado comercialmente ao Dolby Atmos), e o MPEG-H, desenvolvido pela Fraunhofer.

E no *call for proposal*, eles tinham essas exigências para você ter um áudio com boa qualidade, tem que ter eficiência de codificação, poder comprimir o máximo possível sem perder tanta qualidade, tinha que ser áudio imersivo, tinha que oferecer interatividade, tinha coisas que eram indispensáveis, tipo conseguir preservar sincronismo entre áudio e vídeo, coisas óbvias, e tinham coisas que caíram na categoria *desirable*, desejável, aí fizeram esse documento extremamente complexo e abriram em todas as áreas para ver quem se candidatava e realmente, na parte de áudio, houve dois candidatos, um terceiro que foi só teórico. Mas dois *codecs* se candidataram para oferecer tecnologia que cumprisse aqueles requisitos no Brasil. Um foi o MPEG-H da Fraunhofer e o outro foi AC-4 da Dolby, que é o nome do *codec* que comercialmente estão chamando de Dolby Atmos. [...] E aí depois desse *call for proposal*, entrou uma fase de testes que várias Universidades do Brasil participaram, por exemplo, a camada de vídeo foi analisada pelo laboratório [Grupo de Processamento Digital de Sinais] da UNB. As camadas de aplicativos foram analisadas metade na Federal de Juiz de Fora e outra metade acho que na PUC Rio de Janeiro, liderado pelo Prof. Marcelo Moreno. E a camada de áudio foi analisada na USP, na ECA, no laboratório de pesquisa da ECA, liderada por dois professores,

pelo Professor Régis Alves Faria e pelo Professor Almir Almas. Essa fase de testes durou mais ou menos seis meses e aí realmente não tinha nenhuma participação das emissoras, dos fabricantes, dos interessados. Quem foi analisar isso foram as Universidades com esse documento em mãos[...]. E depois dessa fase de testes, chegaram-se a conclusões, então todas as Universidades que estavam fazendo testes deram seus resultados para o Fórum e isso não era um negócio normativo obrigatório, era um negócio de aconselhamento, a escolha seria do Fórum, dos representantes da indústria. No caso do áudio, que é o que nos interessa, foi unânime a escolha do MPEG-H. O MPEG-H cumpriu todos os requisitos, o Dolby AC-4 não conseguiu cumprir a maioria. Então, no sentido de votantes foi catorze a zero, todos votaram por isso (Ferrante, 2025).

Todo esse processo se estendeu por cerca de cinco anos, atravessando o período da pandemia, e culminou em uma decisão técnica e coletiva da escolha dos novos padrões televisivos no Brasil. O Fórum SBTVD, responsável por coordenar os debates e oferecer diretrizes ao Governo Federal, apresentou sua recomendação com base nos estudos e análises realizadas e no consenso entre os diversos atores envolvidos, como emissoras, fabricantes de equipamentos, além das próprias desenvolvedoras dos *codecs* avaliados. Ainda que Fórum tenha um caráter consultivo, sua composição plural e o rigor do processo dão peso institucional às recomendações emitidas. Nesse sentido, embora tecnicamente não vinculantes, tais orientações foram acatadas pelo Governo como referência para a formulação das normas que regulam a nova geração da televisão aberta no país e o MPEG-H será o padrão adotado por todas as emissoras quando a TV 3.0 for oficialmente implementada no Brasil, uma transformação no panorama da transmissão sonora nacional.

Com a participação no Fórum SBTVD e a escolha do formato MPEG-H para a nova TV brasileira, Ariel acabou se tornando consultor da Fraunhofer no Brasil e, em conjunto com outros dois profissionais, promove treinamentos nas emissoras, apresentando a transição para profissionais da área. Um dos desafios dessa etapa é adaptação dos custos e investimentos por parte das emissoras e dos fabricantes, já que, segundo Ariel, acaba ficando um ‘jogo de empurra, empurra’. “O fabricante não quer implementar o *codec* porque fala que não tem nenhuma emissora transmitindo e a emissora não quer botar no ar porque fala que ninguém tem uma televisão que vai ler esse conteúdo” (Ferrante, 2025). O governo brasileiro tem sinalizado em eventos com o Presidente Lula e Ministros a intenção de lançar a TV 3.0 ainda em 2025 ou no início de 2026, de maneira gradual, de forma semelhante à mudança da TV analógica para digital. A intenção é que parte do país já tenha acesso ao novo formato para a Copa do Mundo de Futebol de 2026, de maneira simbólica, uma vez que a mudança completa deve ocorrer ao longo de uma década.

A gente está nesse momento dando treinamento para quem trabalha em emissora, para eles começarem a produzir, entender o funcionamento, não quer dizer que do dia para a noite vai estar todo mundo produzindo áudio imersivo 3D. Muito

provavelmente eles vão continuar estéreo também, mas com a diferença de que vão adicionar esses objetos, que no MPEG-H é chamado objeto, para oferecer interatividade. Então, mesmo conteúdo estéreo, mas que de repente as pessoas em casa comecem a ter televisores ou *tablets*, celulares que decodifica MPEG-H, mesmo estéreo a pessoa já vai ter ali uma opção de desligar uma voz, de trocar a língua, de adicionar um audiodescrição (Ferrante, 2025).

Com a escolha do MPEG-H para o novo padrão televisivo brasileiro e as possibilidades oferecidas, Ariel acredita que o formato deve ganhar relevância no país, principalmente fora do universo cinematográfico, como em séries e principalmente na produção televisiva, como conteúdo jornalístico, esportes, novelas, que muitas vezes já são incluídas em plataformas de *streaming* no mesmo dia da exibição em canal aberto. “Filme, série eu acho que vai continuar sendo o Dolby Atmos. Eles estão muito consolidados nisso e funcionam muito bem, agora para o mundo do *broadcast* mesmo, *Over the air*, que eles falam, transmissão pelo ar, TV aberta, o MPEG-H é uma realidade” (Ferrante, 2025). Para conteúdos ficcionais, como telenovelas, Ariel aponta uma funcionalidade que tem atraído interesse no meio televisivo que é a possibilidade de aumentar os diálogos, funcionando como uma espécie de amplificação vocal. A ideia é que o público possa assistir à versão tradicional e, caso prefira, ative um recurso identificado com algo como “mais voz”, ou outro nome, pois cada emissora pode definir o seu. Ao acionar esse botão, os diálogos passam a se destacar na mixagem, ficando em torno de oito decibéis mais altos.

O som consumido na televisão está concorrendo com o mundo real, né? Então você está concorrendo ali com a panela de pressão, um cachorro latindo, concorrendo com outras telas. Todo mundo com o celular na mão, o tempo inteiro olhando pro celular. Então, existe essa preocupação, é sempre uma preocupação do entendimento dos diálogos, da inteligibilidade do que está sendo dito na televisão e o MPEG-H tem essa facilidade. Lógico que em uma novela não vai ter a opção de desligar a voz do personagem, mas haverá a opção de aumentar todas as vozes um degrauzinho para cima (Ferrante, 2025).

Em São Paulo, a *TV Cultura* e a *TV Globo* estão transmitindo em MPEG-H de forma experimental e a tendência é de ampliação do número de emissoras e afiliadas seguindo o novo padrão da TV 3.0. Os dispositivos com capacidade de reproduzir o conteúdo nos novos formatos ainda não chegaram em grande escala no país, mas, segundo Ariel, tanto a LG quanto a Samsung já desenvolveram protótipos, principalmente pois o MPEG-H já é o *codec* de áudio padrão na Coreia do Sul desde 2017, ou seja, os televisores fabricados por lá já vem com o suporte para o formato e, por mais que não seja oficialmente divulgado, entende-se que essas fabricantes têm a capacidade técnica de ativar esse recurso nos aparelhos já vendidos por meio de uma atualização de *firmware*.

Na *SET* do ano passado, aquela feira que tem em Agosto, de tecnologia, a gente estava com uma TV da LG, uma da Samsung e uma da Hisense, que é um fabricante chinês, que está entrando no Brasil já com MPEG-H. Eu estava sintonizando o que estava no ar pela TV Cultura, decodificando em MPEG-H. Então assim, está muito perto de ter aparelho no mercado, acho que a Hisense até já está vendendo no mercado no Brasil. A Samsung e LG ainda não, mas muito perto de ter, claro que é igual a outra virada, você não troca de TV a cada seis meses, né? Acho que no meio disso vai ter essas caixinhas, essas setup boxes que decodificam. Ou quando for conteúdo que você vai consumir com o celular com o *tablet* aí é via *software*, você não precisa trocar de aparelho, é uma atualização no *app* da *Globoplay* e vai estar decodificando o MPEG-H, então a gente está realmente nessa nesse momento de viradas, eu acho que esse ano fora Hisense, talvez ainda em 2025 já possa ter coisa no mercado (Ferrante, 2025).

Carlos Klachquin, representante da Dolby na América Latina, preferiu não emitir opinião na entrevista sobre o tema da televisão 3.0 e a escolha dos formatos de áudio, explicando que é uma discussão comercial com a qual não está diretamente envolvido. “Obviamente a ideia é que se utilize esse *codec*, mas não são as únicas opções, enfim, é uma discussão que tem a ver com questões industriais, comerciais e de padronização também” (Klachquin, 2025). Ele ainda lembrou sua participação durante a transição da televisão analógica para a digital, quando integrou uma equipe que discutia o uso do *codec* AC-3, o mesmo adotado nos DVDs, como padrão de compressão de áudio multicanal. Na ocasião, o sistema escolhido no Brasil, e também adotado pela maior parte dos países latino-americanos (com exceção da Colômbia), incorporou um padrão desenvolvido pela Dolby, baseado no *codec* HE-AAC (*High Efficiency AAC*).

4.5 Perspectivas futuras

O horizonte de utilização das tecnologias sonoras imersivas e orientadas a objetos no audiovisual indicam um cenário de continuidade no avanço técnico, acompanhado por desafios relacionados à adoção, padronização e integração nos fluxos de produção. Ao analisar a incorporação dessas ferramentas de maneira geral e a partir das declarações dos entrevistados, é possível perceber que o cenário é de expansão, especialmente em contextos como da televisão digital de nova geração e das plataformas de *streaming*, em que os requisitos por qualidade e personalização da experiência sonora ganham relevância. A consolidação dessas tecnologias depende também de uma articulação entre diferentes setores da indústria, incluindo fabricantes de equipamentos, desenvolvedores de *software*, emissoras e produtores de conteúdo. Além dos fatores técnicos, aspectos comerciais e regulatórios têm um

papel decisivo no ritmo e na forma como esses formatos se tornarão parte do cotidiano da produção e da recepção audiovisual.

Ariel se mostra mais conservador ao refletir sobre o modo como essas ferramentas vão se desenvolver na mixagem sonora, avaliando que o uso das tecnologias imersivas ainda está em fase de aprendizado e desenvolvimento. Segundo ele, não há uma estética consolidada ou um conjunto de convenções amplamente aceito para a mixagem em formatos como o Dolby Atmos. Enquanto o 5.1 já conta com diretrizes mais estabelecidas, embora ainda sujeitas a variações conforme o estilo de cada mixador, no caso do Atmos, a percepção é de que o campo permanece aberto, com decisões criativas bastante influenciadas por preferências individuais. Há profissionais que utilizam a camada superior apenas para determinados elementos sonoros, enquanto outros a empregam de forma mais constante, por exemplo, para ambientes. Dessa forma, a padronização do uso desses recursos ainda parece distante. “Hoje, eu acho que a gente ainda está aprendendo, eu acho que ainda não tem uma estética pronta, feita” (Ferrante, 2025). Outro ponto levantado pelo mixador é a questão da necessidade deste tipo de sonoridade imersiva de acordo com o tipo de produção. “Um *Ainda Estou Aqui* [2024], acho que é o maior exemplo nosso, é um que sobrevive bem sem uma tecnologia dessa, ela talvez não fosse agregar tanto a narrativa, a experiência, quanto um filme como *Avengers* [2012, 2015, 2018, 2019] possa ter se utilizar esse tipo de plataforma (Ferrante, 2025)”.

A questão orçamentária também é apontada como um limitador para o desenvolvimento de produtos nacionais com esses formatos, pensando principalmente em diferenças com produções internacionais, mesmo que do ponto de vista tecnológico, os equipamentos e *softwares* utilizados sejam os mesmos, a questão se concentra no tempo disponibilizado para realizar o trabalho de design sonoro e mixagem.

Então eu acho que primeiro é o tipo de produção. E claro, chover no molhado, o orçamento. Com uma ficção bem remunerada no Brasil, eu tenho quatro semanas de mixagem, em boas condições. Nos Estados Unidos, eles têm quatro semanas para cada pré-mixagem, quatro semanas de pré-mix de diálogo, quatro semanas de pré de efeito, quatro semanas de pré de música e daí quatro semanas de mix final. Eu faço uma semana para cada um, uma de diálogo, uma de efeito, uma de música, uma de final, com sorte quando o filme tem dinheiro, porque o normal tem sido fazer os filmes em três semanas. Então, a gente usa as mesmas ferramentas do ponto de vista de tecnologia, a gente não está atrás, só que a gente tem menos infraestrutura e menos tempo e dinheiro, que nesse caso é sinônimo. Se a produção tiver dinheiro para oito semanas de mixagem ou para vinte semanas de edição de som, vai ter um produto diferente do que vai ter em seis semanas de edição do som, né? Seis semanas de mixagem vai soar diferente do que três semanas de mixagem. Eu acho que a gente não tem problema de tecnologia defasada, não tem problema de mão de

obra. Hoje em dia os recursos que eu uso e a formação que eu tenho é basicamente a mesma que eu acho que um mixador lá em Hollywood deve ter (Ferrante, 2025).

Segundo Ariel, em média, os projetos finalizados em Atmos tendem a custar de 20% a 30% a mais que os finalizados em 5.1. “Eu entendo que envolve uma tecnologia mais elaborada, eu vou ter que investir mais em equipamento, eu tenho que ser um profissional um pouco mais atualizado para trabalhar com isso, então eu entendo que isso tem que custar de 20 a 30% a mais” (Ferrante, 2025). Ariel chama atenção ainda que a limitação orçamentária nos projetos não acontece apenas na pós-produção sonora, a realidade que ele vive, mas em todas as etapas, como no som direto que enfrenta problemas desde uma visão limitada de produtores, que tentam enxugar o orçamento com menos diárias de técnicos, a até a compra de equipamentos, que são precificados em dólar.

Eu acho que tem toda uma coisa estrutural, mesmo que vai refletir lá no produto final. [...] Existe também uma questão cultural da geração mais antiga, que talvez não dê tanta atenção para o som quanto deveria. Eu nunca fui um cara de *set*, o que eu estou falando é o que eu escuto dizer, mas dependendo do *set*, se o técnico de som pedir para refazer o *take* ele vai tomar um esporro, vai tomar um não, vai ficar um clima ruim para cima dele. Ele falou “ó esse *take* não valeu para mim, vamos fazer outro”, o cara vai tomar um chega para lá. Agora se um fotógrafo falar isso tudo bem, claro, o fotógrafo é o rei. Então tem uma questão cultural aí, infelizmente. A gente ouve essas coisas do *set* e isso reflete no produto que vai chegar, não tem como, por mais que na pós a gente reconstrua as coisas, chama o ator para dublar, fazer ADR, mix, se tem ferramentas mais modernas e tal, nada supera (Ferrante, 2025).

Por fim, Ariel confessa acreditar que a mixagem orientada a objetos não exerce uma influência significativa sobre a narrativa de um filme. Embora reconheça o valor de um bom som para a qualidade geral da obra, ele destaca que esse tipo de tecnologia não representa uma mudança estrutural na forma como as histórias são contadas e, por mais que possa enriquecer a experiência sensorial, não altera substancialmente o impacto narrativo da obra. A presença ou ausência de elementos imersivos não modifica o sentimento ou a compreensão da cena de forma determinante. “Como o *Ainda Estou Aqui* [2024][...] ‘Ah, se aquele helicóptero no começo do filme, que a Fernanda tá lá na água, se aquele helicóptero tivesse passado na camada superior, ia dar um sentimento...’ Não. É uma pena, mas não” (Ferrante, 2025). Ainda assim, conforme investigaremos adiante através da análise fílmica, o som imersivo, principalmente utilizado através do ponto de escuta, parece promover emoções e sentimentos de maneira mais considerável do que em versões multicanaís convencionais.

Baseado em sua experiência, Ariel observa ainda uma tendência de crescimento no uso da mixagem orientada a objetos tanto no cinema quanto nas plataformas de *streaming*,

mesmo que seja uma expansão lenta, sem aquele receio inicial de que a adoção dessas tecnologias fosse uma exigência absoluta para o mercado de trabalho.

No começo, pouquíssimos projetos exigiam que fossem feitos, por exemplo, em Dolby Atmos, e essa porcentagem está aumentando. Então eu acho, é o palpite meu, que dentro de algum tempo realmente quase tudo produzido para as plataformas – hoje em dia, quem pede Atmos é a Netflix, a Amazon e a Disney – eu acho que a tendência é que isso domine, mas a passos pequenos, não sei te falar em quanto tempo (Ferrante, 2025).

Carlos Klachquin aponta que atualmente adota uma postura mais cautelosa ao tentar prever o futuro das tecnologias, especialmente diante das transformações constantes no meio sonoro e audiovisual. O engenheiro compartilhou um dado recente que considerou surpreendente, pois, segundo informações recebidas na semana anterior à conversa, uma rede internacional de serviços de dados indicou que o volume de dados trafegados por redes sociais superou o das plataformas de *streaming*. “As pessoas assistem mais *TikTok*, *Facebook*, esses conteúdos gerais para o próprio usuário; estão trafegando mais informação nesse ambiente do que a plataforma de *streaming*” (Klachquin, 2025). Para Klachquin, esse dado evidencia uma transformação significativa nos hábitos de consumo audiovisual e levanta questões relevantes sobre o futuro da produção profissional. Além disso, por ser de uma geração anterior, acaba observando com certo distanciamento o impacto das redes sociais, mas entende que há aspectos preocupantes, principalmente os relacionados a uma desordem informacional, ausência de filtros e controle de conteúdo por faixa etária. Ainda dentro do contexto da produção sonora e audiovisual, ele destaca que, ao se falar de tecnologias como o Dolby Atmos, trata-se de um mercado profissional de compra e venda de conteúdo, cuja lógica parece estar sendo superada pelo fluxo massivo de dados das redes.

Klachquin observa, ainda, que o cenário atual é marcado por rápidas transformações tecnológicas, tanto em *hardware* quanto em *software*, citando ferramentas como *ProTools*, *Blackmagic* e *Nuendo*. Essas mudanças ocorrem em ritmo tão acelerado que qualquer análise sobre o momento presente já nasce com prazo de validade.

Inclusive, dentro da própria tecnologia Atmos existe uma velocidade de atualização e de reformulação muito intensa, muito intensa e não é exclusivo de Dolby isso, isso está acontecendo em todos os lugares. Então, eu te diria que as nossas análises têm que ser cientes de que tem um prazo de validade, não é como na Idade Média em que você esquece e esse livro ficava no monastério por séculos. Já não é mais assim (Klachquin, 2025).

O aproveitamento doméstico dos dispositivos de melhoramento sonoro, como *Home Theaters*, é um fator limitador para Klachquin, pois, apesar do avanço e da sofisticação dos recursos atuais, a instalação e o uso doméstico desses sistemas ainda podem ser complexos para o público geral, especialmente para quem adquire os equipamentos sem contar com suporte técnico profissional. Nesse contexto, tecnologias capazes de calibrar automaticamente o sistema com base nas características do ambiente representam uma tendência relevante, como os aparelhos do tipo *soundbar*.

Dolby está investindo muito na área doméstica, sugiro você fazer um pequeno exercício, entrar em dolby.com, na página Dolby, que também se transforma toda a semana muda, mas você vai ver que tem um grande peso o ambiente doméstico, você vai encontrar a maior parte é de *Home* e vai encontrar um pedacinho de cinema que, noutra época, foi contrário, normalmente, Dolby estava no cinema. E o peso que tomou o ambiente doméstico indica uma tendência e não se vai reverter facilmente e isso tem a ver, pura especulação minha, não é de Dolby, sou eu Carlos, mas imagino que o cinema vai encontrar formas de adaptar-se a isto, eu estou suspeitando de que o cinema vai passar a ser uma experiência especial. [...] A experiência do cinema, embora em casa você possa ter tecnologias fantásticas, 4k, Dolby Atmos, é uma experiência diferente, experiência de sair de casa, de encontrar outros de experiência coletiva. Isso é o valor que tem o cinema, é um valor próprio do cinema, que não dá para substituir, seria como uma espécie de paralelo, embora todo mundo tenha em casa uma cozinha, você não dispensa um restaurante de vez em quando. Por que? Porque a experiência é diferente, o cinema tem essa potência. Eu estou pensando que possivelmente o cinema irá apontar salas sofisticadas, você vai buscar uma coisa diferente e isso torna a experiência do cinema diferente (Klachquin, 2025).

Klachquin ressalta também os avanços em tecnologias de imagem, como o Dolby Vision, cuja capacidade de contraste e brilho amplia significativamente a experiência visual tanto em salas de cinema quanto em aparelhos domésticos compatíveis. Essas transformações tecnológicas são vistas por ele como parte de um processo contínuo no qual empresas como a Dolby desempenham papel central, promovendo inovações, como o Dolby Atmos e o Dolby Vision. Essas tecnologias estão sendo gradualmente disponibilizadas também para o ambiente doméstico, refletindo a crescente presença do HDR e do Atmos em conteúdos distribuídos por plataformas de *streaming*. No entanto, apesar do entusiasmo com os avanços técnicos, o entrevistado reforça que a essência do cinema permanece na capacidade de contar boas histórias. Recursos tecnológicos são ferramentas que ampliam a experiência, mas não substituem o núcleo da prática cinematográfica, que é envolver o espectador por meio da narrativa.

Dito essas coisas todas de técnica, embora eu esteja atirando no meu próprio pé, eu prefiro ser honesto. Um bom cinema é uma boa história bem contada, essa é a essência do cinema. Vender a ideia, colocar quem está assistindo nesse outro mundo

que você está criando. E o resto ajuda, ajuda sim, mas a essência é essa, não pode perder a questão básica que é isso, é como contar uma história. Me conta uma história, como criança, me conta uma história que eu gosto... Me engana que eu gosto. É isso que fazemos, precisamos disso, por isso é fantástico poder criar essa fantasia. E Dolby continuará, digamos, com todas inovações, é uma empresa de engenharia, com muito pessoal, muito talentoso, desenvolvendo projetos de coisas variadas. Eu sou um implementador, não estou na parte de desenvolvimento, mas tento fazer pontes, como estou fazendo agora parte dessa tarefa de tentar de alguma forma entender um pouco do que está acontecendo (Klachquin, 2025).

As entrevistas realizadas tornaram possível a reflexão sobre um momento de transição e reconfiguração das práticas sonoras, marcado pela convivência entre diferentes formatos e expectativas de mercado. A introdução de tecnologias de mixagem orientadas a objetos, como o Dolby Atmos e MPEG-H, não implicou uma ruptura imediata ou total com paradigmas anteriores, mas vem promovendo uma expansão das possibilidades criativas e técnicas dentro de um campo ainda em busca de estabilização estética. A ausência de convenções bem estabelecidas para a nova tecnologia, citada por Ariel, evidencia que o processo de incorporação envolve não apenas equipamentos, mas negociações subjetivas entre práticas, formações e concepções individuais.

Ao mesmo tempo, a crescente demanda por conteúdos imersivos nas plataformas de *streaming* e a aposta da indústria em dispositivos, como *soundbars* e sistemas binaurais, indicam um deslocamento importante em direção ao ambiente doméstico como espaço privilegiado de escuta. O consumo audiovisual passa a ser cada vez mais fragmentado e personalizado. A experiência coletiva, associada ao cinema, passa a coexistir com formas de fruição mais individualizadas, ainda que tecnicamente sofisticadas.

Esses elementos sugerem que a centralidade do som no audiovisual contemporâneo está menos vinculada a uma revolução pontual e mais a uma série de ajustes graduais nos modos de produção, circulação e escuta. A tecnologia, longe de ser neutra, participa dessas transformações como mediadora, um vetor de escolhas estéticas, mas continuamente dependente da mediação humana para se tornar linguagem. Mesmo diante de mudanças tecnológicas relevantes, é na articulação entre forma e conteúdo que se produzem os sentidos e as experiências que definem essa prática artística.

Em síntese, os desafios enfrentados na implementação de tecnologias de mixagem orientada a objetos não se restringem à adoção de ferramentas novas ou à formação técnica, envolvem uma reconfiguração ampla do modo como o som é concebido, executado e entregue ao público, considerando questões estéticas, operacionais e contextuais. Os relatos de Ariel Ferrante e Carlos Klachquin mostram que, embora a promessa de flexibilidade e personalização seja real, ela vem acompanhada de tensões significativas entre intenção

criativa e imprevisibilidade do ambiente doméstico ou, ainda, entre o ideal da imersão e as limitações práticas dos dispositivos e dos cronogramas de produção. A partir disso, pode ser relevante observar como essas tecnologias vão se consolidar (ou não) na rotina de produção e recepção audiovisual nos próximos anos, e de que maneira esse processo moldará não só a técnica, como também a escuta contemporânea.

5 EXPERIÊNCIAS SONORAS NO CONTEXTO DOMÉSTICO

Conforme abordado nos capítulos anteriores, a consolidação das plataformas de *streaming* como espaços de distribuição e fruição audiovisual tem implicado em transformações significativas nos modos de produção, circulação e recepção das obras. Entre essas transformações, destaca-se a crescente adoção de tecnologias de mixagem orientada a objetos sonoros, como Dolby Atmos e MPEG-H, impulsionada por uma demanda técnica e mercadológica por conteúdos otimizados para sistemas de reprodução doméstica avançados, como *soundbars*, televisores com suporte imersivo, configurações de *home theater* e principalmente dispositivos móveis, como *tablets* e celulares. Nesse contexto, observa-se um investimento progressivo na exploração de soluções sonoras que permitem tanto a espacialização tridimensional do som quanto a oferta de diferentes possibilidades de personalização da escuta. Este capítulo propõe uma análise fílmica orientada pela escuta crítica da pesquisadora, com o objetivo de examinar como tais tecnologias vêm sendo incorporadas às estratégias narrativas, estéticas e criativas no audiovisual.

Neste capítulo, busca-se observar como esses sistemas têm sido empregados em obras lançadas diretamente para plataformas de *streaming*, considerando tanto as possibilidades técnicas oferecidas pelos formatos de áudio orientado a objetos quanto suas implicações narrativas e estéticas. A análise será conduzida a partir da escuta atenta de determinadas sequências de diferentes filmes, com ênfase na espacialização sonora, na relação entre som e ponto de escuta e na construção de ambiências. O objetivo é identificar de que maneira a adoção dessas tecnologias impactam as estratégias de mixagem e como essas decisões se articulam com a linguagem fílmica. Pretende-se produzir uma análise comparativa da escuta através de fones de ouvido em sistemas com áudio baseado em objeto, como Dolby Atmos, e em sistema comum doméstico, como o padrão estéreo. O dispositivo escolhido é um *tablet* Samsung S6 *lite*, que possui de maneira nativa a opção de habilitação para Dolby Atmos em suas configurações, podendo ser ligada e desligada em simples toques a partir do menu de acesso rápido. Essa possibilidade permite que uma mesma cena seja analisada repetidamente nos diferentes tipos de sistemas sonoros que o aparelho oferece, levando a uma análise comparativa dos formatos de áudio possíveis. Além disso, serão utilizados fones de ouvido do tipo circumaural, modelo JBL Quantum 200, vendido como apropriado para jogos e áudio imersivo, apesar de nenhuma configuração especial além das espumas isolantes por fora da orelha. Explicitar esses detalhes técnicos da escuta para a análise fílmica com som imersivo auxilia na percepção das possibilidades sonoras disponíveis e gera um melhor entendimento

sobre em quais condições foi realizada a pesquisa, uma vez que diferentes dispositivos, mesmo que com configurações espaciais, podem gerar impressões subjetivas diferentes.

Diferentemente de abordagens analíticas centradas na investigação aprofundada de obras ou cineastas específicos, esta pesquisa parte do interesse na evolução das tecnologias sonoras e nas formas como vem sendo sua implementação tanto no circuito cinematográfico quanto nos ambientes domésticos de exibição. A análise filmica, nesse contexto, configura-se como instrumento metodológico para observar, a partir da escuta atenta do produto audiovisual, em que medida se concretizam as promessas de espacialização sonora anunciadas pelas tecnologias baseadas em objetos.

Atualmente, é possível observar a incorporação progressiva do chamado áudio espacial em catálogos de diferentes plataformas de *streaming*, sobretudo em produções originais. Netflix e Apple TV+ destacam-se como os aplicativos que oferecem maior número de títulos com suporte ao áudio orientado a objetos, utilizando o padrão Dolby Atmos, ainda restrito aos planos de assinatura mais avançados. O serviço HBO Max também disponibiliza conteúdos com marcação Atmos, em sua maioria filmes oriundos do circuito cinematográfico. A Disney+ inicialmente adotava o padrão DTS:X em conjunto com o selo IMAX *Enhanced*, oferecendo uma experiência audiovisual expandida, e atualmente segue as demais plataformas e utiliza a versão Dolby Atmos. Já a Amazon Prime Video não explicita as especificações sonoras de seus conteúdos, adotando apenas o selo UHD (*Ultra High Definition*), que se refere à resolução de imagem em 4K, e HDR (*High Dynamic Range*) para a variação de preto e branco.

A seleção dos filmes analisados neste capítulo considerou, portanto, lançamentos originais distribuídos diretamente por plataformas de *streaming* e que apresentassem, de forma declarada, configuração de áudio baseada em objetos, permitindo verificar como tais recursos sonoros são explorados em obras concebidas para o consumo doméstico. Para isso, serão analisadas três obras que se destacam pelo uso expressivo do som: *Okja* (Bong Joon-ho, 2017), *Nada de Novo no Front* (*Im Westen nichts Neues*, Edward Berger, 2022) e *A Sociedade da Neve* (*La Sociedad de la Nieve*, J. A. Bayona, 2023). As escolhas contemplam diferentes gêneros e abordagens narrativas, possibilitando observar como a mixagem orientada a objetos é mobilizada em distintos contextos dramáticos. Em *Okja*, a análise se concentrará na experimentação da novidade tecnológica, na representação acústica da criatura e nos contrastes sonoros entre natureza e cidade. Em *Nada de Novo no Front*, será explorada a construção da paisagem sonora da guerra e o uso do ponto de escuta para acentuar a subjetividade dos soldados. Já em *A Sociedade da Neve*, a atenção recairá sobre a utilização

dos silêncios vocálicos, das ambiências e dos sons corporais em um uso narrativo e sensorial para a criação de um ambiente de suspensão e sobrevivência.

Para além de toda a bibliografia tratada nos capítulos anteriores acerca da temática do ponto de escuta e das possibilidades tecnológicas do uso de ferramentas de mixagem orientada a objetos sonoros no ambiente doméstico, a análise filmica pretende caracterizar as sonoridades de acordo com parâmetros sonoros cânones, abordados também no livro *A imersão sonora no cinema*, de Rodrigo Carreiro, Débora Opolski e Rodrigo Meirelles (2023), no qual citam a frequência como as vibrações por segundo, resultando em sons graves, médios ou agudos; a amplitude como a intensidade sonora; o timbre como a combinação de frequências que confere textura ao som; e a densidade relacionada à quantidade de elementos sonoros simultâneos e suas intensidades (p. 88-90). Aqui também será utilizado o conceito do ponto de escuta imersivo para relacionar os momentos em que é possível ampliar a percepção do espectador para além da audição, despertando sentimentos e emoções.

Além disso, também será levado em conta as recomendações dadas por Leonard Brett (2018) sobre a espacialização sonora com o objetivo de alcançar a imersão por parte do espectador. No capítulo “*Applications of Extended Multichannel Techniques*”, do livro *Immersive Sound The Art and Science of Binaural and Multi-Channel Audio*, o autor propõe uma série de diretrizes práticas e conceituais voltadas à criação de experiências sonoras imersivas por meio de sistemas multicanais. A premissa que orienta sua abordagem é a integração entre clareza narrativa e espacialidade perceptiva, sustentada por princípios psicoacústicos e estratégias de mixagem adaptadas às tecnologias contemporâneas. Uma das recomendações diz respeito à preservação da inteligibilidade do diálogo, ressaltando que, independentemente da complexidade do sistema ou da quantidade de canais disponíveis, a clareza da fala deve permanecer como prioridade composicional, algo já discutido anteriormente também neste trabalho. Para isso, é comum a mixagem do diálogo no canal central ou em posições fixas do campo frontal, evitando conflitos com a música ou efeitos e ambiências que possam comprometer a percepção. Já os canais *surround*, segundo o autor, devem priorizar os sons ambientes, podendo, inclusive, ampliar uma ‘percepção visual’ fora de quadro a partir da espacialização sonora. “Os efeitos sonoros também oferecem um meio de ampliar o tamanho aparente da cena, por meio de um posicionamento frontal exagerado à esquerda/direita, ultrapassando aquilo que está visualmente presente na tela” (Brett, 2018, p. 347)¹⁰⁶.

¹⁰⁶ Tradução livre da autora. Texto original: “Sound effects also provide a way to stretch the apparent size of the scene, by exaggerated frontal left/right placement, exceeding that which is visually present on screen”.

Brett sugere ainda a aplicação de técnicas como o *delay panning*, ou *pan* com atraso temporal, em que pequenas defasagens entre canais criam a sensação de movimento ou reforçam a localização de uma fonte no espaço. Esse procedimento, associado à variação de timbre e intensidade, amplia as possibilidades de posicionamento, sobretudo em sistemas com maior número de canais. A dimensão vertical, segundo o autor, deve contemplar sons que se movem em eixos diagonais ou verticais, como pássaros, chuva ou ainda movimentação de aeronaves. A presença sonora vertical pode ser reforçada ainda por meio da aplicação de reverberações com componentes direcionais voltados para o eixo vertical ou pela utilização de respostas impulsivas capturadas em ambientes tridimensionais.

Leonard Brett (2018) discute também a organização da mixagem em sistemas orientados a objetos, propondo que os elementos centrais à narrativa, como o diálogo e a trilha musical principal, sejam fixados nas chamadas camas sonoras (*beds*), isto é, o conteúdo semelhante ao mixado em canais, enquanto os objetos móveis ou ambientes dinâmicos sejam alocados como objetos independentes. Essa separação favorece tanto a clareza quanto a personalização da escuta, permitindo que elementos periféricos sejam adaptados conforme o sistema de reprodução do usuário. A criação de envolvimento sonoro é outro ponto recorrente em sua análise. A distribuição de fontes ao redor do ouvinte, com atenção à coerência de reverberação e à difusão sonora, amplia a sensação de presença e continuidade espacial.

As diretrizes propostas pelo autor, ainda que elaboradas com foco prioritário em ambientes de exibição específicos para cinema, podem servir como referência inicial para a análise que se segue. No entanto, é necessário considerar suas implicações e limitações quando transpostas para o contexto doméstico, especialmente no caso da escuta realizada por meio de fones de ouvido, cuja espacialização envolve outros parâmetros perceptivos e técnicos, como as Funções de Transferência Relacionadas à Cabeça (HRTF), suas indicações de escuta de acordo com a anatomia individualizada de cada ouvido e o modelo ‘médio’ seguido por *softwares* e *sound designers* para a mixagem binaural. Essas observações permitem discutir não apenas a aplicação técnica das ferramentas de áudio imersivo, mas também os modos como essas tecnologias são incorporadas aos discursos audiovisuais, colaborando para a criação de atmosferas sensoriais que dialogam com o conteúdo dramático e a proposta estética de cada obra.

