

Fábio Wanderley Janhan Sousa

**Mixagem de áudio multicanal no formato
surround 5.1 e sua transcrição para sistemas
binaurais**

**UFMG
Belo Horizonte
2010**

Fábio Wanderley Janhan Sousa

**Mixagem de áudio multicanal no formato
surround 5.1 e sua transcrição para sistemas
binaurais**

**Dissertação apresentada ao
Programa de Pós-
Graduação em Música da
Escola de Música da
Universidade Federal de
Minas Gerais para a
obtenção do título de
Mestre em Música.**

**Linha de Pesquisa:
Sonologia**

**Orientador:
Prof. Dr. Sérgio Freire**

**Escola de Música da UFMG
Belo Horizonte
2010**

S725m Sousa, Fábio Wanderley Janhan

Mixagem de áudio multicanal no formato *surround* 5.1 e sua transcrição para sistemas binaurais / Fábio Wanderley Janhan Sousa.
--2010.

191 fls., enc.; il.

Acompanha um Digital Video Disc.

Dissertação (mestrado em Música) – Universidade Federal de Minas Gerais, Escola de Música.

Orientador: Prof. Dr. Sérgio Freire Garcia.

1. Acústica – estudo. 2. Sistemas binaurais – estudo. 3. Som – estudo. I. Título. II. Garcia, Sérgio Freire - III. Universidade Federal de Minas Gerais. Escola de Música.



Universidade Federal de Minas Gerais
Escola de Música
Curso de Pós-Graduação em Música

Dissertação defendida pelo aluno **Fábio Wanderley Janhan Sousa** em 20 de agosto de 2010 e aprovada pela Banca Examinadora constituída pelos Professores:

Prof. Dr. Sérgio Freire Garcia
Orientador
Universidade Federal de Minas Gerais

Prof. Dr. Régis Rossi Alves Faria
Universidade de São Paulo

Prof. Dr. Fernando de Oliveira Rocha
Universidade Federal de Minas Gerais

Agradecimentos

Em primeiro lugar gostaria de agradecer aos meus pais por sempre apoiarem as minhas escolhas e por me fazerem a pessoa que sou hoje. Do meu pai, Janei Janhan, herdei a persistência, força e coragem para enfrentar os desafios que ultrapassei e os que ainda virão. Da minha mãe, Walkiria, herdei a paciência, inteligência e compreensão que me fizeram chegar até aqui de forma consciente e bem sucedida sem deixar lacunas ou falhas para trás. Com certeza pertence a eles a maior parte do mérito do meu sucesso.

Gostaria de agradecer também à minha irmã, Diana, que sempre esteve perto nos momentos em que precisei e sempre foi uma ótima companheira. À minha namorada, Juliana, agradeço, além do companheirismo de mais de uma década, por me completar de forma a estabelecer o equilíbrio que qualquer um precisa para alcançar sucesso e felicidade. Agradeço também à minha vó, Josefina, e ao meu primo Wagner, pela amizade desde sempre.

Agradeço aos meus professores que, durante todo o meu curso de graduação e mestrado, me orientaram muito bem nesse longo caminho de descobertas. Em especial agradeço ao professor e meu orientador Sérgio Freire, por sempre acreditar em mim, pela amizade e orientação, e aos professores Fernando Rocha e Régis Rossi pela colaboração e participação na defesa dessa dissertação.

Agradeço em especial ao professor e ex-diretor da Escola de Música, Lucas Bretas, pelo apoio desde o meu terceiro período da graduação às minhas atividades no estúdio de gravações da escola como aluno e professor do Centro de Extensão da Escola de Música (CENEX). À professora Margarida Borghoff agradeço pela confiança, apoio e amizade que construímos durante as gravações do projeto Viva Música e do selo da escola. Agradeço também ao pessoal da rádio, em especial ao Judson, ao Elias, ao Paulo César e ao Pacífico, que tiveram parte na minha educação através da parceria nos programas de rádio. Ao pessoal do almoxarifado, em especial ao Sérgio e ao Gilson, agradeço pelo suporte e amizade desses últimos cinco anos.

Agradeço às amizades que surgiram no CEFET-MG, ao Matheus, à Flavinha, ao Thiago, à Camila e à todos os outros, e que até hoje me proporcionam momentos de prazer e troca de idéias mesmo que sendo

momentos um pouco raros. Agradeço às novas amizades que encontrei na Escola de Música, muitas vezes originadas de uma gravação, mas que marcaram momentos muito agradáveis no estúdio ou fora dele.

Gostaria de agradecer também a todos os meus ex-alunos do curso do CENEX que, desde 2007 são uns dos maiores responsáveis pelo meu aprendizado. Por último mas não menos importante gostaria de agradecer a todos os que presenciaram a defesa da minha dissertação e compartilharam desse momento de conquista muito importante e que apenas dá início a muitas outras.

“Nature has given man one tongue, but two ears,
that we may hear twice as much as we speak”
(Greek philosopher Zeno)

Resumo

O tema principal deste trabalho é o processo de mixagem no formato *surround* 5.1 e sua transcrição para um sistema binaural. O termo transcrição aqui utilizado indica que o formato binaural foi obtido a partir dos seis canais gerados inicialmente na mixagem multicanal 5.1.

A primeira parte desse trabalho é dedicada a uma revisão teórica no que diz respeito à espacialização do áudio, iniciando-se com a percepção sonora espacial nas perspectivas física e psico-acústica, seguida das possibilidades oferecidas pelo processo de digitalização do áudio e as diversas técnicas de mixagem e microfonação. A análise específica do sistema *surround* 5.1 e suas técnicas de gravação são assuntos de uma segunda parte do trabalho.

O terceiro capítulo se inicia ilustrando, através de obras musicais e estratégias cinematográficas que exploram a espacialização do áudio, as diferentes abordagens estéticas adotadas no processo de mixagem. Em seguida, essa terceira parte apresenta e discute alguns casos, realizados pelo autor, de gravação e mixagem de concertos musicais tradicionais (baseados em uma abordagem de ambientação) e de música contemporânea (que exploram o espaço e a movimentação) em *surround* 5.1. Também são descritas tentativas de se transcrever o material obtido na mixagem multicanal em 5.1 para um formato binaural, utilizando-se o objeto spat~ (MAX/IRCAM) e *plugins* similares.

Por último é feita uma comparação entre as características espaciais observadas em ambos os formatos e são apresentadas algumas observações que dizem respeito às vantagens e desvantagens da transcrição realizada para o sistema binaural, baseadas em questões práticas, audição pessoal e em comentários de professores e colegas.

Palavras chave: Mixagem de áudio, Multicanal, 5.1, Surround, Binaural, spat~, MAX/IRCAM.

Abstract

The mixing processes in surround 5.1 and its transcription into binaural formats are the main subject of this paper. The term transcription used here indicates that the binaural format is built on the set of six channels delivered by the surround mix.

The first part of this work is devoted to a theoretical review of spatial audio, starting with the human spatial perception of sounds in both physical and psychoacoustical perspectives, followed by the possibilities offered by digital audio and diverse miking and mixing techniques. The specific analysis of surround 5.1 systems and its recording techniques comes next.

The third chapter firstly discusses diverse aesthetical approaches to the mixing process, based on works of music and cinema post-production strategies that explore spatial audio, the varied aesthetical approaches to the mixing process. Then it presents and discusses some case studies based on the author's practice of recording and mixing traditional concert music (based on the ambience approach) and contemporary music (that explores the surrounding space and movement) in surround 5.1. Tentative efforts to transcribe these multichannel surround mixings into the binaural format using the spat~ from MAX/MRCAM and similar plugins are also described.

At last, a comparison between the perceptible spatial characteristics of both formats is done and some conclusions about the advantages and disadvantages of the binaural transcription are given, based on practical issues, personal listening experience and comments of colleges and teachers.

Keywords: Audio mixing, Multichannel, 5.1, Surround, Binaural, spat~, MAX/MRCAM.

Sumário

Introdução.....	1
Cap. 1 - Escuta dos espaços sonoros acústicos e eletroacústicos.....	5
1.1 - Bases físicas e psico-acústicas da percepção espacial auditiva.....	6
1.2 - Possibilidades e aplicações musicais da espacialização.....	28
1.3 - Sistemas de reprodução de áudio.....	35
1.3.1 – O Sistema Mono.....	37
1.3.2 – O Sistema Estéreo.....	38
1.3.3 – O Binaural.....	44
1.3.3.1 – Síntese de <i>surround</i> 3D.....	56
1.3.4 – O Ambiophonics.....	61
1.3.5 – O Sistema Quadrifônico.....	67
1.3.6 – O Sistema Multicanal 5.1 e suas variações.....	68
1.3.7 – O Ambisonics.....	72
1.3.8 – <i>Wave Field Synthesis</i>	77
Cap. 2 – O sistema <i>surround</i> 5.1: entre o ideal e o viável.....	79
2.1 – Histórico e desdobramentos.....	79
2.2 - Métodos de gravação e mixagem em <i>surround</i> 5.1.....	91
2.2.1 – Formatos de Microfonação.....	91
2.2.2 – O controle panorâmico no estéreo e no <i>surround</i>	111
2.2.3 – A questão do monitoramento.....	125
Cap. 3 – Estudo de casos: análises, práticas de mixagem em 5.1 e transcrições para o sistema binaural.....	133
3.1 - Recriação de espaços e movimentos em gravações <i>surround</i> 5.1: apreciação estética de trechos de filmes e produções musicais.....	133
3.2 – Mixando em 5.1: a teoria na prática.....	152
3.3 – Transcrevendo para o sistema binaural.....	161
3.3.1 – O objeto spat~ do MAX/MSP e o KLT3D – transcrevendo o áudio em 5.1 para outros formatos.....	168
3.3.2 – Observações a partir da escuta.....	175
Conclusões e futuras pesquisas.....	177
Bibliografia.....	179
Anexo: DVD.....	191

Introdução

Em seu livro sobre o rádio de 1933¹, ARNHEIM, ao discorrer sobre a imagética do ouvido, já nos chama a atenção para o fato de que, na percepção do mundo através dos nossos sentidos, a audição é a única que está sempre ativa. Apesar de muitas das coisas que compõem o mundo não produzirem som algum, os sons que chegam aos nossos ouvidos, vindos das mais diversas fontes sonoras nos trazem informações suficientes para reconstruir um mundo inteiro, denominado pelo autor de “mundo acústico”². A partir desse conceito o autor define o que seriam as artes aurais como aquelas que dependem do tempo para se estabelecerem e coloca a música, assim como o cinema, o rádio e o teatro, nessa classificação.

Mais recentemente, BREGMAN (1990), autor importante no desenvolvimento da psico-acústica, ao abordar a análise de “cenas sonoras”, faz referências a dois tipos de questões, relacionadas entre si: a primeira é perceptiva, relacionada com o modo pelo qual o nosso sistema auditivo é capaz de criar uma imagem do mundo que nos cerca através da percepção dos sons. A segunda é de caráter ecológico, e tem a ver com o modo pelo qual o meio ambiente cria e modifica os sons e, conseqüentemente, o mundo a nossa volta³.

Essas preocupações com a capacidade da nossa percepção em criar uma imagem de mundo advinda da escuta não se limitam a conceituações teóricas sobre os meios de comunicações ou a pesquisas sobre a cognição humana, mas também acompanham o desenvolvimento dos sistemas de gravação que surgiram com o desenvolvimento da eletro-eletrônica no início do século XX. Tendo como peça fundamental o fonógrafo e seus sucessores e, com ele, a possibilidade de se capturar e manipular os sons, esses sistemas possibilitaram a criação de inúmeros produtos e processos voltados para a música, assim como a possibilidade de recriação de espaços sonoros.

¹ Escrito originalmente em alemão em 1933 e publicado pela primeira vez em inglês somente em 1936.

² ARNHEIM, 1986, p.23.

³ BREGMAN, 1990, p.1.

Desde o início do desenvolvimento dos sistemas de gravação tentou-se imprimir nas gravações aspectos espaciais do som, ou seja, características que nos fizessem distinguir as posições das diversas fontes sonoras. Inicialmente, nos sistemas de reprodução monofônicos só se conseguia controlar os parâmetros referentes à distância de determinada fonte sonora, o que foi ampliado no sistema estereofônico para a representação de direções em um plano horizontal de 180°.

O desenvolvimento da espacialização do áudio teve seqüência com o sistema quadrifônico, que não obteve êxito devido à ausência de potencial comercial. A respeito desse sistema podemos citar o que diz Schafer em sua obra “A afinação do mundo”:

“Quando, depois da Segunda Guerra Mundial, o gravador fez incisões em um possível material gravado, podia-se cortar qualquer objeto sonoro e inseri-lo em qualquer novo contexto desejado. Mais recentemente, o sistema de som quadrifônico tornou possível uma paisagem sonora de eventos sonoros estacionários ou em movimento de 360 graus, o que permite simular no tempo e no espaço qualquer som do ambiente, como também permite a completa transposição do espaço acústico. Qualquer ambiente sonoro pode agora transformar-se em qualquer outro ambiente.” (SCHAFFER, 2001, p. 134).

Com as novas ferramentas oferecidas pela gravação sonora, o cinema foi o primeiro a explorar o espaço sonoro de uma forma mais ampla. Com a apresentação multicanal do filme Fantasia da Disney em 1939⁴, uma proposta de se reproduzir sons posicionados nos mais diversos lugares (numa tentativa de se aproximar da complexidade da audição humana) mostrou-se um pouco ousada e muito problemática face aos recursos da época, principalmente em relação à comercialização desse formato.

Apesar disso, o áudio do cinema continuou se desenvolvendo e, hoje, a reprodução multicanal, principalmente no formato 5.1 e a espacialização tridimensional do áudio se tornaram possíveis e muito utilizadas pela indústria cinematográfica devido, principalmente, ao desenvolvimento do processamento

⁴ RUMSEY, 2001, p.12.

digital do som e ao barateamento dos sistemas multicanais de reprodução de áudio.

Esse novo panorama nos abre possibilidades de uso desse sistema também para a música que, por um motivo puramente comercial se limitou ao sistema estéreo de reprodução durante mais de cinquenta anos. Com a popularização do formato *surround* 5.1 e o barateamento do seu sistema de reprodução, favorecendo o uso doméstico do mesmo, vemos a expansão das possibilidades de produção musical mantendo o mesmo nível de comercialização do sistema estéreo e nos proporcionando uma maior liberdade criativa e uma maior imersão no ambiente sonoro⁵.

Uma das grandes preocupações dos sistemas de gravação e reprodução anteriores estava centrada na reprodução da qualidade tímbrica e dinâmica em uma gravação, parâmetros que atualmente não estão mais em questão devido ao grande desenvolvimento dos últimos 40 anos, onde a relação sinal-ruído, a distorção harmônica e a resposta em frequências já não são problemas reais para a maioria das pessoas. A qualidade técnica dos sistemas de gravação sonora nos possibilita, atualmente, realizar uma gravação capaz de capturar toda a faixa dinâmica e espectral que podemos perceber⁶.

Uma vez que essa qualidade já foi colocada em questão em diversos outros trabalhos, que questionam também as nossas limitações de escuta⁷, e que a qualidade de reprodução espacial do som é, então, possivelmente o único fator restante a ser atacado nessa busca para a qualidade final de reprodução sonora⁸, partiremos logo de questões referentes à espacialização do áudio, nos direcionando a um estudo mais abrangente no que diz respeito às possibilidades criativas dos novos sistemas *surround*.

Neste trabalho faremos, no primeiro capítulo, um levantamento, fundamentado na bibliografia existente, de como se dá, no processo de escuta, a localização espacial de fontes sonoras a partir de princípios físicos e psico-acústicos e, principalmente quais os fatores envolvidos nesse processo e suas variáveis. Logo em seguida veremos do que os diversos sistemas de

⁵ RUMSEY, 2001, p.IX.

⁶ Idem, 2001, p.X.

⁷ Dentre os quais MEYER, 2007.

⁸ RUMSEY, 2001, p.X.

reprodução sonora são capazes e quais os seus principais problemas com relação à reprodução do espaço sonoro.

No segundo capítulo, focando mais no sistema *surround* em 5.1, estudaremos as diversas ferramentas disponíveis para uma produção de áudio nesse sistema, passando pelos formatos de microfonação, questões de monitoramento e formas de controle panorâmico.

No último capítulo, após realizar uma breve análise de algumas produções musicais e cinematográficas observando detalhes estéticos adotados por elas, descrevemos alguns trabalhos de gravação e de mixagem em 5.1 do presente autor, que aplicam os conceitos e ferramentas discutidas no capítulo anterior. Em seguida, buscamos, através de uma transcrição do material trabalhado em sistemas 5.1 para apenas dois canais (sistema binaural), manter a espacialização possibilitada pelo formato multicanal valendo-se de um sistema mais acessível, o já muito bem estabelecido formato estéreo, no entanto direcionado para fones de ouvido. Para tanto utilizamos tanto as gravações analisadas quanto as produzidas durante a pesquisa. Com o material obtido dessa transcrição fazemos observações e mencionamos questões a respeito da compatibilidade entre ambos os formatos a partir da audição dos mesmos e de comentários de colegas.

Capítulo 1

Escuta dos espaços sonoros acústicos e eletroacústicos

O som, como todos nós sabemos, é transmitido pelo ar como ondas longitudinais de pressão que se expandem a partir da fonte sonora e perdem potência ao se propagar. Um som se propaga nas mais diversas direções e, quando encontra algum obstáculo, interage com o mesmo sofrendo reflexões e difrações. Normalmente ocorrem combinações de todos esses processos originando um espectro sonoro proveniente dessa interação do som com os objetos, e que varia de acordo com as propriedades físicas de cada um. Essas interferências construtivas e destrutivas de todas as ondas resultantes criam uma rica mistura acústica que se enriquece mais ainda quando existem várias fontes sonoras. A natureza dessa interação pode variar também com o ângulo de incidência do som nesses materiais⁹.

Um dos principais obstáculos encontrados pelo som é o ouvinte, que recebe várias ondas sonoras chegando em diferentes instantes e de diferentes direções. O som direto, que não sofreu nenhuma interferência antes de chegar ao ouvinte, é portador de informações referentes à posição da fonte sonora em relação ao mesmo, e o som indireto, proveniente das reflexões do som nos outros objetos do ambiente, é portador de informações referentes ao ambiente em que o ouvinte se encontra e à posição da fonte em relação aos outros objetos¹⁰.

Em decorrência da interação das ondas sonoras com o ar, através do qual elas se propagam, ocorre a perda de altas frequências proporcionalmente à distância, como resultado da absorção das ondas pelo vapor de água existente no ar. Quando tentamos fazer com que um som se pareça real ou plausível ao manipulá-lo espacialmente, temos que entender o seu

⁹ MALHAM, 1998.

¹⁰ KENDALL, 1995.

comportamento no mundo real tanto no que diz respeito aos fatores que modificam o seu timbre quanto ao funcionamento da nossa percepção espacial do mesmo¹¹.

O som, de forma geral, é percebido espacialmente pelo ouvido humano principalmente com relação a dois aspectos, sua posição no espaço e seu tamanho. Com relação à posição no espaço, principal objeto desse trabalho, o som, na natureza, possui informações nas três dimensões: horizontal, vertical e profundidade, e essas informações são usadas para se perceber sons provenientes de todas as direções. Obviamente utilizamos mais a visão para identificar fontes sonoras à nossa frente e a audição para fontes sonoras localizadas atrás ou encima, estas últimas, apesar de se apresentarem mais difíceis de serem simuladas, também se valem de alterações em determinados parâmetros que veremos a seguir.

1.1 – Bases físicas e psico-acústicas da percepção espacial auditiva

Todo som que percebemos se apresenta para nós de forma tridimensional, ou seja, proveniente de todas as direções, de cima, de baixo, da nossa frente, de trás e dos nossos lados.

Em ambientes sonoros, “cenários sonoros¹²” ou “paisagens sonoras¹³”, percebemos diversas fontes sonoras, cada uma delas com sua própria posição e seus diversos atributos. Paisagens sonoras mais complexas nos fazem perceber certa mistura de sons, dificultando a separação entre os elementos e resultando no agrupamento de fontes sonoras. Em alguns casos essas misturas de sons nos trazem informações gerais sobre o ambiente e nos fazem perceber o espaço de uma forma generalizada sem uma forte noção de direcionalidade ou localização.

Podemos distinguir claramente, no entanto, um som produzido em um ambiente aberto de um som produzido dentro de uma sala, principalmente pela ausência de reflexões. Em ambientes fechados, as reflexões têm uma função

¹¹ MALHAM, 1998.

¹² BREGMAN, 1990, p.1.

¹³ Termo definido por Schafer como “qualquer campo de estudo acústico. Podemos referir-nos a uma composição musical, a um programa de rádio ou mesmo a um ambiente acústico como *paisagem sonora*.” (SCHAFFER, 2001, p.23).

muito importante, a de nos dar informações sobre o tamanho da sala, a distância dos objetos em relação ao ouvinte e às outras superfícies da sala e, inclusive, informações sobre o material de cada superfície. Muitas vezes pode ocorrer também de as reflexões serem tão próximas do som direto que se torna complicado identificar informações espaciais sobre a localização da fonte sonora e essas passam a compor o próprio timbre do som. De acordo com as características desses sons, compostos de som direto e de reflexões, costuma-se categorizá-los entre fonte sonora e ambiente, este último capaz de criar a noção de espacialidade que não está relacionada a uma fonte sonora específica¹⁴. Schafer faz uma interessante analogia entre o que percebemos auditivamente e visualmente, e que nos ajuda a compreender o trabalho realizado durante o processo de mixagem de qualquer produção fonográfica:

“É realmente possível que alguns termos empregados na percepção visual possam ter equivalentes na percepção auditiva. (...) De acordo com os psicólogos da Gestalt, que introduziram a distinção, a figura é o foco de interesse, e o fundo, o cenário ou contexto. A isso foi acrescentado, mais tarde, um terceiro termo, *campo*, significando o lugar onde ocorreu a observação. (...) A *figura* corresponde ao sinal ou marca sonora. O *fundo* corresponde aos sons do ambiente à sua volta – que podem, com frequência, ser sons fundamentais – e o *campo*, ao lugar onde todos os sons ocorrem, a paisagem sonora. (...) Se quisermos perseguir o tema figura/fundo em termos de percepção auditiva, deveremos fixar os pontos em que uma figura acústica é abandonada para tornar-se um fundo não percebido, ou quando um fundo surge subitamente como figura – um evento sonoro, um marco sonoro, uma experiência acústica notável ou vital. (...) Os termos *figura*, *fundo* e *campo* fornecem uma estrutura para organizar a experiência.” (SCHAFER, 2001 p. 214 e 215).

Para perceber e reconhecer a posição dos sons no ambiente utilizamos diversos mecanismos de escuta. A nossa percepção do som no espaço é baseada na interpretação de inúmeros parâmetros que são extraídos do campo sonoro que nos circunda. É possível, no entanto, que esses parâmetros se mostrem ambíguos ou conflituosos devido à complexidade dos diversos campos sonoros. Nessas circunstâncias pode acontecer da direção e a

¹⁴ RUMSEY, 2001, p.2.

distância de um som, seja ele gravado ou sintetizado, não corresponder à sua distância e direção reais. Devemos observar que, ao fazer esse julgamento, o cérebro atribui um valor para cada um desses parâmetros de acordo com sua aparente ambigüidade e é esse fator que nos possibilita construir campos sonoros artificiais simplificados, mas convincentes¹⁵.

Segundo MALHAM (1998), mesmo com o advento da tecnologia digital e do crescimento do poder de processamento dos computadores, ainda não somos capazes de recriar um campo sonoro original, ou reconstruí-lo, devido à tamanha complexidade e extensão espacial ocupada pelo som que ouvimos, mas podemos sim tomar uma pequena área e aproximá-la do que nos pareceria real.

De uma forma generalizada o ouvido humano percebe a localização de uma fonte sonora através de uma combinação entre as diferenças de tempo (ITD – *interaural time difference*) e nível (ILD – *interaural level difference*) sonoro entre os dois ouvidos e do espectro sonoro do som que chega a cada um deles de forma diferente (HRTF – *Head Related Transfer Functions* – Funções de Transferência Relativas à Cabeça)¹⁶.

Quando uma onda sonora chega ao ouvinte existem também outros dois fenômenos acústicos dependentes da frequência. As altas frequências são refletidas e as frequências mais graves, normalmente abaixo de 1500 Hz, são difratadas e circundam o ouvinte¹⁷. Devido a esse fenômeno podemos dizer que o som emitido pelas mais diversas fontes sonoras tendem a ser mais direcionais de acordo com o aumento de sua frequência e a serem mais omnidirecionais quanto apresentam frequências mais graves.

Segundo KENDALL (1995) e PLACK (2005), tomando a cabeça do ouvinte como referência, podem-se determinar três planos ortogonais, a saber: planos médio, horizontal e frontal. Qualquer evento sonoro equidistante aos dois ouvidos está localizado no denominado plano médio. No plano horizontal não possuímos informações referentes à elevação da fonte sonora, já o plano frontal possui informações referentes à lateralidade.

¹⁵ MALHAM, 1998.

¹⁶ RUMSEY, 2001.

¹⁷ KENDALL, 1995.

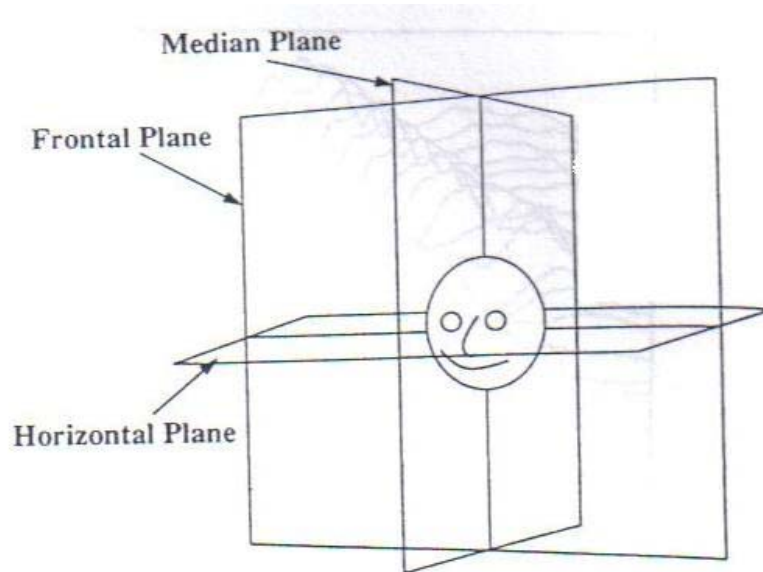


Figura representando os planos médio, frontal e horizontal. Fonte: KENDALL, 1995.

Se as fontes sonoras não são eqüidistantes aos dois ouvidos, a sua posição passa a ser definida por duas grandezas angulares (elevação e azimuth) e uma grandeza escalar (distância).

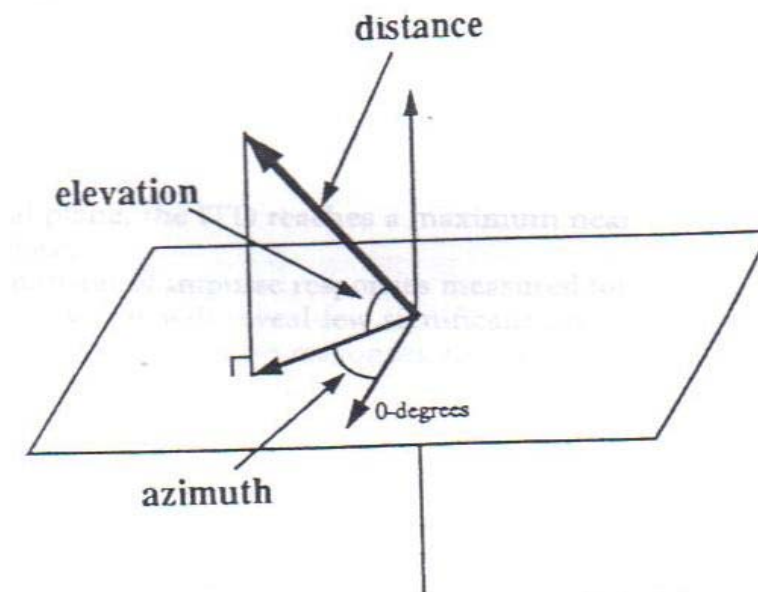


Figura representando as grandezas angulares e escalar de posicionamento de fontes sonoras. Fonte: KENDALL, 1995.

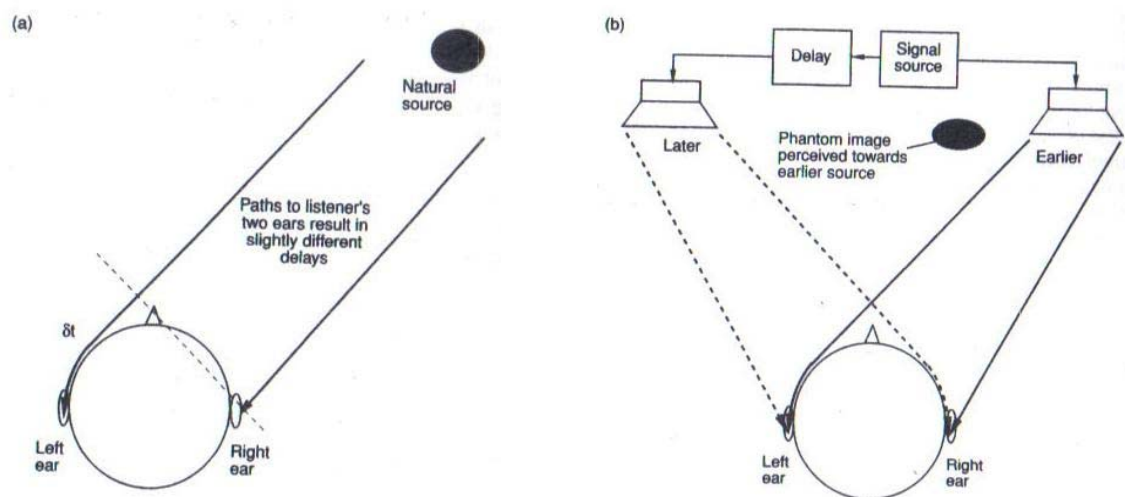
Ainda segundo KENDALL (1995), existe uma grande diferença na resolução em que uma pessoa pode definir a posição de uma fonte sonora nas dimensões verticais e horizontais. A maior resolução se encontra na dimensão horizontal, onde podemos especificar mudanças de posição com uma

resolução de 2° ou menos para sons frontais, este ângulo aumenta para 10° na resolução de sons nas laterais e 6° para sons atrás. Já a resolução da nossa percepção na dimensão vertical é extremamente baixa, de 9° na nossa frente e aumentando à medida que passa sobre a nossa cabeça para até 22°.