5.1 *Okja* (2017)

Okja (2017), dirigido por Bong Joon-ho, foi a primeira produção original da Netflix a vir com a opção de som em Dolby Atmos. A história articula crítica social, ficção científica e melodrama para abordar de maneira caricata os efeitos do capitalismo globalizado sobre os vínculos afetivos e os modos de vida tradicionais. O filme narra a trajetória de Mikha (Seo-Hyeon Ahn), uma jovem que vive com seu avô nas montanhas da Coreia do Sul. Desde seus quatro anos, a personagem convive com Okja, um “superporco” geneticamente modificado, criado como parte de um experimento global conduzido pela corporação Mirando. A empresa, sob a liderança midiática de Lucy Mirando (Tilda Swinton), seleciona 26 fazendas em diferentes países para criar espécies de “superporcos” por um período de dez anos, com o objetivo de descobrir o exemplar mais eficiente para alimentação em larga escala. Okja, criada livremente e com afeto por Mikha e seu avô, é considerada o animal mais bem-sucedido do projeto.

Ao final do período de criação, representantes da Mirando chegam à fazenda para levar Okja a Nova York, onde será exibida como peça central de uma campanha de marketing sobre alimentação sustentável. Mikha, inicialmente enganada sobre o destino do animal, descobre que Okja será submetida a testes científicos e eventualmente abatida. Movida pela urgência de resgatar a amiga, ela inicia uma busca que a leva de Seul aos Estados Unidos, enfrentando obstáculos institucionais, violência e manipulação midiática. Durante sua jornada, Mikha se depara com o grupo ativista ALF (*Animal Liberation Front*), que pretende denunciar os abusos da Mirando a partir de uma operação secreta. No entanto, os interesses do grupo nem sempre coincidem com o desejo de Mikha de simplesmente recuperar Okja. A narrativa avança entre cenas de ação, momentos de tensão emocional e crítica às estruturas do consumo e da indústria alimentícia.

O arco final do filme se desenvolve dentro das instalações industriais da Mirando, onde Mikha presencia o funcionamento da cadeia de produção de carne e as consequências do projeto de biotecnologia. Após confrontar a diretoria da empresa, Mikha consegue resgatar Okja, comprando sua liberdade com uma peça de ouro maciço em formato de porco herdada de seu avô. O filme encerra com o retorno de Mikha e Okja para a montanha, em contraste com a continuidade da lógica de exploração global. Nesse percurso, o filme propõe uma crítica ao agronegócio, ao poder corporativo e à estetização da violência, ao mesmo tempo que enfatiza os vínculos emocionais e éticos entre humanos e animais. A estrutura do filme alterna entre cenas de tom íntimo e pastoral e momentos de ação frenética em cenários

urbanos, compondo um mosaico que mescla crítica ecológica, denúncia de práticas corporativas e reflexão sobre a espetacularização da violência.

A construção sonora do filme auxilia na articulação desses registros narrativos e afetivos. O uso das possibilidades sonoras com a tecnologia de mixagem orientada a objetos ainda é incipiente, com poucas experimentações, exatamente por ser uma das primeiras produções voltadas ao mercado doméstico com Dolby Atmos. Ainda assim, *Okja* demonstra um controle do espaço acústico, seguindo a ideia levantada por Leonard Brett (2018, p. 333), de que um dos primeiros instrumentos que os *sound designers* utilizam na mixagem para um sistema imersivo é o de “espalhar o som”, neste caso, para gerar empatia, tensão e estranhamento, especialmente em cenas de fuga e contenção, como a perseguição em Seul e no laboratório da Mirando, no qual as paisagens sonoras marcam ambientes densos e dinâmicos, contrastando com as ambiências dos sons da montanha.

A trilha musical de *Okja* é composta por Jung Jae-il, músico sul-coreano que também assina o som de outros filmes de Bong Joon-ho, como *Parasita* (2019) e obras com relevância de público, como a série *Round 6* (*Squid Game*, 2021-2025), do criador Hwang Dong-hyuk. A trilha se destaca pela variedade estilística e pelo modo como dialoga com as diferentes nuances sonoras do filme, oscilando entre aventura, drama e sátira. Jung emprega instrumentos acústicos, percussões marcantes e temas melódicos que reforçam tanto a inocência da protagonista Mikha quanto o caráter épico e crítico da narrativa. O design de som foi realizado por Choi Tae-young e Dave Whitehead, também colaboradores frequentes do diretor, e contribui para a expressividade dramática e para a construção de atmosferas sensoriais que ampliam o envolvimento do espectador com a narrativa.

Um dos aspectos destacáveis da trilha musical é a capacidade de construir climas que acompanham as transformações emocionais da história. Em momentos de leveza e ternura, há arranjos quase lúdicos, já em sequências de tensão ou denúncia, a música adquire tons mais graves e incisivos. Nesse sentido, a experiência sonora do filme de trazer a mixagem baseada em objetos para o ambiente doméstico evidencia as transformações estéticas e tecnológicas no campo sonoro do audiovisual contemporâneo e sua articulação com os modos de circulação digital. Ainda que o filme não explore tão criativamente os recursos do áudio imersivo em Dolby Atmos, a experimentação, em *Okja*, se destaca pelo modo como utiliza o som em consonância com sua proposta narrativa e política.

Os momentos de vocalização de Okja, com uma mistura de sons de porco, elefante e outros elementos mais digitais, ajudam a estabelecer sua humanidade simbólica e o vínculo afetivo com Mikha, agindo não apenas como um efeito realista, mas também como parte de

uma estratégia de empatia, ativando no espectador um reconhecimento emocional que transcende a linguagem verbal. Na maioria das vezes, o ponto de escuta das cenas é construído de modo a manter a perspectiva subjetiva de Mikha como eixo central. Em muitos momentos, o som de Okja é enfatizado na mixagem como se filtrado pela escuta da protagonista, principalmente em momentos quando ela interage com o animal, em que a vocalização surge em primeiro plano sonoro, de maneira clara e próxima, mesmo que outros sons do ambiente estejam presentes, reforçando o elo afetivo e perceptivo entre as duas personagens. Em cenas de separação ou sofrimento, como na captura de Okja ou os momentos de confinamento, a vocalização adquire um caráter mais doloroso, e o ponto de escuta pode se deslocar, ora mantendo a perspectiva de Mikha à distância, ora se aproximando da própria experiência sensorial do animal. Essa alternância funciona ampliando a carga emocional da cena e aproximando o espectador que acaba por partilhar dessa aflição vivida pelas personagens.

A espacialização do som é utilizada de forma pontual para acentuar tanto a dimensão épica da jornada da protagonista quanto os contrastes entre natureza e indústria, inocência e crueldade. Logo na sequência de abertura nas montanhas da Coreia, há destaque para os sons ambientes, como vento, vegetação, insetos e pássaros, que são inseridos de maneira espacializada, situando o espectador dentro do habitat natural de Okja. Esse som ‘respirado’ ajuda a criar um vínculo afetivo e imersivo entre o público e o ambiente idílico em que vivem Mikha e sua companheira animal. Em contraposição, a cena da fuga pelas ruas de Seul apresenta uma construção de camadas sonoras mais densas, com ruídos urbanos, passos, freadas, vozes e gritos que atravessam os canais superiores e laterais, sugerindo o caos em movimento. Na sequência da entrada da fábrica da *Mirando Corporation*, os sons se tornam cada vez mais metálicos, industriais e opressivos, com o uso de reverberações e sons mecânicos que reforçam o contraste com os ambientes naturais anteriores e evidenciam como a mixagem sonora pode ser utilizada não apenas para conferir verossimilhança às cenas, mas também para impor uma visão ambiental e ética.

5.1.1 Mikha e Okja na montanha

A sequência de abertura de *Okja* apresenta as personagens Mikha, uma jovem de 14 anos, e a superporca Okja e suas vivências em um ambiente natural e isolado, localizado nas montanhas sul-coreanas. Trata-se de uma introdução construída com delicadeza visual e sonora, na qual a dimensão afetiva do vínculo entre as personagens é intensificada pela

integração entre som, imagem e movimento harmônico e natural. A câmera acompanha a dupla enquanto percorrem trilhas, atravessam riachos e buscam alimento. A ausência quase total de sons vocálicos é compensada por uma sonoridade comunicativa entre Mikha e Okja, composta por grunhidos e resmungos emitidos pelo animal e respondidos com entonações suaves da menina, sugerindo uma relação de entendimento através do som.

A paisagem sonora que envolve a cena é densa e detalhada. Há presença constante de sons naturais, como vento, água corrente, pássaros, movimentação de folhagens, em camadas sobrepostas que constroem um ambiente auditivo coeso. Na versão com Dolby Atmos, essa construção sonora se dá de maneira mais ‘aberta’, com experimentações ainda comedidas. Os sons agudos do canto dos pássaros surgem em faixas superiores, enquanto os passos de Mikha e Okja ressoam em faixa inferior, sobre o solo, com variações de intensidade e reverberação de acordo com o tipo de terreno, terra seca, pedras, vegetação úmida. A intensidade sonora das vocalizações do animal é modulada conforme sua proximidade com a câmera, o que inicialmente parece prezar mais pela fidelidade, contribuindo para a construção de um som diegético verossímil e emocionalmente envolvente. Esses grunhidos iniciais que tecem a relação entre as personagens começam a se desenvolver posteriormente para o campo da inteligibilidade, promovendo um entendimento de diálogo entre as duas, não apenas um efeito sonoro animalesco. A distribuição dos sons em diferentes eixos do campo auditivo cria uma sensação de profundidade espacial que contribui para situar o espectador não apenas como um observador externo, mas como parte do mesmo espaço percorrido pelas personagens. O ponto de escuta se alinha ao ponto de vista diegético, especialmente quando a câmera adota planos subjetivos ou acompanha Mikha em seus movimentos mais próximos da superfície da floresta.

A composição sonora da cena é complexa e contínua, balanceada para evitar a saturação do espaço auditivo. A trilha musical extradiegética é incorporada de forma fluida e equilibrada, alternando momentos de ausência e outros de presença discreta, sem se sobrepor aos sons ambiente. Em termos de densidade, há momentos de maior concentração de elementos, como quando Mikha corre pela floresta ou desliza com Okja encosta abaixo, e outros mais rarefeitos, como na sequência em que ela se deita sobre o animal, marcando uma queda na intensidade geral do som.

Na versão estéreo, a mixagem mantém os principais elementos sonoros, mas há uma perda na precisão do posicionamento e nos detalhes das camadas sonoras perceptíveis. As ambiências permanecem audíveis e coerentes com a imagem, mas tendem a ser percebidas como um plano sonoro mais homogêneo. A ausência de canais superiores e de maior

separação horizontal reduz a sensação de profundidade e imersão narrativa. Ainda que o estéreo preserve os conteúdos fundamentais para a construção da cena, a escuta não se aproxima da mesma forma da perspectiva das personagens, funcionando mais como uma ilustração do ambiente do que uma experiência. As vocalizações de Okja e os sons naturais mantêm seu timbre e função narrativa, mas perdem parte de sua textura espacializada. O balançar das folhas e o som da água corrente, por exemplo, soam mais planos e menos dinâmicos, o que minimiza, mesmo que parcialmente, a construção de um ambiente auditivo imersivo.

A cena com a mixagem espacializada permite que o som seja percebido como algo que ocorre em torno do espectador, e não apenas diante dele. Em *Okja*, esse tipo de escuta se associa diretamente à proposta estética do filme, enfatizando o vínculo afetivo entre humano e animal por meio de uma aproximação sensorial com os espaços que compartilham. A diferença entre os dois formatos de escuta não altera a compreensão narrativa da cena, mas modifica significativamente sua experiência, sobretudo em termos de imersão e relação emocional com o ambiente. O Atmos intensifica a sensação de presença e pertencimento ao mundo fílmico, enquanto o estéreo cumpre uma função mais descritiva, mantendo a inteligibilidade, mas oferecendo menor envolvimento físico e sensorial com a diegese.

5.1.2 O resgate de Okja no centro urbano de Seul

A sequência ocorre a partir dos 42 minutos de filme, no momento em que Mikha tenta resgatar Okja em meio a uma perseguição pelas ruas de Seul, após a superporca escapar do caminhão da corporação Mirando com o auxílio dos ativistas da ALF. A narrativa adota uma estrutura de ação em ritmo acelerado, com múltiplos elementos simultâneos, como veículos em alta velocidade, gritos, explosões, atropelamentos, quedas e impacto físico dos corpos dos personagens, tudo isso intercalando momentos de hiper sonoridades e abafamento súbito, que ressaltam a desorientação dos envolvidos e se estendem por túneis, avenidas e passarelas. Mikha, desesperada para alcançar Okja, atravessa cruzamentos, escala barreiras de contenção e é arrastada pelos excessos da metrópole. Pela perspectiva da construção sonora, a sequência apresenta certa complexidade, combinando camadas de efeitos, sons ambientes urbanos, trilha musical e vocalizações.

O design sonoro da cena é composto por múltiplas camadas, as ambiências sonoras urbanas incluem o som contínuo do tráfego, buzinas distantes, vozes indistintas da multidão e o eco urbano característico de passagens subterrâneas, com a inserção de efeitos sonoros

pontuais, como pneus derrapando, colisões metálicas, gritos próximos e o impacto de corpos caindo ou sendo empurrados. A própria movimentação de Okja, seus passos pesados, a respiração alterada e os rosnados de esforço compõem os *hard-effects* da cena e reforçam sua fisicalidade. A trilha musical extradiegética, por sua vez, surge em momentos pontuais para intensificar o ritmo e sustentar emocionalmente o drama da perseguição, com instrumentos de percussão e timbres sintéticos que seguem a cadência dos eventos visuais.

No sistema Dolby Atmos, o som dos deslocamentos de Okja e Mikha interage com a arquitetura do espaço urbano representado, com a presença de reflexos sonoros quando os personagens entram nos túneis, e uma sensação concreta de abertura quando emergem em vias largas. Mesmo nos momentos de maior sobreposição sonora, a mixagem mantém e preserva a inteligibilidade dos diálogos principais, como a voz de Mikha, por exemplo, claramente audível e reverberante quando ela grita por Okja no meio do caos e dentro do túnel. A presença de frequências graves reforça os momentos de tensão, principalmente quando Okja fica encurralada, sua vocalização ganha o protagonismo sonoro e as demais sonoridades desaparecem momentaneamente, criando um efeito físico de pressão sonora sobre o espectador. As vocalizações da superporca e seu timbre grave e modulado também transmitem dor, medo e exaustão.

Já na versão em estéreo, embora os principais elementos sonoros se mantenham presentes, acabam perdendo parte de sua articulação espacial. Os sons de impacto permanecem eficazes, especialmente quando Mikha é empurrada ou quando objetos caem no asfalto, porém, soam mais centralizados. O mesmo ocorre com a ambiência urbana, que perde profundidade e dispersão lateral. Os sons reverberados dos túneis são simplificados, e a densidade da multidão se dilui num campo sonoro mais plano. A trilha musical funciona narrativamente, mas não tão integrada ao ambiente como na mixagem imersiva. Além disso, a ausência de canais verticais limita a percepção de altura, especialmente interessante de ser escutado em momentos como nas passagens por baixo dos viadutos e nos túneis da cidade. O espectador compreende o que está acontecendo, mas com menor sensação de estar ‘dentro’ da ação. Em Atmos, o espectador consegue se perceber inserido no centro do espaço acústico da cena, compartilhando o ponto de escuta subjetivo próximo ao da personagem. Quando Mikha cai e tudo se torna momentaneamente confuso, há um tratamento na mixagem que reduz as frequências mais altas e ‘amortece’ os sons, simulando um estado de desorientação auditiva e semi-inconsciência. No estéreo, essa nuance é menos evidente, a sensação de abafamento existe, mas sem o mesmo grau de verossimilhança.

Embora o estéreo mantenha a narrativa compreensível, é na versão Atmos que o som se torna um vetor ativo de significação, capaz de expandir o espaço fílmico e aprofundar a experiência sensorial do espectador. A espacialização de sons ambientes, efeitos, trilha musical extradiegética e tratamentos com o ponto de escuta imersivo, sensibilizando emocionalmente o espectador com a fuga de Okja e o desespero de Mikha, reforçam a ideia da mixagem espacializada como recurso narrativo e estético, exemplificando as contribuições que o som do filme conferem ao ritmo e à atmosfera dramática.

5.1.3 Laboratórios de testes Mirando

Após ser capturada nas ruas de Seul, Okja é transferida para um centro de controle nos Estados Unidos, onde encontra diversos outros animais da mesma espécie, mas com diferentes tamanhos, variações e deficiências, todos presos em divisórias metálicas semelhantes a gaiolas. Okja é levada a um galpão sujo e escuro onde é submetida a procedimentos forçados, que vão além de exames clínicos, pois há uma imposição de acasalamento com outro superporco.

A câmera revela Okja amedrontada, com a movimentação limitada, enquanto escuta vozes humanas indistintas e ecos metálicos ao fundo. Sua respiração ofegante e suas vocalizações roucas sugerem uma angústia crescente, antecipando a chegada do Dr. Johnny Wilcox (Jake Gyllenhaal), um cientista-celebridade que encarna a convergência entre o espetáculo e a crueldade. Ao adentrar o laboratório, o personagem performa manipulando instrumentos cirúrgicos com entusiasmo e traz um superporco macho que copula com Okja. Tal situação é revelada por meio de uma câmera oculta acoplada ao dispositivo no ouvido de Okja, que transmite as vocalizações angustiantes do animal, captadas pela ALF. Visualmente marcada por contenção e crueldade, a cena evita a exposição direta do ato de violência, priorizando a escuta como canal primário da narrativa. A protagonista não está em cena, e a ação se desenrola sob o ponto de vista dos integrantes da ALF, que assistem e escutam por meio de um monitor de vigilância.

O som do vídeo reproduzido no computador da ALF é o único canal de acesso à situação. Há uma escuta indireta por parte do espectador, que ouve através de uma representação auditiva fragmentada e distorcida, em que os efeitos, vocalizações, ruídos metálicos de contenção e movimentação dos corpos são comprimidos e levemente saturados, simulando as limitações técnicas da gravação. O ponto de escuta do espectador, nesse caso, é semelhante ao dos ativistas da ALF, que escutam o sofrimento de Okja e, em determinado

momento da cena, pedem para que a transmissão seja silenciada. Ao escutarmos como os personagens e compartilharmos de seus sofrimentos, nos aproximamos do chamado ponto de escuta individual, caracterizado por Høier (2013) como aquele que acontece em situações subjetivas, nas quais o personagem vivencia uma situação física ou psicológica singular, geralmente atrelado a cenas com intensa carga emocional. Além disso, através da escuta habilitada em Atmos, atingimos o ponto de escuta imersivo quando sentimos a angústia e o desespero enfrentado por Okja, compartilhando a escuta com os ativistas.

Na versão em Atmos, mesmo com a origem sonora limitada à tela do monitor, há a separação entre o espaço diegético da ALF e o espaço da violência remota. Os sons reproduzidos no monitor parecem recuados no campo sonoro e ganham uma textura abafada, com timbres definidos que tornam as vocalizações de Okja reconhecíveis. No sistema estéreo, a distinção entre os espaços se mantém, mas de maneira menos evidente. A vocalização de Okja, sua respiração tensa e seus grunhidos são perceptíveis, mantendo a carga dramática da cena. O timbre do som comprimido emula um vídeo de baixa fidelidade, o que contribui para o efeito de verossimilhança e torna o desconforto palpável, mesmo sem a espacialização mais complexa.

5.1.4 Mikha invade as indústrias Mirando

Na parte final de *Okja*, Mikha consegue se infiltrar nas instalações industriais da Mirando com o objetivo de resgatar a superporca Okja. Guiada por membros da ALF, ela percorre corredores metálicos e setores de vigilância. A câmera acompanha seu deslocamento lateral, enquanto ela testemunha o funcionamento de uma linha de produção de abate e processamento de carne. Ao longo de seu percurso, são reveladas diferentes etapas da industrialização dos superporcos, como o transporte automatizado dos corpos, o abate mecânico, a separação dos cortes e o empacotamento final. Ainda que o filme evite o choque visual explícito, a *mise en scène* oferece uma visão da escala e da impessoalidade do sistema. O momento culmina com a identificação de Okja entre os animais confinados, marcando o ápice emocional da sequência.

Logo que Mikha cruza o portão e começa a circular entre as alas da indústria, os sons diegéticos ganham protagonismo através de motores, compressores, ruídos mecânicos e zumbidos de máquinas que compõem a paisagem sonora da sequência, marcada por sons repetitivos e agressivos que sugerem a cadência automatizada da indústria. O ruído constante das máquinas estabelece um pano de fundo sonoro denso e ininterrupto. Há impacto, choques

mecânicos, sons de fluidos cortando a carne e mecanismos pneumáticos em funcionamento, que criam uma textura sonora carregada e impessoal. A quase ausência de diálogos e a rarefação de sons humanos reforçam a alienação do espaço, destacando a lógica automatizada do sistema de produção. Interessante ainda perceber que as poucas vozes distinguíveis falam em espanhol, não em inglês, como no restante do filme, podendo indicar que o trabalho realizado no abatedouro é feito por imigrantes, possivelmente da América Latina. O ritmo das engrenagens e esteiras imprime um compasso repetitivo e maquinal que ressoa com a frieza do processo. A intensidade sonora desses elementos varia de acordo com a distância da câmera e o ponto de escuta que segue o da protagonista. Quando Mikha observa os animais sendo mortos, a presença sonora se acentua com sons de guinchos, choques elétricos e armas de pressão, combinados a ruídos corporais abafados, evocando o fim da vida animal sem recorrer ao choque visual explícito. A escuta prioriza a materialidade das ações, evidenciando o sistema mecânico desumanizado que domina o espaço.

Na comparação entre a experiência auditiva com o sistema Dolby Atmos e a escuta estéreo, é perceptível uma diferença na espacialização e na densidade sonora. Com o Atmos, os sons industriais adquirem profundidade tridimensional. O ponto de escuta de Mikha é frequentemente assumido como o do espectador, com os sons surgindo de diferentes localizações conforme ela se move pelo espaço. O zumbido das esteiras, os estalos metálicos das guilhotinas e os ruídos abafados dos corpos em queda ganham direções distintas, reforçando a arquitetura opressiva da cena. A escuta não apenas acompanha o campo visual, mas expande sua percepção do espaço ao redor, sugerindo que aquilo que não se vê ainda assim está em funcionamento. Já na versão estéreo, embora os sons principais sejam audíveis e a narrativa não perca sua compreensão, há uma redução na percepção da espacialidade e da textura sonora. O cenário industrial se torna mais plano, menos sensorialmente envolvente, o que reduz, ainda que parcialmente, o efeito imersivo da cena. Ainda assim, a sequência mantém sua força dramática e crítica, principalmente pela escolha de sons realistas e não estilizados que traduzem a violência institucional.

Importa também notar que nesta sequência não há trilha musical extradiegética até o momento em que Mikha faz a proposta de troca de Okja pela peça de ouro, o que torna o som ainda mais direto e realista. A exclusividade da escuta ‘real’ da personagem, sem filtros emocionais trazidos pela música, reforça o efeito de testemunho e coloca o espectador em uma posição mais crítica e desconfortável, alinhada ao olhar de Mikha. O som não apenas descreve o espaço, mas ajuda a construí-lo sensorialmente, ampliando a percepção de

desumanização e reforçando o contraste entre o olhar sensível de Mikha e a lógica impessoal da indústria.

5.1.5 Considerações

A análise sonora de *Okja* permite reconhecer sua importância como um marco inicial na adoção do sistema Dolby Atmos por plataformas de *streaming*, especialmente pela Netflix que, à época, ainda consolidava sua presença na produção de longas-metragens originais. A escolha por utilizar a tecnologia de áudio orientado a objetos revela uma intenção em explorar recursos imersivos que transcendam o padrão multicanal tradicional, oferecendo uma outra experiência de escuta. Diferentemente de produções posteriores com domínio mais consolidado dessas tecnologias, *Okja* apresenta um uso ainda pontual e exploratório do Atmos. A espacialização é ativada de forma seletiva, como na cena em que Mikha tenta romper a barreira de vidro na sede da Mirando. Inicialmente, o impacto do corpo contra o vidro parece ineficaz, e o som enfatiza essa resistência com a colisão provocando um ruído seco e um silêncio ressonante subsequente. No entanto, a vibração do copo sobre a mesa, acompanhada de um som sutil e grave de ressonância, antecipa a ruptura iminente. Quando a porta enfim se despedaça, o som da quebra é amplificado com grande riqueza de textura, os estilhaços se espalham em múltiplas direções, e a espacialização sonora reforça a intensidade do gesto. Nesses trechos, o design sonoro realiza uma mediação entre espectador e imagem, articulando frequências sonoras variadas, intensidades distintas, texturas e ritmos que contribuem para a densidade emocional da narrativa.

A tecnologia de áudio imersivo parece favorecer a construção de um espaço auditivo tridimensional, no qual o espectador é inserido de modo a compartilhar, ainda que de forma introdutória, das emoções e sensações das personagens, como ocorre nas vocalizações de *Okja* ou na ambientação claustrofóbica dos laboratórios, através do ponto de escuta imersivo. A densidade sonora é também modulada em função da dramaticidade das cenas, em que momentos mais intensos são marcados por maior complexidade de camadas, enquanto cenas de introspecção ou solidão da personagem Mikha tendem a adotar uma abordagem mais silenciosa, com sons ambientes mínimos e ritmo desacelerado. Em contraste, a reprodução estéreo, embora mantenha a inteligibilidade dos diálogos e a continuidade narrativa, compacta a espacialização sonora e reduz a percepção da direção e da profundidade dos sons. Ainda assim, demonstra eficácia ao sustentar o vínculo dramático entre espectador e narrativa, reforçando que a arquitetura sonora de *Okja* representa uma etapa importante de transição

para a adoção mais ampla de tecnologias imersivas por plataformas de *streaming*, servindo como ponto de partida para as demais investigações sobre o amadurecimento desses formatos no audiovisual digital.

5.2 *Nada de Novo no Front* (2022)

O filme *Nada de Novo no Front*, dirigido por Edward Berger, é uma produção original da Netflix alemã, de setembro de 2022. Adaptado do livro *Nada de Novo no Front* (1929), de Erich Maria Remarque, e do filme *Sem Novidade no Front*, de 1930, dirigido por Lewis Milestone, a obra acompanha o alistamento e a trajetória de jovens soldados no *front* ocidental, evidenciando a brutalidade do conflito e a desconstrução dos ideais nacionalistas que inicialmente motivaram a entrada no exército alemão. A narrativa é conduzida com ênfase na experiência sensorial da guerra, priorizando a construção de atmosferas que reforçam o esvaziamento da subjetividade diante da máquina bélica.

Nesse contexto, a mixagem realizada via Dolby Atmos, opera como elemento da *mise en scène*, sobretudo na criação de paisagens sonoras que reproduzem a espacialidade caótica do campo de batalha. Sons como explosões, artilharia distante, ruídos de trincheira e respirações ofegantes se movimentam pelo espectro audível, muitas vezes substituindo a música como vetor emocional. A organização sonora reforça o colapso físico e psicológico dos personagens, oferecendo um exemplo de como o áudio imersivo pode ser integrado de forma narrativa e expressiva a uma estética de guerra que evita o heroísmo e privilegia a experiência corporal e fragmentária da violência. Ao eleger o som como componente ativo da encenação, o filme permite investigar de que modo tecnologias como o Atmos reconfiguram a percepção espectral em obras concebidas para plataformas de *streaming*, aproximando o ambiente doméstico de uma experiência sensorial mais densa, historicamente associada às salas de cinema ou equipamentos robustos como *home theaters*.

O filme acompanha Paul Bäumer (Felix Kammerer), um adolescente alemão que alista-se voluntariamente com os colegas de escola para lutar pelo Império Alemão na Primeira Guerra Mundial, tomado por ideais nacionalistas e a promessa de honra e glória. Logo após o treinamento básico, ele é enviado para o *front* ocidental, onde a realidade da guerra revela-se brutal e desumanizante. Nas trincheiras da França, Paul testemunha o horror da guerra, com bombardeios constantes, combates corpo a corpo, fome, doenças e mortes aleatórias. A narrativa não segue uma estrutura heróica tradicional, ao contrário, expõe o

esvaziamento progressivo da identidade e da esperança de Paul diante da violência impessoal do conflito.

Paralelamente, o filme mostra os bastidores diplomáticos onde oficiais e políticos alemães negociam o cessar-fogo com a França, evidenciando o contraste entre a realidade vivida pelos soldados e a lógica distante da hierarquia militar e política. À medida que o tempo avança, os amigos de Paul vão morrendo um a um, até que ele permanece quase sozinho, exceto pela amizade construída com Stanislaus “Kat” Kaczinsky (Albrecht Schuch), soldado mais velho e experiente, que se torna o principal aliado e mentor de Paul nas trincheiras. Kat representa uma espécie de elo de humanidade em meio ao colapso moral e físico ao redor, guiando Paul em estratégias de sobrevivência e fornecendo uma relação afetiva fundamental em meio à desintegração emocional do protagonista. Sua morte, ao final do filme, marca uma virada definitiva na trajetória de Paul, simbolizando a perda total de qualquer amparo emocional ou racional na guerra.

Nos minutos finais do conflito, mesmo após o acordo ser assinado, um general ordena um último ataque, numa tentativa de manter uma honra militar ilusória. Paul morre nesse ataque final, pouco antes das 11h do dia 11 de novembro de 1918, quando oficialmente há o cessar-fogo, após a assinatura do Armistício de Compiègne, representando a rendição alemã e o fim da Primeira Guerra Mundial. A cena final mostra um novo recruta recolhendo a placa identificadora de Paul, insinuando a repetição do ciclo de violência e banalização que a guerra impõe.

Nada de Novo no Front adota uma abordagem sensorialmente imersiva da guerra, em que o som ocupa papel significativo na construção da atmosfera de desolação e violência. Desde o primeiro momento em que Paul e seus amigos ingressam no exército, o som funciona como agente de transformação subjetiva, no qual a substituição do entusiasmo juvenil pelo ruído ensurdecido das trincheiras marca o início da desconstrução de suas identidades.

A trilha musical, criada por Volker Bertelmann (Hauschka), foi a vencedora do Oscar da categoria em 2023 e destaca-se pela presença de timbres dissonantes, com o tema de abertura composto por três notas graves distorcidas que se repetem como uma espécie de motivo fúnebre. O compositor utilizou na criação das sonoridades o harmônio, um instrumento semelhante a um órgão, sem os tubos, para os sons graves. “Depois usei os violinos em forma solo, para o tema principal, destacando os momentos de vulnerabilidade. E compus um tema para o violoncelo que representa a amizade entre os companheiros”

(Bertelmann, 2023)¹⁰⁷. Esses elementos são recorrentes ao longo do filme e associados a momentos de tensão e morte iminente, criando um padrão sonoro que se sobrepõe à lógica narrativa e confere unidade estética ao caos visual. Além disso, Bertelmann explica que tentou se desvencilhar dos instrumentos de metal e sopro, comuns nas trilhas de filmes de guerra.

O design de som, por sua vez, explora as camadas do campo de batalha com um detalhamento que é auxiliado pelo uso de tecnologias de mixagem orientada a objetos, uma vez que o ponto de escuta do espectador é situado dentro do espaço acústico da guerra, ouvindo o ricochete de balas que atravessam diferentes direções, as explosões que reverberam pelo solo e os gritos dispersos dos soldados em meio à lama. A mixagem do filme, assinada por Lars Ginzler, Viktor Prášil, Frank Kruse e Markus Stemler, foi concebida com suporte para Dolby Atmos e se vale desta tecnologia para construir uma ambiência sonora que acentua tanto a brutalidade quanto o esvaziamento emocional da guerra. É possível notar a utilização da espacialização sonora para além do realismo técnico, isto é, há a participação do som na construção da subjetividade dos personagens, conforme analisaremos a seguir. O uso dos canais superiores é comedido, mas pontual, como na presença de bombas que sobrevoam o campo e as trincheiras, no deslocamento de veículos de guerra e do trem com os oficiais, mas, principalmente, nas sequências de chuva. Há também uma alternância entre momentos de densidade sonora, explosões, gritos, tiros e o silêncio, que enfatiza o impacto psicológico e físico das batalhas.

A trilha musical de Bertelmann se entrelaça com a camada de efeitos e ambientes, reforçando a tensão e o desamparo. Há uma escolha por destacar sons orgânicos e diegéticos, como respiração, tremores, passos no barro, ampliando a dimensão sensorial da experiência. Mesmo em meio ao caos sonoro da guerra e da trilha musical, o filme privilegia a inteligibilidade dos diálogos, algo comumente utilizado nos filmes clássicos, assim como visto no segundo capítulo deste trabalho. Para isso, há um controle da faixa dinâmica, permitindo que os sons tenham o impacto desejado sem comprometer a clareza vocálica.

5.2.1 Abertura

A sequência de abertura do filme se desenrola inicialmente em uma floresta coberta por neblina, onde o silêncio e a ambientação natural criam uma expectativa sensorial na

¹⁰⁷ Tradução livre da autora. Fala original (em vídeo): “Danach habe ich die Geigen solo eingesetzt und sie für das Hauptthema verwendet, um die verletzlichen Momente hervorzuheben. Und ich habe ein Cellothema komponiert, das die Freundschaft zwischen den Kameraden repräsentiert”.

trama. Na versão com Dolby Atmos, a mixagem se caracteriza por uma construção sonora mais detalhada em amplitude, ou seja, parece haver uma intensidade sonora maior e uma variação de timbres diferenciada. Logo nos primeiros segundos, o som do vento que atravessa as árvores é perceptível não apenas nas laterais do campo estéreo, mas também na dimensão vertical. O som parece deslocar-se do alto para baixo, atravessando o espaço auditivo de modo contínuo, sendo possível observar a utilização do eixo de altura.

Ao fundo, escutam-se folhas farfalhando, galhos estalando e pássaros em intervalos dispersos. Esses elementos sonoros são distribuídos em pontos específicos do espaço ao redor do espectador, sendo possível, por exemplo, perceber um pássaro cantar fora de quadro na borda superior esquerda, seguido de um som súbito vindo do chão à direita, um trovão, acompanhado pelo clarão imagético na mesma direção. Essa densidade observada na espacialização, mesmo em momentos de ‘calmaria’ na história do filme, contribui para a construção de um campo auditivo que simula a escuta *in loco* da floresta e aponta para um uso semelhante às recomendações de Brett (2018) sobre o *delay panning* no áudio imersivo.

Esse recurso auxilia na criação de uma escuta mais densa em ambientes complexos, como em florestas, pois explora a sensibilidade do ouvido humano às diferenças temporais mínimas entre canais sonoros. Pequenos atrasos na chegada do som são interpretados pelo sistema auditivo como variações espaciais, ampliando a percepção de movimento e localização no campo sonoro. Em uma paisagem florestal, esse procedimento permite sobrepor camadas acústicas que simulam a dispersão natural de sons, como o canto de pássaros que parece deslocar-se de um ponto a outro, ou o farfalhar de folhas que ressoa de direções múltiplas ou, ainda, a chuva que se distribui de maneira difusa sobre o espaço. Ao associar tais defasagens a variações de timbre, intensidade e reverberação, o *delay panning* reforça a impressão de tridimensionalidade e aproxima a experiência de escuta das características multidirecionais do ambiente natural. Em um paralelo imagético, daria para compará-lo ao efeito visual da paralaxe, em que diferenças mínimas de deslocamento entre planos próximos e distantes criam a ilusão de profundidade na imagem. Do mesmo modo, no domínio sonoro, as defasagens temporais entre canais geram um campo auditivo mais complexo, no qual os elementos parecem ocupar diferentes distâncias e direções em relação ao ouvinte.

Nos planos seguintes, o uso do silêncio também se destaca. A mixagem em Atmos não preenche constantemente o espaço com som, havendo uma valorização dos vazios acústicos que funcionam como plano de fundo para a emergência pontual de sons localizados. Essa

alternância entre silêncio e atividade sonora acaba amplificando a tensão e antecipando o colapso da calma natural.

Essa ruptura se dá com a chegada gradual dos sons da guerra, vindos de fora do campo visual, assustando os animais aninhados. Na versão em Atmos, é possível escutar detonações distantes, abafadas e graves, que se aproximam lentamente. Um estrondo parece vir do canto inferior esquerdo traseiro, outro do alto à direita, criando uma trajetória sonora que antecipa o movimento narrativo. Esses sons não apenas preenchem o espaço, mas se movem por ele, produzindo a sensação de que o espectador está mergulhando naquela história, uma paisagem sonora imersiva e em transformação.

Na versão sem Atmos, estéreo com fones de ouvido, a espacialidade da sequência é mais restrita. Embora os mesmos sons estejam presentes - vento, folhas, animais, explosões - sua distribuição no espaço acaba sendo comprimida no plano horizontal. O vento é percebido movendo-se lateralmente, de esquerda para direita, os sons de galhos e passos parecem surgir no mesmo plano auditivo, perdendo a especificidade posicional e diminuindo uma variedade de frequências perceptíveis na versão com objetos sonoros. A mixagem parece menos densa e com menor intensidade, a escuta torna-se frontalizada, com menor distinção entre planos sonoros. Assim, os sons da guerra que se aproximam mantêm sua função narrativa, mas perdem em escala e direcionalidade. O espectador continua acompanhando a narrativa, mas sem o mesmo grau de envolvimento sensorial.

Em termos perceptivos, nessa sequência de abertura a versão em Dolby Atmos parece ativar o espaço de escuta como um elemento narrativo em si, integrando o espectador ouvinte às paisagens sonoras que acompanharão a narrativa nos instantes de trégua e os excessos sonoros que virão com os momentos de batalha, enquanto a versão convencional estéreo deixa o som mais como uma apresentação do ambiente e da narrativa. Ambas cumprem sua função dramática e diegética, mas o realismo acústico da floresta e a ameaça crescente da guerra são experiências distintas em cada sistema.

5.2.2 Chegada dos soldados à trincheira

A sequência da chegada da tropa à trincheira sob a chuva se estende por alguns minutos e é marcada por uma construção sonora que reforça, desde o início, a desumanização progressiva dos personagens e a degradação dos corpos no campo de batalha. Inicialmente, os soldados recém-chegados são transportados por caminhões em direção à linha de frente, mas o trajeto é interrompido quando os veículos são requisitados para transportar feridos. A partir

desse ponto, os jovens precisam seguir a pé sob condições adversas. O grupo de quatro amigos, que até então compartilhava um imaginário idealizado da guerra, depara-se com uma realidade brutal e inesperada: hostilidade por parte de superiores, sons constantes de tiroteios ao fundo, ausência total de conforto, temperaturas baixas, escassez de alimentos e recursos limitados. Durante o deslocamento, Paul Bäumer é forçado a manter a máscara de gás por tempo prolongado, como punição por ter demorado a colocá-la durante um treinamento anterior, ao tentar ajudar um de seus companheiros.

A sequência inicia com o deslocamento dos soldados sob chuva intensa que domina a paisagem sonora. A frequência sonora predominante é composta por ruídos agudos e repetitivos das gotas batendo em diferentes superfícies, como tecido, metal e barro. Além disso, ao se aproximarem da trincheira, o grupo é confrontado com a lama e os sons do ambiente escavado, passos afundando no barro, respirações ofegantes, o ruído metálico das armas, tiros e granadas, mesmo que ainda fora de quadro. No sistema Dolby Atmos, é possível perceber gotas localizadas no eixo superior do campo auditivo, sugerindo a chuva caindo diretamente sobre o espectador, compartilhando a sensação desconfortável dos personagens.

O som aqui possui alta densidade, com múltiplos elementos sonoros ocorrendo simultaneamente, mas sem se sobreporem de maneira caótica. O movimento dos personagens é acompanhado por respingos, vozes abafadas, sons de pano encharcado que se deslocam pelo campo auditivo lateral e traseiro, além da chuva constante e dos sons variados das metralhadoras. O ruído da lama sendo pisada, misturado ao atrito dos uniformes molhados e o tilintar distante de artefatos metálicos, forma um timbre denso que remete à deterioração e à precariedade do cotidiano das trincheiras. O ambiente sonoro não é estático, respondendo à movimentação do grupo e à progressiva inserção dos jovens naquele espaço de guerra.

Por volta dos 19 minutos de filme, os jovens soldados chegam à trincheira de destino. O espaço está completamente alagado e o general repreende severamente um subordinado pela falta de contenção da inundação, ordenando que todos comecem a retirar a água utilizando seus próprios capacetes. Nesse momento, o general remove a máscara de gás de Paul Bäumer e acontece uma breve conversa entre o protagonista e Kattczinsky. Kat adverte sobre a severidade do comando e informa que um ataque de grandes proporções é esperado para aquela noite.

Enquanto a chuva permanece encharcando o espaço da trincheira, os soldados obedecem à ordem de esvaziá-la com seus capacetes. Um dos jovens, que anteriormente demonstrava maior entusiasmo com o alistamento, expressa arrependimento ao reconhecer

que suas expectativas foram completamente frustradas pela brutalidade da experiência. Ele afirma não sentir mais as próprias mãos, e Kat sugere que as aqueçam, colocando-as dentro de suas roupas íntimas, gesto que acaba simbolizando a necessidade de adaptação imediata a condições extremas. A partir disso, Kat passa a ocupar uma posição central na dinâmica do grupo, atuando como uma figura de orientação e proteção, ainda que o número de pessoas vá se reduzindo gradualmente à medida que os combates se intensificam.

Para fins de análise da construção sonora, a sequência em questão foi dividida em quatro momentos distintos: (1) da chegada dos soldados à trincheira até o primeiro diálogo com Kat; (2) do final desse diálogo até o início da retirada da água com os capacetes; (3) o diálogo entre os amigos recém-chegados enquanto realizam a drenagem; e (4) o trecho do momento em que aprendem a aquecer as mãos com seus próprios corpos.

Na primeira parte, o design sonoro enfatiza os elementos que estruturam o ambiente de guerra e autoridade, com sons de chuva constante, vozes de comando, explosões e disparos. Não há, nesse trecho, sons diegéticos associados aos passos sobre a água ou ao movimento de drenagem, mesmo que os personagens já estejam em ambiente alagado. A mixagem dá prioridade ao componente vocal, à inteligibilidade e ao ambiente externo, reforçando o contraste entre expectativa e realidade que se impõe aos soldados recém-chegados.

Após a interação inicial entre Paul Bäumer e Kat, na segunda parte, há uma mudança na hierarquia sonora. Os sons dos passos sobre a água tornam-se audíveis, assim como o ruído metálico dos capacetes sendo utilizados para a drenagem da trincheira. A paisagem sonora se intensifica, a chuva ganha protagonismo e os ruídos relacionados à água assumem uma posição central na construção da cena. Os disparos e explosões continuam presentes, mas perdem evidência visual, reforçando o caráter atmosférico da guerra como algo difuso e permanente.

O terceiro trecho é marcado pelo diálogo entre os amigos, em que os elementos vocais passam a ocupar o primeiro plano da mixagem, priorizando a inteligibilidade. A chuva e o som da água sendo retirada continuam presentes, mas agora se reconfiguram como pano de fundo sonoro. A cena reforça o impacto emocional causado pela constatação do abismo entre a expectativa idealizada da guerra e a brutalidade concreta do cotidiano nas trincheiras. O jovem mais entusiasmado com o alistamento pede desculpas aos companheiros, relatando que não consegue sequer sentir as mãos. Ao sugerir que coloquem as mãos na cueca para aquecê-las, Kat oferece não apenas uma solução pragmática, mas insere o grupo em um novo

momento, o de lutar pela sobrevivência não apenas contra o inimigo, mas também contra o frio, a fome e as condições adversas da trincheira.

Após aprenderem a esquentar as mãos, há uma breve alteração na ambiência sonora, com a inserção de um efeito extradiegético, que pontua simbolicamente a quebra de expectativa dos jovens soldados em relação à realidade encontrada. O som da chuva continua presente, constante, em segundo plano sonoro, o ruído da drenagem cessa e os sons de disparos e explosões desaparecem, mesmo que não haja qualquer indicação de trégua através da imagem. O silêncio diegético acaba por reforçar o esgotamento e a desilusão.

A diferença entre os modos de escuta, com e sem Dolby Atmos, torna-se evidente nessa sequência, principalmente por conta da chuva e sua característica verticalizada. Na reprodução estéreo, o som da chuva aparece como um elemento unificado e indistinto, funcional à atmosfera de frustração, mas sem a diferenciação espacial. Já na escuta com suporte a Atmos, a chuva se transforma em um objeto acústico cuja textura se distribui tridimensionalmente no campo auditivo, incluindo o eixo vertical, intensificando a sensação de esgotamento e quebra de expectativa. Os soldados que aguardavam com ansiedade pela oportunidade de lutar pelo país encontram um ambiente insalubre, alagado, envolto pela derrota, pelo frio e fome. Suas primeiras tarefas não são atirar contra o inimigo ou defender fisicamente seus ideais, mas sim usar seus equipamentos de proteção para retirar a água que inunda a trincheira. O ambiente sonoro deixa de ser um simples complemento narrativo e se converte em parte da experiência subjetiva do personagem. Nesse sentido, o ponto de escuta imersivo do espectador aproxima-se da escuta e dos sentimentos de Paul Bäumer, afundado em um espaço saturado por sons e emoções molhadas, frias e desoladoras.

Um momento técnico particularmente interessante ocorre por volta dos 19 minutos e 50 segundos quando um soldado, fora do campo visual, chama por Kaczinsky. Kat está de costas, e o espectador escuta como se alguém o chamasse por trás. A espacialização do diálogo, nesse caso, ainda que sutil, posiciona a voz no eixo posterior do campo auditivo, utilizando a mixagem para simular a direcionalidade da escuta do personagem. Trata-se de um breve, mas significativo exemplo da tentativa de aplicar espacialização não apenas a sons ambientes ou efeitos, mas também às vozes.

O modo de escuta estéreo também dá destaque a elementos específicos que não recebem tanta evidência na versão Atmos. Aos 19 minutos e 45 segundos, por exemplo, durante a conversa entre Kat e Paul, é possível escutar um disparo à direita do campo sonoro, com imagem correspondente, já que um soldado reage ao impacto do tiro. Esse *hard-effect*, embora também presente na mixagem Atmos, mas acaba adquirindo uma menor relevância,

pois é possível perceber com maior minúcia diversas outras sonoridades, principalmente o som da chuva de maneira verticalizada e, com isso, o som do tiro acaba se tornando mais um em meio a todo o contexto de guerra. A percepção da cena, nesse caso, acaba sendo alterada de acordo com o sistema de reprodução.

5.2.3 Tanque avança sobre a trincheira

Após avançarem em combate contra o exército francês, os soldados alemães conseguem tomar uma das trincheiras inimigas. Depois de intensos confrontos corpo a corpo e trocas de tiros, eles encontram um *bunker* que armazenava mantimentos da tropa adversária. O espaço contém uma variedade incomum de alimentos para o contexto vivido por Paul, Kat e os demais soldados, com queijos, diferentes tipos de pães, carnes enlatadas e produtos de charcutaria, itens muito distintos da escassez enfrentada até então. Diante dessa abundância momentânea, os combatentes passam a se alimentar com avidez e entusiasmo. No entanto, esse instante de aparente alívio é interrompido quando começam a perceber vibrações sutis no ambiente, que provocam o chacoalhar de talheres metálicos e o ranger das estruturas do abrigo, indicando a aproximação de uma nova ameaça. A ameaça ainda não é visível, mas o som antecipa sua presença. A construção sonora nessa passagem é marcada pela introdução de um grave contínuo e progressivamente intensificado, cuja frequência baixa não apenas se ouve, mas se sente fisicamente, sobretudo em sistemas com suporte a áudio imersivo.

A sequência inicia com sons abafados de respiração, movimentações discretas dentro do espaço e um silêncio tenso que domina o plano sonoro. Os soldados permanecem calados, alertas, enquanto talheres pousados sobre superfícies metálicas começam a vibrar levemente. Essa vibração opera como um índice físico da aproximação de algo massivo e inicialmente invisível, e mostra como a intensidade sonora crescente e o uso de baixas frequências pode ser empregado para construir dramaticamente a percepção do espaço e do perigo. O ponto de escuta imersivo do espectador consegue se localizar tão próximo dos personagens que é necessário pensar ao mesmo tempo que eles sobre a origem daquele ruído e qual ameaça ele representa naquele momento, o som precede o campo visual, criando um tipo de escuta encarnada, centrada no corpo do personagem.

Por meio de sistemas compatíveis com Dolby Atmos, esse som grave é posicionado de forma espacializada e difusa, ressoando não apenas à frente ou aos lados do espectador, mas também abaixo e em torno, simulando um deslocamento vibracional real de um tanque se aproximando via terrestre. O ouvinte sente o som como se ele emergisse do próprio ambiente,

produzindo uma experiência em que o som torna-se físico, ressoando no corpo e nos objetos e não apenas nos ouvidos, algo próximo ao descrito por Chion (2009) como o “ouvido de um lugar”, uma vez que as paredes ou tapumes do *bunker* ‘ouvem’ como os personagens e espectadores. A percepção dos soldados, atentos e silenciosos, é compartilhada com o espectador que, ao captar as vibrações, sente a aproximação do tanque como algo experienciado em primeira pessoa.

Já na reprodução estéreo essa camada subterrânea do som acaba ficando comprimida. A vibração dos talheres continua audível, mas perde parte de seu potencial sensorial e do efeito de imersão física. O som ainda cumpre sua função narrativa de indicar a aproximação da máquina de guerra, mas deixa de operar como elemento corpóreo e arquitetônico do espaço cênico. A percepção auditiva, nesse caso, permanece mais frontal e bidimensional, reduzindo a capacidade da escuta se alinhar subjetivamente à perspectiva dos soldados. O ponto de escuta, mesmo que mantido como próximo dos personagens, não detém a mesma força do utilizado com mixagem imersiva. O chamado “ultracampo” (Kerins, 2011, p. 92), isto é, o ambiente sonoro tridimensional do mundo diegético, não consegue ser alcançado no estéreo da mesma forma como é sentido com a inserção de objetos sonoros e canais de altura.

À medida que o veículo avança, os efeitos sonoros, como o atrito das lagartas, o motor pesado e o impacto sobre o solo encharcado, são posicionados no campo auditivo com deslocamentos perceptíveis ao redor e acima do ponto de escuta do espectador, simulando o avanço do tanque em direção aos soldados. Em determinados momentos, o som do veículo parece envolver o ouvinte por completo, o ponto de escuta imersivo consegue se aproximar da experiência auditiva e sensorial do personagem de Paul, criando a sensação de que o espectador ocupa o mesmo espaço físico emocional do personagem. Há instantes em que o ruído da engrenagem parece vir de trás ou das laterais superiores, sugerindo que o veículo passa rente ao corpo ou acima da posição do ouvinte.

Na mixagem estéreo, o som do tanque permanece audível e imponente, mas seu deslocamento no espaço é menos preciso e a sensação de profundidade é reduzida. Os sons se distribuem de forma mais uniforme nos canais esquerdo e direito, o que diminui a percepção da ameaça iminente. A escuta permanece eficaz em termos narrativos, mas perde em intensidade sensorial; elementos como os estalos metálicos, a compressão da lama sob o peso do tanque e os gritos dos soldados se comprimem em um campo sonoro mais homogêneo. É possível perceber também uma diferença na densidade sonora entre as duas versões de escuta. Enquanto a mixagem imersiva consegue manter os múltiplos sons da cena, como disparos, gritos e movimentos corporais, preservando suas distinções, a versão estéreo tende a

compactá-los em uma massa sonora, minimizando a apreensão das nuances dramáticas. O uso de frequências graves que, na mixagem Atmos, ressoam com mais impacto físico, acaba suavizado na reprodução estéreo, diminuindo a sensação de peso e ameaça imposta pela máquina. A sequência ainda se vale de pausas estratégicas na trilha extradiegética para acentuar o pânico e aproximar-se de uma experiência ‘realista’, como no momento em que Paul assiste ao tanque atropelar alguns soldados. Essas breves pausas funcionam como quebras de ritmo, destacando a iminência da morte ou do colapso psicológico dos personagens.

5.2.4 Embate entre Paul Bäumer e soldado francês

A sequência em questão ocorre quando Paul Bäumer, isolado no campo de batalha após um confronto com o exército francês, cai em uma cratera provocada por explosões e resolve manter-se refugiado nela até o cessar do combate. Nesse espaço restrito e coberto por lama, Paul é surpreendido por um soldado francês que também cai após outro disparo, iniciando uma luta corpo a corpo. Paul apunhala o adversário repetidamente e, ao perceber que ele está agonizando, entra em crise emocional, tenta estancar o sangramento, oferecer água e pedir desculpas, mas os gestos são em vão e logo o soldado acaba falecendo.

Toda a sequência é marcada por sons abafados do ambiente fora da cratera e com um destaque aos ruídos corporais e vocais dos dois soldados, reforçando a tensão dramática do momento, algo semelhante ao ouvido do personagem, descrito por Chion (2009), que é quando há a escuta de seus próprios sons, seja sua respiração ou batimentos cardíacos. Trata-se de uma cena-chave para o desenvolvimento psicológico do personagem, em que a violência da guerra se sobrepõe às convicções ideológicas, revelando o impacto humano e subjetivo do combate. Os sons externos, como explosões, tiros e ordens militares, permanecem audíveis, mas atenuados, como se filtrados pelas bordas da cratera. Essa espécie de uso de um campo acústico periférico reforça a ideia de que o mundo exterior está momentaneamente suspenso para Paul, cuja atenção se volta ao conflito íntimo que se desenrola no espaço físico e psicológico da cena. Há também um contraste entre sons corporais, como respiração e engasgos, e as ambiências, intensificando a percepção subjetiva do protagonista e humanizando sua angústia.

Na versão com Atmos, o ponto de escuta imersivo se alinha novamente à perspectiva de Paul e a espacialização do som tenta reproduzir o confinamento físico da cratera, com uma textura sonora abafada e os sons como a respiração ofegante, o impacto das mãos na lama, o

atrito do tecido encharcado e os gemidos do soldado francês ferido são distribuídos ao redor do ouvinte, gerando uma sensação de imersão claustrofóbica. Em contraponto, na versão estéreo, a cena preserva sua força dramática, mas sofre uma compressão. A percepção de profundidade entre os planos próximos, como os gemidos e a respiração, e os sons mais distantes, como tiros e explosões, tornam-se mais uniformes, atenuando o contraste e marcando a transição entre ação e questionamento sobre a guerra em si.

Além disso, há uma valorização das ambiências e efeitos sonoros por não haver trilha musical extradiegética. Dessa forma, o timbre de certos sons, como o barulho da faca rasgando a roupa, o som da lama deslocada pelas mãos ou o gorgolejo do sangue na garganta do soldado francês, acaba sendo mais detalhado na versão imersiva, contribuindo para a crueza emocional da cena. Sua duração prolongada, sem cortes sonoros abruptos, favorece também a criação de um ritmo lento e doloroso, sustentado pela continuidade dos sons de sofrimento e pela hesitação de Paul. Por fim, a forma como a densidade sonora evolui ao longo da cena, de uma sobreposição de sons intensos de combate para uma redução quase total da ambiência externa, é perceptível em ambas as versões, mas, na mixagem Atmos, essa transição se realiza de maneira mais fluida, oferecendo ao espectador uma escuta que se aproxima de uma vivência corporal do personagem.

5.2.5 Considerações

A análise de *Nada de Novo no Front* evidencia como a mixagem orientada a objetos, no caso Dolby Atmos, contribui para uma experiência de escuta mais próxima da vivência subjetiva dos personagens, promovendo uma maior imersão e densidade emocional. Ainda que o foco tenha se concentrado em cenas específicas, existem outras sequências do filme que revelam igualmente a técnica da construção sonora. O deslocamento do trem em direção à França e ao acordo de cessar-fogo, aos 57 minutos, mostra uma aproximação sonora que acompanha a imagem e parece querer ‘atropelar’ o espectador, algo semelhante ao sentimento descrito pelos primeiros filmes ainda silenciosos, como em *A Chegada do Trem na Estação* (1896), dos irmãos Lumière. Antes do ‘atropelamento’, a imagem corta para um plano localizado no para-choque da locomotiva e o som acompanha a trajetória do trem sobre os trilhos. Já o som da copa das árvores, a 1h56 do filme, na iminência da morte de Kat, reforça o contraste entre a aparente serenidade da natureza, o anúncio do cessar-fogo e os resquícios de violência e a brutalidade da guerra, que permanecerão por gerações.

O filme se beneficia da finalização em Atmos presente nas versões disponibilizadas em plataformas de *streaming*, o que permite uma espacialização precisa e dinâmica. Essa mixagem não apenas intensifica a sensação de imersão, como também contribui para a percepção subjetiva de Paul diante da guerra. A progressiva perda dos companheiros de Paul é acompanhada por uma modulação do ambiente sonoro, no qual ruídos metálicos, reverberações secas e texturas graves passam a dominar a cena, indicando a crescente insensibilidade e alienação do protagonista. Quando, ao final, Paul é morto nos minutos que antecedem o cessar-fogo, o som é tratado com um rigor que esvazia qualquer traço de grandiosidade. Não há música triunfal, apenas o ruído seco do golpe metálico, reforçando o absurdo do gesto final e a futilidade da guerra.

Essas escolhas sonoras demonstram que, com uma escuta atenta e interessada, é possível perceber as múltiplas camadas e direções do som propostas pela mixagem Atmos, tornando audível a complexidade emocional das cenas. O ponto de escuta imersivo permite que o audioespectador se aproxime da escuta dos personagens, principalmente do protagonista Paul Bäumer, não apenas ouvindo o que eles ouvem, como também compartilhando, ainda que de forma mediada, os estados de frustração, medo e cansaço. Ao favorecer a espacialização do som em sintonia com a perspectiva dramática da narrativa, a tecnologia amplia os recursos técnicos da mixagem, contribuindo para uma escuta sensível que potencializa o envolvimento afetivo com os temas explorados pelo filme.

Ainda que o uso de tecnologias, como o Dolby Atmos, proporcione uma paisagem sonora mais envolvente e uma maior integração com o ponto de escuta imersivo dos personagens, é importante destacar que a versão estéreo mantém a integridade narrativa do filme. A espacialidade sonora é menos expansiva e, conseqüentemente, atinge com menos impacto os picos emocionais da trama. Ainda assim, o conteúdo permanece compreensível, com clareza nos diálogos e equilíbrio entre os elementos sonoros. O Atmos amplia a sensação de presença e realça camadas sonoras muitas vezes imperceptíveis na mixagem multicanal tradicional, operando como um elemento adicional na *mise en scène*, por isso, mesmo que o estéreo ainda mantenha qualidade técnica e estética, é possível que haja alguma perda em relação aos sentidos e emoções que a história queira transmitir.

5.3 A Sociedade da Neve (2023)

“Neste filme, o silêncio é tudo”, é o que diz Matías Recalt, intérprete de Roberto Canessa, no documentário/*making of A Sociedade da Neve - Quem fomos na montanha*

(2024). O filme *A Sociedade da Neve*, dirigido por Juan Antonio Bayona, é mais uma produção original Netflix, baseada na história real do acidente aéreo ocorrido em 13 de outubro de 1972, quando um avião da Força Aérea Uruguaia (FAU) caiu em uma região remota da Cordilheira dos Andes, a caminho do Chile, transportando 45 pessoas, entre elas, jogadores de rúgbi do time *Old Christians* e seus acompanhantes. A partir da queda, inicia-se uma luta pela sobrevivência em condições extremas, que duraria 72 dias, marcada por fome, frio, acidentes naturais e questões morais. Baseado no livro homônimo (*La Sociedad de la Nieve*, 2008), de Pablo Vierci, o filme se organiza em torno da narração póstuma de Numa Turcatti, um dos sobreviventes que não resistiu até o resgate. A partir de sua perspectiva, a obra resgata os dias de sobrevivência nas condições adversas da cordilheira, destacando os dilemas éticos, a solidariedade entre os integrantes e os limites da condição humana.

Inicialmente, a esperança de resgate é sustentada pelas notícias transmitidas por um rádio portátil que informa sobre as buscas empreendidas pelas autoridades. No entanto, à medida que os dias se acumulam sem sinais de aproximação, o grupo passa a confrontar a gravidade real da situação. A convivência na situação crítica, com temperaturas congelantes, ferimentos, desidratação, fome e o trauma psicológico, acabam por impor reorganizações internas, tanto físicas quanto afetivas. Os laços de amizade e companheirismo se fortalecem para a manutenção da sanidade e do senso de humanidade em meio ao caos. A morte de colegas, em decorrência dos ferimentos ou de avalanches, reforça o tom trágico da experiência e desafia os limites morais dos sobreviventes. Em um dos momentos mais delicados, após longas reflexões em grupo, eles tomam a decisão de se alimentar dos corpos dos mortos para preservar suas próprias vidas, um ponto de inflexão ético retratado de maneira sensível, evitando o sensacionalismo comum a outras produções¹⁰⁸ que abordaram a mesma história e propondo uma escuta respeitosa acerca das escolhas necessárias diante das condições inóspitas.

¹⁰⁸ A história do acidente nos Andes já foi retratada diversas vezes, através de documentários, filmes ficcionais e episódios de séries investigativas. O filme de ficção *Supervivientes de los Andes*, dirigido por René Cardona, é uma produção mexicana que dramatiza os eventos da queda e a luta pela sobrevivência, embora com abordagem mais centrada no impacto sensacionalista do acontecimento. Em 1993, houve o lançamento do filme estadunidense *Alive* (no Brasil, *Vivos* ou *Sobreviventes*), dirigido por Frank Marshall. Baseado no livro homônimo de Piers Paul Read, a obra oferece uma representação dramática e emocional dos dilemas enfrentados pelos sobreviventes, especialmente no que diz respeito ao canibalismo e à resistência física e psicológica. No mesmo ano, foi lançado o documentário televisivo *Alive: 20 Years Later*, dirigido por Jill Fullerton-Smith, que traz os próprios sobreviventes refletindo sobre a experiência, duas décadas após os acontecimentos. Em 2007, o documentário *Stranded: I've Come from a Plane That Crashed on the Mountains*, dirigido por Gonzalo Arijón e coproduzido entre França e Uruguai, é composto por entrevistas com os sobreviventes e reconstituições visuais, adotando uma abordagem humanizada e ética que procura compreender o impacto do trauma coletivo. Além dessas produções principais, o acidente também foi explorado em programas de televisão, como *Seconds From Disaster*, *Mayday* e *I Shouldn't Be Alive*, oferecendo leituras mais técnicas e investigativas do acidente.

Com a interrupção definitiva das buscas oficiais anunciada pelo rádio, o grupo compreende que a única chance de sobrevivência depende da própria ação. Mesmo debilitados, Fernando Parrado, Roberto Canessa e Antonio Vizintín se voluntariam para realizar a travessia dos Andes em busca de ajuda. A decisão é precedida por uma organização coletiva do grupo, em que os recursos são redistribuídos, rotas são debatidas e os jovens recebem instruções improvisadas de orientação e proteção contra o frio. A jornada arriscada dura mais de dez dias e culmina com o encontro de um camponês chileno que, ao perceber a gravidade da situação, comunica às autoridades e é iniciado o resgate dos 14 sobreviventes que permaneceram na montanha.

O título faz referência à ideia de que, naquele ambiente extremo e desumanizado, formou-se uma sociedade fundada na empatia, na interdependência e na tentativa de preservar a dignidade diante do impensável. Essa perspectiva é reforçada pela construção sonora e pela narração de Numa, que atravessa toda a obra como um convite à escuta dos acontecimentos e à memória dos que se foram, conferindo ao filme um tom reflexivo e memorialista, enfatizando a coletividade dos dias vividos após o acidente.

Com uma estética que privilegia o realismo, o filme lança mão de uma abordagem sonora que contribui decisivamente para a imersão do espectador. A mixagem realizada em Dolby Atmos constrói uma paisagem sonora que reproduz os elementos naturais, como ventos cortantes, avalanches, neve sob os pés e silêncio rarefeito, criando um espaço auditivo que simula tanto a vastidão geográfica quanto o confinamento físico e psicológico dos personagens. A espacialização dos sons naturais provoca uma sensação de presença e vulnerabilidade, com destaque para os momentos silenciosos, interrompidos apenas por ruídos sutis, marcando o esgotamento e a desesperança.

Em cenas como a do impacto do avião contra a montanha e as sucessivas avalanches, a mixagem posiciona os sons ao redor e acima do espectador, ampliando a percepção de movimento e ameaça. O uso da camada superior contribui para acentuar o colapso estrutural da aeronave e o deslocamento abrupto da neve, criando uma experiência sensorial para além das bordas imagéticas. Os sons vocálicos mantêm a inteligibilidade esperada, mesmo quando os personagens estão soterrados ou em meio a nevascas.

A trilha musical, composta por Michael Giacchino, adota uma presença discreta e emocional, pontuando momentos-chave com sutileza. A composição equilibra momentos de silêncio com estruturas atonais e épicas, criando tensão e densidade emocional. A construção sonora acompanha o arco emocional dos personagens, iniciando com composições mais contidas, marcadas por teclado melancólico e dissonante, que refletem a tensão e a fragilidade

das primeiras etapas da viagem. Em situações de maior intensidade dramática, como o impacto da avalanche e as travessias na neve, a música assume uma dimensão mais percussiva, incorporando tambores de candombe¹⁰⁹ e coros em língua Mapuche. Essa escolha estilística liga os personagens à ancestralidade e à força simbólica da Cordilheira, conectando-os à paisagem. No desfecho do filme, a trilha musical migra para harmonias em tonalidade mais aberta, utilizando guitarras e vozes femininas a fim de sublinhar a recuperação da esperança e a reconstrução dos laços comunitários. O design de som, assinado por Jorge Adrados, Oriol Tarragó e Marc Orts, ajuda na construção da atmosfera de isolamento e sobrevivência. A equipe, vencedora do Prêmio Goya, dedicou atenção especial aos ruídos, como o som da fuselagem, a respiração e os passos na neve, gerando uma presença sonora que acentua a vastidão da Cordilheira, mas evitando excessos e privilegiando sons contidos na paisagem sonora construída do ambiente.

A sequência do acidente foi construída com base em sons captados de forma realista, incluindo gravações físicas dentro de protótipos de aviões e materiais metálicos em contato com a neve, conferindo verossimilhança e uma textura acústica crua à sequência. As cenas acabam sendo o momento mais intenso, considerando o design de som no filme, uma vez que articula sons realistas de motores em falha, vibrações internas da aeronave, colisões metálicas e estalos de gelo se partindo, elementos que se alternam com silêncios abruptos e estrategicamente inseridos, intensificando a percepção de choque. Além disso, a combinação de graves profundos com frequências agudas intermitentes acentua a sensação de tensão e vertigem ao longo da sequência. No interior da fuselagem, há a reconstrução desses sons abafados e espacializados, contribuindo para a sensação de desorientação dos passageiros e ajudando a ‘inserir’ o espectador como mais um sobrevivente, diretamente no centro do acontecimento.

Outro momento de destaque sonoro é o da avalanche que atinge o grupo durante a noite. O desenho da sequência se inicia com um som quase imperceptível ao longe, que gradualmente cresce em intensidade e complexidade até se transformar em um ruído forte que preenche todo o espectro sonoro do filme. A sobreposição de sons abafados da neve em movimento, gemidos e gritos cria uma ambiência sonora fechada e opressiva que comunica a sensação de falta de ar e o desespero dos sobreviventes soterrados.

Já a travessia final realizada pelos passageiros é construída sonoramente com uma abordagem mais minimalista e introspectiva. Há um silêncio vocálico predominante e o

¹⁰⁹ O candombe é uma manifestação cultural afro-uruguaia, com raízes na música, dança e tradições de povos africanos escravizados levados para o Uruguai.

protagonismo de ambiências, como o vento, o ranger da neve sob os pés e a respiração, que ressaltam o isolamento, a desesperança e a exaustão física dos personagens. Nesses trechos, o ponto de escuta imersivo aproxima o espectador da vivência interna dos sobreviventes. Além disso, há o uso da trilha musical extradiegética, marcada por tons graves e instrumentos de sonoridade étnica. Essas sonoridades, verificaremos adiante, parecem ser intensificadas pelas possibilidades oferecidas na mixagem orientada em objetos sonoros, como se o espectador estivesse compartilhando o mesmo espaço físico dos personagens.

5.3.1 O acidente nos Andes

A sequência do acidente aéreo ocorre por volta de 13 minutos do filme e inaugura a tragédia central da narrativa. A bordo do avião da Força Aérea Uruguaia, os jogadores e acompanhantes do time de rúgbi se encontram inicialmente em um ambiente diegético marcado por conversas, expectativas e pequenos ruídos típicos da cabine de passageiros, como vozes dispersas, zumbido contínuo dos motores, vibrações de baixa frequência do voo. Essa primeira camada estabelece uma textura sonora fluida e pouco densa, marcada por baixa amplitude e por timbres metálicos e abafados, típicos de uma cabine pressurizada.

À medida que a aeronave adentra uma zona de turbulência nos Andes, o clima sonoro se transforma, com sons diegéticos de origem mecânica, como vibrações estruturais, variações súbitas de rotação do motor, turbulência e ruídos atmosféricos atingindo o primeiro plano sonoro e ampliando a densidade do ambiente. Os sons do motor aumentam em frequência e intensidade, e a textura sonora torna-se mais espessa, incorporando estalos metálicos, oscilações estruturais e variações de pressão interna. A espacialização com a mixagem orientada a objetos, no caso, em Dolby Atmos, destaca a tridimensionalidade desses sons, com o ranger da estrutura surgindo pelas laterais, a trepidação dos assentos sendo reproduzida com graves em localização inferior, e os alertas e gritos percorrendo o campo sonoro. A paisagem sonora deixa de ser contínua e se fratura em blocos de eventos acústicos de curta duração e grande intensidade organizados em camadas simultâneas, compondo um ritmo espesso e fragmentado que ocupa todo o espectro auditivo com sons de ruptura, o arrastar das cadeiras, ossos se quebrando, objetos perfurando passageiros, gelo sendo cortado e bagagens colidindo com as paredes internas da aeronave e com as pessoas envolvidas no acidente, culminando na colisão com a montanha. O impacto rompe a fuselagem, que se parte enquanto corpos são lançados, gritos se sobrepõem e o vento se infiltra na paisagem sonora. A sequência imediata da colisão termina com a parte dianteira do avião deslizando pela neve até

parar, deixando os sobreviventes em um silêncio paralisante após o desastre. Com a parada da aeronave, há um *fade out* na imagem e o som cessa por alguns segundos. Esse breve vazio aumenta o contraste com a sobrecarga sonora anterior e reposiciona o ponto de escuta imersivo do espectador que se aproxima da própria escuta e das sensações dos personagens, não apenas de forma localizada, mas também com o estado subjetivo dos sobreviventes.