Podemos definir os principais parâmetros espaciais de localização de uma fonte sonora como se segue:

A) ITD (*interaural time delay* – atraso de tempo entre os ouvidos):

Qualquer som que esteja situado no plano sonoro médio, vertical em relação à cabeça do ouvinte, terá sua onda sonora chegando aos dois ouvidos simultaneamente. No entanto qualquer movimento dessa fonte sonora que saia desse plano fará com que as ondas sonoras cheguem primeiro a um deles e, depois, ao outro. Esse parâmetro de definição espacial do som é conhecido como “tempo de atraso entre os ouvidos” (ITD). Segundo BEGAULT (1994), a menor diferença que pode ser percebida depende da natureza do som, podendo variar entre 5 μ s e 1.5ms. A figura abaixo, extraída de RUMSEY (2001), ilustra esse fenômeno da forma como acontece na natureza (a) e como ele pode ser reproduzido em um sistema estéreo, baseado também no princípio do efeito de precedência, que veremos mais a frente (b):



Figuras ilustrando o posicionamento de fontes sonoras em (a) ambientes naturais e (b) através de alto-falantes. Fonte: RUMSEY, 2001.

B) Efeito de precedência:

Basicamente, o julgamento do ouvinte a respeito da origem do som e consequentemente da localização da fonte sonora é predominado pelo som que chega primeiro aos seus ouvidos, através do caminho mais curto e direto. Esse efeito é denominado *efeito de precedência* ou *lei da primeira frente de ondas*.

STERN, BROWN e WANG (2006) definem como efeito de precedência o fenômeno de que a primeira frente de ondas e os parâmetros direcionais que chegam com o som direto têm mais influência na nossa percepção de direcionalidade do que o som refletido.

Segundo RUMSEY (2001) e WALLACH (1973), este efeito se mostra de forma mais intensa quando relacionado a sons transientes do que a sons contínuos. Em sistemas de reprodução onde existem pelo menos duas fontes sonoras em diferentes posições emitindo duas versões de um mesmo som, talvez com a amplitude ou o tempo ajustado para transmitir informações direcionais, ambos os ouvidos percebem todos os alto-falantes e o cérebro tende a localizar a fonte sonora baseado no tempo de atraso que chega na primeira frente de ondas: a fonte vai parecer vir da direção de onde o primeiro sinal saiu. Uma forma de efeito de precedência muito conhecida e mencionada por RUMSEY é a referida como efeito *Haas*.

WALLACH (1973), em seu estudo a respeito do fenômeno de que em salas reverberantes ouvimos apenas um único som e não a longa seqüência de sons sucessivos e separados como ocorre fisicamente, afirma que a repetição de estímulos semelhantes e em rápidas seqüências nos leva a uma experiência sonora única que é qualitativamente não muito diferente da experiência de um único estímulo.

As experiências relatadas no seu trabalho dizem respeito tanto às interpretações referentes à amplitude quanto ao tempo. Um alto-falante mais próximo envia o sinal de áudio tanto mais rápido em relação ao outro quanto em maior intensidade e ambos os fatores contribuem para a localização espacial de fontes sonoras.

WALLACH apresenta a generalização de que se dois sons são emitidos próximos um ao outro, em uma seqüência próxima (entre 1ms e 5ms para sons com características transientes e entre 1ms e 40ms para sons mais

complexos), eles serão ouvidos como um único som, e sua resultante sonora será percebida espacialmente na posição do primeiro som emitido. Essa localização proveniente do primeiro som percebido é denominada pelo autor de *efeito de precedência*.

Entre as observações feitas em seu artigo está o fato de que *clicks* são mais precisamente localizados do que sons contínuos ou, talvez, eles constituam eventos discretos e possuam como resultado uma representação diferente no sistema nervoso. Outra observação apresentada no artigo diz respeito ao fato de que o intervalo de tempo em que ocorre a fusão de dois sons sucessivos se dá até 35ms (dependendo da característica do som em questão), e essa fusão começa a ser desfeita completamente depois dos 70ms, no entanto tais intervalos podem variar e são determinados, além de suas características transientes, também pela reverberação presente no ambiente.

Segundo o autor, o efeito de precedência pode ser facilmente demonstrado quando os sons apresentam certa descontinuidade ou características transientes. Sons contínuos ou ruídos uniformes obviamente não apresentam tal efeito pois não há como definir precedência. *Clicks*, assim como a fala ou música para piano podem ilustrar o efeito satisfatoriamente bem.

Na pesquisa de WALLACH, ele cita as considerações de STEINBERG e SNOW (1934) e BOER (1940), que afirmam que “a percepção normal de direção sonora depende principalmente das diferenças de amplitude” e, assim, descartam as diferenças de tempo, uma vez que, para diversos engenheiros, essas diferenças de tempo podem ser reduzidas a diferenças de fase. Ele menciona também o trabalho de SHAXBY e GAGE (1932), onde eles mencionam que o efeito de intensidade que ocorre em uma diferença de intensidade de 3dB corresponde a uma diferença de 5ms de tempo. Estabelecendo uma relação numérica entre essas diferenças e a variação angular do posicionamento da fonte sonora, 3dB corresponderiam a uma variação de 0,5° no plano horizontal, no entanto no trabalho de STEINBERG e SNOW o sujeito da pesquisa relatou uma variação de 17°, concluindo que existem mais parâmetros na localização de fontes sonoras além da diferença de amplitude.

No artigo de WALLACH, as experiências comprovam a existência do efeito de precedência, no entanto de forma relativa e não absoluta. Em estudos anteriores relatados no artigo, BÈKÉSY (1930), afirma que se o intervalo entre os dois pares de estímulos era superior a 1ms, há a supressão do segundo par. Em intervalos menores do que 1ms há combinações ou indicações de localização duvidosas. Em outros estudos citados, LANGMUIR (1944), afirma que o efeito de precedência é ilustrado de forma semelhante àquilo que acontece em uma simples experiência em salas com ecos e reverberações onde o ouvinte presta mais atenção à voz do pregador do que na sua repetição vinda em uma fração ou um segundo atrasada.

Concordando com o fato de o tempo não ser o único fator que define o posicionamento de uma fonte sonora, WALLACH afirma que a intensidade dos dois sons apresentados para medir o efeito de precedência também tem influência sobre a localização. Se o segundo som apresentar uma intensidade superior ao primeiro de cerca de 15dB, ele sobrepõe o efeito de precedência e supõe que, se o segundo som se apresenta com menor intensidade, o efeito de precedência permanece. Outra importante observação do autor diz respeito ao fato de que o efeito de precedência acontece se os dois sons envolvidos são relativamente similares. Segundo BÈKÉSY (1930), mencionado por WALLACH, sons distintos mantêm a individualidade de suas localizações.

Para concluir, o autor diz que o efeito de precedência nos auxilia no entendimento da nossa capacidade de localizar sons e tem sua principal importância na explicitação da localização de fontes sonoras em ambientes muito reflexivos, além de se tornar mais inteligíveis espacializações em sistemas estéreo onde sons reproduzidos em apenas dois canais recriam ilusões de fontes sonoras posicionadas em mais de duas direções.

C) ILD (*interaural level difference* – diferença de intensidade entre os ouvidos):

Também muito referido como IID (*interaural intensity difference*), pode ser explicado como o fenômeno de que sons provenientes do lado direito da cabeça, por exemplo, chegarão diretamente ao ouvido direito, no entanto sofrerão difração em torno da cabeça para chegar ao ouvido esquerdo. Sua amplitude será reduzida com relação à intensidade recebida no ouvido direito, em parte devido à absorção do som pela cabeça e em parte devido à maior

distância percorrida. Esse outro parâmetro é denominado “diferença de intensidade entre os ouvidos” (ILD).

Ambos os fatores ITD e ILD, importantes na definição da posição de uma fonte sonora, se fundamentam no fato de que toda onda sonora chega aos dois ouvidos do ouvinte de formas diferentes em relação ao tempo de chegada e à intensidade, respectivamente, e também no fato de que se modificam de acordo com a frequência da onda sonora que chega aos nossos ouvidos.

Segundo KENDALL (1995):

“(...) experimentos demonstraram que as diferenças interaurais, que são as diferenças nos sinais acústicos que chegam simultaneamente nos dois ouvidos, são extremamente importantes na percepção espacial do som; o ILD e o ITD possuem impactos significativos na percepção sonora em faixas de frequências específicas. Acima de 1500 Hz ocorre um efeito de sombra acústica pela cabeça e a definição de localização sonora é dominada pela diferença de amplitude ou intensidade do sinal dos dois ouvidos (ILD)¹⁸. Abaixo de 1500 Hz a cabeça não é um obstáculo significativo na propagação do som e a localização espacial de uma fonte sonora é predominada pela diferença de tempo entre os ouvidos (ITD)”. (KENDALL, 1995)

Também segundo KENDALL (1995), o ILD e o ITD afetam somente a percepção de lateralidade de uma fonte sonora, ou seja, a percepção de uma fonte no eixo esquerdo/direito dos ouvidos. Somente com as ILD e ITD o ouvinte não pode determinar onde um evento acústico acontece, se acima, em baixo, na frente ou atrás. Esse fenômeno também é mencionado por CHENG e WAKEFIELD (2001), ao afirmarem que esses parâmetros funcionam somente para sons dentro do azimute e que existem inúmeras posições sobre as curvas de iguais distâncias em relação à cabeça onde se tem o mesmo valor de ILD e ITD. Ambos afirmam que essa ambigüidade de localização a respeito da lateralidade, denominada de *cone de confusão*, pode ser desfeita com a utilização dos perfis complexos das HRTFs.

Segundo RUMSEY (2001), na vida real, ambas as diferenças entre tempo e intensidade trabalham em conjunção para produzir uma imagem clara e bem definida do espaço sonoro. Ele afirma que numerosos estudiosos

¹⁸ Nota do autor: talvez esse seja o motivo pelo qual a maioria dos panorâmicos do sistema estéreo se baseie na diferença de amplitude, como veremos em capítulos posteriores.

investigaram os diferentes valores de intensidade em dB que são necessários para compensar determinados milissegundos de diferença de um atraso binaural e parece que isso depende da natureza do som percebido, como já mencionado por WALLACH (1973). Também segundo ele, até certo ponto essa possibilidade de relação existe; no entanto quando o atraso chega ao seu máximo, onde os dois sons começam a ser percebidos separadamente, essa relação deixa de existir.

RUMSEY cita um outro autor, YOST (1971), que afirma que informações transientes possuem mais informações direcionais e o conteúdo grave desses transientes são mais importantes para a localização do que o conteúdo mais agudo. Outros autores citados, WHITWORTH e JEFFRESS (1961), afirmam que existem dois mecanismos para a localização, um deles baseado no conteúdo grave dos transientes (abaixo de 1500Hz) e outro dependente da intensidade do sinal.

RUMSEY menciona também uma pesquisa de MARTIN (1999) sobre o sistema *surround* 5.1, onde ele descobriu que as diferenças de tempo e amplitude necessárias para se criar imagens fantasma¹⁹ dependem de apenas dois alto-falantes, entre os cinco, que estiverem envolvidos. Eles concluíram que a diferença de amplitude entre os canais proporcionam imagens fantasma mais estáveis do que os atrasos temporais. Também mencionado por RUMSEY, RATLIFFE (1974), ao conduzir experimentos similares, desta vez com sistemas quadrifônicos chegou à conclusão de que as imagens fantasmas baseadas em diferenças de amplitude entre pares de caixas eram pobremente localizadas e as fontes pareciam pular rapidamente da frente para trás ao invés de passar de forma suave pelos lados.

D) Percepção de distância:

Segundo MALHAM (1998), os quatro principais sinais indicadores de distância de uma fonte sonora são:

- A razão entre o som direto e o som reverberado. Em ambientes com uma reverberação significativa, a energia do campo reverberante varia para todas as combinações de posições entre ouvinte e fonte sonora, o que significa

¹⁹ Percepção auditiva de uma fonte sonora localizada em algum ponto entre dois alto-falantes.

que, se uma fonte sonora possui determinado nível de reverberação e este aumenta, isso indica um afastamento da mesma, se a reverberação diminui indica uma aproximação. Segundo o autor, esse fator em particular é o que nos impossibilita de posicionar uma fonte sonora estática mais perto do ouvinte do que o alto-falante mais próximo.

- O padrão direcional e o atraso das primeiras reflexões das superfícies do ambiente variam de acordo com a posição do ouvinte e da fonte sonora.

- Progressiva atenuação de altas frequências com a distância que se deve à absorção das mesmas pela atmosfera.

- A redução da intensidade com a distância, devido ao crescimento da área abrangida pelas ondas frontais quando esta se move para longe da fonte sonora.

Ainda segundo MALHAM (1998), tomar a intensidade sonora como um sinal determinante de distância pode ser muito duvidoso a partir do fato de que testes realizados com pessoas em câmaras anecóicas mostraram erros muito grandes nessa estimativa. Isso se dá, pois, como vimos anteriormente, esses não são os únicos meios pelos quais percebemos o espaço sonoro.

O autor lamenta a dificuldade de se trabalhar experimentalmente com sinais captados na cavidade peitoral e com mecanismos de condução óssea, pois acredita que esses fatores (principalmente a cavidade peitoral) podem ter papel importante na percepção de diretividade de frequências graves. HAMILL e PRICE (2008) são outros autores que citam a condução sonora através dos ossos não somente da cavidade peitoral, mas também do crânio. Eles afirmam que, principalmente quando a intensidade sonora é muito elevada, o crânio e outros ossos do corpo podem vibrar com força suficiente para movimentar o fluido do ouvido interno, estimulando as células sensitivas²⁰.

MALHAM (1998) afirma que, frente a isso, o senso comum de que não podemos determinar com exatidão a direção de fontes sonoras emitindo frequências muito baixas, onde a diferença de fase entre os ouvidos se torna muito pequena, pode ser válido apenas para a escuta através de fones de ouvido. E, segundo ele, isso pode ter implicações nos sistemas de difusão,

²⁰ HAMILL e PRICE, 2008, p.173.

onde o grave se apresenta através de um número limitado de *subwoofers* ou em obras eletroacústicas reproduzidas por fones de ouvido.

E) Distância crítica, primeiras reflexões e reflexões tardias:

Em ambientes fechados, uma porção do som irradiado por uma fonte sonora é absorvida pelas superfícies e pelo ar e outra porção é refletida de volta ao ambiente criando um campo sonoro difuso conhecido também como som ambiente. A figura abaixo, extraída de KENDALL (1995), ilustra o som direto e suas reflexões:

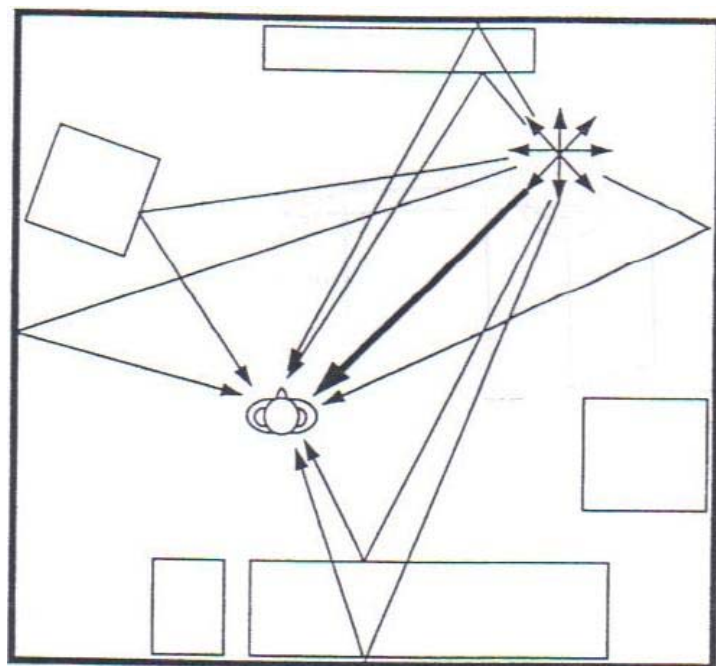


Figura representando o som direto e suas reflexões em ambientes fechados. Fonte: KENDALL, 1995.

RUMSEY (2001) entende como “distância crítica” a distância entre o ouvinte e a fonte sonora onde o nível de som refletido é igual ao nível de som direto e que pode ser calculado a partir do nível de absorção da sala (constante definida pelos materiais empregados na construção das superfícies da mesma) e o fator de diretividade da fonte sonora.

Segundo o autor, as reflexões têm a capacidade de modificar a nossa percepção das fontes sonoras. Às primeiras reflexões são atribuídas a sensação de tamanho e espaço de uma sala, já as reflexões tardias se relacionam com a sensação de espacialidade e envolvimento. Parâmetros tais

como a distância e a profundidade de uma fonte sonora são mais facilmente percebidas em ambientes fechados e estão relacionadas à variação entre o som direto e o som refletido. Devido a isso ele afirma ser interessante conhecer a distância crítica de uma sala na definição das técnicas de gravação²¹.

Em outra parte de seu livro, na qual detalha mais a influência das reflexões na percepção espacial do som, RUMSEY afirma que as reflexões, que ocorrem entre 50 e 80 ms após o som direto têm efeito principalmente sobre a percepção dos parâmetros relativos à amplitude e profundidade de determinada fonte sonora. Reverberações que ocorrem no período acima de 20ms podem gerar certa coloração no timbre se muito acentuadas. E as que ocorrem após 80 ms tendem a contribuir para a sensação de envolvimento ou amplitude espacial (*spaciousness*) do ambiente. Mencionado pelo autor, estudos de GRIESINGER (1997 e 1999) afirmam que reflexões entre 50 e 150 ms são problemáticas na reprodução sonora por comprometer a inteligibilidade e clareza. Apesar delas existirem em espaços reais, GRIESINGER considera importante minimizá-las em reproduções que utilizam reverberações artificiais, concentrando-se no período logo após os 50 ms para simular profundidade e a dimensão da fonte sonora, e acima de 150 ms para criar a noção de amplitude espacial.

RUMSEY (2001) chama a atenção também para o fato de que, em ambientes não-reflexivos, é possível para o ouvinte identificar a distância de uma fonte sonora somente ao se compará-la com uma referência. Já em ambientes reflexivos, a razão entre som direto e reverberado está diretamente relacionada à distância da fonte sonora. Próximo ao ouvinte ocorrem também consideráveis modificações no espectro das HRTFs.

F) HRTF (*head related transfer functions* – funções de transferência relativas à cabeça):

As ondas sonoras que alcançam os dois tímpanos do ouvinte são, na realidade, resultantes da interação da onda original com o corpo do ouvinte. Dessa forma o tamanho da cabeça do indivíduo e o formato da sua orelha são fatores que implicam numa resposta em frequência diferente para cada um dos

²¹ RUMSEY, 2001, p.7.

ouvidos, o que altera o som percebido de acordo com a posição onde a fonte sonora se encontra. Esse conjunto de propriedades pode ser medido e capturado pelas denominadas funções de transferências relativas a cabeça (HRTFs), que são individuais e dependentes da direção da fonte sonora.

CHENG e WAKEFIELD (2001) definem as HRTFs como “filtros que imitam a filtragem dependente da direcionalidade do ouvido externo humano”. Segundo eles, de uma forma geral, a fonte sonora é percebida espacialmente próxima ao ouvido no qual a primeira frente de ondas chega primeiro, onde um grande ITD se traduz como um maior distanciamento da mesma do eixo frontal da cabeça do ouvinte. No entanto, em frequências acima de 1500Hz, o comprimento de onda começa a se equivaler ao diâmetro da cabeça do ouvinte, a cabeça começa a fazer um efeito de sombra no ouvido contrário ao som ouvido, e o ITD se torna uma informação ambígua. Nessas frequências mais altas o ILD gerado pelo efeito de sombra da cabeça se torna importante para a percepção do ângulo de azimute.

O autor complementa que as HRTFs, por serem parâmetros que resumem a filtragem acústica dependente de direção resultantes da interação do som com o torso, a cabeça e o ouvido externo (orelha), são úteis para filtrar um som monoaural e transformá-lo em binaural estéreo, que vai simular a espacialização sonora como no ambiente original.

Segundo CHENG e WAKEFIELD (2001), as HRTFs resumem ambas as informações do ITD e do ILD: os atrasos são codificados no espectro das fases do filtro e as informações referentes à intensidade estão relacionadas ao nível geral do filtro. Por essa razão, alguns autores, como MALHAM (1998), se referem à esse fator como HRFR – *Head Related Response Frequency* (resposta de frequência relativa à cabeça). Segundo ele, para posições onde o ILD e o ITD não demonstram diferenças no sinal recebido pelos dois ouvidos, como acontece no plano médio (sentido vertical) ou quando o ouvinte apresenta pouca ou nenhuma audição em um dos ouvidos, esse se torna o principal mecanismo de percepção espacial do som quando o mesmo não envolve o movimento da cabeça. Esse é também um dos dois mecanismos que nos fazem perceber uma fonte sonora posicionada na nossa frente ou atrás de

nós²². Abaixo uma figura extraída de CHENG e WAKEFIELD (2001) que ilustra o funcionamento das HRTFs em um espaço virtual:

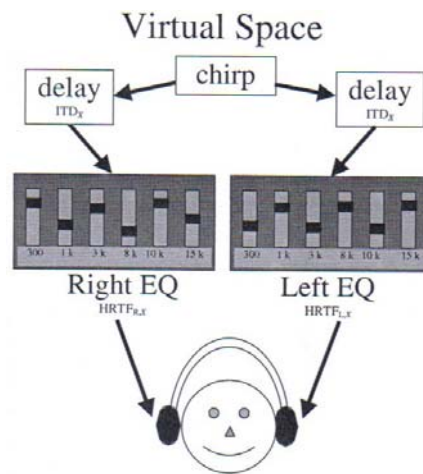


Figura representando o funcionamento das HRTFs em ambientes virtuais. Fonte: CHENG e WAKEFIELD (2001).

Diversos fatores acústicos adicionam certa complexidade e riqueza às HRTFs, um exemplo disso é o claro pico de amplitude na região de aproximadamente 3000 Hz, decorrente da ressonância do canal auditivo, e as fortes atenuações provenientes das interferências construtivas ou destrutivas da interação do som direto com o corpo do ouvinte. O movimento da fonte sonora ou do ouvinte em relação a esta também é um fator modificador da HRTF.

Quando a distância de um evento sonoro se altera, a HRTF também se modifica, no entanto em pequenas proporções se a fonte se localiza em uma distância superior à 2m da cabeça do ouvinte, distância a partir da qual seu comportamento é praticamente plano. Devido a isso, KENDALL (1995) afirma que HRTFs medidas a mais de 2m da cabeça podem ser utilizadas com sucesso na simulação de eventos mais distantes desde que as informações do ambiente também estejam presentes. Abaixo, um gráfico com a resposta em amplitude e em fase representativo de uma HRTF medida com uma *Dummy Head* Kemar, para 19 posições no plano horizontal. A curva mais abaixo se refere à medição a 0° de azimuth, ou seja, na frente do manequim e, a curva mais acima, se refere à parte de trás do manequim (180°).

²² MALHAM, 1998.

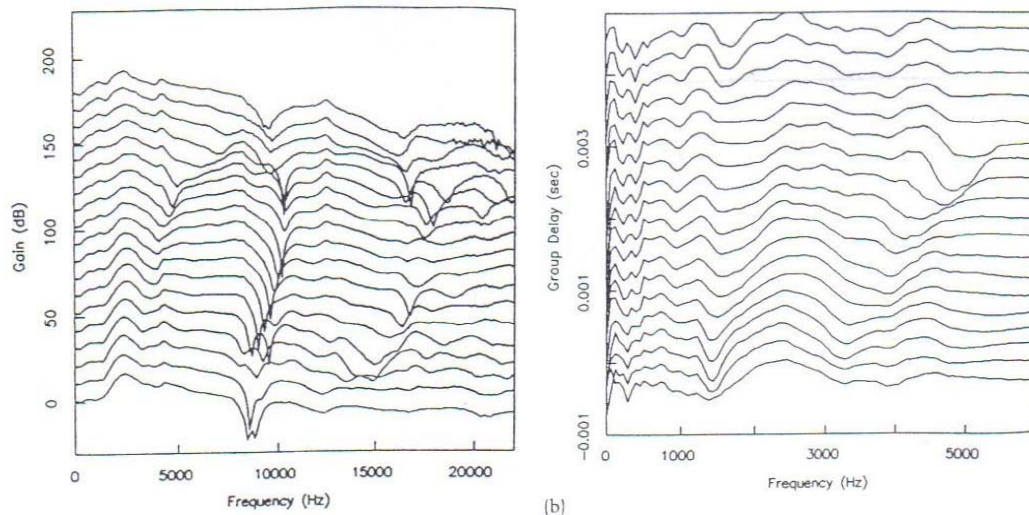
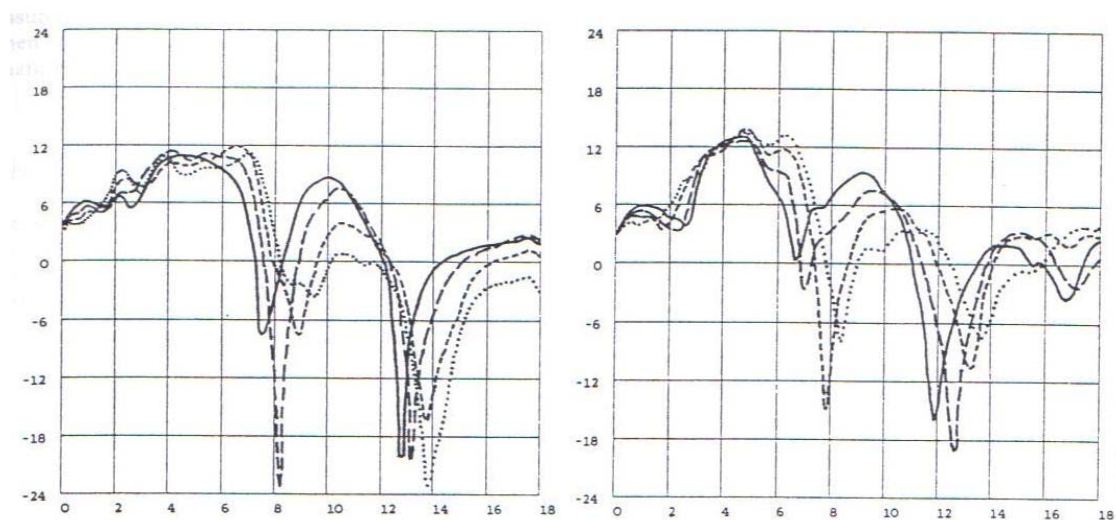


Gráfico com resposta de amplitude e fase de uma HRTF medida com uma *Dummy Head* Kemar. Fonte: KENDAL, 1995.

Pesquisas que fazem uma comparação entre HRTFs medidas em diferentes indivíduos revelam que os parâmetros espectrais não são compatíveis, ou seja, as amplitudes medidas nas HRTFs individuais variam tanto numa visão macro quanto em seus detalhes, o que permanece é uma tendência levemente similar entre elas. O que sugere que, apesar dos indivíduos possuírem cabeças de diferentes tamanhos e orelhas de diferentes formatos, o processo pelo qual as HRTFs são construídas é o mesmo. Apesar disso, a diferença de fase entre os ouvidos será muito afetada pelo tamanho da cabeça do indivíduo e pela distância entre as orelhas. Abaixo um exemplo de HRTFs de dois indivíduos, medidas no plano frontal. Cada gráfico apresenta linhas referentes às medições em 0° (—), 10° (—), 20° (- - -) e 30° (....).



Gráficos ilustrando diferenças nas medições de HRTFs de dois indivíduos. Eixo horizontal: frequências em KHz; eixo vertical: amplitude em dB. Fonte: KENDALL, 1995.

Segundo KENDALL (1995), as HRTFs podem ser medidas de três formas distintas, a primeira delas com um microfone bloqueando a entrada do canal auditivo, a segunda considerando-se o canal auditivo com a utilização de um tubo de prova, e a terceira com a captação da posição dos tímpanos utilizando-se uma *Dummy Head*.

O autor afirma que, a respeito das HRTFs individuais ou generalizadas, é sabido que o julgamento sobre a elevação e a diferenciação entre sons frontais e traseiros são mais prejudicados quando se utiliza HRTFs não individualizadas. KENDALL (1995) afirma também que resultados de diversas pesquisas na área demonstram que os indivíduos costumam localizar fontes sonoras melhor com suas próprias HRTFs; alguns indivíduos possuem HRTFs superiores e que podem melhorar a capacidade de localização de outros e que para isso o tamanho da cabeça de ambos devem ser semelhantes; e que uma melhor localização pode ser alcançada com DTFs²³ sintéticas, cujos detalhamentos diferem das medidas de HRTFs.