A escuta com a tecnologia imersiva possibilita uma percepção amplificada desse colapso progressivo. A espacialização não apenas amplia o campo sonoro, mas orienta o ouvinte a partir de diferentes pontos de escuta, ora internos, como se o espectador estivesse no centro da fuselagem, ora externos, como nos momentos em que partes do avião se desprendem. As frequências graves auxiliam na construção da sensação de instabilidade, com ruídos vindos do piso da cabine e das laterais que acompanham o ranger da estrutura e o impacto contra o gelo, gerando uma vibração física e auditiva. A camada superior do áudio, na qual geralmente são dispostos sons mais agudos, como o som da ventania, o deslocamento de bagagens e objetos metálicos, contribui para uma verticalização do campo sonoro. Com isso, é possível identificar uma diferenciação de timbres entre os sons massivos e pesados do impacto e os sons cortantes e leves que coexistem no mesmo plano, preservando a inteligibilidade. O ruído dos escombros e da neve deslizando acompanha os movimentos laterais e frontais da imagem, com o *delay panning*, preconizado por Brett (2018), e amplia a percepção do deslocamento da fuselagem após a ruptura da aeronave. Esse tipo de espacialização insere o espectador no eixo da narrativa sensorial, em que a duração dos eventos sonoros acaba sendo tratada de forma a estender a dramaticidade cênica, incluindo momentos de suspensão sonora entre os choques, que funcionam como pausas dramáticas e que, em *Atmos*, adquirem um silêncio profundo, quase tático.

Na versão estéreo, esses elementos permanecem presentes, mas com redução da densidade e da precisão posicional. A ausência de canais superiores torna a distribuição sonora mais compacta, reduzindo a percepção da dispersão dos eventos. Os sons tendem a se sobrepor em um campo frontal centralizado e homogêneo, principalmente nas passagens de maior complexidade sonora. A amplitude continua impactante, e os sons mantêm sua intensidade emocional, a textura se torna mais uniforme e a percepção de profundidade mais limitada, o que acaba dificultando a separação entre alguns elementos, como os sons metálicos da fuselagem e os estalos agudos das quebras estruturais. Ainda assim, a versão estéreo mantém a coerência narrativa e emocional da sequência. O ritmo sonoro, construído pela alternância entre ruídos intensos, estalos abruptos e silêncios curtos, segue eficaz, mas

perde parte de sua expressividade. A presença de sons diegéticos fundamentais, como gritos, estalos e impactos, conservam o vínculo afetivo com a narrativa.

Nos momentos imediatamente posteriores ao acidente, o estado de choque, desorientação e vulnerabilidade dos sobreviventes é evidenciado pelo uso de um silêncio relativo, uma vez que os sons diegéticos são reduzidos a sonoridades mínimas de respiração, gemidos e estalos intermitentes da fuselagem danificada, criando uma ambiência rarefeita, de baixa densidade, que contrasta com a intensidade sonora da sequência anterior. Esse esvaziamento auditivo contribui para evocar uma suspensão temporal e reforça a percepção do trauma imediato. À medida que os personagens recobram a consciência, a espacialização dos sons tenta recriar o ambiente fechado e fragmentado da fuselagem destruída. Os gemidos são localizados em pontos distintos do campo sonoro, permitindo ao espectador reconhecer a distribuição espacial dos sobreviventes mesmo antes de vê-los. O som das placas metálicas rangendo, da neve sendo movida ou pressionada contra a estrutura, e das vozes sobrepostas em diferentes estados de desespero e confusão, constrói uma textura sonora marcada por timbres secos e metálicos, com frequência média e baixa, sugerindo o colapso da estrutura física e emocional. Na mixagem Atmos, os sons são organizados com maior precisão espacial, favorecendo uma escuta tridimensional que posiciona o espectador no interior da fuselagem. Há uma alternância entre momentos de maior amplitude, quando as reações são mais intensas (gritos, choros, pedidos de ajuda), e momentos de intensidade sonora baixa, quando predominam respirações e ruídos discretos. Isso contribui para a variação de ritmo e para a construção de uma progressão dramática auditiva. Já na finalização estéreo, esses elementos continuam presentes, mas os sons parecem emanar de um eixo frontal mais homogêneo, o que reduz a profundidade da ambiência e a nitidez das camadas sonoras simultâneas. Ainda assim, a trilha mantém os aspectos narrativos essenciais, como a oscilação entre silêncio e ruído, a emergência dos sons corporais, como tosses e respirações ofegantes, e a inserção sutil de elementos extradiegéticos, como sons ambiente filtrados, que ajudam a marcar a passagem do choque à compreensão da catástrofe.

A análise desta sequência mostra que a proximidade com os personagens não se dá apenas pela imagem ou narrativa, mas também pela escuta que simula a forma como os personagens ouviram e sentiram o impacto, redefinindo, de certa forma, a experiência perceptiva do espectador. O uso de recursos sonoros na ilustração do impacto evidencia como a mixagem constrói uma experiência auditiva que reforça os sentidos de urgência, desorientação e violência física do evento. Através da manipulação de frequência, timbre, textura sonora e dos pontos de escuta, a espacialização sonora permite não apenas

acompanhar o que os sobreviventes escutam, mas simular seus estados emocionais e perceptivos. Ambos os formatos entregam emoção e significado narrativo, mas a versão estéreo acaba oferecendo uma escuta mais linear, enquanto o Atmos consegue amplificar o efeito subjetivo da cena, funcionando como uma extensão sensorial do corpo dos personagens e aprofundando a imersão do público.

5.3.2 Avalanche

No documentário *A Sociedade da Neve - Quem fomos na montanha*, o diretor J. A. Bayona comenta que a coisa mais impressionante do Vale das Lágrimas, local onde o avião da FAU caiu, é o som, “é um lugar inorgânico, onde não se ouve nada e depois tem as rajadas de vento a uma altura de 4.000 metros; elas são muito fortes”. É nessa variação entre o silêncio diegético e a violência física e sonora da montanha que se dá a sequência da avalanche.

Antes do deslocamento da neve, os sobreviventes estão acordados dentro da fuselagem do avião parcialmente soterrado pelo gelo acumulado após dias de acidente. Em um momento de descontração, os jovens se envolvem em um jogo de declamações, versos e rimas de humor, criando um breve espaço de leveza dentro da experiência extrema que vivem. A trilha de áudio é inteiramente diegética nesse momento, com a inteligibilidade destacando os sons vocálicos e as risadas discretas, compondo uma ambiência intimista, marcada por densidade sonora leve e predomínio de frequências médias, com ritmo calmo e fluidez suave, revelando o cuidado em retratar esse raro momento de comunhão com naturalismo e proximidade. Após alguns versos, todos os sobreviventes riem e aplaudem a tentativa fracassada de rima de Numa Turcatti. Durante um aplauso mais entusiasmado, um ruído grave e progressivo, inicialmente indistinto, de textura densa e timbre opaco, começa a emergir no plano de fundo, vai ganhando corpo e amplitude, deslocando-se de forma crescente no espectro sonoro, sendo escutado pelos personagens, assim como uma sensação imagética e sonora de trepidação da fuselagem modificando toda a ambiência da cena.

Em Dolby Atmos, esse som é espacializado com verticalização e profundidade, percorrendo os canais superiores e traseiros de maneira a sugerir o deslocamento maciço da neve vindo do topo da montanha. O ponto de escuta imersivo do espectador se localiza dentro da fuselagem, em perspectiva próxima à dos personagens que também demoram alguns segundos para identificar a ameaça. A densidade sonora cresce abruptamente quando a avalanche atinge a fuselagem, com uma sobreposição de sons metálicos contra o gelo, a ‘parede’ de malas que protege contra o frio sendo jogada contra os personagens, gritos

abafados e principalmente o impacto da neve soterrando os sobreviventes e quase a totalidade dos destroços. Mesmo que esta parte tenha uma duração menor que dez segundos, é possível observar a mudança de ritmo e uma ‘inundação’ do plano diegético com elementos de difícil distinção individual, mas que evidenciam a desorientação e o colapso emocional do momento, reproduzindo a experiência caótica vivida pelos personagens. Na versão estéreo, os elementos fundamentais da cena são preservados, mas o campo sonoro é menos articulado espacialmente. O deslocamento da avalanche é percebido mais como uma parede sonora crescente do que como um movimento espacial complexo. O colapso é audível, mas reduzido em sensação de envolvimento físico do espectador. A textura torna-se compactada e a distribuição dos gritos ou o deslocamento de estruturas acabam perdendo em clareza.

Após a invasão da neve, há um *fade out* total com duração de dez segundos, como se o espectador tivesse sido também soterrado e tentasse lutar, nesse período, para emergir da massa de neve que o cerca. Esse vazio sonoro e imagético reforça a interrupção brusca da vida e amplia o efeito emocional. Nos segundos finais da tela preta, é possível escutar sons abafados que remetem a uma escavação interna, pequenos ruídos diegéticos, como o rangido da fuselagem pressionada pela neve, partículas desabando e, principalmente, as respirações, em diferentes intensidades, timbres e ritmos, que assumem o primeiro plano sonoro. Algumas são rápidas e ofegantes, indicando pânico ou esforço físico; outras mais lentas, quase inaudíveis, refletindo exaustão ou estado de choque. O timbre das respirações, por vezes seco e curto, ou abafado e prolongado, compõe a textura sonora, marcando a transição para a percepção do desastre.

Em seguida, há a confirmação imagética com um dos personagens conseguindo sair da neve, emitindo um grito ainda abafado, com uma vocalização que soa distante e filtrada pelo gelo ao redor. Em Dolby Atmos, essa vocalização é reproduzida como se o som viesse de baixo ou por trás, proporcionando uma simulação do ponto de escuta imersivo dos demais sobreviventes que ainda estão soterrados e suas tentativas desesperadas de emergir da massa de neve deslocada. Pequenos ruídos, como o atrito de mãos removendo neve, ou o som de utensílios se movendo acidentalmente, ganham relevância ao preencher o espaço sonoro rarefeito. Na versão estéreo, esses elementos ainda estão presentes, mas soam mais centralizados. Há manutenção da carga emocional e o impacto da cena continuam eficazes, mas a percepção do espaço e da profundidade sonora é menos pronunciada. A ambiência criada pelo Atmos, que simula o peso da neve acima da fuselagem e a distância entre os corpos, é reduzida no estéreo a uma camada mais plana de som, ainda que tecnicamente bem executada.

A sequência se desenvolve mantendo a perspectiva interna dos sobreviventes soterrados. Após o colapso da massa de neve sobre a fuselagem, a câmera permanece dentro do espaço fechado, escuro e confinado, acompanhando os primeiros movimentos das pessoas que gradualmente recobram a consciência e tentam entender a dimensão do ocorrido. A iluminação é mínima, com foco em brechas de luz exterior, o que reforça visualmente a sensação de clausura, amplificada pela camada sonora. Em relação à textura, a cena é composta por elementos de baixa densidade e timbres opacos que remetem à compressão espacial provocada pelo soterramento. A amplitude dos sons é controlada, priorizando frequências médias e graves, o que contribui para construir uma sensação de confinamento. A interação entre os personagens ocorre de maneira fragmentada, com gritos por nomes, perguntas sobre quem está vivo, ordens improvisadas. O ritmo da cena é conduzido por essas tentativas de comunicação e movimentação em um ambiente reduzido e instável. À medida que os sobreviventes retomam a consciência e compreendem a gravidade da situação, a cena ganha uma maior complexidade sonora. Começam a escavar com as mãos, e os ruídos da movimentação da neve se intensificam gradualmente, gerando um crescendo de densidade acústica. O atrito entre o tecido das roupas e os movimentos dos itens soltos no interior do avião, como malas, além do metal e da neve, tornam-se elementos audíveis do design sonoro com timbres ásperos e irregulares. As vozes, ainda abafadas, tornam-se mais presentes conforme os personagens rompem as barreiras de neve e estabelecem contato uns com os outros.

A progressão da sequência enfatiza também o esforço coletivo para reorganizar o interior do avião e salvar quem ainda possa estar vivo, muitos sobreviventes são retirados dos escombros e começam a ajudar na retirada de outros, quando outro som grave e progressivo começa a ser escutado, revelando uma segunda onda da avalanche. Novamente há um *fade out* na imagem e o silêncio toma conta da cena, dessa vez por menos tempo. Com o retorno do campo visual, é possível entender que houve a formação de um ‘bolsão’ de ar pela neve e os sobreviventes ficaram juntos. As respirações ficam ofegantes e o ar rarefeito e, com o auxílio de uma barra de ferro, eles começam a escavar uma abertura para a saída. A iluminação fica por conta de um isqueiro aceso que também ajuda na confirmação visual de que o oxigênio voltou a circular graças à abertura dessa pequena escotilha. A partir dessa estabilização, os sobreviventes começam a contabilizar quem não resistiu às avalanches e quem ainda se mantém vivo. Na versão Atmos, há uma diferenciação espacial entre as vozes dos personagens em profundidade e altura, permitindo uma percepção, ainda que simbólica, da posição de cada um dentro do espaço cênico. Essa precisão na escuta aproxima-se da escuta

interna dos personagens, intensificando a sensação de confusão e urgência. Já na versão estéreo, o som tende a se agrupar, diminuindo a distinção espacial entre os elementos.

Não há trilha musical até o final da sequência que acontece com a volta da narração de Numa. O foco permanece na construção de uma *mise en scène* sonora que se vale dos silêncios intermitentes, da densidade baixa e dos elementos diegéticos para transmitir o impacto físico e emocional da situação. A progressiva introdução de sons mais definidos, como o rompimento da neve, o ar entrando e gritos de socorro, marcam a passagem do colapso para a sobrevivência e funcionam também como elemento de transição para a próxima etapa da narrativa. Esta cena evidencia o uso da tecnologia sonora imersiva na criação de camadas de significados por meio da espacialização e também sua utilização como um dispositivo narrativo capaz de mediar a experiência sensorial e emocional do espectador. Mais do que ouvir como os personagens, o ponto de escuta imersivo do espectador está tão próximo que ele acaba por se tornar um sobrevivente daquela história, conseguindo vivenciar a angústia do soterramento congelado através dos sons abafados, dos gritos dos colegas e da tentativa de salvamento com as próprias mãos de quem não consegue se movimentar.

Assim, a sequência da avalanche representa um bom exemplo de como as tecnologias de mixagem orientada a objetos podem ser utilizadas de maneira criativa, contribuindo para um envolvimento sonoro que ultrapassa a mera ilustração da imagem. A tridimensionalidade da espacialização, a manipulação minuciosa da densidade, do timbre e o uso expressivo das dinâmicas de silêncio e ruído transformam a escuta, reforçando a experiência emocional e física do espectador.

5.3.3 Travessia final

A caminhada que atravessa a Cordilheira dos Andes rumo ao Chile, realizada inicialmente por Fernando Parrado, Roberto Canessa e Antonio Vizintín, traça o arco final do filme e um ponto de virada, no qual os próprios sobreviventes precisam buscar meios de serem encontrados, uma vez que as buscas oficiais haviam sido encerradas. A sequência até o encontro com o camponês chileno dura cerca de 10 minutos, alternando os momentos de caminhada, as noites frias e a espera angustiante dos sobreviventes nos destroços do avião. Na despedida entre os amigos, há um crescendo de música extradiegética, tambores e cânticos que remetem novamente ao ritmo do candombe, composto por notas graves, longas e espaçadas que pontuam o sofrimento físico e moral da travessia. Os instrumentos utilizados têm timbres orgânicos que remetem à ancestralidade da experiência de sobrevivência que ali

se desenha, conferindo um caráter quase ritualístico à caminhada. A música permanece nas cenas seguintes, alternando momentos em primeiro e segundo plano, já que as vozes e os sons ambientes de escavação, caminhada e o ruído do vento constante também estão presentes. Nos planos gerais que contrastam o tamanho da montanha e vastas paisagens brancas em relação aos corpos exaustos dos sobreviventes, a música se mantém soberana e única, enquanto, em planos mais fechados e em detalhes, há a presença das demais sonoridades diegéticas.

A sensação de esperança e expectativa de chegada ao cume de uma montanha acaba sendo decepcionante, pois, quando eles o alcançam, acabam vendo uma imensidão de outras montanhas pela frente, revelando uma breve frustração sobre a caminhada. Nesse momento, a música extradiegética cessa e ficam apenas as sonoridades da montanha, o vento torna-se gradualmente mais presente, atingindo picos de frequência sonora mais aguda que acentuam a aridez do local e a dificuldade da jornada. Diferentemente das sequências anteriores, marcadas por uma densidade sonora elevada, como a do acidente ou a da avalanche, aqui o design de som aposta na rarefação como estratégia narrativa, em que o silêncio vocálico e a economia de elementos constroem, paradoxalmente, uma experiência imersiva. A diegese é preenchida por fragmentos sutis, como o vento que corta o espaço com frequência aguda, o ranger da neve sob as botas e o som intermitente das respirações profundas e exaustas, construindo uma textura sonora frágil, mas emocionalmente carregada. A cena se sustenta numa escuta detalhista, em que a matéria sonora parece refletir o esgotamento dos corpos e o peso do tempo que se arrasta na paisagem branca. A textura sonora, apesar de rarefeita, é marcada por uma alta fidelidade nos detalhes, em especial, na versão com tecnologia imersiva, como, por exemplo, quando o som do vento se desloca lateralmente ou quando as passadas dos personagens soam de acordo com o movimento da câmera.

Após alguns dias de caminhada, Vizintín retorna ao local do acidente, deixando Parrado e Canessa seguindo sozinhos, com a intenção de preservar suprimentos por mais dias. Os dois sobreviventes seguem sua caminhada e o som mantém a mesma perspectiva, com música extradiegética em planos abertos e a inserção de ruídos corporais e deslocamento na neve em enquadramentos mais fechados. A progressão da sequência inclui momentos em que os personagens param para descansar ou trocam palavras curtas. Nesses instantes, o som das respirações e o cansaço vocal adquirem protagonismo. O timbre da voz revela o desgaste físico, com vozes mais roucas, pausadas e perda de projeção. Já a duração e o ritmo ajudam a estabelecer o tempo psicológico da travessia, marcado pela resistência e pela esperança silenciosa. Enquanto avançam entre as montanhas, a paisagem vai se modificando, tornando-se menos branca e ganhando o tom amarronzado das montanhas. Ao final da

travessia, quando os dois sobreviventes avistam um rio já sem gelo e a presença de animais, os primeiros sinais de civilização, a trilha musical ganha maior presença e amplitude, com harmonia crescente que sugere uma resolução narrativa. Essa mudança também é perceptível na mixagem, com maior densidade e preenchimento do campo sonoro. Trata-se de uma estratégia eficaz de transição que simboliza não apenas a superação física do desafio, mas também a reconexão com o mundo exterior. No Atmos, essa transição acaba sendo mais impactante, pois há uma expansão no campo auditivo, com maior percepção de altura e profundidade, em especial, nos momentos em que o vento se mistura com as camadas da trilha musical. A alimentação escassa, pouco nutritiva e o esforço físico geram náuseas e vômitos, mas a caminhada que seguiu por dez dias e 38 km culmina com o encontro com o camponês chileno, após 71 dias do acidente.

A mixagem realizada em Atmos também posiciona os sons nas cenas de forma a reproduzir a percepção espacial dos personagens. O vento, por exemplo, não é apenas uma ambiência constante, mas um corpo sonoro dinâmico que se move lateralmente, horizontalmente e, por vezes, parece vir de cima, aumentando a sensação de exposição dos personagens ao ambiente hostil, acentuando a vulnerabilidade dos corpos diante da natureza e contribuindo para a criação de uma sensação de desequilíbrio físico e emocional que extrapola o imagético. Já a trilha musical extradiegética, quase sempre sustentada por instrumentos de timbre grave e longas reverberações, surge como contraponto emocional, oferecendo uma camada de introspecção e esperança. Em determinados momentos, seu aparecimento parece acompanhar uma mudança de ritmo no andar, como se a música modulasse a própria cadência da caminhada. Em momentos de pausa, quando os personagens se sentam para descansar ou se protegem entre as pedras, a trilha musical tende a dividir espaço com os ruídos, com sons como o ranger de dentes, a respiração entrecortada e a neve se desfazendo sob o peso do corpo, evidenciando a intensidade do esforço. A densidade é mínima, mas construída com timbres abafados e pouca reverberação, indicando a perda progressiva de forças e a proximidade do colapso físico.

Na versão estéreo, embora os mesmos elementos estejam presentes, a profundidade espacial e a separação entre as camadas sonoras são mais limitadas. A direção do som, especialmente do vento e das respirações, torna-se menos precisa, concentrando-se nos canais laterais sem a possibilidade de verticalização. Ainda assim, o estéreo mantém a integridade narrativa e preserva a proposta estética da sequência, tornando a escuta eficaz para o entendimento das cenas. A principal diferença está na maneira como o Atmos permite uma maior flutuação entre os planos sonoros, em especial, no tratamento das ambiências e dos

sons de baixa amplitude que emergem com mais nitidez. Ainda assim, as sonoridades permanecem inteligíveis e a trilha musical continua eficiente na construção do campo emocional. O que se perde, principalmente, é o grau de imersão e a equivalência mais direta entre o ponto de escuta imersivo do espectador com o dos personagens.

Essa sequência marca um deslocamento em relação à construção sonora do restante do filme, pois, em vez de intensificar a brutalidade do ambiente por meio da densidade e da amplitude, como nos momentos de confronto ou catástrofe, opta-se aqui por esvaziar o campo acústico, conduzindo o espectador por uma escuta quase meditativa. O som não dramatiza o esforço dos personagens, mas o acompanha, quase em suspenso. O pouco diálogo e o predomínio de sons naturais aproximam a escuta do espectador de uma experiência sensorial direta, na qual cada ruído carrega peso narrativo. O som, nesses momentos, acaba agindo como um agente subjetivo, capaz de tornar audível o esgotamento, a solidão e, por fim, a esperança que percorrem cada passo de Parrado e Canessa rumo à sobrevivência.

5.3.4 Considerações

Em *A Sociedade da Neve*, o som é concebido como elemento estruturante narrativo e espacial. A mixagem sonora opera criativamente ao construir atmosferas que ampliam a percepção emocional e espacial dos eventos narrados. A escuta é moldada para transitar entre planos subjetivos e objetivos, alternando entre a percepção interna dos personagens, por meio de sons abafados, respirações e silêncios densos, e a ambiência hostil da cordilheira, marcada por ventos cortantes, neve em atrito e ruídos pontuais. Essa elaboração permite que o espectador acesse os estados psicológicos dos sobreviventes, como o desamparo, a alucinação ou o esgotamento, não por explicitação verbal ou visual, mas por meio de texturas sonoras, manipulação de timbres e variações de amplitude. A criatividade do design de som se evidencia, portanto, na sua capacidade de narrar e sugerir afetos, ausências e presságios, instaurando uma escuta sensível que amplifica a experiência cinematográfica como um todo.

Ainda que o foco desta análise tenha se concentrado em sequências específicas, outras cenas do filme também se destacam sonoramente. Por exemplo, com o passar dos dias, há uma transformação gradativa na densidade sonora, marcada pela acentuação dos silêncios vocálicos que funcionam como elementos dramáticos, sugerindo o esgotamento físico e psicológico do grupo. Ou ainda no momento do resgate, quando os sobreviventes são levados de helicóptero e o fechamento da porta da aeronave é acompanhado por uma súbita interrupção do som ambiente. Esse silêncio atua como elemento narrativo, sinalizando não

apenas o término da jornada nos Andes, mas também o encerramento simbólico de um capítulo extremo na vida dos sobreviventes. Da mesma forma, as diversas sequências passadas no interior da fuselagem, cuja estrutura metálica, curva e fechada favorece a reverberação e o abafamento dos sons e são tratadas com uma espacialização que simula essa escuta circulante, reforçando o confinamento físico e psicológico vivido pelos personagens. Tal estratégia de construção sonora contribui para acentuar a dimensão sensorial e psicológica da experiência de sobrevivência, promovendo uma escuta imersiva que aproxima o espectador dos estados subjetivos vividos pelos personagens. Nesse contexto, a fuselagem do avião ultrapassa sua função de cenário e assume um papel de caixa de ressonância da tragédia, amplificando cada som como testemunho das tensões e fragilidades humanas em situação extrema.

Essas escolhas demonstram como o design de som do filme, especialmente quando apoiado por tecnologias imersivas, amplia as possibilidades expressivas da construção sonora, oferecendo uma escuta envolvente, densa e capaz de traduzir nuances afetivas e perceptivas que potencializam o impacto narrativo. O contraste com a versão estéreo, ainda que eficiente em transmitir a essência da narrativa, ressalta os limites de reprodução espacial e a perda de certas sutilezas acústicas. Assim, a análise comparativa evidencia o papel do som como agente estruturante da experiência audiovisual contemporânea e reafirma sua centralidade em propostas cinematográficas que buscam explorar a sensorialidade como forma de aproximação com o drama humano.

6 CONSIDERAÇÕES FINAIS

A trajetória desta tese parte de uma curiosidade em compreender o som, suas tecnologias e possibilidades de espacialização e imersão no cinema, além de seus desdobramentos em ambiente doméstico, não apenas como elemento técnico ou acessório narrativo, mas como um campo estético, artístico, histórico e cultural que estrutura experiências de escuta. Ao longo dos capítulos, buscou-se construir um diálogo entre a história do som sincronizado, os modelos de fidelidade e inteligibilidade que marcaram sua evolução, o ponto de escuta, seu conceito e categorizações, a proposta conceitual do ponto de escuta imersivo as tecnologias contemporâneas de mixagem orientada a objetos e as práticas profissionais e filmicas que dão materialidade a tais transformações. Essa conclusão não se limita, portanto, a retomar os pontos já tratados, mas pretende enfatizar as articulações e tensões que emergem desta investigação e projetar suas possíveis implicações para os estudos de som e para a prática audiovisual.

Um primeiro aspecto destacável é a centralidade da noção de ponto de escuta. A bibliografia discutida mostrou como diferentes autores buscaram delimitar o modo como o espectador se inscreve na cena sonora, seja através da perspectiva da inteligibilidade, da percepção sensorial e afetiva ou, ainda, por meio de abordagens ligadas à teoria do cinema e da estética sonora. O percurso histórico e tecnológico analisado nos capítulos seguintes demonstrou, contudo, que as condições materiais de produção e recepção reconfiguram constantemente essa noção. O ponto de escuta não deve ser entendido como uma categoria fixa, mas como um dispositivo móvel, atravessado tanto por escolhas criativas quanto por imposições técnicas e econômicas. Ao analisar as tecnologias sonoras orientadas a objetos, é possível entender a complexidade do conceito quando relacionado à espacialização tridimensional, à possibilidade de adaptação da mixagem às saídas disponíveis e à mobilidade das fontes sonoras no espaço, que sugerem uma experiência em que o espectador não apenas recebe o som, mas o habita de modo imersivo.

Essa imersão, no entanto, não se limita ao cinema, e um dos objetivos do trabalho foi entender de que forma é possível criar esse ambiente imersivo fora das salas de exibição, traçando um diálogo entre a sala de cinema e o ambiente doméstico na cronologia das tecnologias sonoras. Historicamente, como se discutiu no terceiro capítulo, as inovações no campo do cinema industrial muitas vezes encontram correspondência nas tecnologias de reprodução voltadas para os ambientes residenciais, com dispositivos até mesmo percorrendo o caminho inverso, como no caso da moviola que, inicialmente, seria um aparelho de

reprodução doméstica e tornou-se essencial na montagem técnica cinematográfica. O estéreo, o *surround* e, mais recentemente, o áudio orientado a objetos mostram que a experiência doméstica é atravessada por um desejo de simular, replicar ou mesmo reinventar a experiência cinematográfica. A percepção estratégica observada com as entrevistas com o mixador Ariel Ferrante e com o representante da Dolby na América Latina, Carlos Klachquin, é que as empresas envolvidas nesses novos formatos, como a Dolby e a Fraunhofer, tendem a indicar um maior orçamento em produtos que facilitem a convergência entre cinema, *streaming* e *broadcast* televisivo, de modo que as fronteiras entre estas esferas se tornam cada vez mais porosas. Assim, pensar o som no cinema implica necessariamente considerar seu trânsito para o espaço doméstico, no qual novos modos de recepção e apropriação se consolidam.

Considerando as questões estéticas e narrativas, a pesquisa demonstrou que as tecnologias orientadas a objetos não apenas ampliam as possibilidades técnicas, mas instauram novas formas de narratividade sonora, além da perspectiva de expansão do sentido artístico do filme, a partir das características imersivas. Nos filmes analisados, recursos como a manipulação do timbre, da densidade sonora e da espacialização produzem efeitos de sentido que não se limitam a ilustrar a imagem, mas instauram atmosferas, conferindo sentido à escuta e modulando a diegese do espectador. Em *Okja*, a exploração de texturas e contrastes sonoros tensiona a experiência emocional; em *Nada de Novo no Front*, o tratamento da paisagem sonora de guerra acentua tanto a violência quanto os silêncios, configurando uma escuta que oscila entre o hiper-realismo e a subjetividade; já em *A Sociedade da Neve*, a espacialização tridimensional é empregada para reforçar a dimensão coletiva da experiência, articulando vozes, ambientes e trilha musical em alta densidade dramática. Essas análises sugerem que a estética do áudio orientado a objetos não é neutra, mas participa ativamente da construção de significados do filme.

A experiência de escuta é ampliada pelo uso das novas tecnologias sonoras que intensificam a imersão do espectador e redefinem sua relação com a espacialidade fílmica. O uso de fones de ouvido, por exemplo, tende a suspender a consciência da arquitetura do ambiente de exibição, neutralizando suas referências físicas e transportando a escuta para um espaço acústico virtual. Mesmo quando projetados para difundir sons a partir do teto ou das laterais da sala, os sistemas de reprodução acabam sendo percebidos de forma distinta do que quando mediados por dispositivos individuais. Há uma percepção de que a imersão, o ponto de escuta e a espacialidade sonora se articulam visando uma redefinição da experiência audiovisual atual.

A voz e os objetos sonoros, quando explorados nesses filmes, instauram uma forma de intimidade com o ouvinte, deslocando o centro da experiência do domínio visual para uma dimensão mais sinestésica e sensorial. O espaço sonoro, nesse caso, não é apenas percebido ou representado, mas torna-se um ambiente habitável, no qual o espectador é envolvido por meio de vibrações, reverberações e densidades acústicas, assumindo a condição de um “corpo sônico”. Esse espaço sonoro não é apenas representado, mas efetivamente habitado, tornando-se um ambiente imersivo que ultrapassa a função ilustrativa da trilha. A escuta, nesse contexto, torna-se performativa, pois o som não se apresenta como mero objeto externo, mas como meio no qual o sujeito está inserido. O propósito subjacente é, portanto, atenuar ou mesmo apagar a percepção da mediação tecnológica, oferecendo ao espectador a ilusão de uma imersão plena em um mundo diegético convincente e sensorialmente tangível.

A espacialização sonora não apenas acompanha a narrativa, mas assume uma função estética que amplia o horizonte sensorial do cinema, mesmo observado em ambiente doméstico. Assim como discutem Elsaesser e Hagener (2018), em *Teoria do Cinema: uma introdução através dos sentidos*, o cinema em seu modo “ouvido” não apenas reproduz sons, como também forma modos de escuta, mobiliza afetos e inscreve o espectador em experiências que excedem a visão, sendo também um princípio estruturador da experiência cinematográfica. O que os autores trazem sobre o experimento do *walkman* é também possível de ser entendido para a utilização dos fones de ouvido na vivência fílmica, em que os dispositivos intra-auriculares criam um espaço interior, no qual os sons podem ser gerados nas próprias cabeças e “a pessoa se torna o epicentro (acústico) do mundo, não importando o quanto sua posição é periférica em relação a outros marcadores (sociais, espaciais) de centralidade ou hierarquia” (p. 172). Compreender essa experiência individual e centralizada da escuta fílmica implica em reconhecer como o som espacializado, reproduzido diretamente nos ouvidos, estabelece um vínculo específico com a realidade construída pelo filme. Nesse processo, o som atua como aliado na constituição da imersão audiovisual, contribuindo para estender essa experiência para além do espaço tradicional das salas de exibição. Mesmo que ainda tratando do advento do som e da imagem digital, Elsaesser e Hagener (2018, p. 173) já citavam como o som pode se tornar “um substituto funcional ou suplementar na experiência cinematográfica”, o que pode ser reafirmado neste trabalho, pois a análise realizada demonstra que o som imersivo, quando experienciado em ambiente doméstico, desempenha um papel central na aproximação do espectador à diegese fílmica, posicionando o ouvinte como alguém que não apenas observa, mas acompanha, sente e vivencia de perto os acontecimentos narrados, como se participasse diretamente da história apresentada.

Embora o conceito de paisagem sonora, formulado por R. Murray Schafer (1977), tenha se consolidado como ferramenta analítica para pensar a relação entre som e ambiente, também recebeu críticas exatamente pela própria nomenclatura que tende a sugerir uma posição de distanciamento, já que o ouvinte pode ser entendido como alguém que contempla o som à maneira de uma paisagem visual. Essa dimensão contemplativa implica um sujeito que observa “de fora”, em oposição a uma escuta que se engaja diretamente com a experiência fílmica. Nas configurações espaciais, no entanto, o som parece operar de forma distinta, já que a imersão sonora não se limita a descrever um espaço, mas cria uma condição de envolvimento, colocando o espectador dentro de um campo de escuta que o circunda por todos os lados. Nesse sentido, o som imersivo contribui para dissolver a ideia de uma paisagem distante, transpondo-a para um espaço sonoro circundante e participativo. A comparação com a escuta do som convencional ressalta esse contraste, pois se este tende a manter uma lógica semelhante à da observação de uma paisagem visual distante, o som imersivo rompe com essa posição contemplativa, transformando a escuta em experiência corpórea e integrativa no interior da diegese.

Conforme tratado no segundo capítulo deste trabalho, em *Audiovisão*, Michel Chion (2008) descreve, dentre outras características, o ponto de escuta como uma analogia auditiva do ponto de vista. Entre suas variantes, destaca-se o ponto de escuta subjetivo, no qual o espectador é convidado a compartilhar a percepção sonora de um personagem específico, ouvindo “como ele” determinados sons que a narrativa privilegia. Essa categoria, porém, ancora-se na lógica representacional do cinema clássico, em que a subjetividade é construída por recursos pontuais, como abafamentos, filtros ou manipulações do timbre, que sinalizam o deslocamento para a escuta de um personagem. Durante a composição da análise fílmica e a observação das possibilidades de uso das tecnologias de espacialização, sobretudo os sistemas orientados a objetos como Dolby Atmos, foi possível aprofundar na necessidade de adequação nominal e proposição de um desdobramento que descreve a proximidade que as tecnologias de imersão sonora proporcionam entre o espectador e a história: o ponto de escuta imersivo. Diferentemente do ponto de escuta subjetivo, em que a audição do espectador apenas imita a do personagem, o ponto de escuta imersivo constitui-se por uma experiência sensorial mais densa, em que a espacialidade sonora envolve o ouvinte de forma tridimensional, produzindo não só a partilha de uma percepção, mas a sensação de uma incorporação. Nesse caso, o espectador não apenas ouve como o personagem, mas sente como ele. Trata-se de uma escuta marcada pela proximidade e pela corporeidade, em que o som circunda o espectador, inserindo-o na ação, como se participasse corporalmente da diegese.

A proposta de um ponto de escuta imersivo intenciona auxiliar na resposta, portanto, de uma necessidade teórica de diferenciar as estratégias tradicionais de subjetivação sonora das novas modalidades de recepção propiciadas pela imersividade tecnológica. Enquanto o ponto de escuta subjetivo mantém-se dentro da lógica de representação narrativa, com a câmera sonora que se coloca “nos ouvidos” de alguém, o ponto de escuta imersivo expande a experiência estética, pois transforma o espaço fílmico em experiência sensorial direta. Essa distinção visa compreender como as tecnologias contemporâneas deslocam o som de uma função narrativa para uma dimensão estética e corporal, na qual a diegese é vivenciada de maneira envolvente e circundante.

A análise fílmica aqui desenvolvida não se limita a apenas verificar a verossimilhança ou a fidelidade da espacialização, mas busca compreender de que forma a construção sonora mobiliza o corpo do espectador e gera experiências de copresença, em que o espaço sonoro circundante cria a sensação de sentir junto com o personagem e de compartilhar corporalmente sua vivência fílmica. A adoção deste termo tem caráter propositivo, visando delimitar um campo conceitual capaz de dar conta das transformações estéticas, tecnológicas e perceptivas que emergem com o uso de ferramentas de áudio imersivo em ambiente doméstico e cinematográfico, evidenciando como as inovações sonoras contemporâneas transformam o lugar do ouvinte na narrativa cinematográfica.

Em *Okja*, por exemplo, a cena da perseguição no mercado de Seul explicita o contraste entre caos urbano e a presença da criatura. A estética sonora aqui vai além da verossimilhança, já que o uso do *delay panning* e de variações de intensidade criam uma sensação de movimento constante que coloca o espectador dentro do labirinto urbano. O ponto de escuta oscila entre uma configuração subjetiva, centrada em Okja e Mikha, marcada por frequências graves que conferem corpo à sua fisicalidade, e uma configuração imersiva, em que a espacialidade tridimensional faz com que a massa sonora da multidão e dos veículos envolva o espectador, instaurando uma escuta incorporada. Nesse deslocamento, o som transforma a experiência estética, de modo que a corporalidade da criatura possa ser sentida pelo ouvido antes mesmo de ser plenamente vista, instaurando uma percepção sensorial em que o fantástico e o hiper-real se interpenetram. Assim, o filme mobiliza o som como recurso artístico que amplia a percepção da cena e tensiona os limites entre realidade e imaginação.

Em *Nada de Novo no Front*, as batalhas são estruturadas pela estética da sobrecarga sonora. Na cena em que o protagonista é lançado à lama das trincheiras, os ruídos metálicos das metralhadoras, os estampidos graves das bombas e os silêncios abruptos, quando a audição se reduz a zumbidos, não apenas traduzem a desorientação da guerra, mas constroem

uma poética da experiência traumática. Essa alternância entre plenitude sonora e silenciamento não se limita a uma interioridade psicológica, mas evoca um ponto de escuta imersivo, também visceral e corpóreo, em que o ouvido do espectador é convocado a partilhar a experiência-limite. A violência da guerra, assim, torna-se não apenas visível, mas sensorialmente tangível. O design sonoro auxilia na construção do ambiente bélico e também o inscreve como experiência estética, aproximando-se de uma escultura sonora do trauma, que dá corpo à memória sensorial da guerra.

Já em *A Sociedade da Neve*, a estética sonora articula silêncio e intimidade como dispositivos narrativos e artísticos. Na cena do impacto inicial do avião, o ruído agudo da fuselagem rompendo e o estrondo grave da colisão se expandem pelo espaço sonoro de forma a preencher o corpo do espectador, instaurando um ponto de escuta imersivo que transborda a simples representação. Em seguida, o silêncio das montanhas assume uma centralidade não apenas de ausência, mas também de presença estética, moldura que confere espessura à experiência do isolamento. Nos diálogos posteriores, especialmente nas cenas em que os sobreviventes dividem alimento ou se agrupam para resistir ao frio, a sonoridade privilegia pequenos sons, como respirações, estalos da neve e murmúrios, que são elevados à condição de estética do mínimo, uma poética da vida em suspensão. Elsaesser e Hagener (2018) sugerem que o cinema, enquanto ouvido, pode dar voz ao inaudível e, nesse caso, o filme confere densidade estética à fragilidade da experiência dos sobreviventes, tornando-os epicentro da experiência sensorial, percebida pelo espectador através desse ponto de escuta imersivo.

Esses exemplos permitem concluir que as tecnologias de mixagem orientada a objetos, ao expandirem a espacialidade, não resultam apenas em uma imersão técnica, mas em uma transformação estética da narrativa audiovisual. O som torna-se um campo artístico em si, capaz de elaborar atmosferas, corporificar emoções e propor ao espectador modos singulares de escuta. Trata-se de um sentido que desloca a primazia da visão para uma experiência sinestésica, na qual o ouvido se torna veículo privilegiado capaz de ampliar noções de afeto, empatia e sensorialidade.

Se, durante a introdução deste trabalho, era perceptível a necessidade da manutenção do público atento e confortável em um lugar comum, sustentando a permanência de padrões convencionais na mixagem do audiovisual *mainstream*, o som imersivo começa a apontar para um deslocamento desta lógica. Conforme demonstrado nas análises filmicas, mesmo em ambiente doméstico, a espacialidade imersiva rompe com a estabilidade perceptiva esperada, retirando o espectador de sua posição confortável e instaurando uma experiência sensorial

mais instável e participativa, na qual a escuta deixa de ser apenas acompanhamento narrativo para se tornar também uma vivência estética.

A ampliação dessas possibilidades imersivas proporcionadas pelas tecnologias sonoras também suscita questões de ordem ética, especialmente quando aplicadas a narrativas de dor, violência ou tragédias. Ao envolver o espectador em um campo sonoro tridimensional, que não apenas reproduz a escuta do personagem, mas o faz sentir como ele, o cinema expande as fronteiras da empatia para o terreno da coparticipação sensorial. Essa potência, embora esteticamente transformadora, exige reflexão sobre alguns limites, principalmente em relação a até que ponto a imersão total na experiência do sofrimento não se converte em uma forma de apropriação emocional ou de banalização da dor alheia. A imersividade, ao intensificar o vínculo afetivo entre espectador e personagem, revela também uma zona de tensão entre a empatia e o desconforto, desafiando os criadores e estudiosos do som a teorizar sobre o papel da escuta como mediadora da experiência fílmica enquanto dispositivo técnico e prática sensível que pode moldar como percebemos e compartilhamos os sentimentos e emoções.

Contudo, o entusiasmo tecnológico não pode apagar os limites e as contradições que atravessam o campo. Os depoimentos dos profissionais entrevistados revelam preocupações práticas com a padronização dos formatos, o custo de implementação e a formação técnica necessária para lidar com a complexidade dos sistemas. Além disso, a democratização do acesso às tecnologias ainda é desigual e falta políticas de incentivo para cinematografias independentes ou periféricas competirem nesse cenário. O ambiente doméstico, por sua vez, também é marcado por desigualdades, já que nem todos os espectadores dispõem do equipamento necessário para usufruir plenamente da experiência sonora prometida pelas empresas. Nesse sentido, a materialidade técnica do som se entrelaça a questões econômicas, sociais e culturais.

Talvez, uma das contribuições desta tese resida justamente em articular esses diferentes planos, como o histórico, técnico, profissional, artístico e estético, com a intenção de compreender o som como experiência complexa e atravessada por múltiplas dimensões. Ao propor uma leitura do ponto de escuta à luz das tecnologias orientadas a objetos, buscou-se atualizar uma categoria central dos estudos de som, expandindo-a para abarcar não apenas o campo cinematográfico, mas também as novas perspectivas de escuta que emergem com o *streaming* e o consumo doméstico. O trabalho evidencia a necessidade de assimilar a escuta como experiência estética e cultural, que reconfigura continuamente a relação entre obra, espectador e tecnologia, para além dos parâmetros técnicos utilizados.

É importante reconhecer que a ênfase no potencial das tecnologias sonoras, sobretudo as imersivas, pode, por vezes, conduzir a uma leitura marcada pelo prazer da descoberta, que tende a superestimar sua capacidade de transformação narrativa. A estrutura de um filme permanece, em grande medida, inalterada, independentemente do formato de exibição, no qual a sucessão de eventos, a construção de personagens e o encadeamento dramático não são substancialmente modificados pela adoção de sistemas de som multicanal ou orientados a objetos. O que se altera, no entanto, é a dimensão estética e artística da experiência, isto é, a forma como o espectador é conduzido a vivenciar essas mesmas narrativas. Esse sentido emerge justamente da reconfiguração da escuta, da ampliação do campo sensorial e da possibilidade de integrar o ouvinte de maneira mais íntima ao espaço fílmico. Dessa forma, a tecnologia pode não criar novas histórias, mas oferece condições para que histórias já conhecidas possam ser sentidas de modo diverso, instaurando camadas de intensidade, corporalidade e empatia que reconfiguram a relação entre espectador e obra. Nesse processo, o papel criativo dos cineastas torna-se central, explorando os recursos oferecidos pelas tecnologias imersivas e imaginando o som não apenas como complemento visual, mas como agente narrativo e sensorial, em que a história não será alterada de acordo com a tecnologia sonora escolhida, mas a densidade estética pode se ampliar e conquistar uma maior aproximação e identificação espectral. Para tanto, é necessário que o design sonoro seja pensado desde as etapas iniciais da produção, articulado à *mise en scène* e ao projeto do filme, e não apenas como elemento adicionado na captação e pós-produção. Assim, a qualidade imersiva não decorre apenas da sofisticação técnica, mas da capacidade criativa de integrar o som à experiência fílmica de forma orgânica, expandindo a densidade sensorial da obra. Reconhecer essa distinção tem por objetivo evitar uma perspectiva apenas tecnicista e, ao mesmo tempo, tentar compreender como a estética sonora pode constituir um campo de invenção artística no cinema contemporâneo.

Por fim, cabe destacar que este percurso abre também novos caminhos de investigação. A lógica do áudio orientado a objetos já se expande para campos como realidade virtual, jogos digitais e ambientes imersivos e interativos, em que a espacialização sonora se torna ainda mais notável. No consumo doméstico, o uso de caixas de som portáteis, tipo *bluetooth*, na ampliação da capacidade sonora de dispositivos como televisores, indicam que o som espacializado tridimensionalmente tende a se expandir ainda mais, colocando o espectador em posições cada vez mais ativas e imersivas. No campo dos estudos de recepção, há um vasto espaço para explorar como diferentes públicos se apropriam dessas tecnologias, seja no cinema, em casa ou em dispositivos móveis, além do entendimento de como percebem

essa diferença sonora entre o estéreo ou o multicanal convencional e a possibilidade tridimensional e imersiva.

Em relação às políticas culturais, torna-se urgente refletir sobre como fomentar o acesso às tecnologias sonoras e como apoiar a formação de profissionais capazes de lidar criticamente com elas. Além disso, a regulamentação das plataformas de *streaming* é necessária e urgente, visando, principalmente, a inclusão de um percentual de produções nacionais em seus catálogos e a contribuição para a indústria audiovisual nacional através do recolhimento de impostos, buscando fortalecer a soberania cultural do país e atrair novos públicos para a nossa produção audiovisual. Finalmente, no campo da crítica cinematográfica e dos estudos de mídia, permanece o desafio de integrar a análise sonora de forma mais consistente, reconhecendo o papel do som na construção da experiência audiovisual.

Assim, mais do que um fechamento, a conclusão desta tese é um convite à continuidade, uma vez que a escuta no cinema e no audiovisual doméstico é um campo em permanente transformação e compreender suas mutações exige atenção às tecnologias, mas também às sensibilidades e às culturas que delas emergem. O ponto de escuta, seja imersivo ou não, como conceito e como experiência, permanece em aberto e em movimento entre salas, lares e mundos virtuais, desafiando-nos a pensar o som como uma força estética e cultural que molda o próprio modo como escutamos, vemos e sentimos o cinema.

REFERÊNCIAS

ALTMAN, Rick. **Sound Space**. In: ALTMAN, Rick (ed.). **Sound Theory/Sound Practice**. 1992. p. 46-64.

A SOCIEDADE DA NEVE. Direção: J. A. Bayona. Espanha: Netflix, 2023. 144 min. Filme. Disponível em: <https://www.netflix.com/title/81252694>. Acesso em: 25 jul. 2025.

A SOCIEDADE DA NEVE: QUEM FOMOS NA MONTANHA. Direção: Manuel Romo. Espanha: Netflix, 2024. 36 min. Documentário. Disponível em: <https://www.netflix.com/title/81710127>. Acesso em: jul. 2025.

AURO Creative. **AuroMax White Paper**. 2015. Disponível em: https://web.archive.org/web/20160506172309/http://www.auro-3d.com/wp-content/uploads/documents/AuroMax_White_Paper_24112015.pdf. Acesso em: 03 jul. 2024.

AURO Creative. **Auro-3D Engine v3.2 White Paper Rev6**. 2023. Disponível em: https://www.auro-3d.com/wp-content/uploads/2023/08/Auro-3D_Engine_v3_2_White_Paper_Rev6.pdf. Acesso em: 03 jul. 2024.

AURO Creative. **Professional**. Disponível em: <https://www.auro-3d.com/professional/>. Acesso em: 04 jul. 2024.

AURO Creative. **Auro Creative**. Disponível em: <https://www.auro-3d.com/creative/>. Acesso em: 04 jul. 2024.

BERTELMANN, Volker (Hauschka). **Composer Volker Bertelmann on the Original Score for All Quiet on the Western Front** [entrevista/vídeo online]. YouTube, fev. 2023. Disponível em: <https://www.youtube.com/watch?v=xFL-62CiZBk>. Acesso em: 17 mai. 2025.

BORDWELL, David. **The Introduction of Sound**. In: BORDWELL, David; STAIGER, Janet; THOMPSON, Kristin. **The Classical Hollywood Cinema: Film Style and Mode of Production to 1960**. New York: Columbia University Press, 1985. p. 301-303.

BRASIL. Agência Nacional do Cinema – ANCINE. **Painel Indicadores do Mercado de Exibição**. Publicado em 26 mai. 2025; atualizado em 25 ago. 2025. Disponível em: <https://www.gov.br/ancine/pt-br/oca/paineis-interativos-1/paineis-interativos-subsecao-mercado-audiovisual/painel-indicadores-do-mercado-de-exibicao>. Acesso em: 26 ago. 2025.

BRASIL. Ministério das Comunicações. **Ministério das Comunicações recebe os primeiros protótipos da antena e do conversor da TV 3.0**. Brasília: MCom, 2025. Disponível em: <https://www.gov.br/mcom/pt-br/noticias/2025/abril/ministerio-das-comunicacoes-recebe-os-primeiros-prototipos-da-antena-e-do-conversor-da-tv-3.0>. Acesso em: 24 abr. 2025.

BRASIL. Ministério das Comunicações. **Lula destaca que o Brasil terá a TV 3.0 para revolucionar a forma de assistir à televisão**. Brasília: MCom, 3 abr. 2025. Disponível em:

<https://www.gov.br/mcom/pt-br/noticias/2025/abril/lula-destaca-que-o-brasil-tera-a-tv-3-0-para-revolucionar-a-forma-de-assistir-a-televisao>. Acesso em: 29 maio 2025.

BRETT, Leonard. **Applications of extended multichannel techniques**. In: ROGINSKA, Agnieszka; GELUSO, Paul (ed.). **Immersive sound: the art and science of binaural and multi-channel audio**. New York; London: Routledge, 2018. p. 333–355.

BUHLER, J.; NEUMEYER, D.; DEEMER, Rob. **Hearing the movies: music and sound in film history**. New York: Oxford University Press, 2010.

CARREIRO, Rodrigo. **A pós-produção de som no audiovisual brasileiro**. João Pessoa: Marca de Fantasia, 2019. Disponível em: https://www.academia.edu/39198372/A_p%C3%B3s-produ%C3%A7%C3%A3o_de_som_no_audiovisual_brasileiro_ebook_PDF_. Acesso em: 12 jun. 2024.

CARREIRO, Rodrigo; OPOLSKI, Débora (org.). **O som do filme: uma introdução**. Curitiba: Editora UFPR, 2018.

CARREIRO, Rodrigo; OPOLSKI, Débora; MEIRELLES, Rodrigo. **A imersão sonora no cinema**. São José dos Pinhais: Estronho, 2023.

CARVALHO, José Jorge de. **Transformações da sensibilidade musical contemporânea. Horizontes Antropológicos**, Porto Alegre, ano 5, n. 11, p. 53-91, out. 1999. Disponível em: <http://dx.doi.org/10.1590/S0104-71831999000200004>. Acesso em: 10 jul. 2024.

CASS, J. L. **The Illusion of Sound and Picture**. *Journal of the Society of Motion Picture Engineers*, n. 14, mar. 1930.

CHANDLER, Daniel; MUNDAY, Rod. **A Dictionary of Media and Communication**. 1. ed. Oxford: Oxford University Press, 2011. ISBN 9780199568758. Disponível em: <https://doi.org/10.1093/acref/9780199568758.001.0001>. Acesso em: 11 jun. 2024.

CHION, M. **A audiovisão**. Lisboa: Texto & Grafia, 2008.

CHION, M. **Film, a sound art**. New York: Columbia University Press, 2009.

COHEN, S. **What is Dolby Atmos Music and how can you experience it?** *Digital Trends*, 2019. Disponível em: <https://www.digitaltrends.com/home-theater/what-isdolby-atmos-music-and-how-to-get-it/>. Acesso em: 04 jul. 2024.

DELLASALA, Gene. **HRTF and Elevated Sound: Dolby Atmos**. Audioholics, 2014. Disponível em: <https://www.audioholics.com/loudspeaker-design/hrtf-and-elevated-sound-dolby-atmos>. Acesso em: 10 jul. 2024.

DOLBY. **Dolby Atmos Enabled Speaker Technology**. 2016. Disponível em: <https://professional.dolby.com/siteassets/tv/home/dolby-atmos/dolby-atmos-enabled-speaker-technology.pdf>. Acesso em: 08 jul. 2024.

DOLBY. **Dolby Atmos Installation Guidelines**. 2018. Disponível em: https://www.dolby.com/siteassets/technologies/dolby-atmos/atmos-installation-guidelines-121318_r3.1.pdf. Acesso em: 08 jul. 2024.

DOLBY. **Dolby Atmos for Sound Bar Applications**. 2018. Disponível em: <https://professional.dolby.com/siteassets/tv/home/dolby-atmos/dolby-atmos-for-sound-bar-applications.pdf>. Acesso em: 09 ago. 2024.

DOLBY. **Dolby showcases global momentum at 2023 CinemaCon**. Disponível em: <https://news.dolby.com/en-WW/225276-dolby-showcases-global-momentum-at-2023-cinema-con>. Acesso em: 19 jul 2024.

DOLBY. **Headphones and Dolby Atmos**. Disponível em: https://professionalsupport.dolby.com/s/topic/0TO4u000000eymdGAA/headphones-and-dolby-atmos?language=en_US. Acesso em: 09 jul. 2024.

DOLBY. **Why Dolby Atmos?** White Paper. 2020. Disponível em: <https://professional.dolby.com/siteassets/content-creation/dolby-atmos/why-dolby-atmos-white-paper.pdf>. Acesso em: 07 jul. 2024.

EIKMEIER, Martin. **O advento do som sincronizado no cinema**. Disponível em: <https://www.spescoladeteatro.org.br/noticia/o-advento-do-som-sincronizado-no-cinema/>. Acesso em: 19 fev. 2021.

ELSAESSER, Thomas; HAGENER, Malte. **Teoria do cinema: uma introdução através dos sentidos**. 1. ed. São Paulo: Vozes, 2018.

EXIBIDOR. **Dolby divulga resultados financeiros do primeiro trimestre do ano fiscal de 2024**. 12 fev. 2024. Disponível em: <https://www.exibidor.com.br/noticias/mercado/13870-dolby-divulga-resultados-financeiros-do-primeiro-trimestre-do-ano-fiscal-de-2024>. Acesso em: 07 jul. 2024.

FILME CULTURA. Rio de Janeiro, 2013, jan/fev/mar, nº 58.

FRAUNHOFER IIS. **MPEG-H Audio: The Future of Audio is Here**. Erlangen: Fraunhofer Institute for Integrated Circuits IIS, 2023. Disponível em: <https://www.iis.fraunhofer.de/en/ff/amm/broadcast-streaming/mpeg-h.html>. Acesso em: 7 abr. 2025.

FRAUNHOFER IIS. **South Korea launches UHD TV with MPEG-H Audio**. Erlangen: Fraunhofer IIS, 2017. Disponível em: <https://www.audioblog.iis.fraunhofer.com/south-korea-uhd-tv-mpeg-h>. Acesso em: 24 abr. 2025.

FERRANTE, Ariel. Entrevista à autora. Belo Horizonte, 25 mar. 2025.

HØIER, S. **The relevance of Point of Audition in television sound: Rethinking a problematic term**. *Journal of Sonic Studies*, v. 3, 2013. Disponível em:

<https://www.researchcatalogue.net/view/252390/252391>. Acesso em: 10 jun. 2024.

KERINS, Mark. **Beyond Dolby (Stereo): Cinema in the Digital Sound Age**. Bloomington: Indiana University press, 2011.

KERINS, Mark. **Home Theater(s): technology, culture and style**. In: MERA, Miguel, SADOFF, Ronald; WINTERS, Ben (ed.). **The Routledge Companion to Screen Music and Sound**. London: Routledge, 2017. p. 388-399.

KLACHQUIN, Carlos. **O som no cinema**. Transcrição da palestra do consultor da Dolby para a América Latina Carlos Klachquin. Seminário ABC – A imagem sonora. Cinemateca Brasileira: São Paulo, 2002. Disponível em: <<https://abcine.org.br/site/o-som-no-cinema/>>. Acesso em: 19/02/2021.

KLACHQUIN, Carlos. **Carlos Klachquin ABC**. Associação Brasileira de Cinematografia. Disponível em: <https://abcine.org.br/socios/carlos-klachquin-abc/>. Acesso em: 5 maio 2025.

KLACHQUIN, Carlos. Entrevista à autora . Belo Horizonte, 15 abr. 2025.

LANDAY, Lori. **The Moviola and Other Analog Film Editing Machines**. In: WOLF, Mark J.P. (org.). **The Routledge Companion to Media Technology and Obsolescence**. New York: Routledge, 2019. p. 136–147.

LEE, Jae-Ho et al. **Quality Enhancement of MPEG-H 3DA Binaural Rendering Using a Spectral Compensation Technique**. *Electronics*, v. 11, n. 9, p. 1491, 2022. Disponível em: <https://www.mdpi.com/2079-9292/11/9/1491>. Acesso em: 25 abr. 2025.

MAXFIELD, J. P. **Some Physical Factors Affecting the Illusion in Sound Motion Pictures**. *Journal of the Acoustical Society of America*, New York, v. 3, n. 3, jul. 1931.

METZ, Christian; GURRIERI, Georgia. **Aural Objects**. *Yale French Studies*, n. 60, 1980, pp. 24-32. Disponível em: <https://doi.org/10.2307/2930002>. Acesso em: 10 abr. 2024.

NADA DE NOVO NO FRONT. Direção: Edward Berger. Alemanha; Estados Unidos: Netflix, 2022. 147 min. Filme. Disponível em: <https://www.netflix.com/title/81586887>. Acesso em: jun. 2025.

NETFLIX. **Dolby Atmos chegando à Netflix**. Disponível em: https://about.netflix.com/pt_br/news/dolby-atmos-coming-to-netflix. Acesso em: 19 jul. 2024.

OJKA. Direção de Bong Joon-ho. Produção de Plan B Entertainment, Lewis Pictures e Kate Street Picture Company. Netflix, 2017. 120 min. Disponível em: <https://www.netflix.com/title/80091936>. Acesso em: jun. 2025.

ONO, Yuya; SAJI, Akira; HUANG, Jie. **Frontal Sound Localization with Headphone Systems Using Characteristics of Each Device**. 2017. University of Aizu, Aizu Wakamatsu, Japan; University of Aizu, Aizuwakamatsu City, Japan. Paper 354. Disponível em: <https://aes2.org/publications/elibrary-page/?id=19305>. Acesso em: 08 jul. 2024.

OXFORD REFERENCE. **Point of Audition Sound**. [s.d.]. Disponível em:

<https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100333623#:~:text=Quick%20Reference,Dictionary%20of%20Media%20and%20Communication%20%C2%BB>. Acesso em: 11 jun. 2024.

PALENCHAR, J. **Positioned as Alternative to Atmos. Twice**, 26 jan. 2015. Disponível em: <https://www.twice.com/news/dtsx-positioned-alternative-atmos-55721>. Acesso em: 04 jul. 2024.

SCHAEFFER, Pierre. **Traité des objets musicaux**: essai interdisciplines. Paris: Éditions du Seuil, 1966.

SCHIFFER, S. **Footsteps, breath and recording devices**: Abandoning a camera-centric construction of point of audition. **The Soundtrack**, v. 5, n. 1, pp. 15-28, 2012.

SERGI, Gianluca. **Knocking at the door of cinematic artifice**: Dolby Atmos, challenges and opportunities. **School of Sound**, 2013. Disponível em: <https://www.schoolofsound.co.uk/read/knocking-at-the-door-of-cinematic-artifice-dolby-atmos-challenges-and-opportunities/>. Acesso em: 04 jul. 2024.

SEXTON, Connor. **Immersive Audio**: Optimizing Creative Impact without Increasing Production Costs. 2017. Avid, Berkeley, CA, USA. Paper 370. Disponível em: <https://aes2.org/publications/elibrary-page/?id=19321>. Acesso em: 08 jul. 2024.

SMITH, Jeff. **Atmos, all around**: A guest post by Jeff Smith. **David Bordwell's Website on Cinema**, maio 2013. Disponível em: <http://www.davidbordwell.net/blog/2013/05/01/atmos-all-around-a-guest-post-by-jeff-smith/>. Acesso em: 04 jul. 2024.

SOCIETY OF MOTION PICTURE AND TELEVISION ENGINEERS (SMPTE). **SMPTE Standards**: Document Index. Disponível em: <https://www.smpte.org/standards/document-index/st#listing>. Acesso em: 07 jul. 2024.

STORM AUDIO. **DTS Audio Solutions for Home A/V Products**. 2020. Disponível em: https://www.stormaudio.com/wp-content/uploads/2020/04/whitepaper_dts_av_one_sheet_11082016.pdf. Acesso em: 04 jul. 2024.

SUN, Xuejing. **Immersive audio, capture, transport, and rendering**: a review. **SIP**, vol. 10, e13, p. 1-24, 2021. Disponível em: <http://doi.org/10.1017/ATSIP.2021.12>. Acesso em: 12 jul. 2024.

THEISEN, Earl. **The story of the Moviola**. **International Photographer**, v. 7, n. 12, p. 367-369, dez. 1935. Disponível em: <https://archive.org/details/internationalpho07holl/page/n367/mode/2up>. Acesso em: 12 mai. 2025.

THOMPSON, Anne. **How Much Does It Cost to Run a Movie Theater?** Disponível em: <https://www.indiewire.com/features/general/how-much-does-it-cost-to-run-movie-theater-1234859340/>. Acesso em: 19 jul. 2024.

TRANÅS, M. W. **Unravelling Point of Audition**. Dissertação (Mestrado em Musicologia). University of Oslo, Oslo, 2017.

TSINGOS, Nicolas. **Object-based Audio**. In: ROGINSKI, G.; GELUSO, P. (orgs.). **Immersive sound: The art and science of binaural and multi-channel audio**. New York; London: Routledge, 2018. p. 244-275.

VIDIGAL, Leonardo Alvarez. **Ponto de escuta e regimes da diegese: ambiguidades na percepção sonora**. In: **SOCINE**, XXVI, 2023, Brasília. Anais de Textos Completos 2023. Brasília: SOCINE, 2023. Disponível em: [https://www.socine.org/wp-content/uploads/anais/AnaisDeTextosCompletos2023\(XXVI\).pdf](https://www.socine.org/wp-content/uploads/anais/AnaisDeTextosCompletos2023(XXVI).pdf). Acesso em: 30 jun. 2024.

VIDIGAL, L. Álvares; NOVAIS, M. de M. **Ponto de escuta e tridimensionalidade sonora na ficção e no documentário. PÓS: Revista do Programa de Pós-graduação em Artes da EBA/UFGM**, p. 224–248, 2020. Disponível em: <https://periodicos.ufmg.br/index.php/revistapos/article/view/21604>. Acesso em: 12/04/2023.

VÖLK, Florian; ENCKE, Jörg; KREH, Jasmin; HEMMERT, Werner. **Relevance of Headphone Characteristics in Binaural Listening Experiments: A Case Study**. 2017. Technische Universität München, Munich, Germany; WindAcoustics, Windach, Germany; Technical University of Munich, Munich, Germany. Paper 9894. Disponível em: <https://aes2.org/publications/elibrary-page/?id=19291>. Acesso em: 08 jul. 2024.

WILLIAMS, Alan. **Is Sound Recording Like a Language?** *Yale French Studies*, n. 60, 1980, pp. 51-66. Disponível em: <https://doi.org/10.2307/2930004>. Acesso em: 09 abr. 2024.

GLOSSÁRIO

AC-3: Codec de áudio desenvolvido pela Dolby Laboratories, também conhecido como Dolby Digital. Suporta até 5.1 canais de áudio e é amplamente usado em DVDs, Blu-rays e transmissões digitais.

AC-4: Codec de áudio da Dolby de nova geração, que oferece suporte para áudio imersivo, compressão mais eficiente e interatividade. Compatível com Dolby Atmos e utilizado em sistemas de TV e *streaming*.

ADM (Audio Definition Model): Modelo de metadados proposto pela ITU (*International Telecommunication Union*) para descrever conteúdo de áudio imersivo, incluindo a posição dos objetos sonoros, características de mixagem e outros parâmetros técnicos.

Ambisonic: Técnica de captura e reprodução de som tridimensional baseada em campos sonoros esféricos, que permite representar espacialmente fontes em todas as direções. Muito utilizada em realidade virtual, jogos e produções experimentais.

Áudio baseado em canais (channel-based Audio): Sistema de representação sonora que distribui o som em canais específicos destinados a alto-falantes ou grupos de alto-falantes no cinema.

Audio beds (camas sonoras): Faixa ou grupo de faixas de áudio fixas utilizadas como base em mixagens orientadas a objetos, sobre as quais os objetos sonoros móveis são adicionados. Em Dolby Atmos, normalmente refere-se a trilhas surround tradicionais (ex: 5.1 ou 7.1.2).

Audio Objects (objetos sonoros): Elementos sonoros capazes de serem localizados e movimentados de forma precisa, renderizados individualmente com metadados que indicam posição, difusão, tamanho e outras características acústicas.

Áudio orientado a objeto (object-based audio): Estrutura de mixagem na qual elementos sonoros individuais (objetos) são manipulados com independência espacial, permitindo maior controle sobre a reprodução em diferentes ambientes e sistemas.

Binaural Audio (áudio binaural): Técnica baseada no funcionamento da audição humana, utilizando dois canais para simular a percepção espacial natural do som.

CGI (Computer-Generated Imagery): Criação digital de elementos visuais, utilizada tanto em efeitos especiais quanto em animações.

BWF (Broadcast Wave Format): Formato de arquivo de áudio derivado do WAV, amplamente utilizado na produção audiovisual. Permite inclusão de metadados como tempo, nome de projeto, e informações de canal.

Cubig Digital Recorder Sound: Sistema de gravação de som direto utilizado no cinema, que permite captura de áudio multicanal com sincronização para pós-produção.

DAW (Digital Audio Workstation): Estação de trabalho de áudio digital. Refere-se a softwares profissionais como Pro Tools, Nuendo, Reaper, entre outros, utilizados para gravação, edição, mixagem e masterização de áudio.

DCP (Digital Cinema Package): Conjunto padronizado de arquivos digitais utilizados para exibição de filmes em salas de cinema. Inclui imagem, som e metadados organizados em pastas e contêineres específicos. É reproduzido através de servidores de cinema digital.

Delay Panning (panorâmica com atraso): Técnica de mixagem que aplica pequenas defasagens temporais entre canais para criar sensação de movimento ou reforçar a localização espacial de uma fonte sonora, intensificando a percepção de profundidade e imersão.

Dynamic Range (faixa dinâmica): Relação entre a maior (mais alta) e a menor (mais baixa) amplitude sonora que um sistema pode reproduzir, expressa em decibéis (dB). Uma maior faixa dinâmica proporciona maior fidelidade e impacto sonoro.

Dolby Atmos: Tecnologia de som imersivo baseada em objetos desenvolvida pela Dolby Laboratories. Permite o posicionamento preciso de sons no espaço tridimensional e é compatível com salas de cinema, dispositivos domésticos e plataformas de streaming.

Dolby Vision: Tecnologia de imagem em High Dynamic Range (HDR) desenvolvida pela Dolby, que proporciona contraste, brilho e gama de cores ampliados, frequentemente utilizada em conjunto com Dolby Atmos.

Downmix: Processo de redução do número de canais de áudio, como de 5.1 para estéreo, preservando o máximo possível da integridade e espacialidade da mixagem original.

Fórum SBTVD (Sistema Brasileiro de Televisão Digital): Entidade responsável pela normatização, desenvolvimento e evolução do sistema de TV digital no Brasil. Envolve membros da indústria, governo e comunidade científica.

HDR (High Dynamic Range): Técnica que amplia o alcance dinâmico da imagem (diferença entre os tons mais escuros e mais claros), aumentando a percepção de profundidade e realismo visual. Presente em sistemas como Dolby Vision.

HE-AAC (High Efficiency Advanced Audio Coding): Codec de áudio de alta eficiência, com bom desempenho em taxas de bits reduzidas. Utilizado em transmissões digitais e serviços de streaming.

Home Theater: Sistema de som doméstico multicanal que busca replicar a experiência audiovisual do cinema em casa. Pode incluir múltiplos alto-falantes, subwoofer, receiver e, em sistemas mais modernos, suporte a áudio imersivo como Dolby Atmos.

HRTFs (Head-Related Transfer Functions): Medidas utilizadas na gravação binaural para representar como cada ouvido recebe sons de diferentes posições espaciais, dependendo das características físicas da cabeça, orelhas, cavidades orais e nasais e torso.

Interactive Audio Bitstream (IAB): Formato de entrega de áudio orientado a objetos padronizado pela SMPTE, utilizado como base para a distribuição de conteúdo imersivo como Dolby Atmos.

Integrated Media Block (IMB): Componente essencial de um sistema de projeção digital que processa e decodifica o conteúdo de um DCP (*Digital Cinema Package*), incluindo áudio e vídeo. Trabalha em conjunto com o servidor de mídia.