RUMSEY (2001) afirma que o som se altera espectralmente dependendo de três fatores: o formato da orelha, reflexões nos ombros e corpo, e da diferença de intensidade devido a uma distancia maior percorrida pelo som entre os ouvidos. Segundo ele, a soma desses efeitos resulta na formação de uma HRTF. Ele menciona o trabalho de BLAUERT (1997), onde o autor afirma que existem algumas faixas de frequências que são marcas de uma percepção espacial quando aparecem com forte atenuação ou ganho. Segundo ele, 8KHz, por exemplo, corresponde à percepção acima da cabeça, as regiões entre 300Hz – 600Hz e 3000Hz – 6000Hz correspondem a percepções frontais, e regiões centradas em 1200Hz e 12000Hz correspondem à percepção traseira.

Segundo RUMSEY (2001), pessoas que participaram de experimentos onde se utilizaram HRTFs de outras pessoas e que tiveram suas orelhas tampadas (inserindo esses sinais diretamente aos canais auditivos) apresentaram suas habilidades de localização reduzidas. No entanto, depois de certo tempo aparentaram se adaptar a essa nova informação. Esse fato tem importantes implicações no processamento de áudio para sistemas binaurais.

²³ Funções de transferência dependentes da direcionalidade.

G) SFRSs (*Spatial Frequency Response Surfaces* – Superfícies de Resposta em Frequência Espaciais):

Essa medição foi desenvolvida por CHENG e WAKEFIELD (2001), para a elaboração de um sistema que simulasse movimentação de fontes sonoras em sistemas binaurais para utilização de compositores de música eletroacústica. Segundo eles, diversos estudos sempre tiveram uma visão macro ao representar as HRTFs como filtros *peak* ou *notches*, em locais particulares da amplitude das respostas em frequência, que se associam ou simplesmente variam com a percepção de azimuth e elevação.

Os autores definem as SFRSs (*spatial frequency response surfaces* – superfícies de resposta de frequência espacial) como representações de HRTFs que conjugam a resposta de todas as HRTFs em uma frequência fixa como função do azimuth e da elevação. São gráficos que indicam quanto de energia cada ouvido recebe como uma função da localização espacial da fonte sonora. Nessa frequência fixa, os dados da HRTF contêm propriedades similares e cruzadas que são contínuas e se modificam relativamente devagar como função da localização espacial. A figura abaixo ilustra a construção das SFRSs, que acontece a partir de dados de uma frequência em particular. Triangulação e interpolação linear também são utilizadas para criar uma superfície contínua:

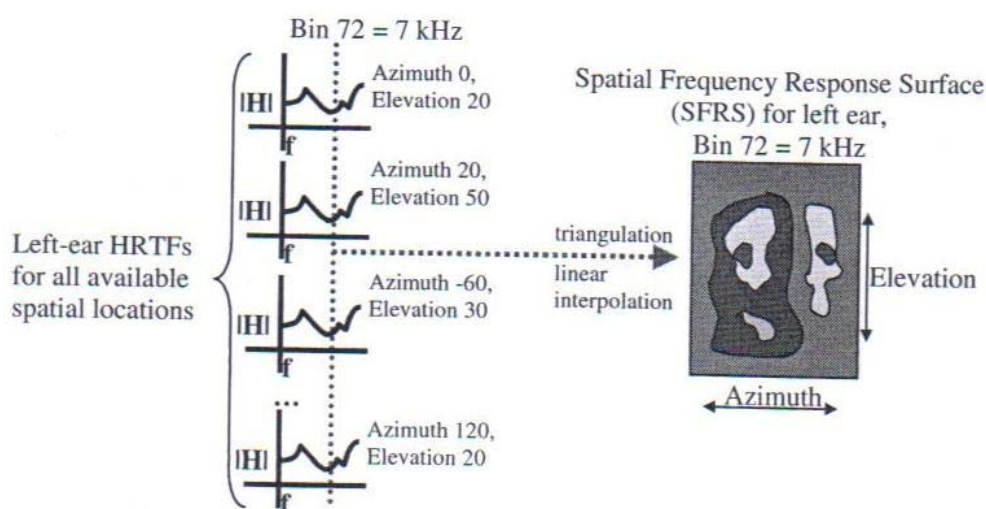


Figura ilustrando o processo de construção das SFRSs. Fonte: CHENG e WAKEFIELD, 2001.

Os autores afirmam que métodos de controle panorâmico são muito eficientes em movimentos horizontais esquerda-direita, e a manipulação de reverberação de forma dinâmica produz ilusões de movimentos referentes à profundidade muito efetivos. No entanto devido à sua necessidade de uma ferramenta para realizar movimentos, inclusive verticais, baixo-cima, e frente-trás, foi necessário o desenvolvimento desse método, realizado no MATLAB e baseado no uso de HRTFs interpoladas e ITDs, para aplicações em sistemas binaurais utilizando fones de ouvido. Segundo eles, o método pode ser utilizado também para posicionar de forma fixa uma fonte sonora em um ponto arbitrário no espaço. Abaixo uma figura, extraída de CHENG e WAKEFIELD (2001), ilustra a interpolação das HRTFs utilizada na criação das SFRSs:

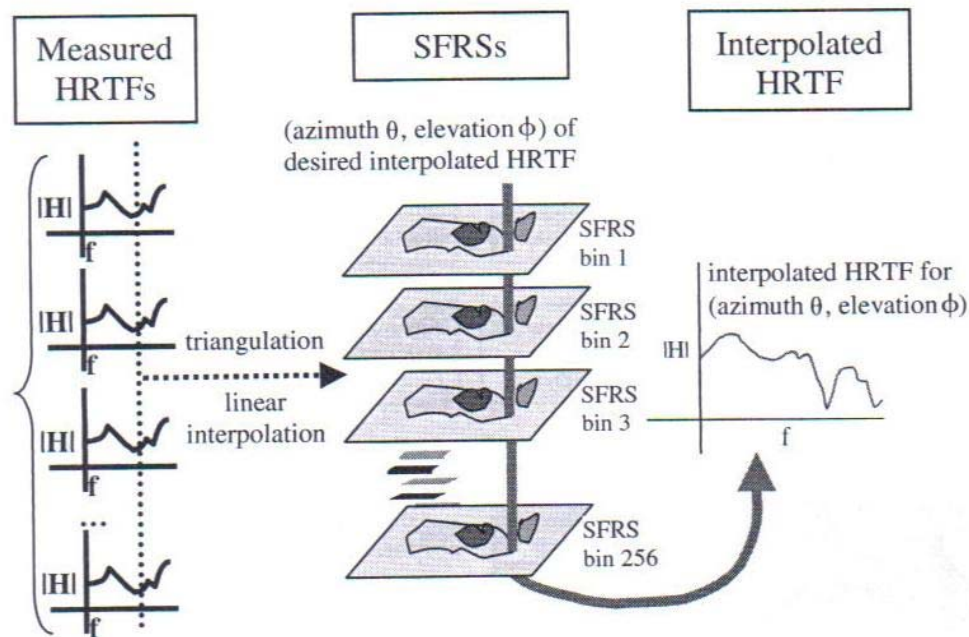


Figura ilustrando a interpolação das HRTFs na criação das SFRSs. Fonte: CHENG e WAKEFIELD, 2001.

Dentre os problemas desse método citados pelos autores estão a falta de “presença” e, o fato de que sons posicionados no plano médio parecerem estar dentro da cabeça do ouvinte. Segundo a pesquisa, movimentos de esquerda-direita são mais convincentes, mas ocorrem confusões entre sons frontais e traseiros, de baixo e de cima e certa dificuldade em simular sons com elevação fora do eixo de 0° . Segundo eles, sons fixos são mais convincentes

em determinadas posições do que em outras, tanto com relação à espacialização quanto à externalização e isso é ilustrado na figura abaixo:

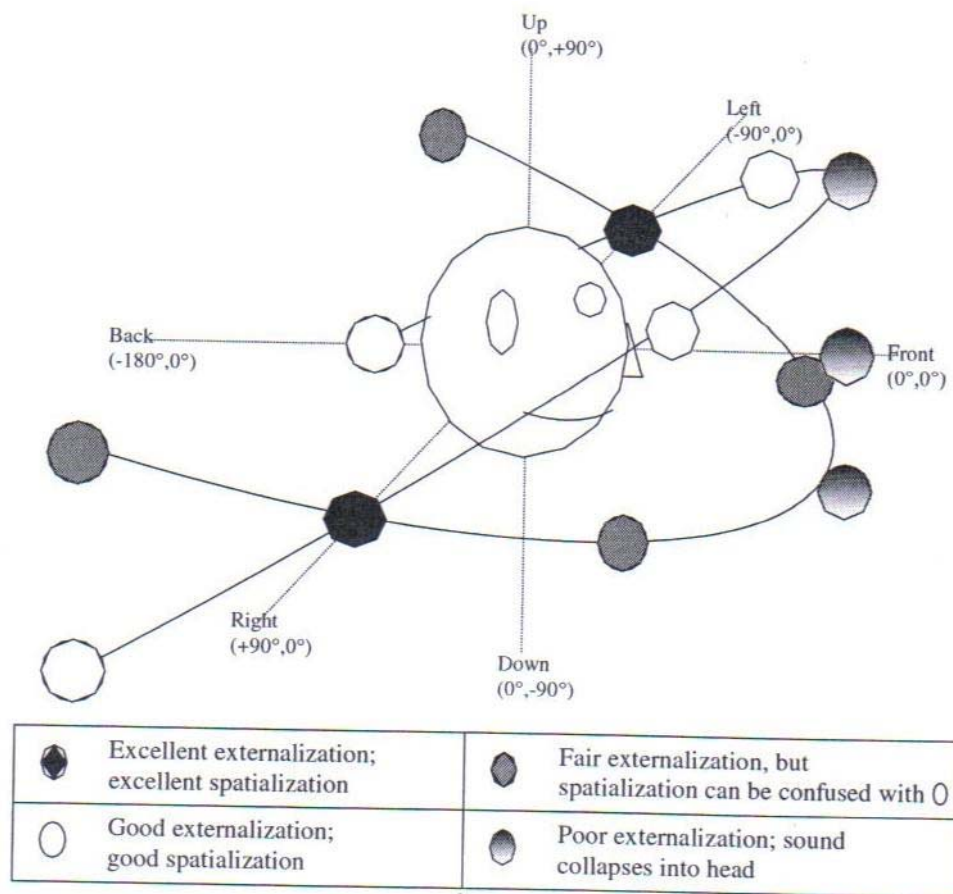


Figura ilustrando as peculiaridades do posicionamento das fontes sonoras. Fonte: CHENG e WAKEFIELD, 2001.

O método de CHENG e WAKEFIELD não foi pensado para simular mudanças com relação à distância da fonte sonora do ouvinte, uma vez que as HRTFs são específicas para ângulos de azimuth e elevação. Uma das vantagens mencionadas pelos autores é a habilidade do sistema em simular fontes sonoras frontais e traseiras, o que não se dá tão facilmente com relação à fontes em baixo e em cima. Ele afirma que inserir certa reverberação ao som antes de espacializá-lo pode auxiliar na percepção de profundidade.

Uma observação muito válida dos autores diz respeito ao fato de que sistemas de controle panorâmico simplesmente alteram o nível de determinado som monoaural em ambos os ouvidos, o que pode gerar certa sensação de falta de equilíbrio, enquanto que sistemas baseados em HRTF, como o

desenvolvido por eles, mantêm o equilíbrio e proporciona uma melhor externalização.

H) Movimento:

Os sons que chegam aos dois ouvidos são modificados espectralmente pelas HRTFs e, além do fato de que em cada ouvido essas modificações serem diferentes, elas sofrem diferentes transformações dependentes do movimento da cabeça do ouvinte e/ou da fonte sonora.

KENDALL (1995) afirma que outro fator determinante para a percepção de direcionalidade de uma fonte sonora diz respeito ao fato de que interagimos com o ambiente ao movimentarmos a cabeça, modificando a forma como o som chega aos nossos ouvidos. Segundo ele, essas mudanças, já incorporadas pelos nossos cérebros, nos auxiliam na percepção da posição das fontes sonoras.

Quando movimentamos a cabeça, tendemos a minimizar o ITD, o ILD e a alterar as HRTF percebidas nos dois ouvidos, e dessa forma, olhar a fonte sonora de frente. Esse movimento é uma importante forma de definirmos posições de fontes sonoras no espaço principalmente em se tratando dos conflitos entre sons frontais e traseiros. Por exemplo, no caso de um som frontal, quando movemos nossas cabeças para a direita fazemos com que esse som chegue primeiro e com maior intensidade no ouvido esquerdo e, quando essa fonte se localiza atrás, ao realizarmos esse movimento será o ouvido direito que será beneficiado.

Sistemas que se utilizam de HRTF para simulação de uma realidade virtual tendem a utilizar de rastreadores de movimento e processamento das HRTFs para simular mais realisticamente o espaço sonoro²⁴.

I) Outros parâmetros:

Outros fatores devem ser levados em conta quando se trata da espacialização sonora em determinada produção musical ou cinematográfica, RUMSEY (2001) chama a atenção, por exemplo, para o fato de que os outros sentidos possuem forte influência na percepção do espaço sonoro,

²⁴ Esses sistemas serão mais detalhados à frente.

principalmente a visão. Segundo ele, a maior parte das pessoas ao ouvirem gravações em sistemas binaurais sem o auxílio de imagens ou qualquer sistema de rastreamento de movimento da cabeça, tendem a localizar fontes sonoras traseiras ao invés de frontais.

Na realidade, simular fontes sonoras em sistemas de áudio com reprodução em fones de ouvido é extremamente difícil, principalmente pelo fato de que é mais simples assumir que, se uma fonte sonora não pode ser vista, provavelmente ela está posicionada atrás. Quando não se utiliza um sistema de rastreamento de movimento da cabeça para corrigir essa ambigüidade entre sinais frontais e traseiros, artifício utilizado na percepção sonora em ambientes naturais, o cérebro tende a assumir imagens traseiras com mais freqüência. De forma semelhante se, ao movimentarmos a cabeça, o cérebro não detecta nenhuma modificação com relação ao ITD e ao ILD, ele deduz que o som ouvido só pode estar vindo do interior da cabeça. Daí também a grande dificuldade de se conseguir externalização em sistemas binaurais para fones de ouvido.

Um outro fenômeno muito estudado relativo à percepção do tamanho de uma fonte sonora, denominado por RUMSEY de ASW (*auditory source width*) e que, segundo ele, parece estar diretamente relacionado a uma medida binaural conhecida por correlação cruzada entre os ouvidos – IACC (*interaural cross correlation*). Estudos mostram que as pessoas tendem a gostar ou preferir grandes quantidades de ASW, ou seja, se interessam mais quando as fontes sonoras parecem ser maiores; no entanto, quando elas aparentam ser excessivamente grandes sua localização se torna difícil e artificial.

Segundo RUMSEY, outros termos tais como “envolvimento” e “amplitude” do espaço sonoro são muito utilizados para descrever propriedades espaciais de um sistema de reprodução sonora e estão relacionados primeiramente às impressões do ambiente reproduzido, sendo resultado do som refletido principalmente vindo das laterais e após 80 ms do som direto. O termo “amplitude” do espaço sonoro – *spaciousness* – é mais utilizado para descrever a sensação de abertura do espaço ou sala onde o sujeito está localizado. Já o termo “envolvimento” – *envelopment* – é um termo similar utilizado para descrever a sensação de imersão em um campo reverberante.

RUMSEY observa também que:

“(...) em algumas circunstâncias de gravação, tais como a captação próxima utilizada em gravações pop, não é desejável capturar qualquer característica espacial uma vez que esse equilíbrio será determinado artificialmente através de panorâmicos e efeitos. Na gravação de música erudita, ou outras formas de gravação onde o espaço original é importante, as técnicas de microfonação serão desenvolvidas para capturar apropriadamente essas características espaciais do ambiente sonoro e as outras etapas do processo até o consumidor serão desenvolvidas para manter essa fidelidade ao sinal.” (RUMSEY, 2001, p.48).

O autor comenta também o fato de que os sistemas ou técnicas de espacialização de áudio de maior sucesso são aquelas que funcionam satisfatoriamente sob condições muito variadas de uso e por uma grande diversidade de ouvintes.

1.2 - Possibilidades e aplicações musicais da espacialização

Segundo RUMSEY (2001), na gravação de música clássica e outras gravações onde o ambiente natural tenta ser reproduzido ou onde um evento ao vivo está sendo retransmitido, é usual dizer que a gravação de alta qualidade e sua reprodução devem recriar a sensação de “estar lá” da forma mais convincente possível. Já outros acreditam que a reprodução em sistemas eletrônicos deve ser uma experiência completamente diferente de uma experiência natural e a recriação de um ambiente “real” não tem sentido algum. Muitos lançamentos comerciais refletem esse último ponto de vista e muitos engenheiros de som ficam entre esse dilema de levar o ouvinte ao ambiente real ou trazer o músico para a casa do ouvinte.

Em outros tipos de gravações musicais e produções cinematográficas não existe um ambiente natural para se recriar e em muitos outros casos, de música pop e cinema ou som para TV, lidamos com situações de completa recriação artificial de um ambiente que não tem nenhuma referência ao natural.

Se a identidade real fosse possível entre o ambiente gravado e o ambiente reproduzido, nas três dimensões e para todas as posições do ouvinte, então seria razoável supor que a habilidade do sistema de reprodução

de criar imagens fantasma de todas as fontes sonoras (inclusive reflexões) seria o único pré-requisito para a fidelidade. Como essa real identidade não é possível, criar e controlar ilusões sonoras dos principais parâmetros subjetivos da percepção auditiva deve ser o primeiro objetivo das técnicas de gravação e reprodução *surround*²⁵.

Um dos principais objetivos deste trabalho é tratar da espacialização do áudio no que se refere à abordagem estética de distribuição do material sonoro, dependente do objetivo da produção fonográfica.

No caso específico do cinema, como se verá em capítulos posteriores, o espaço sonoro se dedica a intensificar efeitos sonoros de impacto para o ouvinte, ou seja, o movimento de algum elemento na imagem corresponde a um movimento também do som produzido pelo mesmo para intensificar a percepção de tal evento. Em explosões, por exemplo, o espaço sonoro é utilizado para intensificar este fenômeno, tal efeito é denominado por CHION (1994) como “extensão do ambiente sonoro” ou “*offscreen trash*”, sobras sonoras que se expandem para além da imagem reproduzida na tela²⁶.

A abordagem do material estritamente musical, sem conteúdo visual, através do sistema 5.1 se mostra um pouco diferente com relação às abordagens sonoras de produções cinematográficas e esse tópico foi abordado pelo palestrante Andrés Mayo, membro da AES (*Audio Engineering Society*), em seu workshop sobre gravação e mixagem em sistemas 5.1²⁷. Sobre as abordagens de mixagem, ele define três possibilidades: a primeira com o ponto de escuta centrado no ouvinte, uma segunda com o ponto de escuta centrado no músico e uma terceira onde não há regras, prometendo infinitas possibilidades até os limites criativos do compositor/ engenheiro de som.

Como exemplo da primeira abordagem mencionada por Andrés Mayo, observamos a maioria das produções que envolvem música erudita, tais como grandes concertos e encenações de óperas, e essa abordagem reflete a tentativa de uma representação fiel do ambiente sonoro representado pelas imagens, com uma orquestra ou palco, com solistas e cenário, bem definidos

²⁵ RUMSEY, 2001, p.9.

²⁶ CHION, 1994, p.84.

²⁷ *Latin American Conference 2009 – Áudio para HDTV e sua evolução*. Workshop de gravação/mixagem em 5.1 realizado de 26-05-2009 à 28-05-2009 em São Paulo.

na frente do ouvinte e a reverberação da sala, aplausos e outros sons da platéia ao seu redor e atrás do mesmo.

Abordagens mais ousadas tendo o músico como ponto de escuta pôde ser observado na visita do presente autor à AES 2009 em São Paulo na competição estudantil - categoria música *surround* - onde um dos estudantes posicionou a orquestra a partir do ponto de vista do oboé tendo os violinos à direita, contrabaixos à esquerda e metais atrás, ao invés da posição tradicional, na tentativa de dar ao ouvinte a perspectiva sonora do oboísta.

Como exemplo da terceira e última abordagem mencionada pelo palestrante, temos diversos exemplos na música contemporânea, principalmente na denominada música eletroacústica, onde a possibilidade de espacialização tridimensional do som oferece ao compositor um novo elemento musical de trabalho. Podemos citar como um outro exemplo dessa abordagem a gravação em DVD-Audio dos seis concertos de Brandenburgo de Johann Sebastian Bach interpretados pela orquestra de câmara de Stuttgart e lançado pelo selo da Tacet Real *Surround Sound*²⁸.

Essa gravação nos remete ao fato de que a espacialização do som não é assunto recente e que antes mesmo do período barroco o espaço já era utilizado como elemento composicional em obras sacras envolvendo diversos grupos instrumentais, diversas orquestras ou diversos coros; e que, na renascença e do início do século XV esta era uma prática conhecida pelo termo *cori spezzati*²⁹, em referência a obras interpretadas por “coros partidos”. Um bom exemplo disso é a obra *Spem in alium* de Thomas Tallis composta para um coro de 40 vozes, para ser interpretada com o coro circundando o público³⁰.

Do período barroco ao romântico, no entanto, esse elemento musical ficou um pouco adormecido exceto por algumas poucas obras, como a Missa segundo São Matheus de Johann Sebastian Bach, para dois grupos de orquestra, coro e órgão e um coro infantil e o primeiro ato da ópera *Don Giovanni* de Mozart no qual o minueto da cena XX é escrito para três grupos instrumentais.

²⁸ Essa produção será descrita com maiores detalhes posteriormente neste trabalho.

²⁹ Termo extraído do dicionário Oxford de Música Online.

³⁰ Essa obra pode ser observada na instalação sonora de Janet Cardiff no Museu de arte contemporânea de Inhotim, em Brumadinho – MG, reproduzida através de 40 alto-falantes.

O compositor Henry Brant, em seu texto à exploração da espacialização musical “O espaço como aspecto essencial da composição musical”³¹, discorre sobre o fenômeno da exploração do espaço em obras de diferentes épocas, que passa pelas *canzonas* de Gabrieli, onde grupos de músicos são dispostos à frente e atrás do salão; pela obra *Unanswered Question* de Charles Ives, com cordas atrás do palco, conduzidas por um outro regente, flautas em uma “caixa” abaixo do centro da sala e um trompete solo posicionado atrás e acima do público; pelo *Réquiem* de Berlioz, com quatro grupos de metais posicionados nos quatro cantos em torno do público; pela *Serenata para quatro orquestras* de Mozart; e pela já mencionada *Spem in Allium* de Thomas Tallis, apresentada com o coro de quarenta vozes dispostos ao redor do público. Brant cita também sua obra *Voyage Four* onde a distribuição espacial dos músicos é planejada, um pré-requisito e um elemento essencial na construção da obra.

Outro autor, que recentemente fez um levantamento interessante de obras que utilizam da distribuição dos músicos no espaço, é THOMAZ, L. F. (2007) em sua dissertação a respeito de sistemas sonoros imersivos. De seu trabalho extraí as tabelas onde ele enumera e nos dá alguns detalhes de algumas obras anteriores ao século XX que abordam o espaço sonoro como elemento composicional:

Algumas composições com espacialização anteriores ao séc. XX			
Compositor	Peça	Ano	Detalhes
W. A. Mozart	<i>Noturno para 4 orquestras</i> , K. 286	1777	Efeitos de eco entre as orquestras
W. A. Mozart	<i>Don Giovanni</i> , K. 527	1787	Três grupos tocam separadamente: no buraco da orquestra, no palco e atrás do palco
L. van Beethoven	<i>Fidelio: Abertura Leonora</i> , Op. 72	1805	Trompete tocado fora do palco
H. Berlioz	<i>Requiem: Tuba Mirum</i> , Op. 5	1837	Quatro conjuntos de metais separados, localizados na disposição dos pontos cardeais, anunciam a entrada do coro
G. Verdi	<i>Requiem</i>	1874	Metais fora do palco
G. Mahler	<i>Sinfonia no. 2</i>	1895	Conjunto de metais e percussão fora do palco

Tabela citando obras anteriores ao séc. XX com seus respectivos detalhes de espacialização. Fonte: THOMAZ, 2007.

³¹ BRANT, 1978.

A partir do início do século XX, o espaço sonoro, elemento musical parcialmente adormecido em períodos anteriores, se torna novamente um elemento de composição entre alguns compositores contemporâneos antes mesmo da atual possibilidade de reprodução de suas obras em mídia digital, e volta aos palcos em obras como a ópera *Prometeo* de Luigi Nono, escrita entre 1981 e 1984, além de obras de Charles Ives, Stockhausen, Varèse, entre outros. THOMAZ (2007) menciona que, em várias dessas obras a platéia é rodeada pelos instrumentos da orquestra, em outras a orquestra se encontra num palco central, rodeada pela platéia e, a posição antes dedicada ao grupo instrumental (na frente do ouvinte) se perde para dar lugar a um envolvimento do público pelo som, o que nos remete ao período renascentista. Ele cita também algumas obras do século XX com detalhes através de uma tabela:

Composições contemporâneas com espacialização

<i>Compositor</i>	<i>Peça</i>	<i>Ano</i>	<i>Detalhes</i>
Charles Ives	<i>The Unanswered Question</i>	1906	Grupos instrumentais espalhados, com as cordas posicionadas fora do palco de frente para madeiras e metais
Edgard Varèse	<i>Intégrales</i>	1925	Utilização da ambiência proveniente do local de execução
Béla Bartók	<i>Music for Strings, Percussion and Celesta</i>	1936	Plano detalhado para a disposição das cordas aos flancos da percussão e do teclado
Henry Brant	<i>Antiphony</i>	1953	5 orquestras separadas no espaço
Ernst Krenek	<i>Pfingstatorium</i>	1955	Primeira peça eletroacústica estereofônica
Karlheinz Stockhausen	<i>Gesang der Jünglinge</i>	1956	Para 4 alto-falantes dispostos em panorama no palco e 1 alto-falante suspenso em cima da audiência
Karlheinz Stockhausen	<i>Gruppen</i>	1955-57	Para 3 orquestras em torno da audiência
Edgard Varèse	<i>Poème Électronique</i>	1958	Peça eletroacústica para fita utilizada no pavilhão Philips, na qual o som se movimenta através de um arranjo de 425 alto-falantes
Iannis Xenakis	<i>Duel</i>	1959	Para 2 orquestras, que respondem entre si executando partes escolhidas pelos regentes

Karlheinz Stockhausen	<i>Carré</i>	1959-60	Para 4 orquestras e 4 coros dispostos em forma de um quadrado
Karlheinz Stockhausen	<i>Kontakte</i>	1960	Primeira composição quadrafônica, criando o efeito de órbita
Karlheinz Stockhausen	<i>Originale</i>	1961	Peça de teatro baseada em <i>Kontakte</i>
Iannis Xenakis	<i>Stratégie</i>	1962	Similar à peça <i>Duel</i>
Henry Brant	<i>Voyage Four</i>	1963	Para percussão e metais no palco, violinos em um balcão lateral, violas e violoncelos em outro, baixos atrás, madeiras e algumas cordas no balcão traseiro e instrumentistas espalhados pela audiência
Iannis Xenakis	<i>Terretektorth</i>	1965-66	Peça para orquestra, com os músicos espalhados pela audiência
Karlheinz Stockhausen	<i>Stimmung</i>	1968	Para 6 vocalistas amplificados por 6 alto-falantes dispostos em torno da audiência
Luciano Berio	<i>Sinfonia</i>	1968-69	Grupo de violinos localizados atrás da orquestra
Pierre Boulez	<i>Domaines</i>	1968-69	Para clarinete solo e 6 grupos instrumentais, em que o clarinetista se desloca pela cena
Henry Brant	<i>Windjammer</i>	1969	Trompa solo parada e várias madeiras movendo-se em rotas predeterminadas enquanto tocam
Gilberto Mendes	<i>Santos Football Music</i>	1969	Para orquestra, alto-falantes em torno da audiência e participação de sons vindos da platéia
Karlheinz Stockhausen	<i>Spiral</i>	1970	Para solistas, com o som amplificado e manipulado espacialmente em uma projeção esférica
Karlheinz Stockhausen	<i>Sternklang</i>	1971	Para 5 grupos em locais distantes em um parque ao ar livre
John Tavener	<i>Últimos Ritos</i>	1972	Peça para orquestra a ser executada em catedrais, na qual as flautas e trompetes posicionam-se em galerias elevadas
Pierre Boulez	<i>Ritual</i>	1974-75	Para orquestra dividida em 8 grupos
Julio Estrada	<i>Canto Naciente</i>	1975-78	Oito metais em forma de cubo ao redor da audiência
Aylton Escobar	<i>Poética II</i>	1978	Para 4 flautas dispostas em losango em torno da audiência
Pierre Boulez	<i>Répons</i>	1981-88	Seis instrumentos solo, orquestra e equipamento eletroacústico
Luciano Berio	<i>Formazioni</i>	1985-87	Duas harpas posicionadas nos cantos do palco, próximas à audiência
Flô Menezes	<i>Parcours de l'Entité</i>	1993	Para flauta, percussão e sons eletroacústicos, com movimentos do flautista pelo espaço cênico
Mauricio de Bonis	<i>Sogno</i>	2001	Para piano, oboé e trompa, com movimentos dos dois últimos pelo palco

Tabela citando obras contemporâneas com seus respectivos detalhes de espacialização. Fonte: THOMAZ, 2007.

Com o sistema *surround*, essa experiência musical de envolvimento pode ser levada ao ouvinte, inicialmente através das salas de cinema, e mais

recentemente através de sistemas *surround* 5.1 instalados nas próprias casas dos ouvintes.