Integrated Media Service (IMS): Sistema que integra as funções do servidor de cinema digital e do IMB em um único equipamento, otimizando a reprodução de conteúdo DCP e facilitando a gestão de exibições nas salas de cinema.

MP4 (MPEG-4 Part 14): Formato contêiner digital amplamente utilizado para armazenamento e transmissão de vídeo e áudio, compatível com diferentes codecs e dispositivos.

MPEG-H: Padrão de codificação de áudio desenvolvido pela Fraunhofer Institute que suporta áudio imersivo, personalizado e interativo, sendo parte da especificação do Fórum SBTVD para o futuro da TV digital no Brasil.

MXF (Material Exchange Format): Formato contêiner de arquivo padrão para intercâmbio de material audiovisual profissional, com suporte a múltiplas trilhas de áudio, vídeo e metadados.

Near Field Audio: Técnica de reprodução de som voltada para curtas distâncias, como em soundbars ou monitores de estúdio, otimizando a percepção espacial mesmo em ambientes pequenos.

PCM (Pulse Code Modulation): Forma não comprimida de representação digital de sinais de áudio analógicos, com alta fidelidade, usada em CDs, transmissões e arquivos WAV.

Pro Tools: Software DAW amplamente utilizado em pós-produção audiovisual, compatível com tecnologias de mixagem imersiva como Dolby Atmos.

Reaper: Software DAW de código proprietário com alta flexibilidade de roteamento e personalização, utilizado em produções independentes e também em workflows com áudio imersivo.

Room Calibration: Processo de calibração acústica de um ambiente com base nas características físicas da sala e dos alto-falantes, para otimizar a reprodução de som. Pode ser automático em soundbars ou feito manualmente com microfones de referência.

Soundbar: Equipamento compacto de reprodução sonora com múltiplos alto-falantes embutidos, projetado para simular som surround ou imersivo em ambientes domésticos.

Streaming: Forma de transmissão de conteúdo audiovisual pela internet, sem necessidade de download completo do arquivo. Plataformas como Netflix, Amazon Prime e Disney+ distribuem conteúdo com áudio imersivo.

Waveform: Representação visual da forma de onda de um sinal de áudio. Auxilia na edição e análise técnica durante a pós-produção.

ANEXO A - ENTREVISTA ARIEL FERRANTE

Raphaella: Vou começar perguntando da formação profissional e de que forma você começou a trabalhar com mixagem e *sound design*?

Ariel: Na realidade, eu venho de uma formação acadêmica até diferente dessa área. De formação, eu me formei em Geografia, sou geógrafo. Mas eu sou filho de músicos, meu pai é músico profissional mesmo, vive disso. Minha mãe era cantora, parou de cantar, mas eu cresci muito envolvido no meio da música. Estudei piano, estudei música desde criança, aquela coisa.

E depois acabei entrando na faculdade de Geografia, fiz o curso todo e lá no final do curso, eu tinha um interesse assim pela área técnica e eu fui fazer um curso de técnico, um curso técnico de áudio, que é um curso de fundamentos de áudio, de acústica no IAV de São Paulo, Instituto de Áudio e Vídeo, uma escola de áudio que tem na Vila Mariana, São Paulo.

Na época era uma das poucas que tinham um curso assim presencial, mais completo assim, é um curso de um semestre, dura um semestre. Você tem aula todo dia, umas três horas por dia, uma carga horária razoável, era o único que tinha assim no Brasil. Então, quando eu fiz esse curso do IAV, realmente eu me identifiquei com essa coisa da parte técnica, da parte operacional e fui trabalhar com música, então antes de trabalhar com cinema, no audiovisual assim, a minha porta de entrada foi pela música. Aí assim que eu acabei esse curso, eu fui trabalhar no estúdio de Estagiário. Um complexo de estúdios grande, em São Paulo, que chamava estúdios Mega. Tinha estúdios Mega no Rio e os estúdios MEGA em São Paulo. Sou de São Paulo, comecei no estúdio mega de São Paulo na parte de música, de gravar artista, banda, de tudo que é tipo.

E esse estúdio, ele era realmente um complexo de estúdios, tinha uma parte de finalização para o cinema. Era uma finalizadora, uma pós-produtora, então tinha a parte de imagem, de correção de cor de telecinagem, ainda tinha bastante película. E tinha lá a parte de finalização de som para cinema e dentro do Mega, quando eu estava lá, eu fiquei um tempo, acabei migrando, tendo um contato maior com áudio para o cinema, para o audiovisual, não é? Então, assim, até então eu vivia muito o áudio, a mixagem para música, para a produção fonográfica, e aí comecei a ter contato lá no Mega com o audiovisual, a parte do cinema, televisão. Inicialmente a contragosto assim era muito jovem, queria trabalhar com música e o meu chefe lá é um cara mais experiente, uma figura saudosa, falou “olha, esses grandes

estúdios de música, eles estão acabando. Tem a crise do mercado fonográfico, aquela crise lá dos anos 2000, o futuro é audiovisual e tal e eu bem resistente no início assim, mas comecei a ter contato e comecei a me dividir: era assistente no estúdio de música e comecei a ser assistente do estúdio de cinema. E aí eu comecei a me identificar comecei a me aproximar do mundo do audiovisual do mundo do cinema por aí, sabe? Então eu não tenho uma formação em cinema, não fiz faculdade de cinema ou de audiovisual. Mas a porta de entrada foi pelo mundo da música e ainda o Mega eu comecei a fazer um pouquinho de cada coisa, da pós de som, de cinema, de filme. Então, fiz um pouquinho de tudo, mas fui me identificando, o caminho que eu fui gostando foi da mixagem, a última etapa do processo, cronologicamente falando. Então foi um pouco isso assim, Raphaela, aí que eu me estabeleci ali, comecei a me identificar mais como mixagem, a focar mais, botar mais energia nisso. Ainda nesse lugar nos estúdios Mega que é, infelizmente está faliu, fechou, não existe mais. Então, foi o meu caminho.

Raphaela: Sim, e isso foi mais ou menos assim. Em que ano que você começou?

Ariel: Eu comecei a trabalhar no Mega em. 2007. E essa migração para o audiovisual começou lentamente em 2008, 2009 mas eu sinto que eu, realmente, tomei esse caminho de falar que é só isso que eu vou fazer, vou me especializar nessa parada em 2010. Então assim, eu comecei a trabalhar no meio, em 2007, como técnico, nessa área técnica, de áudio, mas eu sinto que eu realmente peguei um assim, vamos dizer, resolvi focar muito em me especializar em uma coisa ali para 2010, assim foi mais ou menos aí...

Raphaela: A partir do momento que você tem que começar a trabalhar com isso, com a mixagem, como que percebe essa ideia da mixagem para objetos? Assim, essa ideia de espacialização sonora, essa ideia é de utilização de ângulos, né? Mais do que só de canais. Essas coisas assim. Foi mais ou menos em que época ou ainda está acontecendo?

Ariel: É, eu acho que já não é de hoje, mas ainda está acontecendo. Porque quando eu comecei a ter contato com esse universo, o padrão que reinava ali, e, aliás, ainda reina, era o 5.1. Então você tem ali uma espacialização, não é? Você consegue distribuir a energia de um objeto sonoro não é de um elemento sonoro, você consegue distribuir energia dele entre os 5 canais e de alguma forma se você quer radicalmente em uma coisa ou você quer mais ou menos entre ali e ali, não tem tanta precisão. Mas é o padrão que dominava, e ainda domina é o 5.1.

E quando a Dolby veio com Dolby Atmos para cinema, salvo engano, 2012, na época ali do *Hobbit*, do *Gravidade*, que a gente começou a ouvir esse tema, pelo menos para mim, que foi uma novidade ouvir sobre mixagem orientada a objetos, né? O *render* lá da Dolby veio com aquele *render* que tem as bolinhas, que talvez até traduza melhor a ideia do objeto, porque você posiciona o objeto no espaço e daí o renderizador calcula dependendo do tamanho da sala, onde e o quanto de energia vai sair de cada caixa, para você localizar o objeto ali, a imagem sonora do objeto. Então aí eu comecei a ouvir esse termo, com Dolby Atmos, eu acho que é 2012, 2013. O que eu sinto assim que aconteceu é que o Atmos era primeiro uma coisa só para cinema, né? E depois foi tendo essa coisa com que eles foram transportando, né pro home vídeo, para o Broadcast. Então eu acho que a gente foi ganhando intimidade com isso, todo mundo começou a botar os olhos nisso. Talvez em 2018, 19 não sei, assim um pouco antes da pandemia, teve um certo momento assim de todo mundo falar “nossa, agora só vamos mixar assim. Quem não mixar dessa forma não vai ter mais trabalho.”, sabe... “Ah, se você quer produzir conteúdo para Netflix? Você só vai fazer assim porque eles não vão aceitar de outro jeito.” Teve um certo exagero do tipo “a o 5.1 morreu, o 5.1 acabou”, mas que não se concretizou, mas nem de longe e todo um hype em torno. Legal, o negócio está andando mesmo, está evoluindo, a gente está cada vez mais em contato com, mas nem de longe chegou nisso. 90% das coisas que eu faço é finalização em 5.1, seja para cinema, seja para plataforma.

O pessoal de música, eu conheço alguma coisa da galera de música também, continua 90% mixando do stereo. Teve o Dolby Atmos Music, é legal, todo mundo empolgado, parece que é o futuro, parece que é o que vai dominar, mas esse esse pânico que a galera ficou 5, 7 atrás, não é bem isso. Tem muitos projetos que eu orço as duas coisas, e falam: “eu quero fazer uma mixagem 5.1. Ah e se eu quiser fazer um Dolby Atmos?” e na maioria, ainda falam pra ir no 5.1 mesmo. Isso vem do cliente, né? Do cliente final assim da plataforma, seja Disney, seja Netflix. Está ganhando espaço, sabe? Está crescendo mesmo. realmente eu acho que é não é uma coisa que vai parar, é uma coisa que vai dominar, mas não é com aquele afinco, com aquele desespero que pareceu que seria, sabe? Eu, na minha experiência pessoal, posso estar totalmente errado.

Raphaela: Cada vez mais uma estratégia comercial né, um pouco uma coisa envolvendo a criação mesmo, né? Assim é muita empolgação e poucos resultados finais.

Ariel: É exatamente assim meio que esse lance do cinema, do entretenimento evoluiu muito assim, né? Ah, vamos inventar uma parada para a galera ir ao cinema. Aí o *home video*, o

consumo em casa chegava junto. Ah, agora tem um negócio diferente, Mano, trazer um pouco para o cinema, eu acho que nem um pouco desse jogo comercial. Ah, as plataformas tão dominando o mundo, o cinema vai morrer? Como é que a gente vai trazer o público? Não é?

Raphaela: E hoje, quais são as plataformas de mixagem orientada a objetos ou sistemas, não sei como que você prefere chamar, que você utiliza assim o dolby atmos e tal, mas que mais que você trabalha? Você tinha comentado comigo sobre o MPEG-H também. Aí se já quiser emendar falando sobre o MPEG-H e como está esse seu envolvimento da difusão do formato? Eu cheguei a ler algumas coisas depois que a gente teve aquela conversa inicial, mas se puder aprofundar aqui comigo, também vai ser bom para usar.

Ariel: Então Raphaela, é assim o que dominou o mercado de pós-produção, vamos dizer assim, é filmes e séries é o Dolby Atmos e ponto. Não tem nem sombra de nenhum outro que chegue perto. Existem outros padrões, você está pesquisando, deve saber melhor que eu, mas o Dolby Atmos dominou. Arrisco dizer que também 90% não sei te falar, uma porcentagem é mais uma esmagadora maioria trabalha em uma estação de trabalho digital que é *ProTools*. Na Europa tem bastante um concorrente bem forte do *ProTools*, que é o *Nuendo*, da *Steinberg*, que é o fabricante alemão. Inicialmente para fazer o Dolby Atmos, você precisava de um software externo à estação, então você tinha *ProTools* como a sua estação e um renderizador e processador da Adobe externo, que era o *Dolby Atmos Renderer*, renderizador do Dolby Atmos. Seja em qual plataforma você tivesse, seja *ProTools*, seja *Nuendo*, ou seja qual for, você tinha esse renderizador externo e já nos últimos anos, as próprias fabricantes integraram esse renderizador dentro da sua plataforma dentro do seu software. *Nuendo* foi o primeiro, a *Steinberg* já integrou o renderizador do atmos dentro do *Nuendo*, eu não sei te falar desde quando 2020, talvez 2021 e no ano passado ou retrasado a *Avid*, fabricante do *ProTools*, também integrou o render da Dolby dentro do *ProTools*. Então, a plataforma que a galera de filme, de pós-produção para filme, para série usa, não plataforma, mas a tecnologia é o Dolby Atmos, isso sem sombra de dúvida, qualquer outra que existe, existem outras, mas são muito incipientes, estão muito longe desse tamanho de mercado que a dolby tem. A Dolby é uma gigante de mercado e muito bons de marketing, vamos dizer assim. E eles são muito grandes, eles têm um histórico muito grande em som pra filme né. A cerimônia do Oscar é no *Dolby Theater* em Los Angeles, olha o tamanho disso. E aí, o que acontece quando eu estou quando eu trabalho para uma série original Disney, Disney Plus é e eles pedem Dolby Atmos, isso tem sido o normal.

A outra tecnologia que eu estou muito em contato e eu desconhecia até 2019, tomei contato com ela em 2019. E é uma coisa que também eu acho que está crescendo é, mas eu não vou nem dizer que é um concorrente, eu acho que são coisas diferentes, é o MPEG-H. O MPEG-H, é desenvolvido pela Fraunhofer, que é um instituto de pesquisa na Alemanha que tem vários institutos, né? Vários braços lá dentro e um deles é o Instituto de Circuitos Integrados Fraunhofer IIS, a sigla em alemão. E eles pesquisam tecnologias de codificação de mídia em geral, mas acho que eles são célebres porque há 25 anos ou um pouco mais eles criaram o MP3, então que é um *codec* que mudou o mundo, né? Basicamente mudou a indústria fonográfica, indústria musical, então o MP3 foi criado pela Fraunhofer e depois disso, várias outras maneiras de codificar áudio e vídeo também, né? O h 264 é deles, o h 265, hoje em dia o AAC, o *enhanced AAC*, que é o AAC melhorado, um monte de dispositivos ao redor do mundo usam essas tecnologias, né? Emissora transmite em AAC, emissora de TV digital, os nossos celulares, né? A gente toca a grande maioria em AAC.

E eu tive contato com essa tecnologia da Fraunhofer que é o MPEG-H em 2019 e é uma tecnologia que eles definem como uma mixagem com objetos. E aí eu fui em contato melhor trabalhando com eles desde então, uma relação próxima com eles. O que muita gente vê como concorrente assim do Dolby Atmos, mas eu não acho. Acho que realmente são universos, são objetivos diferentes. O MPEG-H é um *codec* de áudio desenvolvido pela Fraunhofer com o olhar para a emissora de TV para *broadcasting*, né? É tanto para conteúdo ao vivo, quanto para o conteúdo preparado não é pós-produzido, que vai ao ar, mas dentro do universo de TV. Qual é o grande lance do MPEG-H? Ele permite áudio imersivo, é um *codec* de áudio imersivo, ou seja, você também tem vários formatos, tipo Atmos que têm 5.1.4, 7.1.4, 9.1.6, tem várias configurações possíveis de áudio imersivo.

Mas o MPEG-H usa é objetos de áudio e eles têm até uma conceituação de objeto, um pouquinho diferente da Dolby para prover interatividade para oferecer possibilidade do usuário final interagir e fazer escolhas no conteúdo de áudio. E de uma forma muito prática, muito simplificada, que eu acho que é o grande trunfo deles para emissoras de TV. Porque quando a gente está falando de uma série da Netflix que vai ser mixada em Atmos, a gente está falando de um processo de pós-produção que você tem muito tempo, é então uma série de ficção complexa eu tenho 2, 3, 4 semanas para trabalhar em um episódio de 40 minutos, eu com uma equipe de 5, 6, 8 pessoas. Então, né? O Dolby Atmos exige uma complexidade que você consegue fazer ali. Conteúdo para a televisão, não tem esse tempo, a televisão está no ar todo dia, não é? Então o cara tem, sei lá, quantos jornais no ar, eventos esportivos,

reportagens televisivas que são gravadas antes mas que os fluxos de produção são muito intensos. E aí que a Fraunhofer tem essa proximidade com as emissoras de TV, porque esse codec que ele é muito prático e muito objetivo para esse universo da emissora de televisão. Então, ele permite fazer áudio imersivo igual o Dolby Atmos, então você tem não somente o *surround*, que está em volta, tem a possibilidade da camada superior, mas o MPEG-H usa os objetos na realidade para oferecer interatividade, para o usuário poder interagir que num filme que você vai ver no cinema, não é um apelo, mas num conteúdo televisivo, acho que o melhor que o MPEG-H traz assim é, por exemplo, num evento esportivo, um futebol no Brasil, se produtor do conteúdo quiser oferecer isso para o espectador, o espectador pode desligar o narrador só quero ouvir a atmosfera do estádio e desligar a narração. Ah, é o sei lá. É um é um clássico lá, um São Paulo e Palmeiras. “Ah, eu só quero ouvir a torcida do Palmeiras. Eu só quero ver a torcida do São Paulo”.

Ou se eu preciso de audiodescrição, eu vou só ligar audiodescrição ou desligar audiodescrição, se eu tiver uma pessoa com necessidade visual. Isso tudo de uma forma muito simplificada. Hoje em dia, as emissoras, quando elas vão transmitir um conteúdo com a audiodescrição, elas precisam transmitir a mesma mixagem 2 vezes, vai a mixagem normal e em outra banda vai de novo a mixagem com a Audiodescrição impressa nela, gravada. Então isso é ineficiente, né? Isso dá mais trabalho e gasta banda, né? Gasta espaço de transmissão da emissora, ou seja, a emissora tem que diminuir a qualidade, quanto que o áudio está ocupando, ela tem que perder qualidade para poder caber naquela banda 2 vezes a mesma coisa. Então, por exemplo, no MPEG-H, usando o exemplo do MPEG-H é uma transmissão só e essa audiodescrição está desligada e tem metadados que estão ali presentes que chegam no usuário, se o usuário aciona, ele entra com aquele canal ali. Então, em vez de transmitir tudo várias vezes, está transmitindo um pacote único, uma vez só, e o usuário vai interagir. Esses exemplos que eu dei, desliga o narrador ou liga audiodescrição, tem vários tipos de interatividade possíveis e está tudo já está sendo transmitido no mesmo pacote descrito com metadados, o receptor recebe ele lê esses metadados e o usuário tem esse controle. Então o MPEG-H. É muito eficiente, muito prático para uma emissora de TV.

E eles usam esse termo de mixagem orientada a objetos Fraunhofer faz questão de falar *Real Objects*, objetos reais, porque é uma concepção diferente do que foi usado ali no Dolby Atmos, mas nesse sentido de que numa cena de MPEG-H, numa mixagem de MPEG-H, por exemplo, se você está transmitindo um documentário jornalístico você tem a cama, a *bed*, que eles falam inglês, a cama, a base de canais lá que está com tudo o que você quis mixar dentro do

seu universo e você tem um objeto que é a voz, então toda a mixagem de vozes, se ele for falado em português, por exemplo, você tem um objeto lá que é a voz em português. E esse objeto, ele vai separado dentro do mesmo pacote, mas separado da cama, com metadados que explicam para o receptor “ó esse canal aqui é um objeto chamado diálogo em português” e você pode mandar um também, que é em espanhol e um em inglês. E o usuário chaveia, troca o que ele está querendo de forma muito fácil e é isso, sem você repetir tudo. Não é igual eu estava dando um exemplo, não é o conteúdo que era muito comum, até um tempo atrás, acho que ainda existe, o conteúdo dublado de um filme e aí tinha a tecla SAP e apertava a tecla SAP e mudava para a língua original. Então é isso que eu estou te falando, eram 2 vezes o mesmo negócio sendo transmitido, dando espaço. E quando pega MPEG-H, você tem uma transmissão única e você vai chaviando, vai *switching*, que eles falam, alternando esses objetos. Esse seria meio que um funcionamento básico do MPEG-H.

E no momento, por que que eu estou trabalhando com isso, por que eu estou tendo contato com isso? Desde 2020, está em curso no Brasil um projeto que é o da TV 3.0, não sei se você chegou a ver alguma coisa disso. Acho que você vai lembrar que 10 ou 15 anos atrás, acho que era o Governo Dilma teve a mudança para a TV digital no Brasil. Eu lembro que teve toda uma campanha em torno disso, o Brasil ia escolher lá tal tecnologia, que era uma tecnologia do Japão. E as TVs analógicas lentamente, a transmissão analógica seria desligada lentamente, que até agora está terminando de desligar para, aos poucos, ir tendo Transmissão digital. Começaram a vender conversor de TV digital, umas caixinhas. Então essa TV digital que foi feita lá por 2010 não sei, foi chamada de TV 2.0. Então a gente tinha TV analógica desde os anos 50 e aí veio a TV 2.0, que era essa virada para o digital, tanto de áudio quanto de vídeo quanto de tecnologias envolvidas. Nos últimos anos, o Brasil está discutindo e todos os países do mundo estão, cada qual no seu tempo, mas é um movimento mais ou menos uníssono de uma nova virada de tecnologia porque o áudio digital e o vídeo digital mudaram muito, as câmeras, tudo mudou muito. A gente tem tecnologias muito recentes, então vai ter uma virada de tecnologia, que está sendo chamado TV 3.0, em vários países está acontecendo. Isso não é uma exclusividade do Brasil.

E quem está discutindo isso? É um fórum, chamado Fórum SBTVD que é fórum do sistema brasileiro de televisão digital, um fórum com um composto bem diverso, composto por representantes das emissoras, representantes dos fabricantes de equipamento, dos fabricantes de *software* de tecnologia, representantes do Governo e é um fórum que veio debatendo quais seriam então as tecnologias adotadas para essa TV 3.0. Ah, como é que vai codificar o vídeo?

Tem meia dúzia de gente aí oferecendo codec de vídeo, então, qual que é o mais eficiente? O que que o Brasil precisa? O que que vai funcionar num território desse tamanho ou o que que vai impactar mais ou menos na qualidade do que é visto na tela? São discussões muito complexas, que entram na engenharia de comunicação, de *broadcasting* e eu não tenho domínio. Mas foi esse movimento aí de mudança de todas as camadas, eles chamam, camada de vídeo, camada de áudio, camada de closed captions, a camada física de transmissão mesmo, tipo de antena, tipo de recetor, como vão reposicionar antenas e fazer as transmissões em cidade grande, em zona rural? É uma discussão longa, complexa, mas indo para o para a nossa parte, eles fizeram, depois de muita discussão, foi redigido coletivamente um documento de alta complexidade técnica assim, da parte de engenharia de telecomunicações, redigido tudo em inglês, porque é para o mercado Internacional, não é só para dentro de casa, então era tudo traduzido que era uma chamada de propostas, *call for proposal*, que era assim “olha quem mundo afora quiser oferecer a tecnologia pro Brasil codificar áudio, está aqui as nossas exigências”. Então eram 15 exigências das quais tinham subitens, assim, no item um tinha ABCDE item 2. ABC... É, por exemplo, para codificação de áudio, vou só falar do áudio, mas isso você pode aplicar para todos, para vídeo, todo mundo passou por isso. E no *call for proposal*, eles tinham essas exigências assim, coisas que eram exigências para você ter um áudio com boa qualidade, tem que ter eficiência de codificação, poder comprimir o máximo possível sem perder tanta qualidade, tinha que ser áudio imersivo, tinha que oferecer interatividade, tinha coisas que eram indispensáveis, tipo que conseguir preservar sincronismo entre áudio e vídeo, coisas óbvias, e tinham coisas que caíram na categoria *desirable*, desejável, então, “ó isso aqui não é uma exigência, mas é desejável que tal coisa”, aí fizeram esse documento extremamente complexo e abriram em todas as áreas para ver quem se candidatava e realmente, na parte de áudio, houve 2 candidatos, um terceiro que foi só teórico. Mas houve 2 candidatos, 2 *codecs* se candidataram a oferecer tecnologia que cumprisse aqueles requisitos no Brasil. Um foi o MPEG-H da Fraunhofer e o outro foi AC-4 da Dolby, que é o nome do *codec* que comercialmente estão chamando de Dolby Atmos. Na realidade, Atmos é um produto da Dolby que em cada plataforma é uma tecnologia diferente, para cinema é de um jeito, o que eles estão usando agora no *streaming* ainda não é a AC-4 é o AC-3, que é o dolby digital antigo melhorado, AC-3 com os objetos adicionais. Enfim, aí eu também não posso entrar muito na área que eu estou fica até indiscreto também se eu fosse falar, porque nesse caso eu me envolvi muito com o outro lado. Mas enfim a Dolby, propôs o uso do AC-4 comumente chamado de Dolby Atmos e a Fraunhofer propôs o uso do MPEG-H. E aí depois desse *call for proposal*, entrou uma fase de testes que várias universidades do

Brasil participaram, por exemplo, a camada de vídeo foi analisada pelo laboratório da UNB, liderado pela professora Débora. As camadas de aplicativos foram analisadas metade aqui no federal de Juiz de Fora e outra metade acho que na PUC Rio de Janeiro, liderado pelo Prof. Marcelo Moreno. E a camada de áudio foi analisada na USP, na ECA, no laboratório lá de pesquisa da ECA, liderada por 2 professores pelo Professor Régis Alves Faria, se não me engano, e pelo Professor Almir Alvares. E essa fase de testes durou mais ou menos 6 meses e aí realmente não tinha nenhuma participação das emissoras, dos fabricantes, dos interessados. Quem foi analisar isso foram as universidades com esse documento em mãos, uma chamada assim. Número 1: Ah, o áudio tem que estar em *sync* com a imagem, coisas óbvias, eles foram analisando e testando todas as hipóteses. Realmente eu estou dando um exemplo muito singelo, porque as coisas tinham níveis bem complexos de engenharia. E depois dessa fase de testes, chegaram-se a conclusões, então todas as universidades que estavam fazendo testes, os órgãos estavam envolvidos deram seus resultados para o fórum e isso não era um negócio, vamos dizer, normativo obrigatório, era um negócio de aconselhamento, mas a escolha seria do fórum seria dos representantes da indústria. No caso do áudio, que é o que nos interessa, foi unânime a escolha do MPEG-H. O MPEG-H cumpriu todos os requisitos, o Dolby AC-4 não conseguiu cumprir a maioria. Então, no sentido de votantes foi catorze a zero, então todos votaram por isso.

Então estou te resumindo essa história que me levou assim 5 anos aí de discussões com pandemia no meio. Esse fórum que eu te falei, o SBTVD ele é um órgão que aconselha o Governo, ele apresentou, falou, olha, por A mais B os estudos apontam isso e todos nós, representantes da indústria, seja Globo, SBT, todas as emissoras, todos os fabricantes, LG, Samsung, outras e as empresas, Dolby, Fraunhofer, todo mundo envolvido, fala que essa é a decisão do Fórum: “A gente acha que no Brasil TV 3.0 tem que ter um vídeo de tal jeito, o áudio de tal jeito, os *closed captions* de tal jeito. Então, a tendência é que o Governo vá adotar isso. Então assim estou dando essa volta para te falar que é o MPEG-H, foi escolhido aqui como o *codec* obrigatório nas transmissões de TV aberta no Brasil e nas transmissões por Internet, por *streaming*, aí é opcional, o produtor também pode querer usar o AC4, mas na TV aberta no Brasil a partir desse ano do ano que vem, quando for ao ar oficialmente a primeira vez a TV 3.0 é o MPEG-H vai ser o código usado por todas as emissoras. Ponto. Não é uma escolha mais, a escolha foi feita.

Então, até estou falando isso para te falar da relevância do MPEG-H que eu acho que ele vai ter no Brasil, fora do universo cinema, série de ficção, mas nesse dia a dia das emissoras de

produzir conteúdo jornalístico, produzir esportes, de produzir essas coisas do dia a dia de uma emissora, até novelas, a transmissão vai ter que ser feita em MPEG-H. Então eu tenho estado muito envolvido com isso, me tornei consultor da Fraunhofer no Brasil, eu e mais 2 colegas no Brasil, então a gente está muito envolvido nos últimos tempos em dar treinamento para os sonoplastas das emissoras, porque é vai ter uma virada de tecnologia, a gente sabe que as coisas levam tempo. Você imagina que ninguém quer gastar dinheiro se não for obrigado a gastar dinheiro, então as emissoras ficam super com freio de mão puxado, fica ali um jogo de quem veio primeiro o ovo ou a galinha, tipo assim, o fabricante não quer implementar o codec porque fala que não tem nenhuma emissora transmitindo e a emissora quer botar no ar porque fala que ninguém tem uma televisão que vai ler esse conteúdo. Fica ali um jogando por outro, porque vai envolver um custo, mas é inevitável, o Governo ano passado já teve alguns eventos que o Ministro das Comunicações deu fez discurso em um evento ano passado em Brasília e deixou claro que o Presidente Lula está apoiando é que quer. Tinha uma meta de que o brasileiro já pudesse ver a Copa do Mundo do ano que vem com essa nova televisão no ar, claro que a gente sabe que isso é muito simbólico, são 10 anos, serão mais 10 anos de igual foi com a TV digital. Mas a gente está nesse momento, quando pega a galera, a gente está dando treinamento para quem trabalha em emissora, para eles começarem a produzir, entender o funcionamento, não quer dizer que do dia para a noite vai estar todo mundo produzindo áudio imersivo 3D. Muito provavelmente eles vão continuar estéreo também, um pouco, mas com a diferença de que vão adicionar esses objetos, que no MPEG-H é chamado objeto, para oferecer interatividade. Então, mesmo conteúdo estéreo, mas que de repente as pessoas em casa comecem a ter televisores ou tablet, celular que decodifica MPEG-H, mesmo estéreo a pessoa já vai ter ali uma opção de desligar uma voz, de trocar a língua, de adicionar um audiodescrição, uma coisa que é tentadora para todo mundo trabalhar em televisão, seria um boost, né? Um ganho a mais nos diálogos, então a ideia também numa novela é oferecer a novela normal, mas quem quiser tem um botãozinho assim, mais voz, aumentar a voz, sei lá o nome que a gente vai dar para isso, cada emissora vai escolher o seu nome. E aí a pessoa aperta aquilo e todos os diálogos se destacam, ficam, sei lá, 8 dBs mais alto. Então, para quem tem dificuldade auditiva ou para quem consome hoje em dia essas coisas no metrô, no trem, no ônibus ou em casa com a janela aberta, tem cachorro do vizinho está latindo... O som consumido na televisão está concorrendo com o mundo real, né? Então você está concorrendo ali com a panela de pressão, fogo, um cachorro latiu, concorrendo com outras telas, né? Todo mundo com o celular na mão, o tempo inteiro olhando pro celular. Então, existe essa preocupação, é sempre uma preocupação do entendimento dos diálogos, da inteligibilidade do

que está sendo dito na televisão e o MPEG-H tem essa facilidade, né? Igual eu te falei. Lógico que numa novela não vai ter a opção de desligar a voz do personagem, mas eu tenho uma opção de aumentar todas as vozes e aí de repente um degrauzinho para cima. Então, é isso assim tentei te chamar um resumo assim do processo não é? Mas as 2 tecnologias que eu tenho contato são essas e não vejo de forma alguma como concorrentes. Eu acho que são universos distintos.

Filme série eu acho que vai continuar sendo o Dolby Atmos. Eles estão muito consolidados nisso e funcionam muito bem, agora para o mundo do Broadcast mesmo, Over the air, que eles falam Transmissão pelo ar, não é? TV aberta é o MPEG-H é uma realidade. Embora não seja assim comercialmente, não é uma coisa que você vai ouvir falar toda hora, não tem esse apelo. Nenhum apelo, eles são péssimos nisso, mas é uma realidade que a gente está ali já trabalhando com as emissoras.

Inclusive em São Paulo, você tem a TV Cultura e a Globo, transmitindo já em MPEG-H de forma experimental, mas está transmitindo.

Raphaela: E aparelho para isso? Já tem no mercado?

Ariel: Pois é, essa é a luta que a gente está, está muito perto de ter. A LG e a Samsung já têm protótipos. Acontece, na verdade, é o que eu te falei, o fabricante protela ao máximo implementar, embarcar uma nova tecnologia. A TV, porque aquilo gera algum custo para ele, não quer dizer que agora eles vão gastar mais dinheiro e tal, porque é normal toda a tecnologia que está embarcada, paga royalties para o dono, é onde o pessoal ganha. Então assim, o fabricante não tem que comprar nada da Fraunhofer, ou eu como usuário, não tenho que comprar nada da Fraunhofer, mas se lá dentro tem um decodificador do MPEG-H, eles precisam implementar aquele código e pagam *royalties* lá de centavos pra Fraunhofer ou para Dolby para usar o *codec*, o decodificador do Atmos. Isso lá em escala de milhões de dispositivos.

O MPEG-H foi escolhido, igual eu te falei, como único *codec* de áudio também na Coreia do Sul, em 2017. Então, desde 2017, os televisores fabricados lá, ou seja, a gente está falando dos 2 gigantes que são Samsung, LG, os maiores, eles já têm MPEG-H. Então assim, a gente não tem acesso a essa informação diretamente, mas existe aí um entendimento de que se eles quiserem, eles conseguem desbloquear isso nos televisores presentes já no mercado via atualização de *firmware*.

Raphaella: Como se já tivesse ali, e precisasse apenas de um desbloqueio, é mas provavelmente vão fazer com que a gente troque.

Ariel: Provavelmente, ele já tem protótipos, na SET do ano passado, sabe a SET, aquela feira que tem em Agosto, de tecnologia. Na SET, a gente estava com uma TV da LG, uma da Samsung e uma da Hisense, que é o fabricante chinês, que está entrando no Brasil já com MPEG-H, tocando em MPEG-H. Eu estava sintonizando já que estava no ar pela TV Cultura, decodificando em MPEG-H. Então assim, está muito perto de ter aparelho no mercado, acho que a Hisense até já está vendendo no mercado no Brasil. A Samsung e LG ainda não, mas assim está muito perto, muito perto de ter, claro que é igual a outra virada, você não troca de TV a cada 6 meses, né? Então, vai levar alguns anos para chegar no momento que você fala, poxa, vou trocar de TV. Acho que no meio disso vai ter essas caixinhas, essas set top boxes que já decodificam, já tem algumas. Ou quando for conteúdo que você vai consumir com o celular com o tablet aí é via software, você não precisa trocar de aparelho, é uma atualização no APP da Globoplay vai atualizar o aplicativo, ele vai estar decodificando o MPEG-H, então a gente está realmente nessa nesse momento de viradas, eu acho que esse ano fora Hisense, talvez ainda em 2025 já possa ter coisa no mercado.

Raphaella: E só para retomar aqui, para deixar bem claro é essa ideia de objetos em MPEG-H seria mais pensando em escolhas do usuário, assim como aumentar a voz, desligar o narrador igual você explicou, do que uma ideia de ângulo sonoro mesmo não é? Principalmente porque a gente está trabalhando para TV, para dentro de casa, é isso?