Segundo RUMSEY (2001), o objetivo primário da maioria das produções comerciais não é a fidelidade espacial ao ambiente sonoro original, apesar de alguns desejarem recriar alguns parâmetros que sejam consistentes com a experiência em ambientes naturais. A possibilidade de espacialização sonora em 360° pode se apresentar muito importante do ponto de vista criativo para alguns engenheiros de som, no entanto a mesma pode ser extremamente difícil de administrar. Em aplicações desse tipo, classificadas como criativa/ artística, não existe uma grande necessidade de representação espacial tão apurada quanto em outras aplicações devido também ao fato de que o ouvinte nem sempre estará posicionado no ponto adequado de escuta.

Em algumas outras aplicações, o principal objetivo de uma representação do campo sonoro está relacionado à interação com a orientação do indivíduo em um mundo recriado virtualmente. Nessas situações é necessário reproduzir parâmetros precisos que definam a localização e o movimento das fontes sonoras assim como da apresentação de um cenário e, às vezes, da interação do indivíduo com ele. Essas aplicações tendem à implementação de uma tecnologia binaural, pois a mesma proporciona um controle preciso de diversos parâmetros do ambiente com o qual o indivíduo se relaciona³². Abaixo uma tabela do autor nos mostra em qual dessas duas categorias cada aplicação se encaixa:

³² RUMSEY, 2001, p.19.

Table 1.1 Categorisation of spatial audio applications

Application	Creative/artistic illusion	Accurate 3D rendering
Classical/live/natural music	•	(•)
Pop music	•	
Radio/TV sound (drama, documentaries, etc.)	•	(•)
Cinema sound	•	
Virtual reality	(•)	•
Computer games	•	(•)
Simulators/control systems (e.g. aircraft)		•
Conferencing/communication systems		•

Tabela com a categorização das aplicações do áudio espacializado. Fonte: RUMSEY, 2001.

1.3 - Sistemas de reprodução de áudio

Segundo MALHAM (1995), desde o início da história das gravações tentou-se incluir nas mesmas informações referentes à localização espacial de eventos sonoros. Um dos grandes problemas que surgiu no decorrer desse desenvolvimento, segundo o autor, foi o da distribuição do material de uma gravação multicanal; mas com o aparecimento da nova geração de mídias digitais como o DVD isso se tornou uma realidade, mesmo que servindo somente a alguns formatos, como o *surround* 5.1, e não suportando outros formatos, como o Ambisonics, por exemplo.

A exploração do espaço sonoro é também a preocupação de muitos compositores e intérpretes de música eletroacústica. Técnicas estéreo para dois canais têm sido largamente empregadas no gênero, mas a espacialização sonora em suas formas mais sofisticadas têm sido restrita àqueles que possuem acesso a recursos técnicos específicos. O desenvolvimento tecnológico recente tem aumentado o acesso à novos sistemas e estúdios equipados para mixagem multicanal de alta qualidade; e novas aplicações para a espacialização do áudio têm sido identificadas: realidade virtual, computação multimídia, filmes, vídeos em geral e jogos eletrônicos têm utilizado efeitos sonoros dessa natureza. Adicionado a isso, o aumento do poder de processamento dos computadores tem levado ao desenvolvimento dessas

áreas ao suportarem a quantidade processamento referente ao trabalho com o áudio multicanal. Todos esses fatores levam a um maior interesse pelas técnicas de espacialização do áudio, possibilitando enorme potencial para o acelerado desenvolvimento da área.

MALHAM (1995) define também duas possibilidades de abordagem que podem ser adotadas pelos sistemas que permitem que os sons sejam posicionados artificialmente no espaço. Eles podem imitar as complexas mudanças de timbre, tempo de atraso e amplitude que ocorrem entre os nossos ouvidos quando um som se move de uma posição à outra, ou eles podem criar *imagens fantasma* através da quantidade de sinal endereçada para os dois ou mais alto-falantes.

A primeira abordagem consiste na utilização de HRTFs e temos como exemplos sistemas como o Qsound, Roland's RSS e o Convolvotron, e todos eles requerem o uso de fones de ouvido. Alguns deles, como uma variação do RSS, utilizam-se do cancelamento dos sinais cruzados entre os ouvidos para poder estabelecer imagens sonoras tridimensionais a partir de dois alto-falantes. Segundo o autor, sistemas que utilizam esse último método mencionado funcionam somente em uma região pequena, denominada *sweet spot*, na qual o ouvinte deve se posicionar corretamente para perceber o efeito da espacialização. Já os sistemas desenvolvidos para fones de ouvido possuem a característica de representarem bem um espaço estático quando o ouvinte utiliza de uma HRTF individual; em contrapartida, se não existe um sistema de rastreamento de movimento da cabeça, o ouvinte perde a referência de posição ao movimentar a cabeça.

A segunda abordagem mencionada por MALHAM, a da criação de *imagens fantasma*, condiz com a maioria das ferramentas disponíveis para a manipulação do áudio nos sistemas estéreo e *surround* 5.1 mais populares.

De acordo com KENDALL (1995), quando nos referimos ao áudio 3D, queremos dizer que o ouvinte ouve sons provenientes de direções que não podem ser reproduzidas pelo formato estéreo convencional. Um sistema de áudio 3D deve exceder as possibilidades de direcionamento sonoro do estéreo tradicional tanto com relação à dimensão horizontal quanto à vertical. De forma semelhante, um sistema de áudio tridimensional baseado em fones de ouvido deve ser capaz de reproduzir fontes sonoras posicionadas fora da cabeça do

ouvinte, tanto na frente quanto atrás do mesmo. Segundo ele, e como será visto mais detalhadamente à frente, a chave para se conseguir tal efeito está na superposição de funções de transferência dependentes da direcionalidade (DTFs). Tais funções devem ser capazes de recriar os complexos parâmetros acústicos que fazem com que o ouvinte perceba o som, e seja capaz de determinar sua posição, em um ambiente tridimensional.

KENDALL também afirma que essas funções capazes de recriar parâmetros acústicos espaciais só passaram a poder ser manipuladas recentemente devido ao grande poder de processamento computacional necessário, e também pelo fato de seu armazenamento não ser possível em mídias comuns como os discos de vinil ou fitas K7, por exemplo, mas somente com a tecnologia de armazenamento em formato digital, em HD, discos compactos (CDs) ou em DVDs.

1.3.1 – O Sistema Mono

Segundo RUMSEY (2001), o primeiro gramofone e sistema de gravação fonográfico surgiu no final do século XIX e era monofônico. Os únicos parâmetros referentes à espacialidade na reprodução monofônica eram a distância e a profundidade, percebidas a partir da reverberação³³. Tal sistema possuía grande influência no posicionamento dos músicos com relação à aparelhagem fonográfica durante o processo de gravação, como podemos observar nas figuras a seguir representando a gravação mecânica e elétrica, ambas monofônicas:

³³ RUMSEY, 2001, p.10.



Foto ilustrando o processo de gravação monofônica ainda de forma mecânica. Fonte: MAXFIELD e HARRISSON, 1926.



Foto ilustrando o processo de gravação monofônica já de forma elétrica. Fonte: MAXFIELD e HARRISSON, 1926.

1.3.2 – O Sistema Estéreo

De acordo com RUMSEY (2001), o primeiro exemplo de transmissão estereofônica se deu por Clement Ader, na exibição de Paris em 1881, quando ele posicionou cápsulas de telefone (microfones) espaçadamente, na casa de

ópera, e reproduziu esse sinal em um par de receptores, mantendo algum realismo espacial.

Nos anos 30, nos laboratórios Bell, se posicionaram diversos microfones omnidirecionais, um ao lado do outro e conectados individualmente a um amplificador e um alto-falante, transmitindo a frente de ondas recebida pelos microfones para um outro espaço, na tentativa de se reproduzir o espaço da sala dos microfones³⁴. STEINBER e SNOW (1934) observaram que apenas três microfones eram relativamente convincentes nessa reprodução, enquanto quando reduzido para dois, fontes sonoras centrais aparentavam mais distantes e o tamanho do palco parecia aumentar. É interessante observar que, apesar do canal central não ter sido adotado em sistemas caseiros, no cinema ele se tornou necessário muito rapidamente, uma vez que o público e o tamanho da imagem eram muito grandes para serem cobertos por apenas dois alto-falantes. No cinema o canal central tem a função de estabilizar uma imagem central para ouvintes fora do eixo da tela e tem sido utilizado desde o filme *Fantasia* da Disney em 1939. Em produções recentes como se verá mais à frente, este alto-falante do centro é utilizado principalmente para a difusão dos diálogos entre os personagens, uma vez que, se colocada nos alto-falantes laterais não causariam a ilusão do posicionamento central dos mesmos.

Ainda na década de 30, Alan Blumlein, no Reino Unido, desenvolve, juntamente com o grupo da EMI, uma forma de reproduzir imagens estéreo apenas através das diferenças de intensidade do som recebido pelos ouvidos. Diferentes quantidades de som reproduzidas por dois alto-falantes montados à esquerda e à direita do ouvinte, formando um ângulo de 60° entre eles, eram somadas de forma que a diferença entre o que os dois ouvidos percebiam poderia criar a ilusão de uma fonte sonora percebida entre os dois alto-falantes. A posição em que essa fonte se encontrava não dependia somente da intensidade emitida pelos falantes, mas também da frequência envolvida nesse sinal, e somava-se a isso a exigência de que o ouvinte deveria se posicionar frente aos falantes para que essa imagem se mantivesse estável. Devido à sua simplicidade, rapidamente esse formato se tornou padrão para as produções musicais³⁵.

³⁴ Detalhes desse procedimento em *Wave Field Synthesis* 1.3.8

³⁵ MALHAM, 2003.

RUMSEY (2001) descreve o sistema estéreo segundo os estudos de Blumlein, onde a configuração do sistema estéreo 2-0³⁶ consiste num triângulo equilátero, onde o ouvinte é um dos pontos do triângulo e os alto-falantes os outros dois. Se configurado de outra forma, as imagens fantasma reproduzidas pelos alto-falantes se tornam mais instáveis e o sistema como um todo mais susceptível aos efeitos de rotação da cabeça do ouvinte.

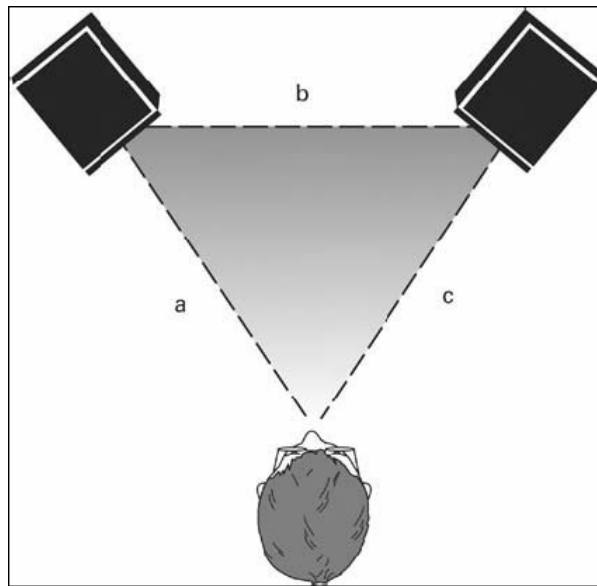


Figura que ilustra o correto posicionamento dos alto-falantes no sistema estéreo. Fonte: HOLMAN, 2008.

RUMSEY afirma também que as melhores ilusões de fontes sonoras são criadas entre os alto-falantes quando o sinal apresentado aos dois ouvidos é semelhante ao que se teria em situações naturais, ou ao menos quando os parâmetros de localização espacial presentes não são contraditórios. Segundo ele, é possível se criar essas ilusões em diversos ângulos entre os alto-falantes utilizando apenas diferenças de amplitude somente quando as diferenças de tempo são inferiores à 1ms.

O autor discute também a implementação de sistemas que posicionam fontes sonoras entre os dois alto-falantes e que consideram somente as diferenças de amplitude ou somente as diferenças de tempo. Passando pela microfonação estéreo desenvolvida por Blumlein e outros estudos, que procuram a relação entre amplitude e tempo para o posicionamento das fontes

³⁶ Segundo RUMSEY (2001), nos padrões internacionais que descrevem configurações de alto-falantes, a nomenclatura utilizada se apresenta na forma *n-m stereo*, onde *n* é o número de canais frontais e *m* de canais traseiros.

sonoras, ele conclui que um sistema de captação e reprodução deve considerar ambas as diferenças de amplitude e tempo tanto na gravação quanto na reprodução dos sinais nos dois alto-falantes para conseguir reproduzir o espaço sonoro de forma convincente.

Diversos métodos de se imprimir um sinal estéreo em discos de vinil de uma forma similar ao proposto por Blumlein foram introduzidos comercialmente no final dos anos 50 e as gravações estéreo se tornaram largamente disponíveis ao público nos anos 60. Diversas gravações monofônicas foram processadas artificialmente com filtros para serem reproduzidas em dois canais. Segundo RUMSEY (2001), o álbum dos Beatles "*Sergeant Pepper's*", de meados da década de 60 foi o primeiro gravado em vários canais (quatro) e foi lançado em versão mono e estéreo. Também nos anos 60 surgiram as primeiras rádios estéreo, proporcionando um alto nível de qualidade de áudio ainda não experienciado pela maioria dos consumidores.

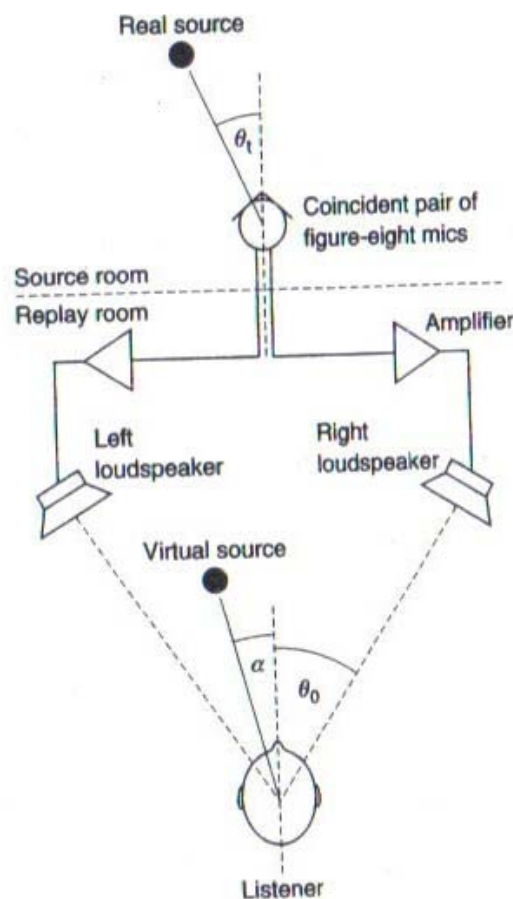
Como todo sistema de reprodução, o sistema estéreo possui suas limitações. A primeira delas é mencionada por MALHAM (1998), quando ele afirma que o sistema estéreo pretende reproduzir somente uma região limitada ao palco sonoro e localizada na frente do ouvinte. RUMSEY (2001) também menciona essa limitação de oferecer imagens e reverberação ao redor do ouvinte somente em um espaço correspondente a um ângulo de 60° à sua frente.

MALHAM (1998) afirma também que imagens fantasma são muito pouco estáveis quando não produzidas próximas aos alto-falantes, o que torna a separação dos mesmos a mais de 60° inaceitável. Segundo ele isso se torna mais grave quanto não se tem os mesmos posicionados corretamente, formando o ângulo de 60° em relação ao ouvinte. Quando se alcança os 90°, idealizados pelo sistema quadrifônico, por exemplo, a formação de imagens entre os alto-falantes se torna impossível. O mesmo acontece com outros sistemas multicanais onde o ângulo formado entre os alto-falantes excede os 60°.

Essa questão, que pode ser perfeitamente aceitável em aplicações cinematográficas, onde o ouvinte tem a atenção centrada na imagem, se torna inaceitável para compositores de música eletroacústica. MALHAM (1995) afirma que os únicos sistemas que tentam resolver essa incompatibilidade são

o QMX, baseado nos trabalhos de COOPER e SHIGA (1972) e o sistema Ambisonics de GERZON (1972), trabalhos esses com fundamentação equivalente³⁷.

Um outro problema mencionado pelo mesmo autor apresentado pelo sistema estéreo é o empobrecimento de sua performance quando o ouvinte se desloca do eixo de dispersão sonora, ou seja, do *sweet spot*. Além disso, ele afirma existir uma dificuldade grande de se estabilizar uma imagem quando o ouvinte movimenta a cabeça, uma vez que ao girar a cabeça o ouvinte está em certos momentos com a cabeça de frente para o alto-falante e em outro paralelamente ao mesmo, o que torna impossível gerar imagens fantasma aparentemente estáveis.



Localização Real X Imagem Fantasma ou localização da fonte sonora virtual na reprodução estéreo. Fonte: RUMSEY, 2001.

³⁷ MALHAM, 1995.

Um outro trabalho que discute os problemas da reprodução estéreo é o de CLARCK, DUTTON e VANDERLYN (1958), da EMI, que, retomando os trabalhos de Blumlein, chegaram a conclusão de que realmente o sistema que utiliza três canais não é econômico para uso doméstico e mesmo que o sistema estéreo consiga representar, apenas através da diferença de amplitude entre os sinais enviados para os alto-falantes, diferenças de fase entre os ouvidos de forma semelhante ao que acontece em ambientes naturais, essa redução para dois canais possui a tendência de criar o efeito de um buraco acústico central. Com isso eles mostraram que o método de Blumlein adaptado por eles não conseguia, somente com o uso das diferenças de intensidade, aproveitar as vantagens de todos os mecanismos da escuta binaural, especialmente do efeito de precedência, e deveria ser mais bem elaborado.

MALHAM (1998), ao descrever o funcionamento do sistema estéreo de uma forma mais técnica, afirma que um som direcionado para o alto-falante da esquerda, por exemplo, com cerca de -18dB em relação ao outro gera uma imagem posicionada no da direita. A medida que a intensidade do sinal desse alto-falante aumenta se percebe um movimento posicional da direita para a esquerda. A área onde esse efeito funciona é pequeno, no entanto maior do que em sistemas transaurais que veremos a seguir.

KENDALL (1995) é um outro autor que menciona essas deficiências e diz que, apesar de o formato de gravação estéreo não nos possibilitar a recriação tridimensional do espaço sonoro, ao ser reproduzido em dois alto-falantes, nos proporcionando somente uma impressão do espaço sonoro posicionada à nossa frente, existe uma outra situação da reprodução estéreo, onde ela é reproduzida através de fones de ouvido e temos o espaço sonoro se formando como se estivesse dentro de nossas cabeças.

RUMSEY (2001), ao citar BAUER (1961), também coloca essa questão, onde o autor citado afirma que sinais estereofônicos desenvolvidos para reprodução em alto-falantes, se reproduzidos em fones de ouvido, apresentariam diferença muito grande de intensidade entre os ouvidos, comparada à situação real, e um correto atraso entre os ouvidos não existiria. Isso resultaria em imagens estéreo artificiais que não possuem a sensação espacial desejada e surgem dentro da cabeça do ouvinte.

Um outro ponto colocado por KENDALL (1995) é o fato de que, quando estamos ouvindo gravações dessa forma (através de fones de ouvido) não podemos interagir com o ambiente sonoro e assim perdemos a oportunidade de atualizar a nossa percepção do ambiente, perdendo também uma importante ferramenta na percepção posicional das fontes sonoras. Tal fenômeno é amenizado quando existem imagens, no caso do cinema, onde a nossa visão nos auxilia nessa percepção espacial.

Uma afirmação interessante colocada por KENDALL (1995) é o fato de que o entusiasmo pela multimídia e pela realidade virtual parece ser parte do desejo cultural de se criar formas artístico / culturais sem a presença de convenções na sua forma de apresentação, fazendo com que experiências sonoras, inicialmente impossíveis no mundo real, sejam mais intensas do que a gravação de áudio e vídeo convencionais de alta fidelidade. Segundo ele, apesar do pop e do rock serem gêneros musicais produzidos em estúdios com o propósito de satisfazer esse desejo, criando um ambiente imaginário, ainda existe a necessidade de uma forma de intensificar essa experiência sonora. Uma parte dessa experiência se baseia na tridimensionalidade do som, não alcançada através do sistema estéreo.

1.3.3 – O Binaural

Segundo MALHAM (2003), há mais de cem anos já se tentava reproduzir um ambiente sonoro correspondente à realidade e essa busca percorreu mais de um século sem muito sucesso devido a limitações tecnológicas. Na década de 20 a primeira gravação para sistemas binaurais utilizando *Dummy Heads*, ou outros microfones desenhados especialmente para isso e pensados para um sistema de reprodução baseado em fones de ouvido, foi adotada por Dr. Harvey Fletcher e sua equipe nos laboratórios Bell. Sua tentativa não foi muito bem sucedida, pois não se possuía, na época, uma mídia capaz de armazenar dois canais de áudio. MALHAM (2003) afirma também que não era possível reproduzir tal gravação em alto-falantes, devido à sua natureza; assim via-se esse sistema muito útil apenas para aplicações domésticas.

Outras aplicações, mais recentes, são mencionadas por KYRIAKAKIS, TSAKALIDES e HOLMAN (1999): entre elas estão os sistemas de teleconferências; a realidade virtual; o entretenimento; o controle de tráfego aéreo, os sistemas de guia e de aviso de perigo para pilotos; a educação à distância; a edição de áudio e vídeo para televisão e filmes; e o auxílio para pessoas com visão prejudicada.

Para KENDALL (1995), o conceito de *binaural* diz respeito à combinação das informações recebidas pelos dois ouvidos (oposta ao monoaural que pressupõe a utilização das informações de cada ouvido de forma independente). Utilizar a palavra *binaural* implica também na utilização de informações interaurais dependentes da frequência, típicas das HRTFs.

RUMSEY (2001) define como áudio binaural “um meio de representar cenas sonoras tridimensionais codificando e reproduzindo os mesmos de forma a oferecer ao ouvinte sinais muito semelhantes aos que seriam observados numa situação natural de escuta” (p.52). Segundo ele, de uma forma generalizada, podemos afirmar que toda reprodução sonora é binaural, uma vez que são ouvidas pelos dois ouvidos do ouvinte; no entanto, o termo estéreo binaural é usualmente reservado para definir sinais que foram gravados ou processados para representar as características temporais e de amplitude das pressões sonoras ao serem reproduzidas diretamente no ouvido humano através de fones de ouvido.

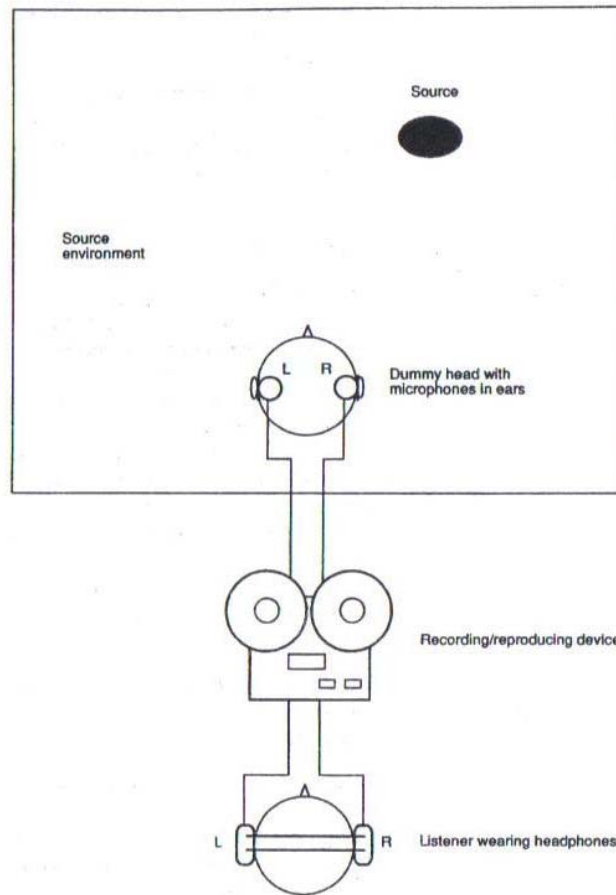
O autor afirma que a reprodução através de fones de ouvido é diferente da reprodução através de alto-falantes uma vez que, através de fones de ouvido, cada ouvido recebe somente um canal de sinal de áudio. A escuta binaural permite que cada ouvido receba um sinal que diferem entre si em relação ao tempo e também à amplitude, em níveis similares aos que resultam do efeito de sombra causado pela cabeça, entre outros.

As gravações binaurais têm fascinado pesquisadores por anos, mas recebido pouca atenção comercial até recentemente. Parte do problema diz respeito ao fato de ser muito difícil trabalhar com tamanha diversidade de ouvintes e fones de ouvido encontrados no mercado, parte dessa dificuldade também diz respeito à incompatibilidade entre fones de ouvido e alto-falantes. O estéreo convencional é aceitável pela maioria das pessoas quando reproduzido por fones de ouvido, apesar do efeito “dentro da cabeça”, já as

gravações binaurais, segundo diversos autores, dentre eles RUMSEY (2001), não são boas quando reproduzidas em alto-falantes sem qualquer tipo de processamento, e as imagens formadas se tornam imprecisas.

RUMSEY afirma que desenvolvimentos tecnológicos recentes têm tornado possível o processamento de sinal necessário para sintetizar sinais binaurais e lidar com a conversão entre fone de ouvidos e alto-falantes a custos razoáveis e de forma mais acessível. Agora é possível recriar os parâmetros sonoros tridimensionais e sintetizar ambientes acústicos virtuais utilizando DSP, e é essa área de simulação de ambientes através de aplicações computacionais que tem recebido maior atenção comercial na tecnologia binaural.

Dentre as aplicações do processamento de áudio binaural o autor cita os simuladores de vôo, jogos eletrônicos de computador, aplicações em realidade virtual, auralizações arquitetônicas e síntese de alto-falantes onde não existe nenhum, como nos sistemas de *home-theaters* virtuais. Uma abordagem óbvia de áudio binaural é posicionar um par de microfones exatamente na mesma posição de um par de ouvidos no ambiente a se captar, e reproduzir esse mesmo sinal através de fones de ouvido. Nesse momento surgem questões como onde deve se posicionar o microfone, se no canal auditivo, na orelha ou na posição do tímpano; que tipo de orelha e cabeça deve ser usada, uma vez que cada indivíduo possui uma diferente; que tipo de fone de ouvido seria o ideal; e se esse sinal poderia ser reproduzido também através de alto-falantes.



Sistema básico de gravação e reprodução binaural. Fonte: RUMSEY, 2001.

Para que uma gravação binaural funcione corretamente, as HRTFs das fontes sonoras ou do ambiente sintetizado devem ser precisamente recriadas nos ouvidos do ouvinte durante a reprodução. Isso significa captar o tempo e o espectro sonoro de forma precisa para cada um dos dois ouvidos. A partir do momento que cada posição da fonte sonora possui uma HRTF específica, podemos supor que tudo o que é necessário é assegurar que o ouvinte ouça corretamente cada uma delas na reprodução.

Entre os problemas dos sistemas binaurais, o autor supracitado cita:

>As HRTFs de cada indivíduo são diferentes, apesar de algumas similaridades, é difícil generalizá-las para fins comerciais.

>Movimentos da cabeça nos ajudam a resolver confusões a respeito do posicionamento das fontes sonoras e são difíceis de serem incorporados em situação de reprodução.

>Pistas visuais são perdidas na reprodução binaural e isso tem forte efeito na percepção sonora.

>Fones de ouvido diferem quanto a sua equalização e método de uso causando problemas de distorções nas HRTFs percebidas na reprodução.

>Distorções como erros nas respostas de fase e frequências na cadeia de reprodução do sinal podem afetar os parâmetros reproduzidos.

Segundo RUMSEY, apesar de serem individuais, as HRTFs podem, até certo ponto, ser generalizadas ao se partir de similaridades comuns, e existem evidências de que podemos, gradualmente, nos adaptar a novas HRTFs, como as utilizadas nos sistemas binaurais, e de que os erros de localização podem diminuir a partir da familiaridade. Estudos de BEGAULT (1991) e MOLLER (1995a) estudam essas generalizações, como o fato de que as diferenças até 8KHz são relativamente pequenas e acima disso pode-se encontrar uma estruturação generalizada para a maioria das direções.

A questão das referências visuais são resolvidas em sistemas de completa realidade virtual onde informações visuais 3D são adicionadas às informações sonoras. Já para a questão da equalização dos fones de ouvido, MOLLER (1995b) sugere a adoção ou configuração dos mesmos com uma resposta em frequências plana. Segundo ele também é quase impossível encontrar um fone de ouvido que reproduza precisamente sinais binaurais e que o timbre seja idêntico à reprodução em alto-falantes.

Como é impossível utilizar uma cabeça humana para realizar uma gravação binaural, diversos métodos tentam simular o efeito de sombra da cabeça na recepção sonora de dois microfones. Gravações que utilizavam um disco entre os microfones, por exemplo, são relativamente compatíveis com a reprodução por alto-falantes.

Dummy Heads são mais utilizadas, assim como manequins de ombros e torso, que consideram algumas importantes reflexões que ocorrem na escuta natural e que podem contribuir na formação das HRTFs; no entanto isso pode ser inconveniente se não compatível com os ombros e torso do ouvinte. A *Dummy Head* Neumann KU100, por exemplo, foi desenvolvida para ter uma boa compatibilidade entre as reproduções em fones de ouvido e alto-falantes.

Este método de gravação estéreo usando dois microfones localizados nos ouvidos de uma *Dummy Head* tem sido o mais popular no meio acadêmico

por muito tempo, devido ao seu potencial de codificar todos os parâmetros espaciais recebidos pelos ouvidos humanos. Quando reproduzidos através de fones de ouvido, essas gravações podem recriar uma marcante sensação de realismo.



Foto de uma *Dummy Head*. Fonte: EARGLE, 2005.

Infelizmente as variáveis desses sinais, particularmente as diferenças entre as cabeças e orelhas utilizadas na gravação e as do ouvinte, assim como as diferentes respostas dos fones de ouvido e seu acoplamento ao ouvido ou distorções que podem ocorrer no percurso do sinal, podem destruir facilmente os parâmetros espectrais e temporais necessários para a correta percepção espacial do som. Além disso, um dos problemas desse sistema, citado pela maioria dos autores, é que o importante artifício do movimento da cabeça que permite ao ouvinte resolver confusões entre sons frontais e traseiros em ambientes naturais não se encontra, inicialmente, presente na reprodução binaural.

RUMSEY (2001) menciona que, a princípio, as gravações binaurais não são compatíveis com a reprodução em alto-falantes, o que pode ser visto como

uma barreira para a sua difusão comercial. No entanto, muito tem sido feito para alcançar tal compatibilidade, o que tem resultado num ressurgimento no interesse pela tecnologia binaural que vem sob o título de “áudio 3D” e “*surround virtual*”. Como atualmente já é possível processar múltiplas faixas de áudio para misturar fontes sonoras e posicioná-las espacialmente usando representações digitais da resposta de um auditório, por exemplo, de forma semelhante, o material proveniente de gravações binaurais também pode ser processado e adaptado para a reprodução em alto-falantes. Essa tecnologia têm sido largamente utilizados em placas de áudio de televisores e computadores para o melhoramento da percepção espacial em dois alto-falantes. Da mesma forma sistemas de realidade virtual e ambientes de jogos de computadores têm se beneficiado muito, inclusive com o uso de rastreamento da cabeça do ouvinte para a incorporação desse movimentos modificando os parâmetros do processamento³⁸. Com relação a esse tipo de reprodução, é esperado que eles funcionem satisfatoriamente somente para uma região limitada de posições de ouvintes, o que será visto adiante sob o título de Ambiphonics.

MALHAM (1998) entende como sistemas binaurais aqueles que são reproduzidos através de fones de ouvido e também aqueles que se utilizam da mesma abordagem para, no entanto, serem reproduzidos através de alto-falantes. Esses sistemas são mais conhecidos como transaurais – termo utilizado por COOPER e BAUCK (1989) e registrado pelos mesmos como marca registrada.

Segundo eles, a reprodução do áudio através de sistemas binaurais utilizando fones de ouvido é, talvez, a forma mais óbvia e correta de se abordar a questão da espacialização do som tridimensionalmente. A duplicação exata do que os ouvidos ouviriam em uma situação real deveria se consistir no melhor meio de reprodução. De certa forma esse sistema o faz, no entanto surgem alguns problemas.

Quando uma paisagem sonora (*soundscape*) é produzida sinteticamente em um sistema binaural, deve-se tratar cada fonte sonora com sua respectiva

³⁸ RUMSEY, 2001, p.14.

HRTF, uma para o ouvido direito e outra para o esquerdo. Essa HRTF pode ser medida através de diversos métodos:

- Individualizada: medida para cada ouvinte.
- Generalizada: medida através da média obtida de diversos ouvintes.
- Medida através de uma *Dummy Head* que, por sua vez foi obtida através da média de diversos ouvintes.
- Sintetizada: calculada através de um modelo matemático

Segundo MALHAM (1998) a abordagem individualizada tem se mostrado mais bem sucedida e é potencialmente capaz de produzir resultados mais próximos à realidade; no entanto, para a maioria das situações, HRTFs generalizadas têm sido utilizadas. Apesar dos sinais individuais de ILD e de ITD serem pouco diferentes entre os indivíduos, a diferença entre uma HRTF generalizada e a individual se torna muito grande, principalmente nas altas frequências. Devido à importância desses sinais para o discernimento entre fontes sonoras na frente e atrás, confusões dessa natureza têm sido muito comuns. Algumas vezes os resultados têm sido um fracasso ao se perceber sons na frente do ouvinte ou atrás do mesmo. MALHAM (1998) menciona esse problema e afirma que ele poderia ser resolvido se a cabeça do ouvinte pudesse ser rastreada e, dependendo da posição, se utilizar de outra HRTF. Se isso fosse feito, sinais relativos à rotação da cabeça do ouvinte serviriam para que ele tivesse um melhor discernimento entre tais sons ao simular as alterações dos diversos fatores que ocorrem devido ao movimento.

Segundo RUMSEY (2001), o rastreamento da cabeça é o meio pelo qual os movimentos do ouvinte são monitorados. Em algumas aplicações essas informações podem ser utilizadas para modificar os parâmetros binaurais que são reproduzidos, trazendo mudanças realísticas nos sinais enviados aos ouvidos. Esse sistema pode ajudar a resolver problemas de confusões principalmente de fontes sonoras frontais e traseiras, como mencionado anteriormente, pois somente através de sutis movimentos conseguimos diferenciá-las. Para autores como HORBACH (1999) o rastreamento da cabeça é um fator crucial no aperfeiçoamento da precisão da reprodução dos sistemas binaurais assim, em seus estudos, ele chegou à conclusão de que o espectro adicional oferecido pelas orelhas é menos importante se comparado ao efeito de rotação da cabeça.



Foto de uma ouvinte em câmara anecóica³⁹ equipada com sistema de rastreamento. Fonte: CHENG e WAKEFIELD, 2001.

MALHAM (1998) menciona também o fato de existir um grande problema de cancelamento de sinais cruzados entre o sinal dos dois ouvidos de um material preparado para um sistema binaural que utiliza fones de ouvido ao serem reproduzidos em alto-falantes. E afirma que uma das soluções para resolver esse problema e obter certa compatibilidade entre esses dois sistemas é enviar o sinal que vai para um alto-falante também para o outro com a fase invertida, diminuindo esse efeito (COOPER e BAUCK 1989). Segundo o autor, esse tipo de sistema exige que a posição e orientação do ouvinte com relação aos alto-falantes seja precisamente conhecida para ótima operação, o que não acontece no mundo real, mas quando bem desenhado e planejado, pode-se obter resultados interessantes como no caso de PCs e televisões que possuem opção de som tridimensional baseado na posição de seus alto-falantes fixos e área de atuação reduzida.

Um outro grande problema mencionado por MALHAM (1998) diz respeito às aplicações dessas técnicas baseadas em HRTFs se aplicadas em situações de concertos, pois seria necessário que se equipasse todo o auditório com fones de ouvido individuais, sistemas de rastreamento de movimento de cabeça e processadores dedicados ao processamento da HRTF.

³⁹ Câmara anecóica – sala tratada acusticamente de forma a não apresentar reflexões.

Uma outra necessidade seria persuadir todo o público a utilizar corretamente os fones de ouvido.

KENDALL (1995) discorre com maiores detalhes a respeito da reprodução estéreo de sons em 3D, onde cada fonte sonora simulada em um sistema de espacialização 3D, assim como suas reflexões, deve ser processada por um par de filtros, as já mencionadas HRTFs, que sejam capazes de capturar as propriedades direcionais de movimento da cabeça do ouvinte. Concordando com o que vimos anteriormente, ele afirma que esses filtros devem mudar sua resposta de acordo com o movimento do ouvinte ou de acordo com as mudanças do ambiente simulado e que se existe mais de um ouvinte, as mudanças devem ocorrer de forma independente para cada um deles.

De acordo com o autor, construir um sistema que dialogue com o ouvinte é, talvez, mais importante do que a compatibilidade com a acústica da real do ambiente simulado, uma vez que sabemos que nossa audição é seletiva com relação às informações que absorve. Os principais problemas no desenvolvimento de sistemas desse tipo, e que já foram observadas no trabalho de CHENG e WAKEFIELD (2001), seriam: a confusão na percepção entre fontes frontais e traseiras, modificação no timbre, diferença entre a percepção dos diferentes ouvintes e as diferenças que ocorrem quando se utilizam fontes sonoras de diferentes naturezas.

KENDALL (1995) menciona um outro autor, MOLLER (1992), que afirma que enquanto a reprodução em alto-falantes e em fones de ouvido possuem similaridades técnicas e objetivas, elas apresentam também diferentes problemas. Entre as similaridades podemos citar o fato de que cada fonte sonora e cada reflexão simulada começa como um simples sinal monofônico que, eventualmente, é distribuído em um par estéreo. Como já mencionado, cada canal estéreo deve ser processado por filtros independentes que modificarão a resposta a partir da intenção de localização da fonte sonora. No final, todos os sinais estéreos das diferentes fontes sonoras são somados formando um sinal estéreo composto para ser reproduzido por alto-falantes ou fones de ouvido. Ele acrescenta que os filtros utilizados nesse processamento são filtros FIR (*Finite Impulse Response*) cuja resposta deve corresponder à uma biblioteca de HRTFs.

Com relação aos diferentes problemas, enquanto o processamento dedicado a sistemas de reprodução em fones de ouvido necessitam de processamento dedicado ao rastreamento do movimento da cabeça do ouvinte, os sistemas dedicados à reprodução através de alto-falantes necessitam de processamento dedicado ao cancelamento dos sinais cruzados, como será visto a seguir.

A respeito das reproduções de sons no formato estéreo tradicional através de fones de ouvido, KENDALL (1995) afirma que uma grande limitação do mesmo diz respeito ao fato de que além do fato de elas criarem a impressão de que os eventos sonoros são originados dentro de nossas cabeças, mesmo com a manipulação dos parâmetros de IID e ITD, as imagens se movem somente da esquerda para a direita dentro de nossas cabeças. Segundo ele, o som 3D deve realizar a externalização dessas imagens, fazendo-as serem percebidas do lado de fora das nossas cabeças e que as imagens que se apresentam na frente não se confundam com as traseiras. Isso tem se mostrado muito difícil de ser realizado através do uso de filtros direcionais associados a fones de ouvido estéreo tradicionais, tendo êxito somente na reprodução de sons laterais e não na de sons frontais e traseiros. A externalização pode ser auxiliada através da apresentação de som ambiente com incoerências interaurais que simulem as propriedades acústicas do campo reverberante. Segundo pesquisa informal, o autor supracitado descobriu que a discriminação entre sons traseiros e dianteiros pode ser otimizada através de modificações nas HRTFs que exagerem as diferenças nessas regiões.

KENDALL (1995) também menciona os sistemas de rastreamento da cabeça (de movimento do ouvinte), que são um grande avanço se combinados com os filtros direcionais dos fones de ouvido. Segundo ele um computador deve receber continuamente as informações espaciais do ouvinte e atualizar a direção dos filtros para manter a posição absoluta da fonte sonora em relação ao ambiente enquanto o ouvinte se move. Isso simula o tipo de interação que acontece na realidade, quando o som permanece fixo no ambiente enquanto a cabeça do ouvinte se movimenta. Segundo o autor, esse rastreamento é um ingrediente fundamental para um sistema de realidade virtual, pois:

“Se a cabeça do ouvinte se move e nada se modifica no tímpano, existe somente um lugar de onde esse som pode estar vindo – do meio da cabeça”. (KENDALL, 1995).

Segundo KENDALL (1995) ainda existem outros fatores que afetam a reprodução em fones de ouvido e uma delas é a questão da utilização de HRTFs individuais, que é mais importante na reprodução em fones de ouvido do que em alto-falantes, e isso ocorre devido à necessidade de externalização nesse formato, sendo mais fácil quando o ouvinte utiliza sua própria HRTF. Outro fator mencionado é a escolha do fone de ouvido, sendo preferível a utilização de fones abertos ou eletrostáticos.

O autor também diz algo a respeito da reprodução de sinais binaurais em alto-falantes e conclui que ambas as reproduções em alto-falantes e fones de ouvido dos parâmetros de direcionalidade sonora apresentam problemas e podem ser executadas com sucesso em configurações de reprodução controladas. Concordando com MALHAM (1998) o autor em questão afirma que a reprodução com fones de ouvido e rastreamento de movimento possibilita a forma mais realista de reprodução, mas também a mais cara e complicada, devido à dinâmica dos filtros e do sistema de rastreamento.

Uma outra questão levantada por RUMSEY (2001) diz respeito ao fato da cavidade interna do ouvido (denominada *concha*) ser a principal responsável pela sensação de externalização (a sensação de que um som emana de fora da cabeça). Se tomarmos o fato de que os fones de ouvido tendem a criar imagens sonoras dentro da cabeça do ouvinte, se houvesse alguma forma de simulação dessa cavidade, talvez fosse possível manipular o nível de externalização.

Segundo RUMSEY também, existe uma diferença entre a percepção auditiva quando os dois ouvidos detectam uma única frente de ondas (fonte sonora única) e quando sons similares chegam simultaneamente aos dois ouvidos vindos de direções diferentes. Isso chama a atenção para o que chamamos de atraso binaural (essencialmente a diferença entre o tempo de chegada de um som entre os ouvidos para um ângulo de incidência específico) e o efeito de precedência (ou lei da primeira frente de ondas). Com relação às reproduções em fones de ouvido e alto-falantes, RUMSEY destaca que os

fonos de ouvido permitem que os ouvidos sejam estimulados de forma independente, já em alto-falantes, devido ao fato de os dois ouvidos receberem os dois sinais provenientes dos dois alto-falantes, o efeito de precedência se torna bastante relevante.

A respeito dos recursos técnicos para uma gravação nesses sistema, KYRIAKAKIS, TSAKALIDES e HOLMAN (1999) mencionam que os microfones para uso em sistemas binaurais devem combinar a funcionalidade da filtragem espacial, os benefícios da localização sonora natural e a inteligibilidade da fala para a escuta binaural. Atualmente isso só é possível sem processamento através da utilização das já mencionadas *Dummy Heads*.

Uma outra aplicação muito utilizada do sistema binaural diz respeito aos estudos de análise computacional de cenas sonoras⁴⁰, muito conhecido no meio acadêmico como CASA (*Computational Auditory Scene Analysis*), onde o foco central é o fenômeno de os seres humanos serem capazes de, mesmo em ambiente sonoros complexos com múltiplas fontes sonoras, além de identificarem diversas dessas fontes, serem capazes de “focarem” sua audição em apenas uma delas. Segundo STERN, BROWN e WANG (2006), nesse complexo processo, um dos elementos estudados diz respeito à localização das fontes sonoras no espaço, que ocorre através de mecanismos binaurais e é essencial para que esse tipo de análise seja realizada.

1.3.3.1 – Síntese de *Surround 3D*

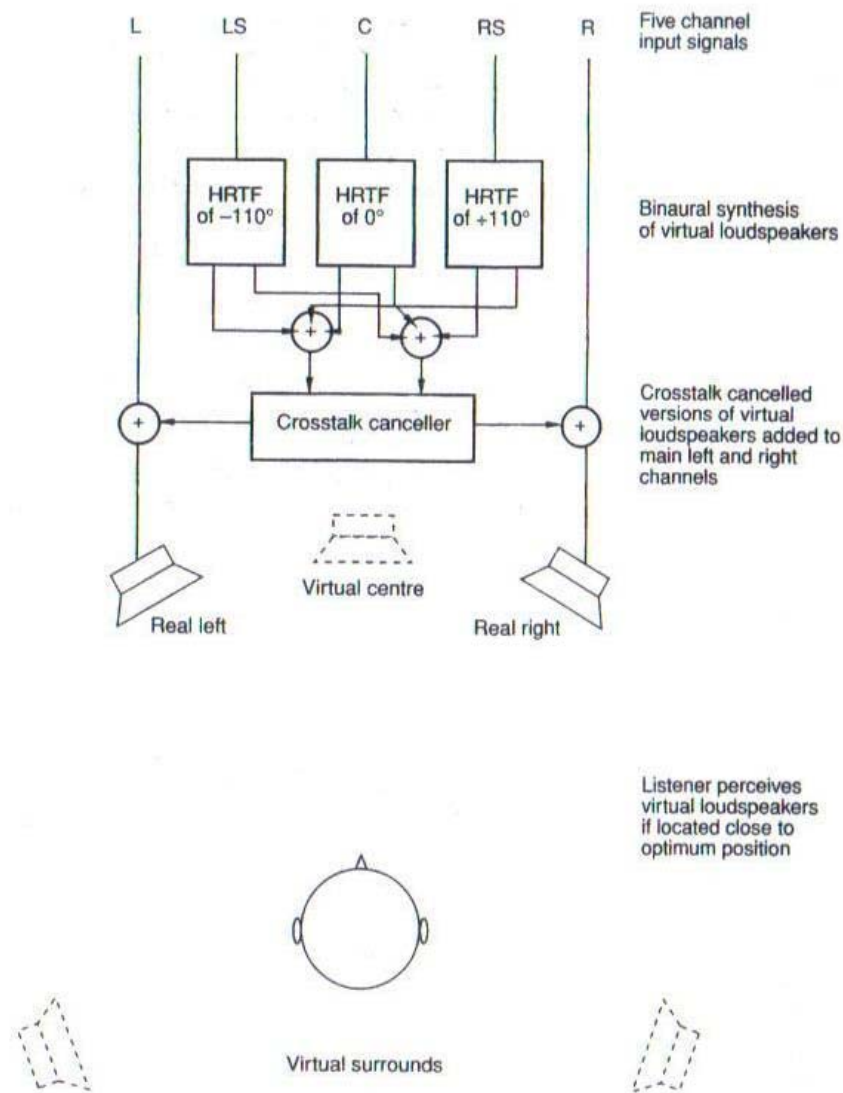
RUMSEY (2001) descreve brevemente os sistemas de *surround* virtual e sistemas virtuais de *home-theatres* (VHT), principal objeto da terceira parte do presente trabalho. Segundo ele, sistemas VHT utilizam princípios binaurais e transaurais para representar de forma virtual os alto-falantes de um sistema *surround 5.1* em ambientes onde não é possível o posicionamento de alto-falantes reais. O canal central muitas vezes também é virtualizado apesar de poder ser obtido através de métodos tradicionais de criação de imagens fantasma do sistema estéreo. O grande desafio desses sistemas é fazer com que sons apareçam nas laterais e atrás do ouvinte; neles, os canais Ls e Rs

⁴⁰ Termo já mencionado nos estudos de BREGMAN (1990)

também são processados para criar fontes virtuais com a HRTF correspondente a aproximadamente 110°. Se utilizado em alto-falantes, segue-se a etapa de cancelamento do cruzamento de sinais, descrito com maiores detalhes na seção sobre sistemas transaurais ou ambiophonics.

O autor afirma que um dos problemas principais diz respeito às alterações percebidas no timbre reproduzido resultante do processo. Tais sistemas são encontrados em televisores e equipamentos *surround* com softwares dedicados a evitar saídas físicas extras. São normalmente otimizados para uma região relativamente pequena de escuta, então o efeito de espacialidade se deteriora gradualmente quando o ouvinte se move para fora do denominado *hot spot* (ou *sweet spot*) e o sinal resultante dos canais traseiros são mais difusos, o que não é grande problema uma vez que o material sonoro desses canais não são precisamente localizáveis.

Uma outra aplicação muito interessante diz respeito ao fato de que, em algumas situações, alguém pode querer monitorar um programa *surround* através de fones de ouvido, e aqui a tecnologia binaural pode ser muito útil. Uma vez que os fones de ouvido possuem dois transdutores e nós possuímos dois ouvidos, o autor afirma que deve haver uma forma de mapear os cinco alto-falantes para sistemas binaurais tanto em reproduções em fones de ouvido quanto em dois alto-falantes.



Virtualização dos canais centrais e do *surround* em sistema de *home-theatres* virtuais. Fonte: RUMSEY, 2001.

HORBACH (1999) descreveu um sistema de auralização baseado em fones de ouvido para monitoração *surround* que virtualiza as posições dos cinco falantes e incorpora um sistema de rastreamento de cabeça para que o ambiente monitorado responda aos movimentos do ouvinte. O sistema de rastreamento fazia atualizações a cada 8,3ms e possuía uma latência de 50ms. Testes conduzidos pelo autor determinaram que *delays* de até 80ms não eram percebidos pelos ouvintes. Em adição aos sinais dos alto-falantes virtuais, o sistema incorpora respostas de impulso da acústica da sala de controle, então os alto-falantes parecem estar em um ambiente natural. Segundo ele, a adição de reflexões realísticas na auralização de fontes sonoras contribui positivamente para a sensação de externalização dessas fontes.

A tecnologia binaural é muito utilizada na síntese da acústica de ambientes virtuais e seu uso tem crescido em sistemas de auralização para design de acústica de ambientes. Podemos citar como exemplo dessa última aplicação as sonatas para teclados de Haydn gravadas pelo intérprete Tom Beghin, e os engenheiros de som Martha de Francisco e Wieslaw Woszczyk, e lançadas em setembro de 2009 em 4 discos *Blue Ray* de alta definição com o título de “*The Virtual Haydn: Perspectives for Scholarship and Performance*”⁴¹. A ambientação de 9 salas de concerto virtuais baseadas em salas originais de época acusticamente medidas, digitalmente mapeadas e, em seguida, precisamente recriadas no estúdio de gravação, permitiu ao intérprete interagir com seu som em tempo real, como se ele estivesse realmente tocando naquela sala, inclusive, segundo o próprio intérprete Tom Beghin, interferindo em sua performance⁴² durante a reprodução e escuta do instrumentista através de fones de ouvido.

RUMSEY (2001) menciona diversos sistemas que têm sido propostos para a reprodução de áudio binaural. Entre eles o Holophonic, patenteado por Hugo Zucarelli em 1980, é mais do que uma precisa recriação de uma *Dummy Head*, onde os microfones estão situados no fim do canal auditivo e uma equalização é feita para remover o efeito de ressonância desse canal. Uma peruca e as cavidades orais também são simuladas para tornar o modelo mais realístico. Este sistema foi utilizado na produção musical de artistas como Michael Jackson no álbum “Bad”, Paul Mc Cartney e Richard Branson, tem sido utilizado também em trilhas sonoras de filmes, música popular e televisão.

Outro sistema mencionado pelo autor, o Qsound, desenvolvido nos anos 80 e início dos anos 90 e que prometia imersão sonora através de dois alto-falantes. Não era de fato um sistema baseado em síntese binaural ou no cancelamento dos sinais cruzados, mas sim baseado nas diferenças de fase e amplitude dependentes da frequência entre os dois canais do sistema estéreo convencional que dava a impressão das fontes sonoras estarem posicionadas em locais diferentes. Segundo a empresa (QLabs) o Qsound é oferecido comercialmente nos formatos de software ou plugins e utilizado por algumas empresas na área de telefonia móvel, produtos de informática, jogos

⁴¹ The Virtual Haydn Project.

⁴² Performing in virtual acoustics – Essay by Tom Beghin.

eletrônicos, filmes, programas de televisão, áudio na internet e distribuição de mp3.

A construção básica do QSound 3D, um dos produtos oferecidos pela empresa, consiste no processamento de um único sinal mono para criar dois ou mais canais. Segundo eles as saídas incorporam diferenças sutis de intensidade, frequência e tempo que, quando reproduzida no sistema específico (fones de ouvido ou alto-falantes) para o qual cada algoritmo foi desenvolvido, o efeito simula parâmetros direcionais que o ouvinte teria em uma escuta natural, criando assim uma localização tridimensional das fontes sonoras. Os sinais processados de cada fonte monofônica é mixado e formatado para o sistema escolhido (fones de ouvido, dois alto-falantes ou sistema multicanal).

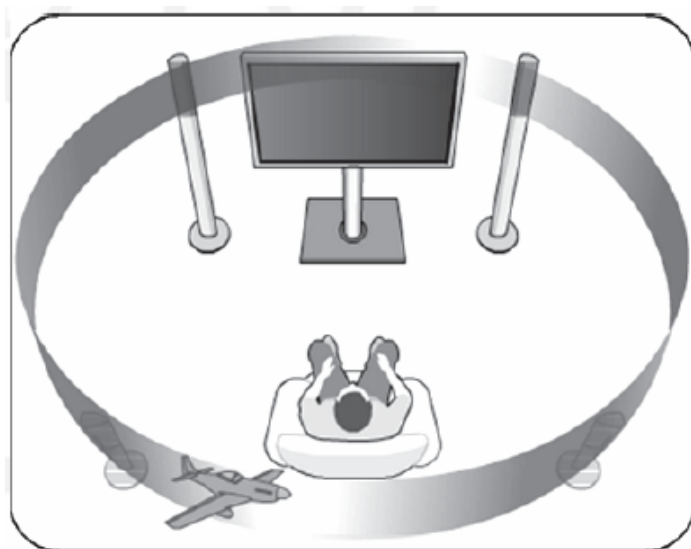
Um sistema mais recente é o Sensaura, baseado na otimização de HRTFs, eficiência computacional e algoritmos de cancelamento dos sinais cruzados para alto-falantes. Desenvolvido pelos laboratórios de pesquisa central do reino unido, a companhia desenvolveu também uma estação de mixagem binaural de áudio 3D, oferecendo sinais que podem ser combinados com a saída de uma *Dummy Head*.

Um outro sistema não mencionado pelo autor mas que segue a mesma linha é o RSS da Roland, desenvolvido para aplicações de reprodução em alto-falantes que utiliza processamento de fase, *delay* e filtragem pra tornar um sinal monofônico em binaural e adicionando um sinal para o cancelamento em fase dos sinais cruzados que normalmente ocorrem na reprodução através de alto-falantes. Através da manipulação das fases dos canais esquerdo e direito, filtragem e nível em tempo real, pode-se fazer com que fontes sonoras se movimentem e, em boas condições de escuta, rodeiem o ouvinte. Muito semelhante ao QSound, se difere por utilizar somente modelagem física ao invés de dados empíricos. Um grande problema do sistema é a sua incompatibilidade com o sistema monofônico. O RSS se apresenta tanto na forma de hardware quanto de software e tem sido de grande utilidade para desenvolvedores de jogos eletrônicos.

1.3.4 – O Ambiophonics

O sistema Ambiophonics, ou sistema transaural, diz respeito à reprodução de sinais binaurais, inicialmente preparados para reprodução em fones de ouvido, que são processados e adaptados para alto-falantes.

Com relação a essa reprodução em alto-falantes, KENDALL (1995) afirma que, do ponto de vista de um sistema de um auditório, a reprodução através de alto-falantes é uma categoria especial de audição de um ambiente, pois as ondas dos mesmos chegam a cada ouvido do ouvinte de duas direções, o que implica num tempo e numa intensidade de chegada diferente para cada falante, como um som direto seguido de uma forte reflexão. Essas diferenças podem simular o efeito de reflexões vindo de uma parede a diferentes distâncias além de diversos outros efeitos que podem ser conseguidos através da interação da intensidade e do tempo entre os sinais dos dois alto-falantes, como o efeito de precedência.



Efeito gerado por um sistema Ambiophonics. Fonte: KIM, 2008.

Segundo MILLER (2002) e diversos outros autores⁴³, esse sistema funciona muito bem apenas para uma região limitada de posições do ouvinte pré-determinada e muito focada, de forma semelhante ao denominado *sweet spot* do sistema *surround* 5.1. KENDALL (1995) menciona também que o

⁴³ KENDALL (1995), KYRIAKAKIS, TSAKALIDES e HOLMAN (1999) e RUMSEY (2001).

movimento do ouvinte em relação aos alto-falantes altera definitivamente a imagem sonora que ele recebe, sendo que somente os ouvintes que se sentam exatamente na linha central do auditório, ou sala de escuta, posicionada de forma eqüidistante dos alto-falantes, recebe o sinal pretendido.

Segundo KENDALL (1995), para uma reprodução tridimensional do som, a posição do ouvinte deve ser fixa e conhecida para que os cálculos de compensação possam ser realizados, ou seja, o sinal processado por HRTFs que chega ao ouvido direito deve ser enviado com a fase invertida ao outro alto-falante para que haja o efeito de tridimensionalidade, e vice-versa.

RUMSEY (2001) também menciona tal processamento, se o que se pretende é que todos os parâmetros 3D da gravação binaural original sejam mantidas na reprodução em alto-falantes. GRIESINGER (1989) propõe métodos de equalização de gravações binaurais feitas em *Dummy Heads* para fazê-las mais compatíveis com alto-falantes. Tais métodos resultaram em um relativo sucesso na reprodução de sons frontais, mas o mesmo não se deu com relação a sons localizados verticalmente e com parâmetros que resolvem confusões apresentadas no plano frente-trás.

KYRIAKAKIS, TSAKALIDES e HOLMAN (1999) estudam mais profundamente a reprodução de áudio 3D através de dois alto-falantes e afirmam que, atualmente, um sistema de imersão sonora deve preservar a imagem sonora com localização das fontes inclusive quando o ouvinte está em movimento.

Segundo eles, diversos estudos realizados com certo êxito para aplicações com fones de ouvido são bastante válidos, mas, no caso de se utilizar os alto-falantes, é necessário eliminar o cancelamento de sinais cruzados que ocorre entre eles. Tal cancelamento, já mencionado por KENDALL (1995), ocorre pois cada alto-falante envia o sinal para ambos os ouvidos, diferentemente das aplicações para fones de ouvido, sendo que o conteúdo de cada sinal é condizente apenas com um deles. A principal limitação dos primeiros sistemas desenvolvidos era o fato de que qualquer movimento do ouvinte que excedesse 75 a 100 mm destruía completamente o efeito espacial reproduzido e, atualmente, se objetiva a maior liberdade do ouvinte sem interferência na reprodução sonora.