Ariel: É isso, exatamente. Mesmo fazendo uma mixagem imersiva, é no caso do MPEG-H, você pode fazer o Pan, né? Mudar o panorama do objeto, de um áudio, de um som. Então você pode passar para a camada superior fazer passando por aqui por ali dentro de “paneador” 3D, qualquer um que faça isso e só pelo fato de um elemento sonoro estar lá, um helicóptero passando em cima de você, isso não quer dizer que ele virou um objeto na concepção do MPEG-H, realmente o que está descrito como objeto quando você vai empacotar uma cena em MPEG, vai fazer uma alteração em MPEG-H, o que vai estar descrito ali como objetos são esses elementos que você vai oferecer algum tipo de interatividade, seja aumentar, abaixar volume, seja ligado, desligado. Esses são os objetos, tá?

Raphaella: E aí a gente provavelmente não vai vir a o MPEG-H em *streaming*, igual a gente vê atmos, né?

Ariel: Provavelmente não. A gente está lutando para que veja, mas eu acho que talvez é no *streaming* da Record que se chama Play Plus, né? A Record tem um *streaming* próprio que está crescendo e no Globoplay eu imagino que a gente possa ver muitos conteúdos jornalísticos, não é? Nos conteúdos mais televisivos, né? Domingo Record, o Fantástico, sei lá, acho que esses conteúdos ou Transmissão de Jogo de futebol de fórmula 1 não é, Sei Lá, cara, quem curte fórmula 1, imagina, você pode escolher qual rádio, de qual equipe você vai ouvir “Ah, eu quero ouvir o rádio da Ferrari, isso aí para quem gosta, é incrível” Então, eu acho que nos *streamings* vai ser mais esse conteúdo televisivo, eu acho, estou dando um palpite aqui, puro achismo. Enquanto que essa coisa mais série e filme, eu acho que a tendência, acho também, não posso te dar certeza, é seguir esse caminho que já vem vindo, desde muitos anos do atmos. É um padrão já, né?

Raphaella: E aí, voltando mais pra ideia de *streaming*, mais pra Atmos, eu queria que você falasse um pouco sobre essas diferenças básicas de uma mixagem voltada para cinema e uma mixagem já voltada para *streaming* para dispositivos móveis, ambientes domésticos. Assim, e se quiser entrar um pouco também já na ideia do binaural. Se o uso de, por exemplo, objetos sonoros para *streaming* é uma realidade ou é só, por enquanto, uma grande possibilidade que que vem sendo vendida, né? O que você pensa disso?

Ariel: Raphaella, eu acho que é assim uma realidade. O que acontece? No Brasil, a gente tem pouca gente fazendo atmos para cinema, a maioria da nossa produção e do nosso parque de exibição é 5.1, mas independente disso mesmo se tivesse, eu acho que, como mixador agora eu estou falando, que vai e mixa e tal, eu sinto que a grande diferença não seria na abordagem de como eu iria trabalhar os objetos ou como eu iria espacializar as coisas. “Ah, se for cinema, eu vou fazer assim. Se for *streaming*, eu vou fazer assado”. Tem alguma diferença. Mas, eu acho que a grande diferença entre os dois não é nem de espacialidade, acho que é muito de range dinâmico, de possibilidade dinâmica, entendendo dinâmica como a diferença entre os altos e os baixos. Eu acho que é o que mais muda assim, né? Porque cada vez mais, e aí já não tem a ver nem com a Atmos, nem com áudio imersivo, acho que é uma coisa do mercado, porque essas telas, essas janelas de exibição estão muito próximas uma das outras, né? Há 10 anos atrás, eu mixava um filme para cinema, ele era lançado no cinema e passava meses ou quiçá um ano para o produtor falar “poxa, a gente vai passar na Globo. Vamos fazer uma masterização aqui para televisão, porque vai ter que tocar na televisão”. E mil coisas mudam, não é? Eu não vou entrar nisso, mas a principal coisa é o range dinâmico. A TV tem muito menos range dinâmico que o cinema. Então, eu tinha muito tempo entre uma janela e

outra, era muito diferente. Agora, Raphaela está muito misturado, muito sim. Eu preciso fazer um filme, por exemplo, tem acontecido o filme original da plataforma. Ele foi concebido o projeto desde sempre dentro da plataforma, um produto que vai para a plataforma, mas quando a gente está finalizando o produto, o produtor vai e fala “olha, a gente quer um DCP”, que é um pacote para cinema. “A gente quer uma cópia para fazer uma estreia em cinema”. Então eu já tenho que entregar um produto só praticamente, a gente dá uma mexida, lógico, mas já não é assim, é muito misturado, se eu fiquei ali 4 semanas mixando um filme para uma plataforma, imagina eu não posso cobrar mais uma semana para ter a versão cinema, se não eles não vão pagar, vão achar que eu estou querendo extorquir, é intolerável isso para eles, do ponto de vista de recurso, de dinheiro, de tempo também.

Então o que a gente está passando assim, é muito isso, é muito emendado uma coisa na outra. Eu entrego a master da plataforma e já entrego junto a do cinema ou mesmo que seja o contrário, o filme para cinema, eu finalizo ele para cinema e na semana seguinte, aconteceu isso recentemente, um filme grande que eu fiz, na semana seguinte, eles queriam uma saída para o Telecine, que era um dos patrocinadores e uma saída para a Netflix, porque dali x meses ia entrar na Netflix. E tudo ali de imediato. Então a gente não está tendo muito esse tempo de fazer muito diferente. Está tendo que dar uma homogeneizada nas coisas assim, o que não é bom, mas a gente é empurrado para isso, mas o que eu mais sinto de diferença sem entrar muito em detalhe, é range dinâmico, é possibilidade dinâmica que um filme de cinema tem que reprodução de cinema tem, que uma reprodução em televisão não vai ter, por mais que “Ah não é TV aberta, é *streaming* tudo bem”, mas a maneira como vai ser consumido, é no celular, no Tablet, com o fonezinho ou em casa numa televisão. As TV tão dessa espessura (pequena), os Alto-Falantes não conseguem reproduzir muito grave, muito médio, grave. Ou Como Eu te falei, você vai concorrer com o mundo real, o cara vai ver a TV com o celular tocando, com o Instagram aberto, com o cachorro latindo. Então eu acho que a grande diferença para mim é a dinâmica, de como colocar as coisas do ponto de vista de dinâmica de alto e baixo, até mais do que espacialidade, sabe? Agora, indo para o consumo mesmo que você falou até do binaural, eu acho que essas ferramentas novas, como o Atmos, é muito interessante a maneira como o codec calcula ali os as espacialidades. E na hora de reduzir isso, do Atmos que pode estar lá multicanal, multipista, multi objeto e isso é reduzido, isso é empacotado no binaural, no estéreo normal ou no binaural. Mas a maneira como isso se dá realmente é muito interessante. Então eu, eu acho, e muitos colegas meus estão fazendo isso, muitos projetos que às vezes nem são para ser Dolby Atmos, eles vão ter sua vida em 5.1 ou

mesmo em estéreo e a gente já mixa é Dolby Atmos, já uso o renderizador para fazer as espacialidades, porque quando ele faz esse re-render, essa re-renderização para binaural, realmente o resultado é muito legal. Você tem um estéreo rico ali, mais espacialidade para as coisas mais definidas espacialmente. Aí entra em termos subjetivos.

Eu entendo que o consumo vai ser esse, ninguém vai sair montando *Home Theaters* de áudio imersivo em casa, uma minoria que faz isso, a gente sabe. Eu acho que o consumo ainda vai ser fone de ouvido, *Tablet*, celular, ou mesmo em Casa, televisão estéreo, que é o consumo. No máximo soundbar, que é uma solução legal da indústria, a indústria viu que não deu certo a febre dos *Home Theater*, ninguém monta aquilo direito, um monte de fio pendurado. O pessoal comprava *Home Theater* 9.1 e botava as 9 caixas em frente a televisão. E aí, o soundbar é uma solução, um objeto único, bota ali na frente e acabou acho uma solução ótima.

Mas eu acho que o consumo vai ser estéreo. Então, realmente o binaural, essa redução não é quando a gente reduz a mix para estéreo usando esses algoritmos mais desses codecs mais novos, que é o caso do atmos, o caso do MPEG, o resultado é realmente legal, você tem um negócio mais rico, mais espacialidade. Eu estou eu arrisco também, tudo aqui no chute, tenho certeza de nada, mas eu arrisco dizer que isso é um futuro promissor.

Raphaella: A ideia de um que então se um mix foi feito em AC-4, atmos e ele vai se ajustar para diversos formatos de sala e tal. É real, então?

Ariel: É real, isso não é só marketing, é real. No MPEG-H também, um dos pilares que a gente chama no MPEG, é a entrega única, o delivery universal, que é o que a Dolby vende e é verdade. Você fez já no formato maior, vamos dizer no mais complexo e você vai tirando as reduções, os subprodutos dali. Isso é real e isso é muito prático para quem está produzindo e realmente entrega o negócio de qualidade. Bem legal. Isso não é só marketing.

Eu não sei se você chegou a ver um vídeo da Dolby quando eles lançaram o Atmos, 2012, que que eles davam exemplo, aquelas matrioskas, aquelas bonecas russas, sabe? É não é, é meio isso, não é marketing puro.

Raphaella: Bom saber, porque às vezes eu falo isso e eu mesmo vou me sentindo desconfortável porque é como se eu tivesse caindo no marketing. Parece muito prático, pra quem trabalha com isso, parece que isso é a grande ideia assim, de praticidade.

Ariel: É agora mesmo, eu estou finalizando um filme complexo, uma ficção bem maluca, assim, um filme de gênero, bem de terror assim, bem maluco. Não posso falar nada além disso, mas em Atmos, e ontem, a diretora me pediu uma saída de um trecho que ela estava em dúvida que ela queria dar uma olhada, estéreo. Então, é muito simples, eu abro o render falo “Ah, eu quero um output estéreo disso aqui”. Embora o filme tenha sido em Atmos, ele dá uma saída estéreo. É muito prático, muito prático.

Raphaela: Você tem algum exemplo assim de filme, principalmente desses, que já saem original Netflix original *streaming*, que você me indicaria assim como um bom filme que utilizou essa espacialidade, esses objetos, essas coisas.

Ariel: Nossa, não ocorre agora.

Raphaela: Não precisava responder agora...

Ariel: Se me ocorrer eu te falo... Assim um que eu acho que foi, mas o filme era feito para isso, foi ali a marca, a Bola na marca do Pênalti foi só o cara chutar é o Gravidade.

Então o Gravidade, se eu não me engano, você deve ter visto a história, o Alfonso Cuarón já tinha mixado em 5.1 e 7.1 e veio a notícia do Atmos e ele remixou o filme, e eu escutei em Atmos numa sala de cinema, foi a primeira sala de cinema com Atmos do Estado de São Paulo, que era em Sorocaba. E era um shopping novo, não é? Então, na época, só shoppings novos estavam tendo Atmos.

Raphaela: BH não tem Atmos até hoje, porque não tem shopping novo aqui.

Ariel: Então, se o cara não quer fechar o cinema, fazer uma reforma de 3 meses para instalar um novo sistema de som, eu entendo o dono do cinema.

Raphaela: Até o ano passado, Minas Gerais tinha uma sala Atmos em Montes Claros e uma outra em Uberlândia e só. (correção, na verdade é Uberaba). E BH nada. Parece que vai abrir um shopping agora, mas assim, eu ainda não sei se vai ser um shopping assim com cinema, minha grande esperança.

Ariel: Em São Paulo, mesmo hoje, São Paulo, que é uma coisa absurdamente grande, São Paulo tem 5 salas Atmos e, sei lá, São Paulo inteiro deve ter uns 900 cinemas, não sei, e tem 5 salas. Então, na época teve isso em Sorocaba, a gente estava em São Paulo, a gente pegou o carro e foi assistir Gravidade, então assistir o Gravidade em Atmos foi uma experiência

absurda, mas o filme é muito propício para isso, não é? O cara fica girando o filme inteiro, é pra isso.

Raphaela: Como meu trabalho tem a prerrogativa de uma análise física, apesar de eu brigar muito com isso, porque apesar de eu não estar envolvida nesta prática, eu sou uma pessoa muito da prática, né? Eu trabalho com técnico de audiovisual tudo, então sou muito dessa coisa e como não tem Atmos aqui para o cinema, a minha ideia é sempre voltada ao *streaming* e por isso eu quero escolher filmes originais das plataformas não filmes que foram para o cinema e do cinema para a plataforma. Eu até fiz um artigo sobre aquele Nada de novo no Front, porque tem muita cena de Chuva, né? Eles estão num bunker assim que é mais baixo e a guerra acontecendo em cima, então.

Ariel: Esse que é aquele alemão que ganhou o Óscar?

Raphaela: Isso é que ele já era a releitura de um filme, mais antigo, baseado num livro também. Então assim, se você tiver alguma ideia depois você me fala...

Ariel: Se me ocorrer eu te falo. Você sabe que um dos produtores desse filme é Brasileiro né?! É um menino brasileiro que mora nos Estados Unidos e tal, mas ele é brasileiro e começou a produzir coisas legais assim. Ele produziu também aquele “Sérgio”, aquele filme do Wagner Moura que ele fez um diplomata da ONU brasileiro foi assassinado quando foi intermediar um conflito no Timor-Leste.

Não lembro se foi no Timor ou já foi em outra missão, acho que já foi em outra missão que explodiram o lugar que ele estava, ele ficou nos escombros e morreu. Sérgio Vieira de Mello o nome dele.

Se Eu não me engano, esse produtor brasileiro que eu te falei, o Daniel, ele produziu esse filme também para Netflix antes do Nada de novo no Front.

Ele estava envolvido e é tipo coisa grande. Eu tenho ele em alguma rede social aqui, Daniel, também não lembro o sobrenome.

Raphaela: Mas aí assim, se algum dia qualquer te ocorreu um filme... Eu fiquei muito impactada com aquele A sociedade da neve, do avião lá que fica na cordilheira e tal. Gostei da perspectiva do filme mais do que os outros que documentaram a mesma história e tal. Tem um som bacana que tem muita nevasca naquele ambiente fechadão ali do avião. Então ele talvez seja uma escolha. Vou dar uma olhada nesse aqui.

Ariel: Sim e é latino-americano, legal não ser Hollywood.

Raphaela: É, eu queria fugir um pouco disso também, apesar de filme brasileiro é sempre uma questão que eu sou questionada em congressos “Ah, porque você não escolhe filme brasileiro” e eu falo, porque é difícil de encontrar um filme que é tá nesses padrões ainda, né? Assim é que.

Ariel: É isso é uma discussão que a gente tem assim. Ah a gente tem o Atmos. E o produto assim? Eu preciso de Atmos em uma comédia romântica na Globo filmes? É extremamente baseado no diálogo, a comédia brasileira, é muito falada, e você tem lá, humoristas geniais, que é fala, fala, fala, fala, fala. Você não precisa de Atmos num filme desse.

E é esse estouro de bilheteria, não é? Hassum foi um fenômeno, o Paulo Gustavo o cara é um gênio daquele negócio, mas nesse tipo de filme por exemplo, tem cabimento eu queria usar uma tecnologia dessa? Agora no “Ainda estou aqui”, eu preciso de Atmos para aquela história.?

E to falando do ponto de vista criativo, narrativo, a gente quer usar a tecnologia, é brinquedo novo. Mas, é isso que você tem que ter um produto que te permita, como O Sociedade da Neve tinha motivo para você ouvir e falar, cara que som, Sei Lá, os sentimentos que aquilo provoca. Assim, não é um preconceito contra o cinema nacional.

Raphaela: É e fica parecendo um pouco né. Principalmente nesses congressos que o pessoal é e tem que ser mesmo, mais protecionista e tal. Aí chegou um professor que ele falou assim: Abrace o Blockbuster. Esquece isso que as pessoas estão falando, a sua pesquisa ela tem que envolver Blockbuster, porque é onde vai chegar primeiro, né? É onde tem mais dinheiro é isso.

Raphaela: Fui ao cinema e assisti aquele Mickey 17, não é esse último do Bong-joon Ho.

Ariel: Sei, é legal?

Raphaela: Sim é, é bem a cara do diretor lá, né? Aquele monte de crítica lá social, mais futurista e eu vi na sala Imax que é o melhorzinho que a gente tem aqui. E eles especializaram diálogo. Eu achei muito doido porque assim é difícil, né? A gente vê produção com diálogo mesmo vindo lá de trás.

Teve 1 hora que eu mesmo caí e virei o rosto. Assim que um diálogo foi tão lá para trás que eu fiz assim, sabe é? E aí eu mesmo fiquei “Não acredito que eu fiz isso”, mas é reflexo, né? E eu achei muito doido.

É um filme futurista, né? Então ele pode brincar com coisas assim, que eu acho que o diretor também gosta. Achei doido o diálogo porque a gente não está acostumado com esse diálogo descentralizado que tinha.

Ariel: Não muito raro. A gente tem receio, às vezes dá vontade, mas assim, sempre tenho um receio de acontecer isso que aconteceu com você de tirar a pessoa da história da pessoa lá para trás e nessa logo se desconcentrar da narrativa. E tenho receio da qualidade da reprodução não é a gente não tem boas salas, na real. Temos boas salas, mas é minoria, a maioria do circuito comercial de cinema, e as condições de reprodução de som não são muito boas.

A gente tira o som do canal central e você botar uma informação relevante no canal *surround* esquerdo e simplesmente perder. A pessoa não vai ouvir, vai perder sentido do filme, então a gente acaba sendo – a gente assim, que está mixando – sendo muito conservador para o filme sobreviver. Porque se você está vendo ali o filme dele, bem ou mal se desse algum problema, numa sala imax não que elas são bem cuidadas, mas se desse algum problema, você tinha um suporte da legenda. No filme em português, você não tem suporte da legenda. Tirar um diálogo ali e você perder aquele diálogo, acabou, você perdeu. Então a gente é muito conservador com isso. É uma pena, às vezes dá vontade de brincar mais.

Raphaela: Acaba que o entretenimento vira trabalho né? rs

Ariel: Nem me diga, cara, difícil eu conseguir desligar assim, eu fico feliz quando desligo, mas geralmente nos primeiros minutos de filme eu estou vendo a mixagem do filme, conseguir desligar e se envolver segue bem difícil.

Raphaela: Como é o direcionamento que você recebe para fazer é, por exemplo, na direção é para fazer essa mixagem. Há uma previsão do que, por exemplo, vai entrar como cama, o que que entra como objeto. Essa decisão é sua? Pensando assim que vai ser uma mixagem em Atmos, como é que é essa decisão?

Ariel: Raphaela, é muito raro, pelo menos com os diretores que eu trabalho é muito raro vir dar direção a esse tipo de pedido. É só em pontos muito específicos assim. “aquela cena tem tal elemento e nossa, eu queria isso girando pela sala”. Às vezes, se o diretor chega nesse ponto, ele é muito específico de uma cena muito específica, via de regra, é uma coisa muito

mais entre o *sound designer* e o mixador do que uma coisa com a direção. Claro que vai enriquecer, às vezes o diretor vai falar “olha, que legal, vocês puseram um som assim, ótimo”. Mas isso cabe muito mais a gente olhar e falar, “pô, será que cabe? Será que vai?” Ou da gente vender o peixe e falar “ó tem uma proposta, chega para o diretor, quero fazer isso, isso dá o resultado”. Aí você gosta “o cara, Ah, gosto, Ah, não, não gosto”, mas vir um pedido da direção é muito raro, muito raro, tem uns que gostam, que são mais ligados nisso, diretores ou produtores que até pedem alguma coisa, tipo, “Ah, eu queria uma coisa aqui mais passando por trás da gente”, faz um comentário assim, mas é exceção, minoria absoluta assim. O normal é que isso fique entre o desenho de som, edição do som e a e a mix mesmo. Seja um diálogo mais interno ali da equipe de som, mas para testar e ver o que que dá certo, do que pedirem pra mim.

Raphaella: É e dentro dessa perspectiva de utilização dessa tecnologia hoje, a ideia segue sendo de utilizar o atmos para “abrir o som”, pouco pra se espacializar mesmo e pouco para direcionar as cenas e colocar efeitos específicos angulados. Então a perspectiva, ela segue sendo essa para abrir o som para ou houve uma diferença significativa na mixagem de Hoje para essa de 2012 de 2015 em relação a essa espacialização, lógico que em relação à tecnologia sim, né, houve diferença, mas eu falo mais no sentido da criação mesmo.

Ariel: Eu acho que a gente ainda está aprendendo assim, sabe? Eu acho que ainda é muito novo. Eu não consigo te falar assim, se ainda essa ideia, não consigo nem te afirmar se essa foi a algum dia. É, eu entendo que sim. A ideia do da imersão, o áudio imersivo, de você ter uma espacialidade legal, mais envolvente, mais imersivo. Mas, eu acho que é isso, o próprio exemplo que eu te dei, o Gravidade, que é um dos primeiros filmes mixado em Atmos, ele é super pontual no uso do negócio. É assim, tem lá um ferrinho que estourou não sei onde e você ouve ele rodando em sua volta, não é uma coisa difusa, não uma coisa geral é uma coisa bem pontual. Claro que é isso, o filme é um prato cheio para isso, mas voltando para Hoje, eu acho que a gente ainda está aprendendo, eu acho que ainda não tem uma estética pronta, feita. No 5.1 já tem umas convenções mais seguras que mesmo assim eu acho que cada mixador tem o seu jeito e eu acho que para o Atmos eu sinto que a gente ainda não está totalmente... Não tem convenções tão rígidas estão bem acertadas no mercado eu acho que está bem aberto assim, acho que cai num subjetivo, cai no gosto, “Ah, eu acho legal fazer de tal jeito. Ah, eu prefiro isso. Ah eu eu só uso a camada superior quando for um elemento específico”, o outro pode falar “não, mas eu sempre boto ambientes na camada superior” Eu ainda não vejo uma coisa 100% resolvida ou bem definida, não é?

Raphaella: E sobre diferenças assim de mixagem de produtos brasileiros para produtos internacionais? A gente falou um pouco aí agora sobre a necessidade, né? Essa questão da comédia e tudo, mas há diferença assim que você percebe, em relação ao orçamento eu imagino que que tenha alguma coisa e tudo, mas tem alguma diferença criativa assim de direcionamento?

Ariel: Não, eu acho que pelo tipo de filme que está sendo feito mesmo. Se eu pegar um “Ainda estou aqui”, acho que é o maior exemplo nosso. É um filme que não exigiria, que não precisa, poderia ter, mas talvez é um que sobreviveria bem sem uma tecnologia dessa. Ela talvez não fosse agregar tanto a narrativa, a experiência quanto um filme, Sei Lá, Avengers possa ter, se utilizar esse tipo de plataforma.

Então eu acho que primeiro é o tipo de produção. E claro, chover no molhado o orçamento, claro. A gente, com uma ficção bem remunerada no Brasil, eu tenho, Sei Lá, 4 semanas de mixagem, uma ficção, em boas condições. Nos Estados Unidos, eles têm 4 semanas para cada pré-mixagem, 4 semanas de pré-mix de diálogo, 4 semanas de pré de efeito, 4 semanas de pré de música e daí 4 semanas de mix final. E que eu faço uma semana para cada um, uma de diálogo, uma de efeito, uma de música, uma de final, com sorte quando o filme tem dinheiro. Porque o normal tem sido fazer os filmes em 3 semanas. Então, a gente usa as mesmas ferramentas do ponto de vista de tecnologia, a gente não está atrás, só que a gente tem menos infraestrutura e menos tempo e dinheiro é a mesma coisa, é sinônimo, nesse caso é sinônimo. Se a produção tiver dinheiro para 8 semanas de mixagem ou para Sei Lá? 20 semanas de edição de som vai ter um produto diferente do que vai ter em 6 semanas de edição do som, né? 6 semanas de mixagem vai soar diferente do que 3 semanas de mixagem. Eu acho que a gente não tem problema de tecnologia defasada, não tem problema de mão de obra. Hoje em dia os recursos que eu uso e a formação que eu tenho é basicamente a mesma que eu acho que um mixador lá em Hollywood deve ter. Mas existe essa diferença do produto, da produção e do tempo, da infraestrutura. Eu acho que talvez essa seja a maior diferença de todas, quanto a gente precisa trabalhar aqui para chegar nos resultados. O que nos é dado ali para a gente conseguir chegar no mesmo ponto. E não é só na mixagem, estou falando da pós-produção, na realidade tudo parte do som direto, a gente tem muitos problemas de som direto, de qualidade do som e não por culpa dos técnicos, ou dos gravadores, ou dos aparelhos ou dos microfones, mas muito por culpa da mentalidade do produtor brasileiro que vai ser querer economizar, às vezes por mentalidade, às vezes porque realmente não tem a verba. Então um filme que às vezes precisa de 40 diárias e o cara conseguiu rodar em 32 diárias, vai ficar tudo mais corrido,

então não vai ter tanto tempo de refazer, takes, né? Aí entra, enfim, eu acho que tem toda uma coisa estrutural, mesmo que vai refletir lá no produto final.

Equipe reduzida, equipamento. Se você é técnico de som direto e você quer comprar um microfone novo, cara, o dólar está 6 pra 1, como que o cara vai conseguir fazer? Existe também uma questão cultural da geração mais antiga, que talvez não dê tanta atenção para o som, nem dê tanto valor para o som quanto deveria. Eu nunca fui um cara de set, né? O que eu estou falando é o que eu escuto dizer, mas dependendo do set, se o técnico de som pedir para refazer o take ele vai tomar um esporro, vai, ele vai Tomar um não, Vai Ficar um clima ruim para cima dele. Por causa do som não é? Ele falou “ó esse tem que não valeu para mim vamos fazer outro”, o cara vai Tomar ou não chega para lá, não é agora um fotógrafo se falar isso tudo bem, claro, o fotógrafo é o rei né. Então tem uma questão cultural aí, infelizmente. Então a gente ouve essas coisas do set e isso reflete no produto que vai chegar, não tem como não é. Por mais que na pós a gente reconstrua as coisas, chama o ator para dublar, fazer ADR ou você mix se tem ferramentas mais modernas e tal, nada supera.

Raphaela: O que você acredita? A mixagem voltada para objetos sonoros influencia tanto assim a narrativa, o filme, a história completa de forma diferente da mixagem 5.1. Stereo. Você acha que há uma grande diferença no produto final enquanto narrativa, enquanto trama pela diferença tecnológica?

Ariel: Não, não, gostaria até, mas não. Se a história é boa precisa ter um bom som, claro, como precisa ter uma boa imagem, mas falar que essa foi a grande mudança, não foi. Como o “Ainda estou aqui”, foi lá e ganhou o Oscar. E também não é termômetro de nada, mas acaba que é. Não deveria ser mais é. “Ah, se aquele helicóptero no começo do filme, que a Fernanda tá lá na água, se aquele helicóptero tivesse passado na camada superior, ia dar um sentimento... Não. É uma pena, mas não.

Raphaela: E quais as perspectivas mercadológicas assim em relação ao uso desse tipo de mixagem no cinema e, principalmente, em plataformas de *streaming* aqui no país?

Ariel: Assim, na experiência que eu tenho, a tendência é ir aumentando. Então, no começo, pouquíssimos projetos exigiam que fossem feitos, por exemplo, em Dolby Atmos, e essa porcentagem está aumentando. Então eu acho, é o palpite meu que dentro de algum tempo realmente quase tudo produzido para as plataformas – hoje em dia, quem pede Atmos é a Netflix, a Amazon e a Disney – eu acho que a tendência é que isso domine, mas a passos pequenos não sei te falar em quanto tempo.

Raphaela: Você sabe se a Apple pede?

Ariel: Eu Não sei, deve pedir, né? É que eu nunca fiz nada com eles, mas deve pedir. E eles não produzem no Brasil, né? Não tem produção local. A Disney, a Amazon e a Netflix produzem muito local.

Mas eu acredito que as séries deles devem ter, né? Acredito que sim, mas aqui no Brasil não, não tive contato assim, mas eu acho que essa fatia vai aumentar, está aumentando, mas é lento, tá? Não é porque nossa uns anos atrás é o que eu te falei lá no início da nossa conversa, foi terrorismo que “então, se você não tiver isso, você não trabalha mais cara”. Já passou uns 7 anos dessa história. São passos pequenos.

Raphaela: Se você puder falar, se você não puder fica à vontade, em relação a valores você falou que às vezes é a produtora pede um orçamento em Atmos um orçamento 5,1, em porcentagem qual seria mais ou menos a diferença?

Ariel: Isso também é uma coisa igual eu te falei da linguagem, eu acho que a gente ainda está aprendendo. Posso falar numa boa. Nos orçamentos que eu tenho praticado, eu entendo que o atmos deva custar de 20 a 30% a mais, pelo menos na etapa de mixagem, se não for globalmente na edição do som inteira, na pós de som inteira, pelo menos a etapa de mixagem eu entendo que tem que custar no mínimo 20% a mais. Mas isso é da minha cabeça, não tem nenhum embasamento, pra te falar por A Mais B, mas eu entendo assim que envolve uma tecnologia mais elaborada, eu vou ter que investir mais em equipamento, eu tenho que ser um profissional um pouco mais atualizado para trabalhar com isso, então eu entendo que isso tem que custar de 20 a 30% a mais. Mas é o que eu te falei a maioria das vezes ainda o cliente, quem está demandando isso, fala, “não pode deixar, não faz”. Então, por isso eu não sei te falar exatamente.

Raphaela: Se já é já a nossas perguntas habituais aqui acabaram, mas é só uma questão que eu pensei. Já aconteceu, por exemplo, de um filme sair em arte, mas um filme nacional, por exemplo. E quando ele vai para uma plataforma, ele tem um *downgrade* assim, ele vai para 5.1. Que você tenha ficado sabendo?

Ariel: Não sei te falar, mas eu acredito que seja o oposto, porque na realidade assim a gente está produzindo quase nada de Atmos para cinema. Quase nada. Tem 2 salas de mixagem que conseguem masterizar Atmos para cinema no Brasil e elas são voltadas para filmes dublados. Não são nem salas assim que estão rodando muito para conteúdo Brasileiro. Se você quiser,

you aluga a sala e vai lá e fecha o seu filme. Mas assim tem muito pouco Atmos para cinema, isso é de 1 ano, 2 anos para cá. Está na realidade, o oposto tem sido mais comum, você produzir Atmos para a plataforma e ter uma saída 5.1 para o cinema.

Raphaela: Ariel, muito obrigada pela entrevista. Vai ser importante demais para mim essa nossa conversa..

Ariel: Por nada, que bom que deu certo.

ANEXO B - ENTREVISTA CARLOS KLACHQUIN

Raphaela: Para além desse discurso da imersividade, como que a Dolby define tecnicamente o Atmos no que diz respeito ao posicionamento espacial dos objetos sonoros?

Carlos: Bem, sinceramente, a própria definição de Atmos é de um sistema que permite criar, transmitir e decodificar som espacial, som imersivo. Utiliza metadados pelos quais, além da informação do áudio, é também transmitida a informação espacial, seu posicionamento desses objetos. Além das camadas, que são basicamente uma herança dos sistemas convencionais 5.1, 7.1. Então, a definição básica é essa, o sistema tem todas as ferramentas e possibilidades para criar, transmitir, estocar né, gravar, transmitir e decodificar.

E tem, eu diria assim, uma característica interessante, além da capacidade de transmitir informação espacial e o posicionamento de objetos, é o fato de você ter um sistema flexível, no qual você pode ter diferentes tipos de ambientes para ouvir, de fones até uma sala com diferentes distribuições de caixas. Existem muitas soluções atualmente e o interessante é que o *decoder*, o decodificador, você instruí-lo sobre qual é seu ambiente de monitoração e você escolhe e digamos que o algoritmo interno, modifica, toma a informação genérica que vem por canais, digamos, de transmissão para adequá-lo ao meio que você tem para monitorar. Então, se tem o sistema é um 2.0, 5.1, 7.1.4, enfim, quaisquer possibilidades. Então, em essência essa é a ideia.

Raphaela: Como essa abordagem orientada a objetos transforma a experiência sonora no ambiente doméstico, considerando dispositivos com capacidades tão distintas?

Carlos: O que chamamos de algoritmo, em Dolby chamamos de *renderer*, que é um conjunto de algoritmos que permitem fazer esta codificação, colocar esta informação. Quem criou conteúdo tinha a intenção de colocar um determinado objeto em determinados lugares e essa informação é transmitida e depois reconstruída da melhor maneira possível, sempre existem comprometerimentos, para o ambiente de escuta, do outro lado, que interessa.

E você mencionou uma coisa que eu gostaria de adicionar à sua informação. Existem, na verdade, duas avenidas principais que são a *Home Theater*, de entretenimento doméstico e a de cinemas, as duas são de Dolby Atmos, mas são, eu diria assim, animais de diferentes pesos. A filosofia subjacente é a mesma, é a primeira definição que eu dei, “é o sistema que transmite informações espaciais”, mas por diferenças técnicas, que depois podemos ampliar um pouco mais, existem diferenças muito grandes no ambiente porque cinema são um espaço

grande necessariamente e doméstico é um espaço reduzido, pequeno, é a sala, é o quarto. Então o número de caixas que você utiliza é diferente, a largura de banda você dispõe para trabalhar em cinema, é muito diferente da que se dispõe para *streaming*, para doméstico e isso obriga a implementar de forma diferente. Ou seja, a filosofia é a mesma, mas os dados como são enviados são diferentes e têm nomes diferentes e a forma de monitorar tanto na produção, quanto como ouvimos são muito diferentes entre os cinemas e o ambiente doméstico. Embora seja muito flexível o Atmos, estas diferenças são necessárias, tem a ver como a mídia trafega. Isso é importantíssimo levar em conta, porque afeta o codec que passa a ser outro e afeta a forma da distribuição de caixas e como se processa essa informação para poder reproduzir.

Raphaela: Essa, inclusive, era minha próxima questão, que era quais são as diferenças básicas na mixagem feita para salas de cinema para essa de feitas em ambientes domésticos.

Carlos: Bem e eu gostaria de ordenar, digamos, então há aspetos, são 2 aspetos. A primeira é a monitoração, como você monitora e como você vai reproduzir isso. A segunda questão tem a ver com os codecs, porque isso afeta como que você vai entregar aquilo depois da mixagem, como você vai finalizar, como você vai empacotar essa informação toda para poder entregar. O cinema tem uma forma e o doméstico tem outra, em inglês, Home Entertainment, Atmos HE, outro Atmos Theatrical, em inglês, vamos chamar aqui de Atmos para cinema e Atmos para doméstico.

A primeira questão da monitoração, embora seja uma questão física muito básica que é o seguinte: em um ambiente pequeno, doméstico, você tem é tempo de reverberação relativamente mais curto, porque a medida que você aumenta o volume de um espaço necessariamente, já que o som tem uma velocidade definida e a acústica tem a ver com as reflexões do som no ambiente que ele limita, a condição de contorno, em matemática chamamos de condições de contorno, são os muros onde se reflete o som. Então, você imagina o seguinte, entrando um pouquinho em questões técnicas, basicamente o que é a acústica, você tem que considerar que cada elemento da superfície, teto, muro, objetos é um refletor, por um lado reflete o som, por outro absorve, e não reflete igual nem absorve igual em toda as frequências, normalmente absorve mais as frequências mais altas, mas às vezes acontece o contrário, depende da substância que você está encontrando nesse muro. Então, quando o som se reflete em cada elemento cada, eu diria assim, infinitos sinais da superfície, vai acontecer vários fenômenos: o primeiro é este que uma parte absorvida e uma parte refletida e essa absorção não é uniforme na frequência, algumas são mais absorvidas que outras e, por consequência, algumas frequências são refletidas com maior intensidade que

outras. Esse é o primeiro elemento. Agora pensa que isto é para cada elemento da superfície, imagina que cada elemento poderia se tornar como se fossem pequenos azulejos, que vai ter essa capacidade de filtragem dependendo da matéria.