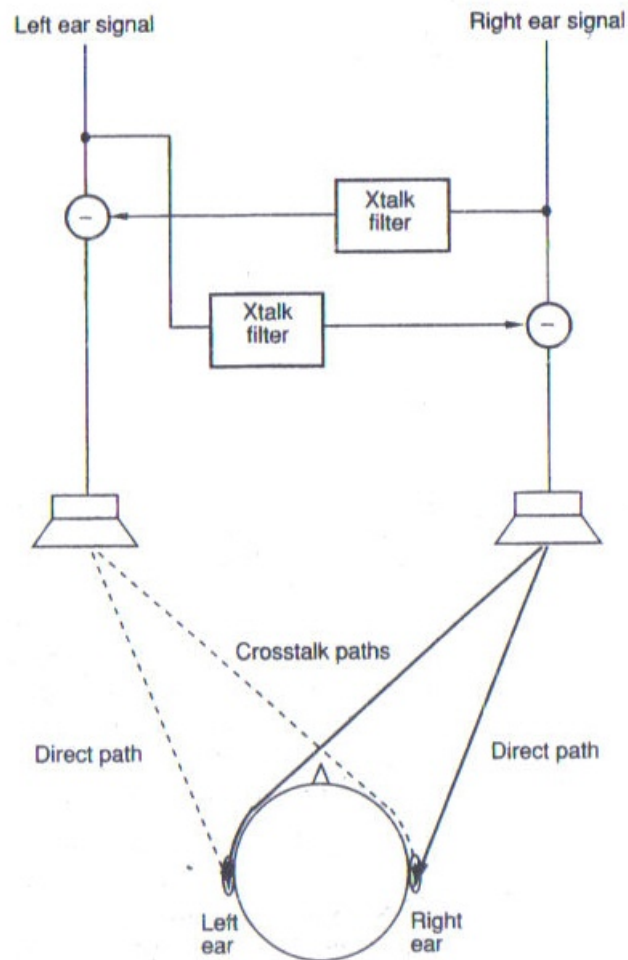


Figura ilustrando o processo de cancelamento dos sinais cruzados, adaptação de um sinal binaural para alto-falantes.

Fonte: RUMSEY, 2001.

RUMSEY (2001) afirma que as situações em que tal processo se mostrou mais eficiente foram em aplicações de computadores pessoais e sistemas virtuais de *home-theatres*. A razão para esse sucesso é o fato de que o ouvinte em tais situações permanece sentado numa posição conhecida em relação ao monitor de vídeo, e os alto-falantes podem ser anexados ao mesmo. Apesar das diversas abordagens com alto-falantes, esse tipo de reprodução inibe a criação de imagens sonoras em uma ou mais regiões, devido às primeiras reflexões do som reproduzido no ambiente sonoro e às assimetrias do próprio equipamento.

De acordo com KENDALL (1995), essa reprodução por alto-falantes, mesmo que limitada à monitoração próxima ou uma região determinada, é mais conveniente do que através de fones de ouvido, mas se aproxima menos da realidade.

KYRIAKAKIS, TSAKALIDES e HOLMAN (1999) mencionam também o fato de que, em um típico sistema de monitoração sonora de um *desktop*, o som é entregue ao ouvinte através de dois alto-falantes que estão tipicamente posicionados de cada lado do monitor de vídeo. Eles afirmam que esse ambiente, combinado com os problemas acústicos naturais das salas pequenas, causa problemas sérios que contribuem para a distorção audível do som reproduzido e do surgimento de primeiras reflexões. Outros problemas tais como a coloração da resposta em frequência, o surgimento de filtros pente e a modificação do timbre também são apontados. Outro problema diz respeito às fortes reflexões provenientes da mesa ou dos suportes para o monitor de vídeo e alto-falantes que resultam em um ganho forte nas regiões médio-graves de frequências.

Como solução para esses problemas KYRIAKAKIS, TSAKALIDES e HOLMAN mencionam a combinação dos diversos elementos psico-acústicos no design do sistema, onde é possível colocar o ouvinte no campo de som direto que é dominante sobre a região de som refletido ou reverberado. A grande vantagem do sistema de áudio dos *desktops*, segundo eles, está no fato de que a posição dos alto-falantes e, na maioria dos casos, do ouvinte, são conhecidos *a priori*, o que pode auxiliar no cálculo da equalização e configuração do sistema como um todo. Outra solução, já mencionada anteriormente, seria o rastreamento do movimento do ouvinte servindo para o ajuste de uma equalização dinâmica.

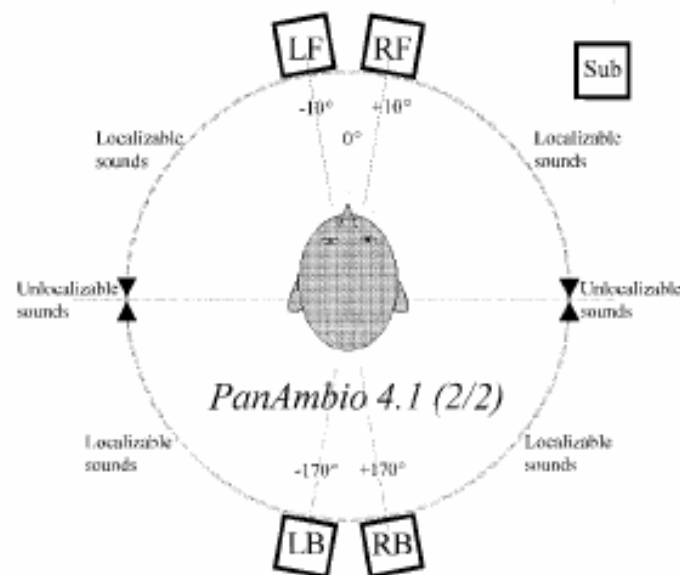
MILLER (2002), em uma comparação entre o sistema Ambiphonics e o *surround* 5.1, menciona o fato de que, apesar de ser uma ferramenta importante para a exploração da imagem sonora além dos limites da tela do cinema, o sistema ITU 5.1⁴⁴ apresenta uma grave deficiência com relação ao fato de ele não abranger o que ele denomina perifonia, campo esférico completo de percepção auditiva humana. A partir disso ele apresenta propostas de sistemas Ambiphonics, baseados em 2, 4, 6, 8 ou 10 alto-falantes que poderiam se aproveitar do atual sistema padrão de distribuição multicanal.

Segundo ele, enquanto o sistema *surround* 5.1 é capaz de reproduzir um ambiente de 60° na sua parte frontal, o Ambiphonics é capaz de reproduzir

⁴⁴ Padrão de posicionamento de alto-falantes para reprodução em *surround* 5.1.

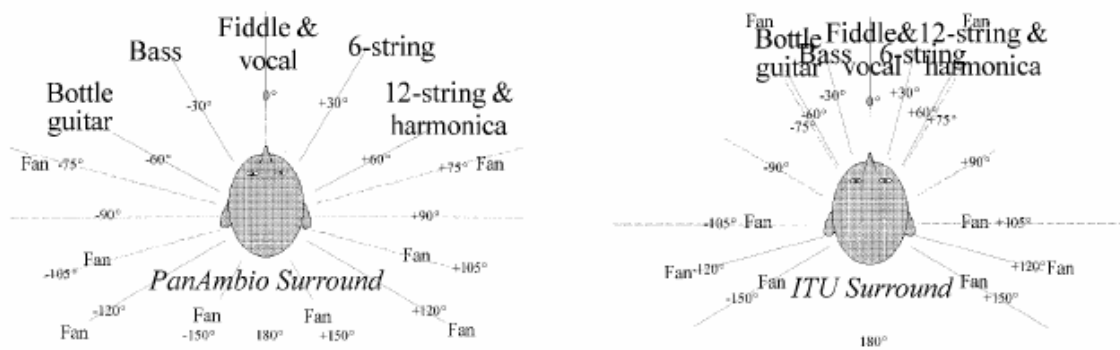
até 150°. Uma das observações feitas pelo autor diz respeito ao fato de que o sistema estéreo poderia também ser otimizado através do Ambiphonics, onde o sinal sofreria um processamento binaural e, em seguida, um processamento de adaptação aos dois alto-falantes. De forma semelhante, uma produção em *surround* 5.1 poderia se utilizar do sistema para monitoração virtual.

Segundo o autor, o sistema Ambiphonics mais simples utiliza dois alto-falantes próximos e cancelamento de sinais cruzados para simular o isolamento obtido pelos fones de ouvido e pode localizar imagens sonoras precisamente entre um ângulo de 120° a 150°. Um segundo par de falantes posicionados de forma semelhante atrás do ouvinte – sistema denominado pelo autor de PanAmbiophonics – resulta na recriação de um ambiente 360° envolvendo o ouvinte.



Sistema PanAmbiophonic 4.1. Fonte: MILLER, 2002.

MILLER (2002) aponta que o envolvimento de 360° em torno do ouvinte proporcionado pelo sistema se apresenta um pouco confuso nas laterais e, em condições muito favoráveis suporta até dois ouvintes em seu *sweet spot*, apesar disso, a grande diferença entre o sistema PanAmbiophonics e o sistema *surround* 5.1 fica no fato de que a precisão de escuta para fontes sonoras nesse novo sistema é muito superior, chegando a ser de 5°.



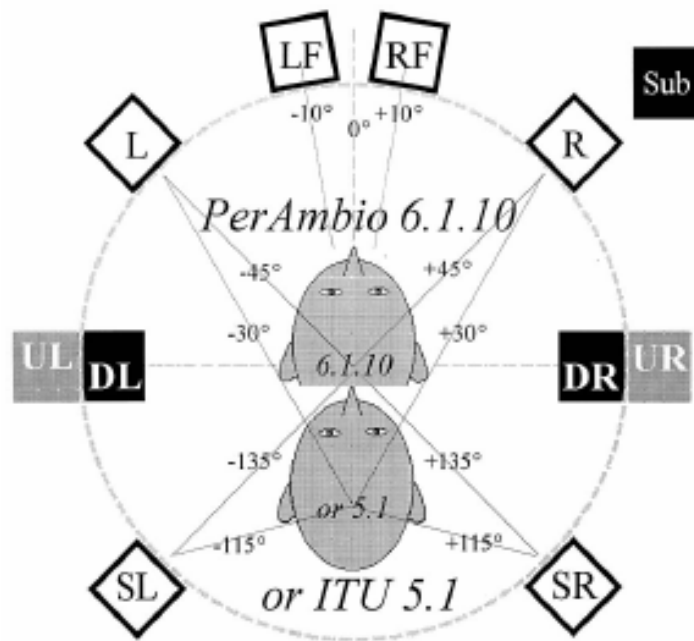
Comparação da espacialização realizada em um sistema PanAmbiophonics e em um sistema *Surround* 5.1.

Fonte: MILLER, 2002.

Uma das grandes vantagens mencionadas pelo autor diz respeito à compatibilidade do sistema em seu formato 4.1 com relação ao próprio 5.1 aplicado em *home-theaters*, uma vez que as alterações se dariam somente no cancelamento do alto-falante central e movimentação dos outros quatro. Gravações realizadas com o sistema PanAmbiophonics pelo autor combinam localização frontal superior aos outros sistemas, assim como uma espacialização e envolvimento mais naturais, além de poder ser reproduzido em um sistema ITU 5.1 padrão.

Uma expansão desse sistema para incorporar informações referentes a altura dos posicionamentos das fontes sonoras seria o apresentado pelo autor como PerAmbiophonics, onde a reprodução de uma esfera completa seria obtida com 8 alto-falantes, com a possibilidade de uma simplificação para seis canais de mídia (que poderia utilizar o sistema 6.1). Esse sistema PerAmbiophonics 6.1.10⁴⁵ também se mostra compatível com o 5.1 padrão e apresenta a possibilidade de ser extensível para um sistema perifônico completo, com os dez alto-falantes, num futuro próximo.

⁴⁵ Sistema composto de seis alto-falantes com um canal para efeitos de baixas frequências opcional e 10 outros alto-falantes que reproduzem um *surround* híbrido de Ambiphonic perifônico e Ambisonics.



Sistema PerAmbiophonics 6.1.10. Fonte: MILLER, 2002.

1.3.5 – O Sistema Quadrifônico

Segundo RUMSEY (2001), o sistema quadrifônico representa uma tentativa fracassada de introduzir o som *surround* para o consumidor. Não utilizava o canal central e era normalmente configurado para um arranjo quadrado de alto-falantes, dois na frente e dois atrás. Havia diversos métodos de codificação com diferentes graus de compatibilidade entre eles.

O autor afirma que, o ângulo de 90° entre os alto-falantes dianteiros se provou problemático devido à falta de compatibilidade com o sistema estéreo tradicional e oferecia um empobrecimento das imagens frontais, normalmente com a formação do indesejado buraco acústico central. A incompatibilidade entre formatos, assim como a dificuldade em se instalar dois alto-falantes adicionais, impossibilitou o formato de se estabelecer e um outro fator importante nisso tudo foi o fato de que muitas pessoas sentiam que a codificação quadrifônica comprometia a integridade do sistema estéreo.

Mesmo tendo sido bastante utilizado por compositores de música eletroacústica durante muito tempo, foi gradativamente sendo abandonado ou substituído por sistemas mais comerciais. Atualmente muitas produções realizadas para esse sistema fazem uma transcrição para o sistema *surround* 5.1, mais acessível, passível de distribuição e comercialização, como é o caso

da produção Immersion⁴⁶, de diversos compositores contemporâneos, mencionada mais adiante.

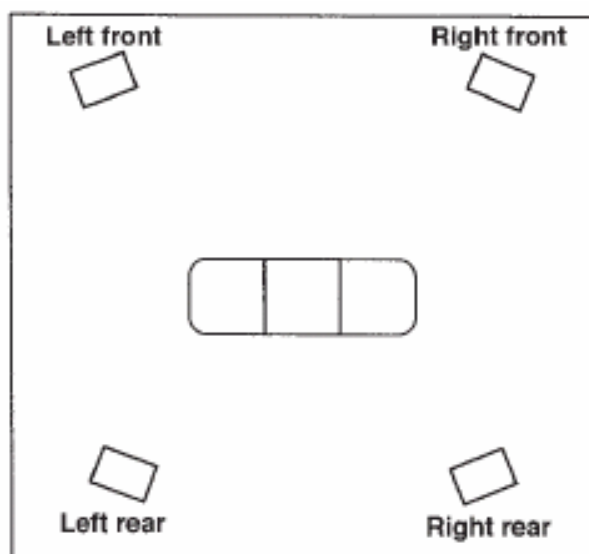


Figura ilustrando o posicionamento dos alto-falantes no sistema quadrifônico no ambiente doméstico. Fonte: EARGLE, 2005.

1.3.6 – O Sistema Multicanal 5.1 e suas variações

Depois dos múltiplos conjuntos de transmissores utilizados por Clement Ader em 1881, na primeira tentativa de se reproduzir um espaço sonoro, do sistema de Harvey Fletcher e a equipe dos laboratórios Bell em 1920 baseado em fones de ouvido e técnicas binaurais, segundo MALHAM (1995), houve também tentativas como a dos próprios laboratórios Bell em desenvolver um sistema baseado em uma “cortina” de microfones e outra de alto-falantes com a intenção de se recriar a frente de ondas originais de um ambiente. No entanto tal sistema era inviável pela quantidade enorme de canais que envolvia. A redução desse sistema para um baseado em apenas três microfones e um conjunto de três alto-falantes foi melhor recebida e utilizada no trabalho de Leopold Stokowski e em diversas gravações estéreo de orquestra até o ano de 1932.

Segundo o autor, houve pouco desenvolvimento na área até o aparecimento do filme *Fantasia* de Walt Disney, em 1939, onde a orquestra foi

⁴⁶ DVD-Audio Imerssion Starkland, 2001.

gravada com 33 microfones em um gravador de nove pistas e reproduzida em três canais distintos. A idéia inicial era de que esse áudio multicanal fosse reproduzido em 90 alto-falantes distribuídos atrás da tela e ao redor do auditório, embora por motivos financeiros isso raramente tenha sido feito na prática.

RUMSEY (2001), quando fala sobre o desenvolvimento do sistema *surround* 5.1, começa afirmando que, além dos efeitos utilizados em *Fantasia*, em 1939⁴⁷, o cinema não incorporou o sistema estéreo até os anos 50, e a utilização do panorâmico para o posicionamento dos diálogos combinados com os elementos cênicos foi abandonado rapidamente em favor do posicionamento do diálogo em um canal central acompanhado de música e efeitos especiais nos canais laterais.

Tal processo também é mencionado por MALHAM (1998), ao dizer que esse sistema primeiramente contou com a adição de um canal central para fixar o diálogo à imagem e melhorar a performance para ouvintes fora do centro. Logo em seguida se adicionaram os canais traseiros para melhorar a difusão, em conjunto com os canais frontais, devido ao tamanho do público. Ele menciona também um detalhe técnico que diz respeito ao sinal que vai para os canais traseiros, que é atrasado em relação aos frontais para que estes não sejam ouvidos antes dos mesmos, técnica muito utilizada atualmente para regulação do sistema de sonorização de grandes shows onde se tem grandes torres de PA principais e algumas torres de *delay* (atrasadas em relação à principal).

Ainda na década de 40 a Warner foi a primeira a introduzir um quarto canal na parte de trás da platéia e a *Century Fox* adotou um sistema similar. RUMSEY (2001) comenta que os formatos multicanais se tornaram mais populares nas décadas de 50 e 60 culminando no chamado *baby boomer*, formato envolvendo múltiplos canais frontais, um canal traseiro e um *subwoofer*. No início da década de 70, a Dolby introduziu o sistema de quatro

⁴⁷ Segundo MALHAM (2003), em 1939, o filme *Fantasia*, utilizou originalmente nove gravadores óticos, sincronizados entre si. Para a reprodução foram utilizados quatro canais distintos e sincronizados com as imagens. Um sistema dedicado controlava a distribuição do som para os múltiplos alto-falantes posicionados ao redor do auditório. Esse foi, provavelmente, o primeiro uso de automação no áudio. No entanto, devido à complexidade do sistema, o mesmo não foi muito utilizado posteriormente e nunca pôde se tornar um padrão para o uso doméstico.

canais *surround* gravados em fitas 35mm. No cinema atual, gradualmente se movendo para formatos digitais, o áudio de sistemas mais convencionais pode apresentar de 5 à 7 canais *surround* e um canal de *subwoofer* para efeitos especiais no registro grave.

Atualmente se vê uma grande diversidade de formatos multicanais que são variações do 5.1, tais como o 6.1, o 7.1, o 10.1, o 10.2 e o mais recente, o formato 22.2, trabalho japonês apresentado por Kimio Hamasaki, membro da NHK (Corporação de Radiodifusão Japonesa), cuja apresentação técnica e experiências auditivas puderam ser presenciadas pelo presente autor em visita a Convenção Latino Americana da AES de 2009⁴⁸.

O sistema *surround* 5.1 só começou a se consolidar com o desenvolvimento dos DVDs e dos sistemas de cinema caseiros, onde encontrou um caminho para dentro das casas dos consumidores como uma forma de intensificar a experiência de se assistir a um filme. Quando a reprodução musical em si aparenta ser uma justificativa insuficiente para reconfigurar a mobília de uma sala de estar, assistir a um filme é considerado uma experiência aproveitável por toda a família, removendo a barreira para a instalação de um equipamento de alta qualidade.

Segundo RUMSEY (2001), o sistema 5.1 consiste em três canais frontais, dois canais traseiros e um canal adicional para efeitos subgraves, onde os canais esquerdo e direito da parte frontal mantém certa compatibilidade com a posição dos dois canais do sistema estéreo. Ele afirma que esse formato define apenas a posição dos alto-falantes, não definindo de forma alguma como o material sonoro deve ser distribuído nos mesmos. Apesar de não existir um método correto de representação espacial para esse sistema, é importante saber que os três canais centrais são utilizados para fontes sonoras primárias que possuem atributos direcionais bem claros, enquanto os canais traseiros e o subgrave pretendem apenas suportar a ambiência, os efeitos sonoros e o som da sala, de forma a intensificar o efeito de espacialização. Muitas pessoas acreditam ser possível definir imagens sonoras nos 360° com esse sistema, embora isso seja extremamente difícil. Apesar de não ser impossível, a criação de imagens laterais não se dá de

⁴⁸ HAMASAKI, Kimio. A geração futura – surround 22.2.

forma tão estável quanto nos canais frontais, por exemplo, e o ângulo grande formado pelos alto-falantes traseiros também faz a definição dessas imagens bastante problemática.

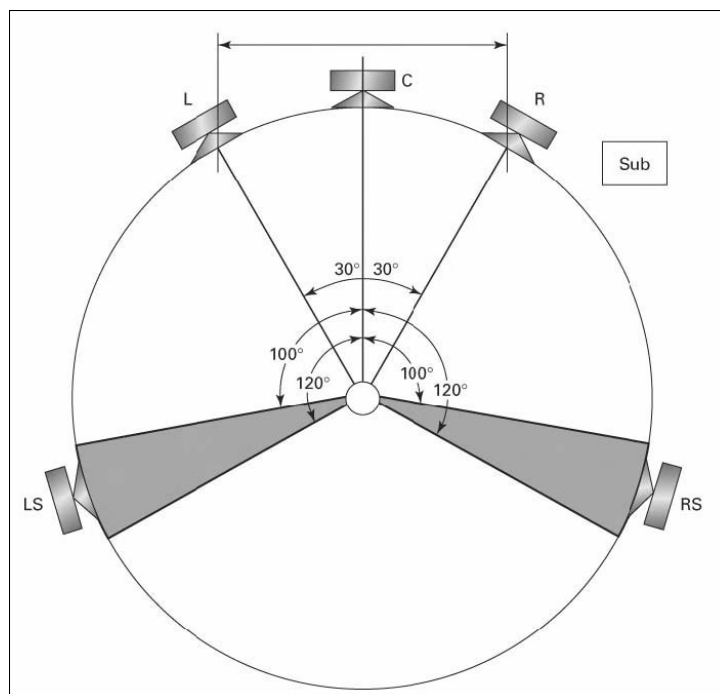


Figura ilustrando o posicionamento correto dos alto-falantes no sistema *surround* 5.1. Fonte: HOLMAN, 2008.

Enquanto a indústria cinematográfica ganha força com a revolução dos sistemas *surround*, a indústria fonográfica se beneficia também. Uma vez que os consumidores possuem sistemas 5.1 instalados em suas casas, eles podem reproduzir lançamentos musicais e transmissões de televisão utilizando o mesmo sistema, talvez até mesmo não acompanhados por imagens. Apesar de alguns engenheiros de som se oporem a tal idéia devido ao fato do sistema 5.1 não ser ideal, ele se mostra uma boa oportunidade para a intensificação da experiência musical de forma espacial para muitas pessoas⁴⁹. Tal afirmação de RUMSEY (2001) pode ser comprovada pela posição de MALHAM (1998):

“Apesar do sistema *surround* 5.1 estar sendo utilizado cada vez mais para aplicações musicais e composições, ele não foi desenhado para tal. Pode-se argumentar que um sistema para tais aplicações deve apresentar uma imagem sonora homogênea e coerente. O sistema desenvolvido pelo cinema não cumpre tal critério, no entanto pode-se contornar tais problemas através

⁴⁹ RUMSEY, 2001, p.18.

de trabalho em estúdio ou utilizando processamento computacional.”
(MALHAM, 1998).

MALHAM (1998) comenta que na indústria cinematográfica é imperativo que a imagem esteja em primeiro plano e que a atenção do ouvinte não se desvie da mesma, o que leva à adoção de sistemas mais focados à frente do ouvinte. Por esse mesmo motivo o sistema 5.1 adotado pelo cinema não é aconselhado para aplicações musicais por privilegiar uma pequena parte (a frente) em detrimento das outras regiões do espaço sonoro, mas ainda assim é muito utilizado para esse propósito. Pode ser muito bem aproveitado, no entanto, para gravações de tradição ocidental, onde se possui um palco muito bem definido à frente do ouvinte não oferecendo uma imersão completa em um outro ambiente sonoro⁵⁰.

Apesar do sistema 5.1 não ser um sistema ideal, RUMSEY (2001) afirma que este é um compromisso que tomamos para manter a compatibilidade com os sistemas estéreos, os sistemas *surround* do cinema e as aplicações para televisão, oferecendo uma real possibilidade para o desenvolvimento do áudio espacial. Muitos têm observado que este sistema não é a melhor opção para a reprodução musical, e este argumento tem certo mérito, embora se torna menos relevante ao considerarmos que não é muito comum que os consumidores instalem dois sistemas de som em suas casas, um para assistir filmes e outro para ouvir música⁵¹.

1.3.7 – O Ambisonics

O sistema Ambisonics foi desenvolvido na década de 70 por Gerzon, Fellgett e Barton. O fundamento principal desse sistema está na abordagem de Alan Blumlein, em seu trabalho, na EMI, Inglaterra, quando conseguiu excelentes resultados na recriação de ambientes sonoros com apenas dois microfones (com padrão polar bidirecional) cruzados sendo reproduzidos em dois alto-falantes. O sistema desenvolvido a partir daí pretende ser uma abordagem hierárquica completa com relação à localização espacial sonora,

⁵⁰ MALHAM, 2003.

⁵¹ RUMSEY, 2001, p.XI.

seu armazenamento, transmissão ou reprodução, que seja igualmente aplicável aos sistemas mono, estéreo, *surround* horizontal ou reprodução completa perifônica que inclui informações posicionais de altura.

Segundo RUMSEY (2001), este sistema pode ser adaptado para qualquer configuração de alto-falantes e possibilita a criação de ambientes sonoros tridimensionais. Permanece como uma ferramenta para o engenheiro de som que acredita na reconstrução vetorial da localização das fontes sonoras com a posição do ouvinte como a chave para uma reprodução espacial de alta qualidade. O autor afirma que, apesar de ser muito utilizado como um formato de representação interna para ambientes sonoros de produtos 3D, a tecnologia não foi aceita no campo comercial principalmente pelas dificuldades em se comunicar suas vantagens e potenciais, ficando restrito a um pequeno grupo de entusiastas ⁵².

MALHAM (1998) acrescenta que, no sistema Ambisonic, o som e seus componentes direcionais são codificados vetorialmente num conjunto de harmônicos esféricos que, no caso da representação de uma imagem 3D completa, são quatro. Esses sinais são conhecidos como *formato-B* e ao decodificá-lo pode-se utilizar um conjunto de alto-falantes em 3D. Os resultados sobre a esfera criada em torno do ouvinte podem ser tão boas quanto um sistema estéreo o faria na frente do mesmo. Tal abordagem pode ser utilizada para formar a base de um sistema que possua uma saída binaural seja ela para fones de ouvido ou alto-falantes (transaural).

Segundo o autor, existem somente dois requisitos aos quais um sistema em Ambisonics deve cumprir. Primeiramente um número mínimo de alto-falantes que, para representar um ambiente em 2D deve ser quatro em forma de um retângulo e, para representar 3D devem ser oito em formato de cubo. O segundo requisito é ajustar o algoritmo que decodifica o sinal de áudio para a configuração e quantidade de alto-falantes instalados.

Também segundo MALHAM (2003), o sistema Ambisonics possibilita ao ouvinte uma sensação quase holográfica de estar no local original da gravação, embora esse formato não tenha sido aceito mais amplamente principalmente devido ao fato de não ser compatível com as necessidades do som *surround*

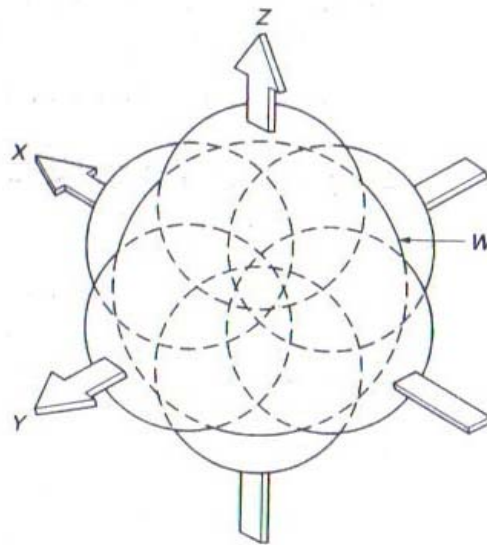
⁵² RUMSEY, 2001, p.17.

da indústria cinematográfica. Segundo ele, o *formato-B* do sistema Ambisonics não é diretamente compatível com o sistema estéreo, no entanto um sistema britânico denominado UHJ permite a codificação das informações horizontais do *formato-B* para o sistema estéreo.

RUMSEY (2001) atualiza essas informações ao afirmar que, apesar de o sistema prever reprodução em arranjos simétricos, ou seja, para quatro, oito, ou mais alto-falantes, já foi apresentado uma forma de decodificação dos sinais desse sistema para sistemas *surround* 5.1, com relativo sucesso.

MALHAM (2003) menciona que o sistema *surround* Ambisonics é essencialmente uma solução tecnológica para resolver o problema da codificação da direção do som (e amplitudes) e da reprodução do áudio em um sistema de alto-falantes que seja prático e no qual o ouvinte possa perceber o som de forma tridimensional. Isso pode ocorrer em 360° somente na horizontal (sistema *pantofônico*) ou de forma esférica (sistema *perifônico*). O sistema codifica o áudio no *formato-B*, que contém três canais para o sistema *pantofônico* e um adicional para o *perifônico*. Esses sinais são codificados a partir da informação de microfones de primeira ordem (cardióides, figura 8, etc) e necessita de mais de quatro alto-falantes dependendo do tipo do sistema (*panto-* ou *perifônico*) e do tamanho do espaço de performance.

Segundo MALHAM (1995), a maior vantagem do sistema Ambisonics consiste na simplicidade de representação de fontes sonoras tridimensionais e algumas das possibilidades de manipulação e geração do campo sonoro são exclusivas desse sistema. Existe também uma possibilidade de se converter esse formato para outros, como o binaural.



Os componentes W, X, Y e Z, do formato Ambisonics B, representam uma componente de pressão omnidirecional e três componentes de velocidades ortogonais (figuras oito) do campo sonoro, respectivamente. Fonte: RUMSEY, 2001.

A fonte de um sinal para o sistema Ambisonics deve ser um microfone próprio, denominado Calrec Soundfield (formato-A), ou sinais monofônicos, provenientes de microfones de primeira ordem, como mencionado anteriormente, devidamente codificados nas quatro componentes do formato-B. O formato-C, por sua vez, consiste em quatro sinais (L, R, T e Q) que são usados para a transmissão e compatibilidade com sistemas mono e estéreo, já os sinais do formato-D são utilizados na distribuição do áudio entre os alto-falantes durante a reprodução e são definidos a partir do layout escolhido.

RUMSEY (2001) afirma que o microfone denominado “*Soundfield*”, ou de campo sonoro, foi desenvolvido para captação de uma perifonia completa no formato A Ambisonics, e é acoplado a um processador que possui saída nos formatos B e D. Alguns decodificadores podem ser desenvolvidos para estabelecer a compatibilidade com o sistema *surround* 5.1, assim como para estéreo, mas a perifonia completa pode ser obtida apenas através da reprodução em um conjunto de alto-falantes tetraédrico capazes de reproduzir informações referentes à componente altura.



Microfones "SoundField" desenvolvidos para captação do sinal no formato B Ambisonics. Fonte: SoundField Ltd.



Microfones "SoundField" com seus respectivos processadores. Fonte: SoundField Ltd.

A partir do entendimento das diversas possibilidades oferecidas pelos variados sistemas de reprodução de áudio oferecidos, seja ele tridimensional ou não, segue-se um aprofundamento maior, focado no sistema *surround* 5.1, incluindo possibilidades de microfonação, particularidades com relação às especificações de sistemas e salas para sua reprodução e uma análise das possibilidades estéticas adotadas por produções musicais e cinematográficas lançadas para tal sistema.

1.3.8 – *Wave Field Synthesis*

Segundo BRANDENBURG (2009), os primeiros experimentos que envolveram os princípios da *Wave Field Synthesis* (síntese por frente de ondas), aconteceram nos laboratórios Bell há mais de vinte anos atrás e utilizavam “cortinas acústicas”, que consistiam em diversos alto-falantes e tinham como propósito reproduzir um campo sonoro com certa fidelidade. A partir desse princípio, implementações utilizando somente dois ou cinco alto-falantes (sistemas estéreo e *surround* 5.1 respectivamente) surgiram como um compromisso entre a fidelidade e os recursos disponíveis. Pesquisas recentes retomam esse sistema, visto a capacidade atual de processamento de áudio e a sua aplicação em grandes salas de exibição, tais como o cinema e ambientes temáticos.

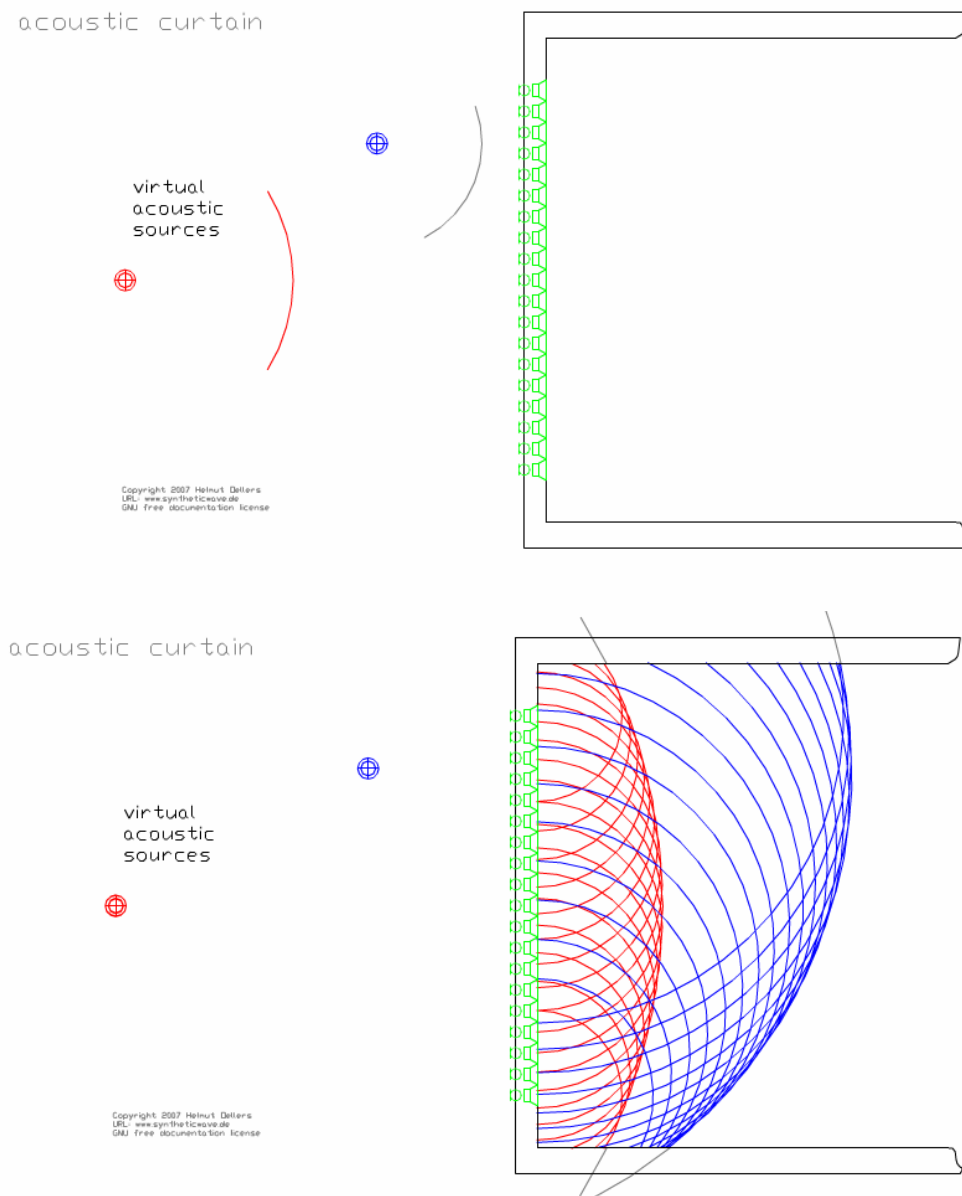
Todos os sistemas vistos até aqui que utilizam um arranjo de alto-falantes compartilham de limitações semelhantes, tais como o *sweet spot* limitado ou a recriação imprecisa de imagens fantasma para o estabelecimento de fontes sonoras entre dois alto-falantes. A *Wave Field Synthesis* se desenvolveu com o intuito de recriar a frente de ondas criada por qualquer campo sonoro através de um conjunto de alto-falantes com certa fidelidade, e tem como objetivo principal aumentar o *sweet spot* oferecido por um sistema *surround*.

O princípio no qual o método em questão se baseia foi estabelecido por Huygens, onde ele considera que os diversos pontos de uma frente de ondas se comportam como fontes sonoras pontuais individuais que podem ser representadas por pequenos alto-falantes próximos entre si e organizados linearmente. Esse conjunto de alto-falantes é projetado para recriar frentes de ondas com fidelidade temporal, espectral e espacial na sala onde se encontra instalado.

Uma das limitações desse sistema está no fato de ele conseguir reproduzir somente um plano horizontal e não representar um *surround* 3D completo incluindo informações referentes à elevação das fontes sonoras⁵³. Efeitos como o *aliasing* podem surgir quando o limite de uma polegada entre os

⁵³ <http://www.holophony.net/Wavefieldsynthesis.htm>

alto-falantes não é respeitado, assim como uma certa distorção da acústica natural da sala de execução e o efeito de truncamento, no caso do sistema não estar envolvendo completamente o ouvinte, que consiste no surgimento de ondas resultantes da interação de outras ondas que não contribuem para o resultado total da reprodução do espaço sonoro e causam distorções no mesmo.



Figuras representando o princípio de funcionamento da *Wave Field Synthesis*. Fonte: <http://www.holophony.net/Principle%20of%20wave%20field%20synthesis.htm>

Capítulo 2

O Sistema *Surround* 5.1: entre o ideal e o viável

2.1 – Histórico e Desdobramentos

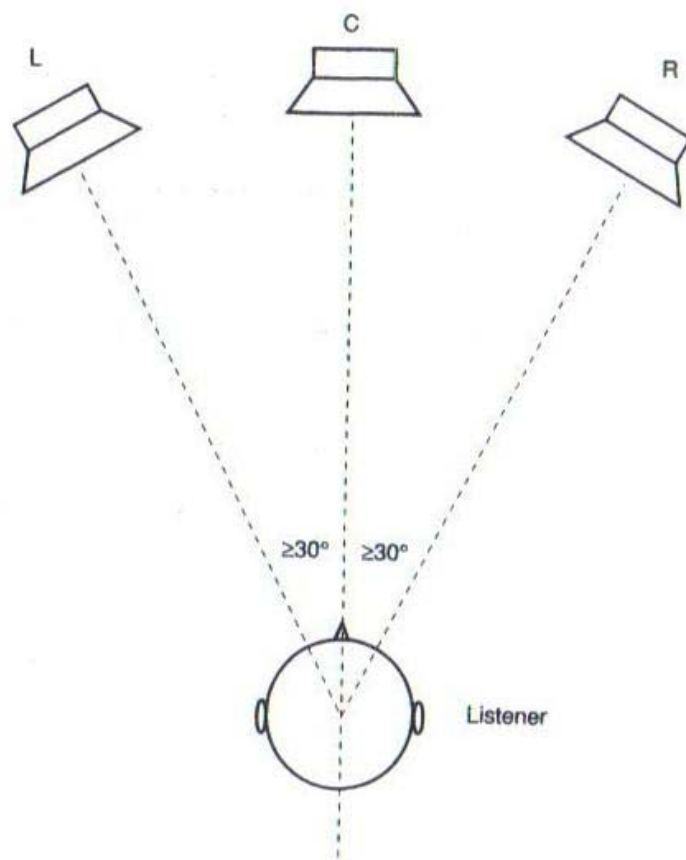
O sistema *surround* 5.1, já abordado no capítulo anterior, nos traz diversas possibilidades com relação à espacialização do áudio num espaço sonoro ampliado com relação ao tradicional estéreo. Apesar de esse espaço sonoro ser reproduzido por alto-falantes de forma assimétrica e com diversos problemas relacionados à estabilidade de imagens fantasma criadas entre eles, esse sistema ainda é a forma mais rápida e simples de se levar ao ouvinte uma experiência sonora mais interessante no que diz respeito à imersão sonora.

Por se apresentar como um sistema ainda muito recente, existem muitas questões referentes à realização da captação e a mixagem para o sistema *surround* 5.1, principalmente com relação à distribuição do material musical em alto-falantes tais como o canal central e o *subwoofer*.

RUMSEY (2001) afirma que a padronização dos sistemas *surround*, assim como de qualquer outro sistema de reprodução de áudio, normalmente consiste em mais do que simplesmente estabelecer as configurações dos canais e a forma como posicionar os alto-falantes; eles fornecem também ao mercado e principalmente ao profissional de áudio algumas noções de como criar ou representar o campo sonoro espacial para o ouvinte.

A nomenclatura padrão internacional de representação de configurações de alto-falantes se apresenta na forma *n-m estéreo*, onde *n* é o número de canais frontais e *m* o número de canais traseiros ou laterais. Com isso, o sistema estéreo com três canais seria representado por 3.0 (três canais frontais e nenhum traseiro) e o *surround* que conhecemos como 5.1, na verdade possui a nomenclatura oficial de 3.2 (três canais frontais e dois traseiros). No entanto devido ao fato da nomenclatura *surround* 5.1 ter sido largamente adotada no meio profissional e pelos consumidores, utiliza-se ela para referência ao sistema 3.2.

O sistema estéreo com três canais (L, R e C) é a base de todos os outros sistemas multicanais e, segundo RUMSEY, têm seus precedentes históricos no desenvolvimento do sistema estéreo de STEINBERG e SNOW nos anos 1930 que utilizava três canais. Devido ao grande afastamento dos canais L e R, resultado do tamanho da tela e à necessidade de se cobrir a grande área onde se localizam os ouvintes, esse sistema estéreo com três canais frontais teve grande aceitação no cinema, onde a presença do canal central possibilita uma maior faixa privilegiada de posições de ouvintes além de fixar os diálogos na região central da tela.

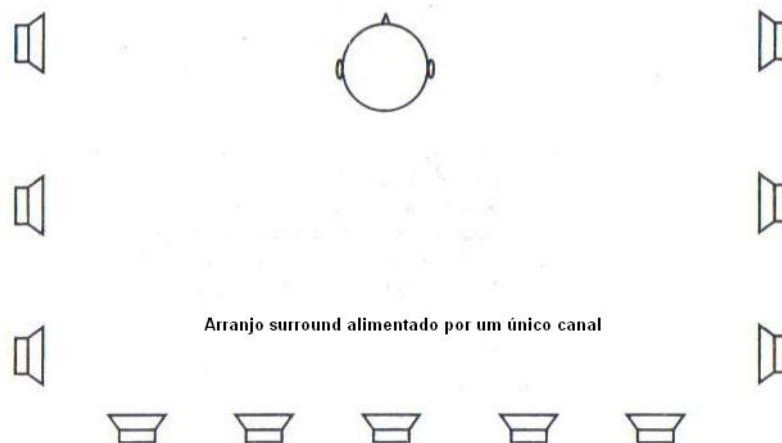
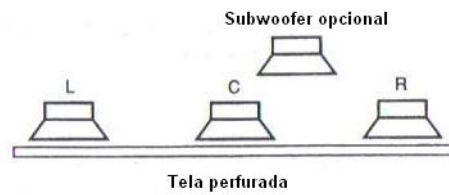


Sistema de reprodução estéreo com três canais espaçados igualmente em frente ao ouvinte com um ângulo de 60° entre os dois extremos. Fonte: RUMSEY, 2001.

Os dois canais do sistema estéreo tradicional somente se consolidaram nos sistemas caseiros por questões de economia e conveniência, e particularmente devido à facilidade de se imprimir dois canais no disco ou fita analógicos ao invés de três. Um problema apontado por RUMSEY (2001) com relação ao sistema baseado em três canais frontais é a inconveniência da

localização do canal central em sistemas domésticos, que muitas vezes é posicionado acima ou abaixo da tela e dos outros alto-falantes, e é construído em menor tamanho para se adaptar melhor a essa situação.

Durante o desenvolvimento do sistema *surround*, um sistema com quatro canais foi desenvolvido com o propósito de intensificar os efeitos sonoros dos filmes. Tal canal adicionado, denominado canal *surround*, em adição aos três canais frontais, recebia material sonoro que era distribuído para um ou mais alto-falantes posicionados atrás e/ ou nas laterais do ouvinte. Inicialmente desenvolvido para aplicações cinematográficas, possibilitava um maior grau de envolvimento do ouvinte, no entanto não havia intenções nem capacidade de se estabelecer imagens sonoras nos 360° ao redor do ouvinte, mesmo porque o material sonoro direcionado para o canal *surround*, seja através de um ou vários alto-falantes, era monofônico e de menor importância com relação ao material exposto à frente do ouvinte. A reprodução deste canal através de vários alto-falantes tinha como objetivo a simples cobertura de uma maior área de audiência.



Sistema em formato 3-1, utilizando um único canal *surround* endereçado para diversos alto-falantes que, em sistemas domésticos, pode ser simulado em dois canais *surround*, intensificando a impressão sonora mais difusa presente na apresentação no cinema. Apresenta também um canal de *subwoofer* opcional. Fonte: RUMSEY, 2001.

RUMSEY (2001) menciona que a maior limitação desse sistema, o canal *surround* monofônico era, no entanto, sua maior novidade. Muitas pesquisas psico-acústicas que se dedicaram a esse objeto como foco sugerem a necessidade de sinais diferentes para se criar alguma sensação de envolvimento, e que os efeitos *surround* poderiam ser melhor espacializados através de dois ou mais canais.

No sistema atual, conhecido como 5.1, a desvantagem do canal *surround* monofônico foi superada pela implementação de um par de alto-falantes traseiros que possibilitam uma ambiência estéreo acompanhando um sinal frontal primário. RUMSEY afirma que:

“Essencialmente os três canais frontais são utilizados para a reprodução de uma imagem sonora frontal convencional, enquanto os canais traseiros/ laterais pretendem somente gerar ambiência suporte, efeitos ou uma impressão da sala. Nesse sentido o padrão não suporta diretamente o

conceito de localização de uma imagem em 360°, apesar de isso poder ser alcançado através de técnicas de gravação ou métodos de processamento de sinais com esse propósito” (RUMSEY, 2001, p.87).

Um grande problema do sistema *surround* 5.1, mencionado por diversos profissionais, diz respeito à sua instabilidade, uma vez que os diversos ângulos entre os cinco alto-falantes fazem com que a estabilidade da imagem sonora reproduzida seja menos confiável em alguns setores do que em outros. Para quem não possui acesso às sofisticadas leis de controle panorâmico ou matrizes psico-acústicas, necessárias para alimentar os cinco canais de forma precisa, pode ser melhor tratar o formato no “estilo do cinema”, com uma imagem frontal bem convincente e dois canais *surround* para efeitos e ambientação. Com essa abordagem, mesmo que não de forma muito precisa, ainda é possível criar ilusões espaciais relativamente convincentes, com bom envolvimento sonoro e localização.

De acordo com os padrões internacionais, os alto-falantes direito e esquerdo devem ser posicionados em aproximadamente 30° em relação ao central, principalmente para manter certa compatibilidade com o sistema estéreo tradicional. Os canais *surround* são posicionados a aproximadamente 110° em relação ao central, de forma que se assume certo compromisso entre a necessidade de direcionar efeitos para trás do ouvinte e fornecer a energia lateral necessária para um bom envolvimento.

O canal de efeitos de frequências graves (LFE – *low frequency effects*) do sistema *surround* 5.1 é um canal de sub-graves com o limite superior de 120Hz. Inicialmente foi desenvolvido para a intensificação de efeitos sonoros de baixa frequência e grandes pressões sonoras, tais como explosões, ruídos e quedas de grandes objetos, que os canais principais não suportam. Não foi desenvolvido com o propósito de reproduzir as componentes de baixa frequência dos canais principais, e sua aplicação é primeiramente direcionada para as produções cinematográficas onde esses efeitos de grande intensidade são amplamente explorados.

A reprodução do filme normalmente é calibrada para não ultrapassar certo limite de intensidade nos canais principais, limitando principalmente a intensidade da região grave do espectro sonoro. A utilização de um canal

dedicado a essa região de frequências permite que ele possa ser calibrado 10dB acima dos canais principais e que se obtenha uma reprodução desses efeitos muito melhor do que a que poderia ser obtido através dos canais principais. No cinema esse canal LFE é reproduzido por um *subwoofer* dedicado, enquanto nos sistemas caseiros o que ocorre é uma filtragem que inclui tanto os canais principais, quanto o canal LFE, que são direcionados para um alto falante com resposta mais grave.

Na mixagem para qualquer aplicação, os sinais graves abaixo de 120Hz não necessariamente devem ser enviados ao canal LFE, a menos que se deseje um efeito específico, como nas aplicações cinematográficas. É muito normal que, em mixagens de música, se utilize todos os canais reproduzindo todo o espectro sonoro, de forma semelhante ao sistema estéreo; nesse caso o canal grave pode ser utilizado somente para dar ênfase a efeitos dramáticos. Uma das razões principais para isso é a de que o canal LFE normalmente é descartado numa redução para o sistema estéreo através de um decodificador, de forma que algum conteúdo pode ser perdido se endereçado somente a ele. Com isso a gravação deve ser satisfatória mesmo sem a utilização deste canal e é importante que qualquer material sonoro de baixa frequência que seja muito significativo para a integridade do conteúdo musical não seja direcionado exclusivamente para ele.

Muitas questões surgem também na utilização do canal central em aplicações musicais, uma vez que nelas não se tem um diálogo para ser apresentado no centro da imagem frontal como ocorre em aplicações com imagem integrada, e os principais problemas da sua utilização dizem respeito à técnica de panorâmico utilizada, à microfonação e à conversão de uma mixagem *surround* para estéreo.

No que diz respeito ao conteúdo musical endereçado aos canais traseiros, os microfones utilizados para a captação de ambiência ou a reverberação artificial obtida através de processamento normalmente são direcionados para eles. Apesar dessa orientação, muito seguida e bem sucedida para diversos estilos musicais, nada impede que eles sejam utilizados para espacialização de material musical primário em tentativas de se estabelecer imagens sonoras nos 360°, como se verá mais a frente neste trabalho.

Alguns engenheiros realizam a mixagem em *surround* e em estéreo simultaneamente, outros fazem duas mixagens diferentes, uma para cada sistema, e preferem utilizar quatro canais ao invés de cinco, prática apontada por RUMSEY (2001) como uma facilidade no momento de estabelecer certa compatibilidade do sistema *surround* 5.1 com o estéreo.

No que diz respeito à movimentação de fontes sonoras, observamos uma maior exploração em aplicações cinematográficas, uma vez que podem acompanhar imagens e dar mais realismo à situação apresentada visualmente, embora RUMSEY observe que essa movimentação pode ser estranha em aplicações musicais uma vez que não necessariamente temos uma imagem para explicar o que está acontecendo com o som.

Segundo RUMSEY (2001), até mesmo a Dolby foi a primeira a atestar que o sistema *surround* 5.1, assim como outros formatos *surround* desenvolvidos para o cinema, nunca foram desenvolvidos para a reprodução de programa musical, apesar de muitos terem se aventurado na mixagem de música nesse sistema e obtido resultados satisfatórios.

A respeito das limitações do sistema 5.1, o autor acima citado afirma que a principal delas está na já mencionada incapacidade do sistema *surround* 5.1 de reproduzir de uma forma precisa imagens sonoras fantasmas nos 360° de cobertura dos alto-falantes. Uma segunda limitação está no fato de que a imagem frontal do sistema é menor do que seria se ele não necessitasse ser compatível com o sistema estéreo tradicional, realidade que ainda deve ser observada no processo de produção. Uma terceira limitação está no fato de que o canal central pode se mostrar um problema para o equilíbrio do sistema para aplicações musicais, uma vez que leis de panorâmico convencionais e técnicas de microfonação não foram otimizadas para três alto-falantes e sim para dois. Uma quarta limitação mencionada pelo autor está no fato de que os canais *surround* (Ls e Rs) estão localizados em uma posição comprometida, levando a um grande buraco na imagem sonora atrás do ouvinte e tornando difícil encontrar locais físicos para os alto-falantes em salas práticas.

No que diz respeito à limitação do *sweet spot* do sistema de reprodução *surround* 5.1, ou seja, a região sobre a qual o sistema funciona efetivamente, também uma limitação muito comentada por diversos autores, uma pesquisa apresentada no 7º Congresso da AES Brasil em 2009, pelos autores Jean-

Marc Lyswa e Alexis Baskind (pesquisadores do IRCAM/ CNSM de Paris), se utiliza de técnicas de espacialização binaurais e transaurais na reprodução multicanal em 5.1.

Na apresentação, eles realizaram uma mostra com trechos de peças eletroacústicas em um sistema *surround* 5.1, onde grande parte do público conseguia observar a espacialização pretendida sem grandes mudanças quando alteravam seu posicionamento ou realizavam movimentos de rotação, laterais ou frontais com a cabeça. O objetivo específico dessa pesquisa é o de minimizar os efeitos desses movimentos na percepção do espaço reproduzido pela mixagem 5.1 e posteriormente elaborar o produto no formato de um *plugin*, inicialmente denominado *transPan* e que se baseia no objeto spat~ do MAX/MSP, que apresenta as diversas opções demonstradas pelos autores. O processo, como exposto pelos próprios autores, consiste em basicamente quatro etapas:

- 1- Captação através de uma técnica *surround* clássica determinada pela própria composição e instrumental a ser captado, objetivando estabelecer uma estrutura espacial.
- 2- Mixagem desse material com o sinal de microfones com captação próxima e processamento corretivo baseado em *delays* e que objetiva uma maior definição dos instrumentos.
- 3- Processamento binaural com processamento de HRTFs para definição dos instrumentos no espaço através da resposta em frequência.
- 4- Processamento transaural para cancelamento de sinais cruzados e compatibilidade com os alto-falantes, realizado com o par de alto-falantes frontais e, posteriormente, com o par de alto-falantes traseiros.

Com relação ao alinhamento de níveis para reprodução no sistema *surround* 5.1, as recomendações ITU e EBU são de -18dBFS para o nível de alinhamento e -9dBFS como máximo de sinal permitido, isso para estúdios de gravação e televisão. Em situações de masterização e operações de som para filmes é muito comum se utilizar toda a faixa dinâmica, ou seja, o nível máximo permitido de 0dBFS. Também é norma para mixagem de música e televisão alinhar todos os canais para níveis iguais tanto na gravação da mídia quanto na

monitoração acústica, embora a transferência do sinal do filme para sistemas caseiros exija uma alteração de -3dB no sinal dos canais *surround*.

Apesar do sistema *surround* 5.1 estar se tornando amplamente divulgado e adotado como padrão pela maioria das instalações, existem diversos outros sistemas e propostas que envolvem mais canais e cobrem uma área maior de ouvintes de forma mais precisa. É razoável assumir que, quanto mais alto-falantes existirem em torno do ouvinte, menor será a necessidade de se criar imagens “fantasmas” para o posicionamento de fontes sonoras, e maior será a liberdade do ouvinte de se movimentar; em contrapartida, maior será a dificuldade de se mixar algo em um sistema com tantos canais.

O formato 7.1, por exemplo, implementado primeiramente pela Sony com o nome de SDDS, – formato 5.1 *surround* adicionado de dois outros alto-falantes, CL e CR, posicionados entre os canais L, R e o central - não é um formato pensado para o consumidor final, mas sim para sistemas de grandes auditórios, onde o tamanho da tela (constituída de material perfurado, o que permite a passagem do som) exige que alto-falantes adicionais sejam instalados para cobrir todo o público.

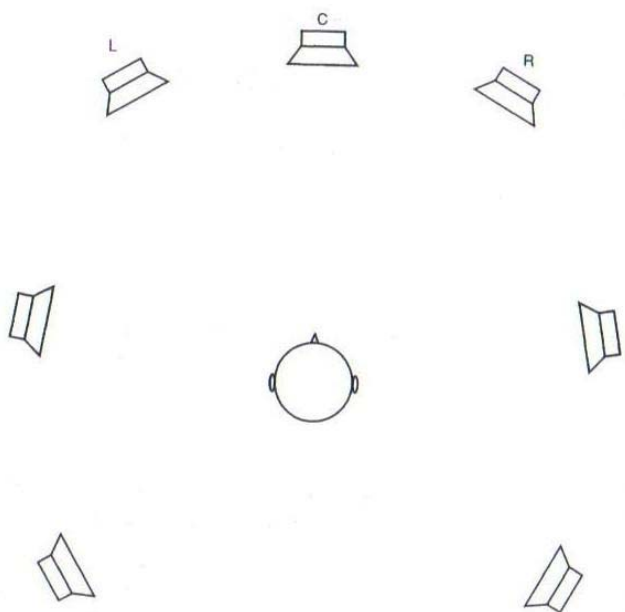
Outro exemplo mencionado por RUMSEY (2001) é o de Tomlinson Holman, que desenvolveu um formato 10.2 onde, ao sistema *surround* básico (5.1), são adicionados dois alto-falantes frontais laterais e um canal central traseiro. São adicionados também dois canais mais altos e um segundo LFE, destinado a promover alguma separação lateral do conteúdo grave, como sugerido por outros autores para aumentar a espacialização das frequências graves.

A empresa Dolby, sempre envolvida no desenvolvimento de sistemas para o cinema, foi a primeira a implementar em mídias de 35mm e 70mm a codificação de mais de dois canais de áudio, respectivamente 4 canais (L, R, C e S) e 6 canais (L, R, C, LC, RC e S), mesmo apresentando sempre um único canal *surround*. O Dolby Stereo, como ficou conhecido, consiste no sistema já mencionado de quatro canais de áudio (L, R, C e S), codificados na fita de 35mm e utilizado largamente na indústria cinematográfica. O Dolby Surround consiste na codificação desses canais de forma a emular, em um ambiente doméstico, os efeitos do cinema, possibilitando que filmes do cinema possam ser assistidos em ambientes domésticos. Já o sistema Dolby ProLogic foi

desenvolvido para incluir um mecanismo que possibilita uma maior separação entre os canais.

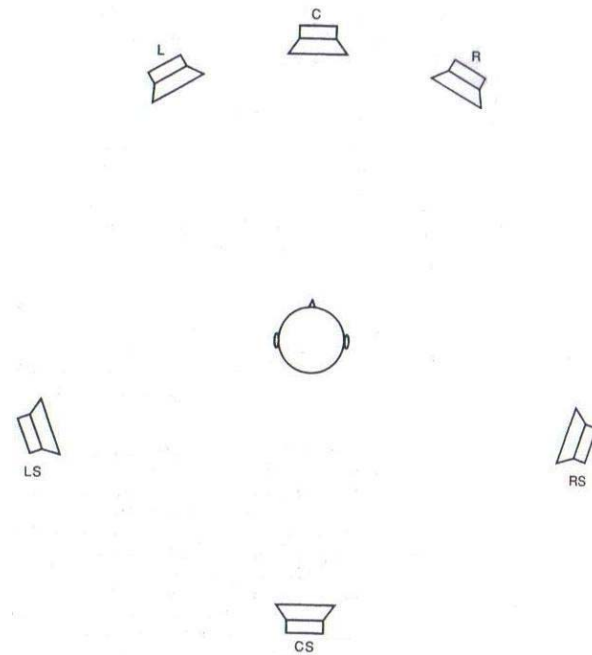
Um outro sistema, também mencionado por RUMSEY (2001), denominado Circle Surround e desenvolvido pela Rocktron, é capaz de codificar canais estéreo *surround* em adição aos canais frontais convencionais, o que faz esse sistema mais apropriado para a reprodução de música do que o da Dolby, por exemplo.