Outro aspecto é que entre o som direto que você recebeu de um alto-falante e o som que se é refletido nesse mesmo alto-falante, nessa superfície, que depois atinge o teu ouvido, existe uma diferença de tempos porque os percursos são diferentes. Existem inclusive reflexões secundárias, um som bateu nesse muro e depois bateu no outro e retornou... Isso tudo cria um ambiente extremamente complexo de sons que chegam no teu ouvido com tempos diferentes e filtrado de formas diferentes na frequência, isto tem um efeito que se acumula, o que você percebe é a soma de cada um de todos elementos de grande quantidade de elementos, vamos chamar assim, com diferentes retardos. Então dá para entender, porque a ciência da acústica é tão complexa, porque cada ambiente tem sua, vamos chamar, função de transferência, ou seja, como esses sons originais que saiu de um alto-falante é modificado e chega ao teu ouvido.

Existe um outro fenômeno, que podemos entrar depois quando a gente conversar sobre binaural que acontece dentro de nosso sistema, tanto orelhas como o ouvido médio, ouvido interno, tudo isso também é afetado, mas isso vamos conversar depois. Então, primeiro aspecto é isso, então, se considera que ambiente doméstico e ambiente de cinema necessariamente são diferentes. E o fato de você precisar, digamos, ter uma plateia, você tem um número grande de pessoas que está assistindo, você precisa de alguma forma gerenciar da melhor maneira possível que cada um destes espectadores receba a informação da melhor maneira possível. Em outras palavras, quando você projeta um cinema, especialmente o *surround* tem suas características, mas todas as vias, todos os alto-falantes, você precisa que, na medida do possível, cada alto-falante atinja toda a plateia. E isso não é possível, digamos assim, fisicamente é muito difícil. Um exemplo, quem está sentado na fileira ao lado do muro, o som das caixas *surround* passará por cima dele, não vai ter a mesma percepção de aquele que está sentado na parte central da plateia. Então é impossível, você tenta fazer o projeto sendo o mais democrático possível, pois é importantíssimo a diretividade das caixas de som, né? Pensa como se fossem luzes, se quer iluminar a própria.

Este problema de alguma forma acontece também em ambiente doméstico, só que é muito menos grave, porque normalmente você precisa iluminar um sofá, uma poltrona ou poucas poltronas. Então aí o tema existe, mas não é tão crítico quanto no cinema. Um elemento que facilita porque espaço menor, as reverberações são menores, o tempo das reverberações são menores, mas também você tem mais o meio ambiente, cinema costumam ser mais

silenciosos, você pretende que seja silencioso e ambiente doméstico você não trata acusticamente a sua sala, então você lida já com uma classe de ruídos diferentes. Apontando todos os elementos que diferenciam o doméstico de cinema. Aí você tem no cinema a necessidade de colocar mais caixas especialmente em canais *surround* porque precisa uniformizar melhor essa distribuição. Isso é um cinema convencional 5.1, 7.1 e em Atmos existem outros compromissos, além desses porque exige uma densidade de caixa maior para poder criar esse ambiente virtual, então temos fileiras de caixa no teto, além das caixas que ficam na parte de trás e na lateral, enfim. Essa é uma parte complexa, mas tem muito a ver com a fórmula com a qual se projeta essas salas.

E em ambiente doméstico as necessidades são diferentes e o número de caixa é bem menor, tão menor que o espaço por cima de você, o que você sente para assistir é tão reduzido que permite implementar ou que conhecido como soundbars, que é muito prático, porque você não tem que fazer muitas integrações e você mais ou menos consegue recriar o ambiente multicanal com um elemento só que é de uma barra de caixas que trabalham, seja por reflexão ou por criar uma imagem virtual e mais ou menos conseguindo ter um resultado espacial. Este o primeiro elemento.

Agora vamos pelo segundo elemento. Em cinema, você transmite o conteúdo, distribui o conteúdo através do cinema DCI, cinema comercial, você distribui esse conteúdo através de discos rígidos ou por download via satélite que não é em tempo real e guarda na biblioteca e você utiliza disco rígido na capacidade relativamente alta. Um filme pode conter, pode ter um tamanho de 150 GB, 250 GB depende se é 2D, 3D, quanto dura, quanto tempo dura o filme, mas está nessa ordem de centenas de gigas. E os sistemas internos que trafegam esta informação, os servidores que têm o cinema para o projetor e do projetor para, digamos, a caixa mágica que é a *Integrated Media Service*, um dispositivo essencial no cinema de DCI que mastiga seus dados e outra forma na forma útil para projetar e ouvir. Assim se chama *Integrated Media Block*.

E essa informação viaja com uma capacidade muito alta, centenas de megabits por segundo, então você tem à disposição um canal de alta velocidade, que necessariamente faz parte da construção do sistema, porque a imagem consome dados em quantidade enorme, realmente grande. Então, tem uma outra questão a mais, que você tem um número de caixas bem maior no cinema e uma resolução espacial maior do que os ambientes domésticos necessariamente. Isso faz com que somado a isso, existem outros elementos especiais que tem a ver com a segurança do conteúdo e isso cria a necessidade de gerar o chamado DCP.

E que aproveito para adicionar um erro bastante comum que as pessoas confundem DCP com DCI. DCI é o padrão para cinema comercial, DCP é o digital cinema package que é o arquivo grande, o arquivo enorme, um conjunto de arquivos com o qual se entrega o filme para o cinema, ele contém tudo, legenda, áudio e imagem e enfim *watermark*, marca d'água, enfim, uma série de elementos de segurança. E quando você passa o arquivo que contém um som dentro do DCP, é um formato chamado MXF é esse é o formato. No caso de som doméstico, formato doméstico você tem 2 formas principais, isso tudo depende de quem é gerenciador e fornecedor, digamos, de conteúdo, a plataforma. Isso é chamado de ADM e ele tem dois formatos de guardar os dados, um é chamado VWF, outro é IAB, são os mais populares e o IAB é o mais popular. Então, quando você está finalizando a tua mixagem e você vai codificar, criar o pacote para entrega, você vai precisar exportar essa informação, depois vamos conversar um pouquinho mais, num formato que empacote tudo isso. E que tenha significado para o receptor, então, esses formatos são ADM, digamos, ADM significa *Audio Definition Model*. E o mais popular de todos é o IAB, que *Immersive Audio Bitstream*, que é outra sigla, que você vai encontrar isso em *renderers*, que vai perguntar como você quer exportar isso. Em cinema quando exportava, entre aspas, e conteúdo, além do Atmos, isto é importantíssimo, também é obrigatório por padrão DCI que seja exportado o áudio em 5.1 ou 7.1, este é um formato cru, é o PCM, uma simplificação a esse, digamos, que qualquer distribuidor de cinema vai entender. Vamos supor um cinema que não tem Atmos, nem sabe o que é Atmos, não vai decodificar mas ele vai buscar essa informação aceita, isso é obrigatório. Então, quando se cria esse conteúdo, se coloca no áudio, no pacote do DCP, se coloca nos dois, no MXF e também no PCM do áudio 5.1. Em formato doméstico, voltando a essa outra parte, você tem estes codecs, esta forma de entrega, e você também tem a possibilidade, outros dois formatos muito importantes para exportar, que se utilizam no formato doméstico, um deles é o famoso MP4 que reúne áudio e vídeo. E o próprio *renderer* Atmos permite exportar também MP4, inclusive pergunta se você tem imagem para associar com este áudio que você está criando. Então você pode criar já o arquivo MP4 que é *embed*, então é importante isso. E tem uma outra forma de exportar que hoje está se tornando, vamos dizer assim, o arroz com feijão, que é o binaural, do qual podemos conversar separado, mas, desculpa me estender tanto, mas é importante entender que essas são as diferenças essenciais.

Quando você cria conteúdo, você precisa criar um ambiente para o qual você está criando, como se você mixa para o cinema, a sala de mixagem tem que ter um volume mínimo para reproduzir aproximadamente as condições de um cinema médio. A sala não tem um padrão?

Espaços, volume diferentes, relações de largura, altura e comprimentos diferentes, Atmos resolve essa história da tecnicidade, mas a parte acústica, você precisa ter um ambiente de monitoração próximo ao que você vai ter depois no cinema, que não é o mesmo que você tem em ambiente doméstico ou em uma salinha de mixagem pequena, o que chamamos de Campos Near Field, você tem caixa muito próximas, então, o som direto que vem dessas caixas, no caso doméstico, é muito dominante sobre o efeito reverberação, porque você está próximo a elas, então, o campo direto é muito dominante sobre o campo reverberante, no caso doméstico, isso cria uma diferença vital, em outras palavras, se você mixar monitorando em campo, *Near Field*, doméstico, não tente levar isso para um cinema porque vai funcionar mal e muito mal. Mixagem para cinema se faz em ambiente com um volume mínimo e com um sistema de monitoração adequado e tem a necessidade de caixa bem maior que para o doméstico. E finalmente você utiliza o código adequado para cinema, que é o que você vai entregar para quem vai criar o DCP. Por isso eu te digo, são duas avenidas que tem aspecto, é muito importante insistir sobre isso, porque eu vejo muito realizador que desconhece esta questão. É importante que seja, digamos, divulgada esta informação que tem que ser uma coisa muito básica. Esse material mixado para o cinema pode ser utilizado, não o codificado, a mix pode ser utilizada para doméstico depois. Aí funciona, funciona bem e isso ainda tem uma experiência, longa experiência da época de Dolby Digital de cinema que depois se colocava esse conteúdo em DVD. Então essa mesma mixagem também funciona bem em casa, mas ao contrário não. Por essas questões acústicas que eu mencionei.

Raphaela: O downmix funcionaria bem, o upmix não tanto, certo?

Carlos: Não funciona.

Raphaela: E você acha que é também por isso que a produção brasileira ainda é tão incipiente assim pra cinema, com Atmos?

Carlos: Tem um fenômeno histórico, eu diria assim. Sim, é incipiente, não é extremamente popular, mas existe, até tem filmes que estão sendo feitos atualmente brasileiros em Atmos, mas para quem produz um filme... Um pouco depende também do gênero, para quem está fazendo um filme onde o principal seria uma mensagem ou o diálogo, então, neste caso, o atmos não importa tanto, poderia de forma artística, mas tem um aspeto histórico clássico que eu vivi desde início dos anos 1990 mais ou menos, quando eu comecei a trabalhar na área de cinema aqui, que era a tecnologia existente na época e nessa época ainda se fazia cinema

mono e se começar a utilizar o que na época se chamava Dolby Stereo, 4 canais, o mundo analógico, Dolby Stereo, 4 canais.

E o que eu vivi foi uma evolução muito interessante, não só na tecnologia, mas também na qualidade do trabalho, técnicos, o pessoal, o artista que criava o som foram todos evoluindo e crescendo de uma forma fantástica. E se melhorou muito as técnicas de captação do som direto, você começou a cuidar muito mais, os processos de laboratório melhoraram muito, laboratório é um processo muito complexo até chegar a uma cópia, são nervos, é uma corrente, tem que funcionar direitinho, senão a corrente parte pelo elo mais fraco, não é?

Então, isso foi evoluindo de forma linda e conseguimos ter uma produção no Brasil muito cuidada. Agora havia um fenômeno e era que essa produção é extremamente cara, o laboratório é caro, a tecnologia 35 mm é cara e tínhamos a produção de som feita em 35mm, porque a mesma película da imagem tinha o magnético, uma capa de elemento magnético onde se registrava o som, analógica. E então tudo isso se sincronizava perfuração a perfuração, todo mundo combinava, sincronizava a imagem e o som utilizando este material magnético 35mm. Mas as máquinas são de custo alto, o que fazia com que existisse uma espécie de funil, ou seja, se você queria finalizar um filme e queria para cinema, 35mm, você ia a um dos estudos de finalização que tinha essa tecnologia, que eram poucos. E finalmente esse material ia para laboratório, no Brasil tinha 2 só, então isto formava uma espécie de funil na qual em cada etapa também funcionava como controle de qualidade, ou seja, mixava em um estúdio adequado para o cinema e havia uma espécie de filtro, se a coisa não estava bem feita, normalmente já se exigia muito mais, como uma evolução natural. Então o Brasil chegou a ter, até a chegada do cinema digital, um pouquinho mais depois, um cinema de produção profissional de altíssimo nível em todas as fases: captação, edição, mixagem, laboratório... Os cinemas tivemos uma evolução diferente e não simultânea, no início eram em sala de rua, muitos reverberantes, equipamentos que não eram de qualidade, digamos assim, não havia muito cuidado com isso, o som era secundário, interessava a ter uma tela e uma caixa de som por trás acabou a história. Isso foi mudando com a chegada dos multiplex, passamos para outro patamar, o que temos hoje é outro mundo, mas isso é uma evolução interessante.

Agora na produção, o que aconteceu? Se democratizou a produção, porque o mundo digital permitiu que você já não precisasse trabalhar mais com 35 mm, trabalhava em um ambiente simples, com investimento menor, não só a produção você democratizou, pois é possível você poder trabalhar em casa para mixar, não é? Mas isso por outro lado tem um calcanhar de

Aquiles, acabou o filtro, acabou o controle de qualidade e também quem faz isso não tem a experiência que tinha o pessoal de som nos estúdios, que sabiam como trabalhar, lidar com esse tipo de conteúdo para cinema. Isso criou um ambiente que temos hoje na qual é necessário reeducar e treinar e conscientizar pessoas sobre esses aspectos técnicos que se deixam de lado. E não só isso na produção, mas também na exibição, especialmente em festivais, é fundamental também cuidar dos padrões, existe um padrão para cinema, é o mesmo padrão. Mudou a tecnologia mas a forma que ouvimos, monitoramos, com a exceção de Atmos, que introduz uma nova tecnologia, o resto está padronizado e nem sempre se segue esse padrão. Então, qual é a consequência disto, a intenção do criador do conteúdo não chega ao público da mesma maneira. É como um pintor de quadros que expõe um quadro num local mal iluminado. Então a essência do que você criou, não chega ao destinatário, então eu diria assim que é importantíssimo, principalmente numa universidade, todo o âmbito educar nesse sentido, de tentar civilizar de novo isso que tínhamos conseguido, né? Porque as salas inclusive melhoraram muito. Festivais meio que se improvisam, mas aí mesmo improvisando, é necessário colocar dentro da pauta para respeitar a obra do criador.

Raphaella: Voltando lá para a nossa parte técnica, vamos falar sobre o binaural? E como que a Adobe tem explorado as possibilidades do render Binaural no contexto atmos virtualizado, o Atmos Home, especialmente considerando essa crescente popularização? Acredito que pelo menos aqui no Brasil a gente tem a presença muito maior do Atmos dentro de casa, em plataformas de *streaming* em dispositivos móveis e tudo o que em cinema... Há um esforço específico para garantir essa experiência espacial dentro desse tipo de escuta, através de fones de ouvido?

Carlos: Bom eu te diria que hoje essa é a área que tem mais pessoas da Dolby no mundo inteiro. O mundo se tornou uma grande aldeia, um fenômeno que acontece aqui são os fenômenos que você encontra na Europa, nos Estados Unidos, imagino que na Ásia, eu nunca estive na Ásia, mas aparentemente é a mesma situação. O binaural, eu não o número de ouvintes, só que eu vou te contar duas outras coisas. A maior parte do esforço da Dolby está justamente na difusão e popularização na escuta binaural de música. É uma realidade, você está na rua, vai encontrar muitas pessoas com fones, vai no metrô, vai encontrar maioria deles assistindo conteúdo, não sei o conteúdo, mas é comum. O binaural, então, é extremamente popular, tem crescido de forma incrível, ele continua crescendo de uma forma incrível. É comum por se encontrar isso em laptops, estou vendo o logo aqui no meu laptop, nos fones também, é extremamente popular, inclusive quando você cria conteúdo já específico para

binaural, você pode monitorar tranquilamente por um sistema binaural [fones de ouvido], não tem um custo como um sistema completo. O binaural parece ser o arroz com feijão atual, o renderer Dolby, Dolby Atmos, você tem como exportar em binaural, se você faz a mixagem, você pode exportar neste formato que mencionei e também pode exportar para o binaural. Ou seja, há todo um suporte de Dolby, inclusive a Dolby incentiva a produção de conteúdo, porque interessa a Dolby que os dispositivos que têm Dolby Atmos tenham conteúdo. Essa é a questão básica. Então, o peso atual sobre ensino, sobre licenciamento tem muito a ver com essas tecnologias domésticas e dentro das tecnologias domésticas, binaural tem um peso enorme. E tem uma outra coisa linda no binaural: você pode utilizar basicamente ou praticamente qualquer tipo de fone. O fone assim não é parte especializada, o truque está no algoritmo que decodifica, no renderer inserido que decodifica segundo o Atmos. E tem um aspecto que não é muito evidente, mas a forma que está construído esse *streaming* de dados, como está codificado em Atmos permite que o conteúdo praticamente possa ser reproduzido para qualquer dispositivo, mesmo que não tenha Atmos, você não vai perder informação, vai perder informação espacial, claro, mas você não vai perder o conteúdo e isso é importantíssimo, sempre tem que cuidar da compatibilidade e da retrocompatibilidade, isso é uma filosofia permanente. Vou te contar um exemplo assim muito rápido. na época da película quando se incorporou do digital 5.1, o formato analógico, antigo, permaneceu, nos filmes, permaneceu de forma muito prática para você não ter uma distribuição de filme, de cópias, Dolby Digital para um tipo de cinema e não Dolby Digital para outro, uma loucura isso, para a produtividade é impensável. E tem outro aspecto, se o sistema digital dava defeito, o sistema diretamente passava para o analógico. Isso é uma filosofia muito inteligente e permitiu, digamos assim, que essas cópias, pudessem tocar em qualquer cinema, mesmo mono, o sistema é compatível até com mono.

No caso de Atmos, isso é uma coisa que algumas distribuidoras não perceberam. Quando eu falei do DCP que vai para os cinemas, lembra que eu comentei que você tem arquivos que correspondem a Atmos e arquivos convencionais que não são Atmos, é a mesma filosofia, uma sala que não tem sistema Atmos vai ignorar por completo esse arquivo de Atmos e vai buscar o PCM 5.1, 7.1, que já sabe que para a prossecução tem que colocar isso. A sala Atmos quando recebe um DCP que não tem conteúdo Atmos, quando não tem conteúdo Atmos vai buscar o conteúdo PCM convencional e o processador entra nesse modo, ou seja, há uma compatibilidade num caso e a compatibilidade no outro caso. No caso de *streaming* orientado a binaural aconteceu uma coisa parecida, ou seja, quando você cria isso, isso você

vai chegar a qualquer um, o dispositivo entenda Atmos ou não, isso é permanente, isso é muito, digamos assim, inteligente no sentido de que o seu conteúdo vai chegar naquele lugar.

Raphaela: E o funcionamento do Atmos segue aquela premissa do HTRF né? Essas funções de transferência da cabeça e o algoritmo, o *renderer* Atmos consegue direcionar o ângulo ali que você consegue trabalhar?

Carlos: Sim, mas queria contar uma coisa que é inevitável e é impossível de resolver, vamos falar assim. Acontece que cada um de nós, cada indivíduo, tem diferenças de forma, digamos assim, assim como o nosso rosto é diferente para cada pessoa, a nossa constituição do ouvido, embora tenha os mesmos princípios, ele não é exatamente igual de uma pessoa para outra. Então, essa função de transferência, que você disse muito bem, mas quem tem ouvido mágico que tem essa função de transferência, porque cada indivíduo vai ter uma função de transferência diferente. Afortunadamente, parece ser que a dispersão, digamos assim, o *sigma* funciona, a diferença não é tão grande assim, mas realmente cada um tem sua própria função de transferência, então essa função de transferência se utiliza de uma transferência média, de um estudo estatístico. Então o que acontece é que algum indivíduo estará mais próximo dessa função de transferência e outro não. O quanto não sei, tem que ver essa curva de suspensão aí, mas aparentemente funciona. Mas é importante considerar isso, que a função da transferência tua e minha muito provavelmente não serão iguais. Não vamos ouvir exatamente igual, mas o mesmo acontece com a mixagem que não é binaural, você tem diferentes ambientes de monitoração em casa, as caixas você pode distribuir de diferentes maneiras, a acústica é diferente, então alguma diferença existirá. Tudo isso, tentando fazer o melhor que seja possível, mas é impossível fazer uma correspondência que seja exata entre a produção e a exibição. Mas existe um objetivo, tentando que se aproxime.

Raphaela: Você falou sobre a empresa incentivar a produção binaural, a produção doméstica e especificamente ainda mais a produção binaural. Existem ferramentas ou iniciativas para democratizar esse tipo de produção? Assim, para produtoras independentes, para criadores de conteúdo... E assim já aproveito para falar sobre a existência de parcerias estratégicas ou políticas de incentivo específicas para a América Latina visando essa disseminação da tecnologia no setor doméstico ou educacional, igual você comentou anteriormente.

Carlos: Existem incentivos sim, vou mencionar um deles: hoje um *renderer* Atmos é muito barato, no caso do ProTools já vem embutido o *renderer*. E se não, se você não utiliza o ProTools, compra o *renderer* por centenas de dólares e a instituição de ensino ainda tem

desconto, isso é bem democrático. A gente está falando de 100, 200 dólares. Tem que ver aí a tabela, mas está nessa ordem, não estou na área comercial, estou na ideia. Então, nesse sentido, o render é extremamente democrático, se você pensa no custo de um plugin, uma coisa assim, o custo do renderer Dolby Atmos é possível.

Você precisa de um hardware para rodar isso. Então, precisa de uma workstation, isso independe da Dolby, e vai precisar de uma interface multicanal porque você passa a ter um sistema binaural ou com caixas, o formato você pode escolher 7.1.4, tira da lista com 7.1.2 que eu tenho a minha crítica particular com o 7.1.2, isso é um erro, um erro. Mas você pode ir para 7.1.6., enfim, tem várias distribuições possíveis. O renderer você explica para ele o que você tem e o renderer trabalha baseado nessa informação, essa parte não é problemática.

Agora você precisa de uma interface que consiga manejar, digamos, até 16 canais. Então você tem várias opções de hardware para isso, isto é questão comercial, estou a entrar nessa área, prefiro ser agnóstico nesse sentido porque você tem que ter a liberdade de escolher. Hoje em dia é muito popular a Matrix Studio, da Avid, porque eles reúnem uma quantidade enorme de possibilidades e de formas de entrar e sair, mas por outro lado, ela é enxuta no sentido que tem tudo o que você precisa. Tem outra opção, possivelmente mais barato que é a Audient Orii, também é uma interface multicanal e deve ter mais por aí, tem um mercado e cada vez que você entra para ver o que que apareceu, muda muito, mas eu diria que hoje está bastante democrático

E no caso de você monitorar com caixas, esta é uma opção que não para de crescer que são as caixas para monitorar, que tem todo o tipo de preço, depende do que você tem possibilidade, digamos assim, tem umas sofisticadíssimas e outras que são minimamente boas, mais econômicas, então eu diria que montar um estúdio para mixar em Atmos, inclusive com monitoração com caixas, não é uma coisa impossível, se monta isso com pouco dinheiro. Só para ter uma ideia assim, você tem opções de entre 2700 e 5000 dólares em interfaces, ou seja, não é uma coisa impossível. E pode prescindir de mesa, inclusive, pode trabalhar direto na tela. Mas como dirigir carro com as mãos, eu prefiro ter um volante.

Raphaela: E aí tivemos a escolha pelo Fórum SBTVD do MPEG-H para o sistema de TV 3.0 no Brasil. Como você e a Dolby enxergam essa coexistência, entre esses 2 formatos no mercado brasileiro em um padrão doméstico? Existe uma interoperabilidade entre o Atmos e o MPEG-H em ambientes híbridos, como o *streaming* e a TV aberta, pensando também que

existem plataformas de *streaming* que vem da TV aberta, né? Há algum movimento da Dolby nesse sentido? Como é que você enxerga essa coexistência que está para vir aí?

Carlos: É, eu nesse momento eu preferia lhe negar essa opinião porque é uma discussão comercial, eu não estou envolvido nisso, não estou tanto dentro das últimas discussões. Eu trabalhei nessa área, ou seja, o teu histórico nisso, mas não atualmente, porque não é meu departamento, estou na área de serviço de conteúdo para cinema. De fato há que aqui no Brasil já se está na televisão digital, inclusive já existem transmissões Atmos. Você não vê isso nos logos, mas existe em alguns eventos especiais, tipo Carnaval, algum evento de futebol. Eu trabalhei quando fizemos a grande transição da televisão analógica para digital, eu integrei uma equipe e a ideia era de que o codec de áudio fosse o AC-3, o mesmo AC-3 que se utilizam nos DVDs, que é um codec de compressão de áudio multicanal. Então nessa época eu trabalhei nisso, não se adotou no sistema brasileiro de produção digital, eu iria latino-americano, por exemplo, todo mundo adotou o padrão brasileiro, menos a Colômbia, mas o padrão que se foi adotado pertence a Dolby, nenhum padrão posterior, High Definition AAC e o AAC High Efficiency, certo? HE AAC essa sigla. Atualmente, eu te diria que é essa parte não é minha, está bastante segmentado, então eu não posso responder sobre como é, obviamente a ideia é que se utilize esse codec, mas não são as únicas opções, enfim, é uma discussão que tem a ver com questões industriais, comerciais de padronização também.

Raphaella: Você comentou agora a pouco que não tem dados sobre as projeções de crescimento de consumo de conteúdos com som doméstico ou binaural, certo?

Carlos: E eu vou te dar um dado interessante, da semana passada. Os dados existem, eu particularmente não estou atualizado. Mas semana passada recebi uma informação muito curiosa, uma rede que fornece serviços de dados de Internet no mundo e tem alguns números que tem a ver, porque eles têm informação de que de forma são utilizados os dados. O volume de dados trafegados para redes sociais superou o volume de dados utilizado pelas plataformas. Isso está dando uma pauta de que as pessoas assistem mais TikTok, Facebook, esses conteúdos gerais para o próprio usuário está trafegando mais informação nesse ambiente do que a plataforma de *streaming*. Isso está dando, a mim me surpreendeu, eu não imaginei que isso ia acontecer.

Raphaella: Acho que eu não fico tão surpresa (rs)... De fato, é assustador, para nós que trabalhamos com isso, que estamos nesse meio, não deixa de ser assustador, mas eu não ficaria tão surpresa, as pessoas hoje são muito viciadas e eu me incluo nessa... Há o momento

de parar e consumir conteúdo de *streaming*, mas acaba que quase em todos os outros momentos, a gente está consumindo esses conteúdos de redes sociais, não é? Então é muito mais tempo dentro de redes sociais do que em conteúdo audiovisual em *streaming*...

Carlos: (risos) Você está mais envolvida do que eu. Como sou mais antigo, utilizo o volante freio. Então não é, digamos assim, me deixo observar, mas eu não me deixo dominar tanto por essa questão das redes sociais agora. Como você mencionou, assusta um pouco porque tem muitos fenômenos que estão saindo de controle. As redes sociais não tem filtro, você não tem censura por idade, então são coisas assustadoras. Antigamente nós tínhamos um monopólio da informação através das agências informativas, hoje virou um caos. E em relação a conteúdos, se estamos falando de Atmos, estamos falando de uma produção profissional, produção na qual o conteúdo se vende e se compra. E se está superando, está passando por cima, eu te diria que até isso seria muito interessante como como outra tese, digamos, de análise, o que inclusive é complicado, porque é muito dinâmico. está mudando a forma extremamente rápida e essas mudanças rápidas nos coloca facilmente, nos tiram do equilíbrio facilmente, porque está difícil avaliar coisas. E isso que a gente falou agora de plataformas, *hardware*, inclusive plataformas na qual se pode criar conteúdo, você vê que não é só ProTools que tem isso, você tem Blackmagic, tem Nuendo, não sei, isso muda de uma forma tão rápida que o que eu possa ter dito agora, pode não valer, porque isso tem prazo de validade, sabe? Inclusive, dentro da própria tecnologia Atmos existe uma velocidade de atualização e de reformulação muito intensa, muito intensa e não é exclusivo de Dolby isso, isso está acontecendo em todos os lugares. Então, eu te diria que as nossas análises têm que ser cientes de que tem um prazo de validade, não é como na Idade Média em que você esquecia e esse livro ficava no monastério por séculos. Já não é mais assim.

Raphaela: Hoje em dia tudo é muito efêmero né... E aí vamos falar de tendências futuras, já que estamos nessa hora. Como é que a evolução do próprio padrão Dolby está sendo planejado? Se você puder falar também sobre o surgimento de novas tecnologias. Eu vi algo sobre aquele Flexconnect, se não me engano, da própria Dolby, que se conectaria com caixas tipo Alexa, aquelas da Amazon, caixas portáteis tipo JBL etc. Em um cenário ideal, como é que você imagina o futuro da Dolby e experiência sonora doméstica daqui a uns 5 ou 10 anos? Vamos pensar no curto prazo, já que tudo muda tão rápido...

Carlos: Eu tenho uma frase: “antigamente eu fazia profecia, hoje eu não tenho mais coragem”. Olha, sim, essas tecnologias são muito interessantes porque o setup de forma doméstica se torna complexo para quem, digamos, compra um equipamento e quer tirar

proveito dele e não pretende chamar um profissional. Então, essa tecnologia é boa porque o sistema reconhece o que tem em volta ou pode eventualmente fazer alguma sugestão e deixa o sistema calibrado, isso é uma tendência forte. Agora, os *soundbars* vieram para ficar. Há um peso muito grande no ambiente doméstico, sem dúvida, isso levaria a um questionamento de qual é o futuro do cinema? E por outro, também diria, sim, que poderia ser uma outra tese.

Dolby está investindo muito na área doméstica, sugiro você fazer um pequeno exercício, entrar em dolby.com na página Dolby, que também se transforma toda a semana muda, mas você vai ver que tem um grande peso em o ambiente doméstico, você vai encontrar a maior parte é de *home* e vai encontrar um pedacinho de cinema que, noutra época, foi contrário, normalmente, Dolby estava no cinema. E o peso que tomou o ambiente doméstico indica uma tendência e não se vai reverter facilmente e isso tem a ver, pura especulação minha, não é de Dolby, sou eu Carlos, mas imagino que o cinema vai encontrar formas de adaptar-se a isto, eu estou suspeitando de que o cinema vai passar a ser uma experiência especial. Lembrando de uma história um pouquinho para trás, quando surgiu o cinema, estou falando do início do século 20, os atores de teatro estavam extremamente preocupados porque diziam que agora iam filmar minha representação e depois vão matar o teatro e não aconteceu. O teatro nasceu há 2500 anos e está aí muito vivo, funciona e é uma experiência diferente no cinema. Quer dizer, são coisas diferentes. A experiência do cinema, embora em casa você possa ter tecnologias fantásticas, 4k, Dolby Atmos, o outro Dolby Atmos, é uma experiência diferente, experiência de sair de casa, de encontrar outros de experiência coletiva. Isso é o valor que tem o cinema, é um valor próprio do cinema, que não dá para substituir, seria como uma espécie de paralelo, embora todo mundo tenha em casa uma cozinha, você não dispensa um restaurante de vez em quando. Por quê? Porque a experiência é diferente, o cinema tem essa potência. Eu estou pensando que possivelmente o cinema irá apontar salas sofisticadas, você vai buscar uma coisa diferente e isso torna a experiência do cinema diferente.

Agora, em última instância, eu diria assim. A televisão não matou o cinema, as plataformas não mataram no cinema e não vão matar, o cinema vai permanecer, como o teatro permanece, porque são experiências diferentes, talvez se modifiquem em algum aspecto que não se sabe, como telas maiores, tipo experiência IMAX ou enfim, salas VIP ou alguma coisa assim, mas ele vai permanecer como arte. E Dolby passou, digamos assim, por transformações anteriormente e continua, continua muito ativa, são transformações grandes de tecnologia e como como esta que você mencionou que permite ajustar domesticamente. Existem ainda outras áreas Dolby que tem a ver com é o High Dynamic Range, HDR em que chama, que se

amplia, Dolby Vision é uma aplicação que você vê na rua, vou te contar outra. Existe na linha de cinema chamado Dolby Cinema que são salas, que tem a marca Dolby, mas aqui não temos ainda, não tem esse programa. São salas que reúnem Dolby Atmos é um sistema de projeção muito sofisticado, com uma relação de contraste que pode chegar a 1.000.000:1. Só para você ter uma ideia, projetores DCI normais, a relação de contraste, ou seja, entre o maior brilho e o brilho residual, quando se fala de preto, pode ser na casa de dois mil. Eu assisti quando estava sendo desenvolvido o sistema a uma demonstração o engenheiro que liderou o projeto disse para colocar o projetor no modo escuro, é um projetor DCI. E ele se colocou na frente da telha e disse que você está vendo minha sombra, é residual. Esse era um 2000:1. E passamos para o outro sistema de projeção de Dolby, hoje se chama Dolby Vision, realmente escuro porque é 1000000:1. Ou seja, o residual é uma milionésima da luz plena. E por outro lado alcança níveis de iluminação enormes muito grande, por exemplo, 3, 4 vezes a iluminação maior, o branco branco completo e este afeta também as cores, um ganho maior.

Isso está sendo disponibilizado, vai ser disponibilizado em breve para os cinemas esse tipo de projeção. Ou seja, não necessariamente numa sala Dolby, qualquer sala vai poder ter acesso a isso, em Atmos e Dolby Vision, em cinema. Se você tem a oportunidade de assistir TV em sua casa e se a TV tem o sistema Dolby Vision, hoje em dia tem muito conteúdo que está em HDR, é muito comum você nos delivery para plataforma encontrar HDR e Atmos, porque entre plataformas existe essa concorrência, de oferecer esses recursos. E é incrível, tecnicamente, que antigamente se perdia em preto ou escuro, uma coisa assim, você tem um lance de cinzas diferente de dias.

Mas dito essas coisas todas de técnica, embora eu esteja atirando no meu próprio pé, eu prefiro ser honesto. Um bom cinema é uma boa história bem contada, essa é a essência do cinema. Vender a ideia, colocar quem está assistindo nesse outro mundo que você está criando. E o resto ajuda, ajuda sim, mas a essência é essa, não pode perder a questão básica que é isso, é como contar uma história. Me conta uma história, como criança, me conta uma história que eu gosto. Me engana que eu gosto.

É isso que fazemos, precisamos disso, por isso é fantástico poder criar essa fantasia. E Dolby continuará, digamos, com todas inovações, é uma empresa de engenharia, com muito pessoal, muito talentoso, desenvolvendo projetos de coisas variadas. Eu sou um implementador, não estou na parte de desenvolvimento, mas tento fazer pontes, como estou fazendo agora parte dessa tarefa de tentar de alguma forma entender um pouco do que está acontecendo.

Raphaela: Carlos, Eu te agradeço muito tanto pela disponibilidade rapidíssima para marcar essa entrevista quanto toda a nossa conversa aqui.

Carlos: Eu também é um prazer, espero ter sido útil. Sorte na tua tese. Gostaria que você depois me mantenha informado como caminha seu trabalho.