A fabricante Lexicon desenvolveu o Logic 7, codificador que poderia ser utilizado no lugar do Dolby Surround e possui algoritmos específicos para música. Desenvolvido para um sistema doméstico, possui sete alto-falantes, sendo dois posicionados nas laterais em adição aos já estabelecidos pelo formato 5.1, com a função de garantir que sinais laterais fossem melhor localizados quando, por exemplo, um efeito que passasse do canal L para a parte de trás, passasse pelo canal lateral.



Posicionamento de alto-falantes para o sistema Logic 7 da Lexicon, apresentando os dois canais laterais adicionais. Fonte: RUMSEY, 2001.

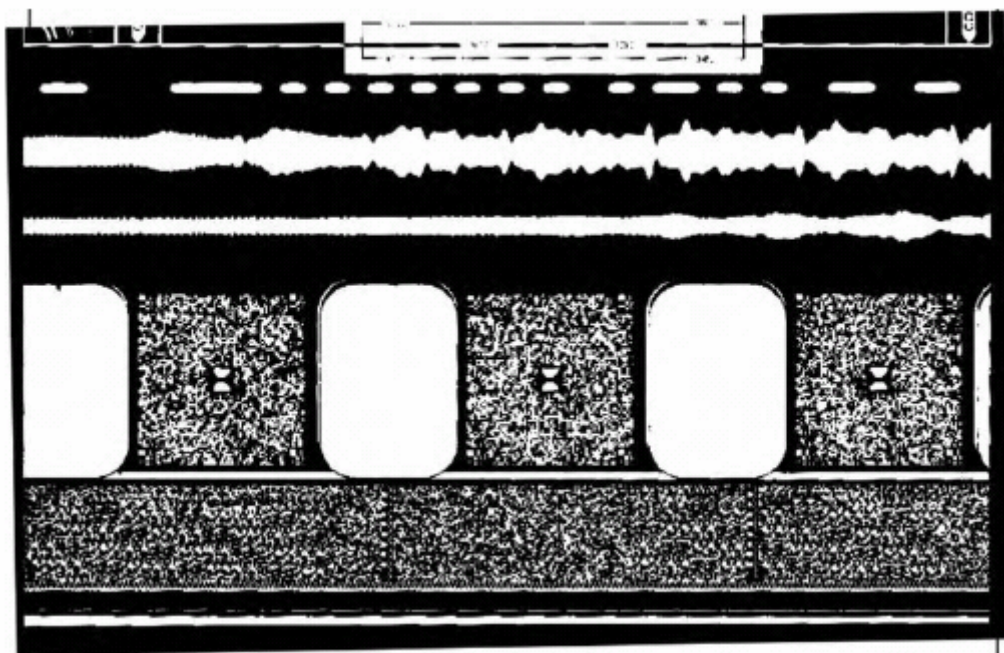
Em 1998 a Dolby, juntamente com a Lucasfilm desenvolveu um sistema denominado de “Dolby Digital – Surround EX”, que, ao adicionar um canal central traseiro ao formato básico do sistema 5.1, possibilita uma melhor localização de sons atrás do ouvinte. A codificação desse canal é uma derivação dos outros dois canais traseiros.



Sistema Dolby EX com o canal central traseiro adicional, originalmente obtido através da codificação dos dois canais *surround*, semelhante ao processo de codificação do Dolby Stereo. Fonte: RUMSEY, 2001.

Devido à grande quantidade de canais dos sistemas *surround*, se todos fossem gravados na mídia em formato PCM, não haveria espaço disponível, por isso eles precisam ser comprimidos.

O formato AC-3 ou Dolby Digital é o utilizado em cinemas e *home-theaters* para a compressão de sinais baseados no padrão 5.1 em 35mm, e possui seus dados armazenados de forma ótica no espaço entre os buracos laterais da fita do filme. Algumas diferenças podem ser ouvidas com relação à qualidade do áudio comprimido e o não-comprimido, no entanto elas são mínimas e são resultado de um compromisso entre qualidade sonora e taxa de transferência de dados.



Fita de 35mm apresentando de cima para baixo, o limite da área destinada ao conteúdo visual, o canal com o time code no formata DTS, os dois canais com o áudio estéreo, o áudio Dolby Digital entre as perfurações e metade dos dados correspondentes ao áudio digital Sony SDDS (a outra metade se encontra fora das perfurações na beirada oposta da fita). Fonte: HOLMAN, 2008.

Segundo RUMSEY (2001), o codificador Dolby Digital pode operar em taxas de amostragem de 32KHz a 48KHz e o canal LFE é amostrado em 240Hz devido à sua limitação de espectro. Uma inversão de fase de 90° é aplicada aos canais traseiros, o que aparentemente aumenta a suavidade do panorâmico dos canais frontais e traseiros e reduz o cancelamento entre os canais traseiros e o central quando decodificado pelo Dolby Surround. É importante que o produtor ou engenheiro de som realize a monitoração através do decodificador para garantir que essas modificações não alterem a intenção de espacialização da gravação. O autor menciona que o codificador Dolby Digital inclui uma variedade de parâmetros operacionais para aumentar a flexibilidade do sistema e adaptar a reprodução às mais diversas situações do ouvinte. Isso inclui normalização dos diálogos, e uma opção de controle de faixa dinâmica para utilização onde o ruído de fundo interfere na audição de toda a faixa dinâmica do material.

Já o codificador DTS, também mencionado pelo autor supra citado, pode suportar taxas de compressão de 32kbps a 4Mbps, se mostrando superior ao Dolby Digital, com até oito canais de áudio e taxas de amostragem até 192KHz.

RUMSEY afirma também ser possível se ter um filme em versão multi-formato, ou seja, apresentar a codificação digital Dolby Digital, DTS, SDDS e ainda uma faixa analógica Dolby Stereo. O codificador da Sony (SDDS), também mencionado, apresenta uma codificação de 5.1 com suporte também para 7.1, no entanto não é muito comum nos lançamentos cinematográficos.

O padrão de sinal digital MPEG é largamente utilizado em aplicações com muita compressão de conteúdo de vídeo e áudio, sendo o MPEG-1 descrito como formato estéreo e o MPEG-2 estendido para multicanais (MPEG-2 AAC e MPEG-2 BC), sendo que o MPEG-2 AAC possui um algoritmo mais sofisticado em relação ao BC. Alguns autores e profissionais da área afirmam inclusive que esse padrão AAC possui maior qualidade em comparação ao AC-3 da Dolby Digital.

Entre os formatos digitais de codificação sem perdas no sinal original especificado para utilização em DVDs-Audio ou mídias de grande capacidade, como o *blue-ray* por exemplo, podemos citar o formato MLP (*Meridian Lossless Packing*), que é um formato que suporta uma taxa de amostragem de até 192KHz e resolução de 24 bits para até seis canais de áudio.

RUMSEY menciona também o THX, que não é um formato de codificação em si, mas sim um sistema desenvolvido para melhorar a qualidade sonora dos cinemas, aproximando a experiência do ouvinte à do engenheiro de som durante a pós-produção. É mais direcionada à acústica do cinema e design dos alto-falantes, melhorando características acústicas, diminuindo níveis de ruído e estabelecendo um sistema particular de alto-falantes e *crossovers*.

2.2 - Métodos de gravação e mixagem em *surround* 5.1

2.2.1 – Formatos de Microfonação

As primeiras questões que surgem quando discutimos técnicas de gravação, seja para sistemas estéreo ou *surround*, dizem respeito ao ponto de vista estético do engenheiro de mixagem e/ ou produtor musical, a natureza do material sonoro gravado e aos meios técnicos necessários para alcançar os objetivos estéticos da produção. A gravação é um processo tanto científico

quanto artístico e não se trata apenas de obter com precisão a reprodução de espaços e fontes sonoras naturais.

Segundo RUMSEY (2001), algumas técnicas de microfonação, consideradas “puristas”, se distinguem por envolver arranjos minimalistas de microfones capazes de capturar as características acústicas na posição do microfone com razoável precisão espacial e tímbrica.

As diversas técnicas de captação estéreo representaram um maior aperfeiçoamento sobre o sistema monofônico com relação à fidelidade da reprodução espacial, pois com elas é possível criar imagens fantasmas entre os alto-falantes e criar uma maior sensação de espacialização na reprodução de um espaço real. Apesar de muitos alegarem que tais técnicas são capazes de envolver completamente o ouvinte, o que se cria é uma ilusão de envolvimento, uma vez que todo o som reproduzido é proveniente apenas de dois alto-falantes que cobrem apenas 60° em frente ao ouvinte. Até mesmo a reverberação, maior responsável por essa sensação de envolvimento, em situações reais, é proveniente também de reflexões atrás dos ouvintes e, quando reproduzida em um sistema estéreo, se originará, assim como os sons diretos, dos alto-falantes em frente ao mesmo.

Já em sistemas *surround* 5.1 temos o campo sonoro representado pelos alto-falantes ampliado para os 360° ao redor do ouvinte, no entanto ainda somente na dimensão horizontal. Até mesmo sistemas *surround* tridimensionais, tais como o Ambisonics, já possibilitam uma grande fidelidade na reprodução, embora ainda exista a necessidade de que o campo sonoro recriado tridimensionalmente seja disponibilizado para múltiplas posições do ouvinte em relação aos alto-falantes.

Abordagens utilizando técnicas binaurais possibilitam tal espacialidade tridimensional, mas restringem a área de posicionamento do ouvinte ou o fazem através da utilização de fones de ouvido. Uma outra observação a ser feita diz respeito ao fato de que o ouvinte raramente se sente na posição ótima de funcionamento do sistema e gosta de se movimentar enquanto ouve.

Diferentemente de aplicações direcionadas a conteúdos musicais, em aplicações de áudio para cinema ou televisão, normalmente são apresentadas ao ouvinte algum conteúdo visual à sua frente que exige sua maior atenção. A maioria dos diálogos se restringe ao centro, efeitos sonoros nos canais

traseiros são utilizados para criar ambientação imersiva e espacial de forma a complementar as imagens e nunca contradizer o que o ouvinte percebe através da visão. Dessa forma lidamos com o paradigma da predominância do espaço frontal, onde sons laterais e traseiros são usados de forma econômica e principalmente na insinuação de um contexto ambiental apropriado.

Como já mencionado anteriormente, esse paradigma da mixagem espacial empregado na gravação será ditado pelas necessidades do material sonoro, pelo enfoque comercial do projeto e pelas preferências estéticas do produtor e engenheiro de som. “Não há regras sobre como as imagens sonoras devem ser criadas ou como o espaço sonoro deve ser usado” (RUMSEY, 2001, p.154).

Gravações de música clássica tendem a aderir à estética de recriação do espaço sonoro com maior fidelidade do que a música pop, por exemplo. Muitas outras gravações, tais como música pop, televisão e filmes, tendem a se utilizar de formas artificiais de equilíbrio que dependem das técnicas de panorâmico e reverberação artificial para reprodução de efeitos espaciais. As limitações criativas de utilização do espaço no formato *surround* 5.1 serão determinadas pela flexibilidade artística da própria produção.

A utilização de técnicas que utilizam sinais provenientes de *Dummy Heads* sem processamento para a reprodução em sistemas estéreo convencionais tem encontrado engenheiros de som adeptos; no entanto outros afirmam que as imagens estéreo obtida através dessas técnicas são desfocadas e imprecisas. Técnicas baseadas em arranjos de pares espaçados de microfones também são muito utilizadas e se baseiam principalmente no fenômeno do efeito de precedência. A precisão de imagens “fantasma” é menor nesses arranjos e a impressão de espacialidade obtida é resultante de cancelamento de fases e filtros “pente”.

RUMSEY (2001) menciona que o autor Griesinger sempre afirmou que o espaçamento dos microfones a pelo menos uma distância equivalente à distância crítica da sala proporciona uma relação adequada entre os microfones para se obter uma boa espacialidade, e essa técnica pode ser muito bem aproveitada para a obtenção de som ambiente em gravações *surround*.

Muitas técnicas propostas para a microfonação em *surround* são extensões das técnicas utilizadas para sistemas com dois ou três canais e pretendem a captação natural da acústica do ambiente, utilizada principalmente na captação de música clássica e outros ambientes naturais. Captações direcionadas ao sistema Ambisonics também entram nesse contexto uma vez que pode ser decodificadas para reprodução em sistemas *surround* 5.1.

As técnicas de microfonação *surround* se dividem em basicamente dois grupos: aquelas baseadas num arranjo de microfones relativamente próximos uns dos outros e aquelas que tratam os canais frontais e traseiros separadamente. A forma do arranjo de microfones sempre se baseia em teorias capazes de gerar imagens “fantasma” com diferentes graus de precisão em um plano horizontal de 360°. O segundo grupo de técnicas normalmente utiliza um arranjo de microfones para a criação de imagens “fantasma” frontais aliadas a formatos de captação da ambiência do espaço gravado. É raro que alguma dessas técnicas preveja a gravação do canal LFE separadamente.

O conceito de utilização de um arranjo ou configuração principal de microfones é mais utilizado por engenheiros europeus do que americanos e esta abordagem européia tende a envolver inicialmente uma técnica de microfonação principal que proporciona uma imagem sonora *surround* de referência e captura os efeitos espaciais do ambiente de gravação de forma satisfatória, e depois aplica diferentes níveis de microfonação próxima, se necessário, para trazer maior definição das fontes sonoras. Técnicas que utilizam microfones próximos como principal fonte sonora, normalmente utilizam reverberação artificial para a intensificação do efeito de envolvimento proporcionado pelo ambiente *surround*.

Diferenças de nível e tempo entre pares de alto-falantes tais como L e Ls ou R e Rs, que se situam nas laterais do ouvinte, são incapazes de gerar imagens sonoras estáveis, embora nessas situações as diferenças de nível sejam mais eficazes. O efeito de um “buraco acústico” ocorre também entre esses canais laterais assim como entre os canais traseiros; por essa razão dentre outras, é mais utilizada a abordagem de se criar imagens estéreo frontais bem estáveis e utilizar os canais traseiros somente para ambientação.

Algumas técnicas de microfonação assim como o sistema Ambisonics têm tido mais sucesso na obtenção de imagens sonoras mais estáveis nos

360° ao redor do ouvinte; no entanto, tais técnicas limitam a região do posicionamento do mesmo.

No grupo de abordagens que separam a imagem frontal da traseira, de uma forma geral, técnicas de microfonação baseadas em microfones omnidirecionais podem ser estendidas para três canais e aplicados na formação dessas imagens frontais, apesar do fato de que uma análise teórica é mais complexa devido ao fato de existirem múltiplas fontes sonoras e do efeito de precedência ser intensificado na reprodução. Experimentos informais do autor RUMSEY (2001) sugerem que uma imagem sonora frontal convincente pode ser obtida através da utilização de uma *Decca Tree*⁵⁴ e direcionando o sinal do microfone central para o alto-falante central ao invés de posicioná-lo entre os alto-falantes dos canais L e R. O espaço entre os microfones direito e esquerdo devem ser diferentes da *Decca Tree* utilizada na gravação estéreo para que haja certa adequação à largura do palco, e as dimensões devem ser ajustadas para compatibilidade com o tamanho do palco e distância das fontes sonoras.

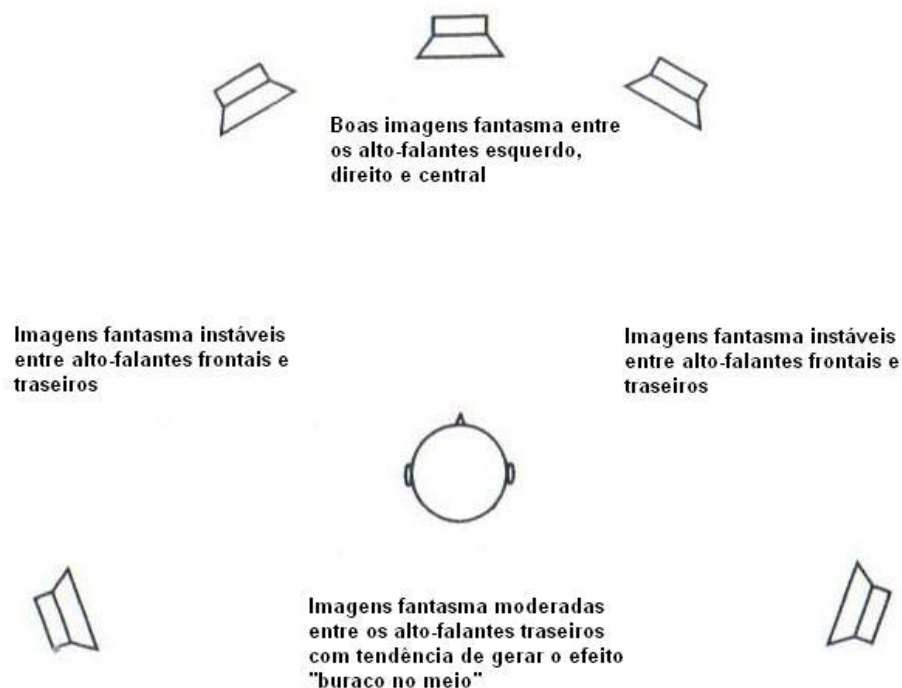
Técnicas de gravação que utilizam pares coincidentes de microfones são problemáticas quando adaptadas para a reprodução nos três canais frontais, pois os padrões polares exigidos para uma correta compatibilidade do sinal com o ouvido do ouvinte para uma ótima localização espacial não são padronizados.

Muitos engenheiros optam por uma microfonação estéreo tradicional adicionada a um microfone central com ganho inferior para a solidificação da imagem central. Tal técnica possui a vantagem de ser mais compatível com projetos que necessitam tanto de uma mixagem em *surround* 5.1 quanto em estéreo, se utilizando de uma mesma microfonação.

Em vista de tudo isso, não há como se escapar do fato de que é mais fácil se criar imagens fantasma onde há alto-falantes próximos e de que as imagens criadas entre alto-falantes que se encontram separados por grandes ângulos são instáveis e criam a sensação do “buraco acústico”. O que observamos a respeito da precisão sonora reproduzida pelos alto-falantes do

⁵⁴ Detalhes adiante.

sistema *surround* 5.1 é que temos mais definição nos canais frontais, definição moderada nos canais traseiros e muito variável nas laterais.



Precisão da imagem sonora representada pelos alto-falantes de um sistema *surround* 5.1. Fonte RUMSEY, 2001.

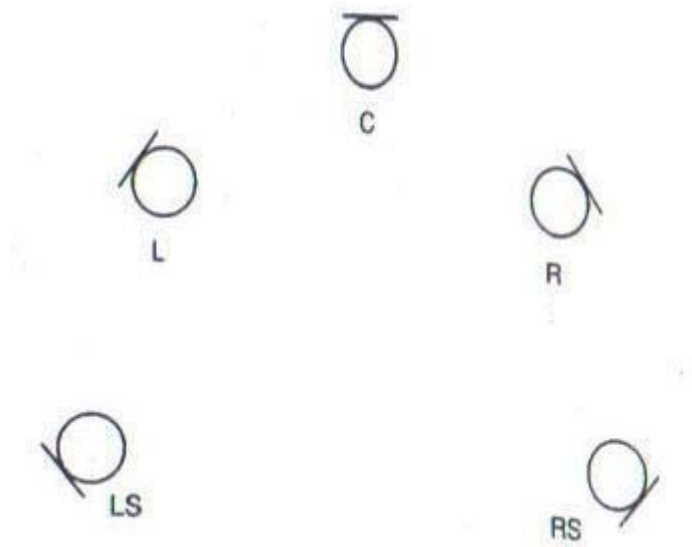
Devemos aceitar também que a maioria dos sistemas caseiros é instalada de forma muito variável com relação à localização e à natureza dos alto-falantes, o que impossibilita a criação de campos sonoros muito precisos em ambientes domésticos. A grande motivação de se utilizar o sistema *surround* 5.1 é a de que ele permite a oportunidade de se criar algo mais rico espacialmente e que funcione bem em uma grande área de posições do ouvinte em relação ao sistema estéreo.



Florian Camerer, da televisão australiana, realizando gravação *surround* de ambientes naturais, utilizadas em documentários. Fonte: HOLMAN, 2008.

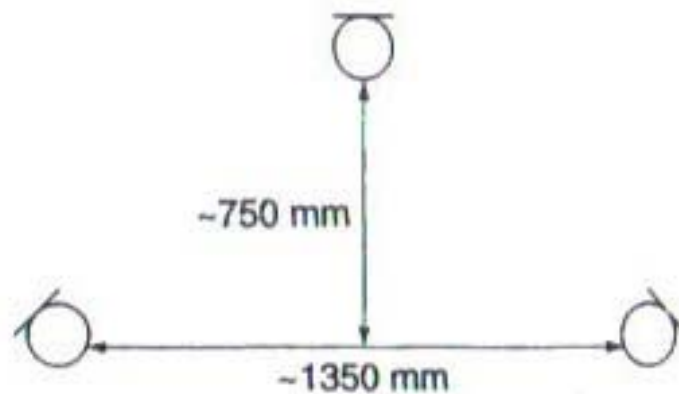
As diversas técnicas de microfonação desenvolvidas para gravação *surround* têm como princípio básico a diferença de tempo e intensidade derivada de cada um dos pares de microfones de um arranjo com cinco deles, em sua maioria cardióides, que se encontram adjacentes entre si. Usualmente se trata cada um desses pares como se ele cobrisse determinada área ou seção com um ângulo determinado no arranjo de microfones e se espera que os sinais destinados aos outros microfones sejam suficientemente menos intensos ou possuam um tempo de atraso grande o suficiente para não interferir na criação da imagem sonora daquele setor.

O espaçamento e ângulos entre as cápsulas dos microfones são baseados nas diferenças de tempo e amplitude necessárias entre pares individuais para a criação de imagens sonoras em determinadas posições.



Layout genérico de um arranjo de cinco canais de microfones baseado em relações de tempo e amplitude. O espaçamento entre os microfones é de 10cm a 1,5m dependendo do padrão polar, ângulo entre eles, uso de *delays* artificiais e regulação de níveis. Fonte: RUMSEY, 2001.

A conhecida “*Decca Tree*” é um dos arranjos mais populares e é baseado em três microfones espaçados, onde o microfone central tem o papel de estabilização de uma imagem sonora central, evitando o famoso efeito de um buraco acústico central do som estéreo. Para fontes sonoras muito grandes, tais como orquestras e corais, um par de microfones omnidirecionais espaçados costuma ser adicionado. Apesar de não apresentar canais dedicados à captação do sinal *surround*, a “*Decca Tree*” ainda é muito utilizada, principalmente associada a outros arranjos para a captação do sinal *surround*, que a completam.

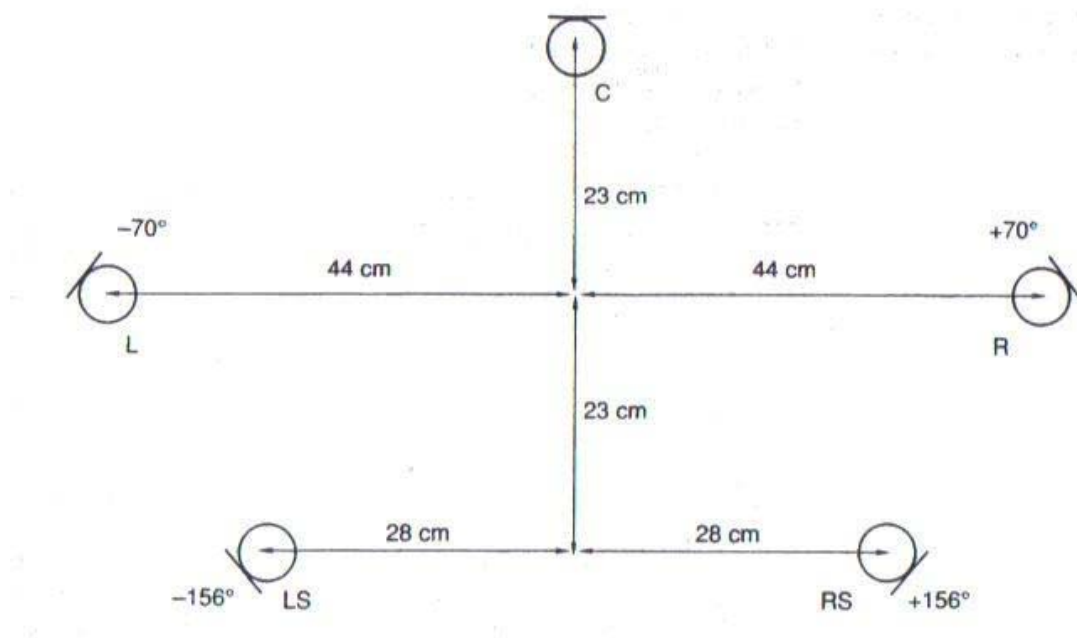


“Decca Tree”, arranjo clássico que envolve três microfones omnidirecionais com o central um pouco à frente dos demais. Fonte: RUMSEY, 2001.

Diversas técnicas baseadas nesses princípios gerais receberam nomes específicos, muitas vezes das pessoas que a desenvolveram ou as tornaram produtos comerciais. O primeiro exemplo de microfonação é a denominada TSRS (*True Space Recording System* – sistema de gravação do espaço real), desenvolvida teoricamente por WILLIAMS e LE DÛ (1999, 2000) e aplicado comercialmente por Mora e Jacques.

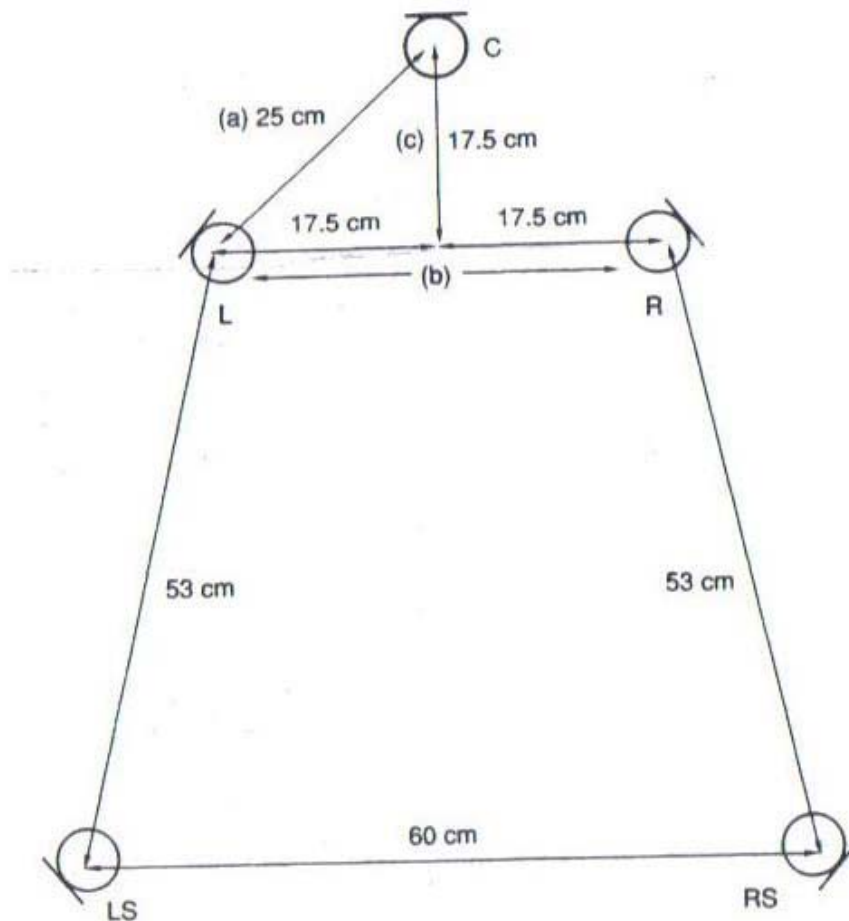
WILLIAMS e LE DÛ descreveram um processo que eles denominaram de “vinculação crítica” entre os ângulos de gravação ou setores cobertos por cada par de microfones em um arranjo. Com o propósito de que cada par de microfones cubra um setor que não se sobreponha a nenhum outro, certas diferenças de tempo e de níveis iniciais devem ser introduzidos entre os canais, através de modificação de posição ou ganho eletrônico, ou uma combinação dos dois. Na configuração proposta abaixo, que segue tais cálculos, os três microfones centrais devem ser atenuados 2,4dB em relação ao par traseiro. Na pesquisa de Mora e Jacques a respeito dessa técnica de posicionamento de microfones, eles concluíram que tal arranjo é bastante compatível com a escuta natural. Segundo THEILE (2000), na utilização dessa técnica, múltiplas imagens sonoras serão criadas, resultado das diferenças entre os pares envolvidos no arranjo frontal e que originam múltiplos efeitos de filtro pente quando combinados para formar uma resultante estéreo. Ele afirma também

que alguns engenheiros de gravação têm obtido mais sucesso ao utilizar microfones omnidirecionais ao invés de cardióides.



Uma dos arranjos desenvolvidos por Williams e Le Dû, baseado em microfones cardioides. Nesse exemplo, os três canais frontais devem ser atenuados 2,4dB em relação ao par traseiro. Fonte: RUMSEY, 2001.

Outro arranjo de microfones, o INA (*Ideale Nieren Anordnung*) ou “Arranjo Ideal de Cardioides”, mencionado por RUMSEY (2001), é descrito por Hermann e Henkels e foi desenvolvido para fins comerciais, onde suas cápsulas podiam se movimentar e girar, e seus padrões polares se alterarem, de forma que diversas distâncias e ângulos podiam ser experimentados. O grande ângulo frontal de 180° implica no fato de que, para que ele seja utilizado como arranjo principal, ele deve ser posicionado muito próximo à fonte sonora, resultando na aproximação das mesmas no centro da imagem reproduzida. Segundo RUMSEY, tal arranjo possui resultados mais interessantes quando utilizado para a obtenção de um som geral captado mais no fundo da sala.

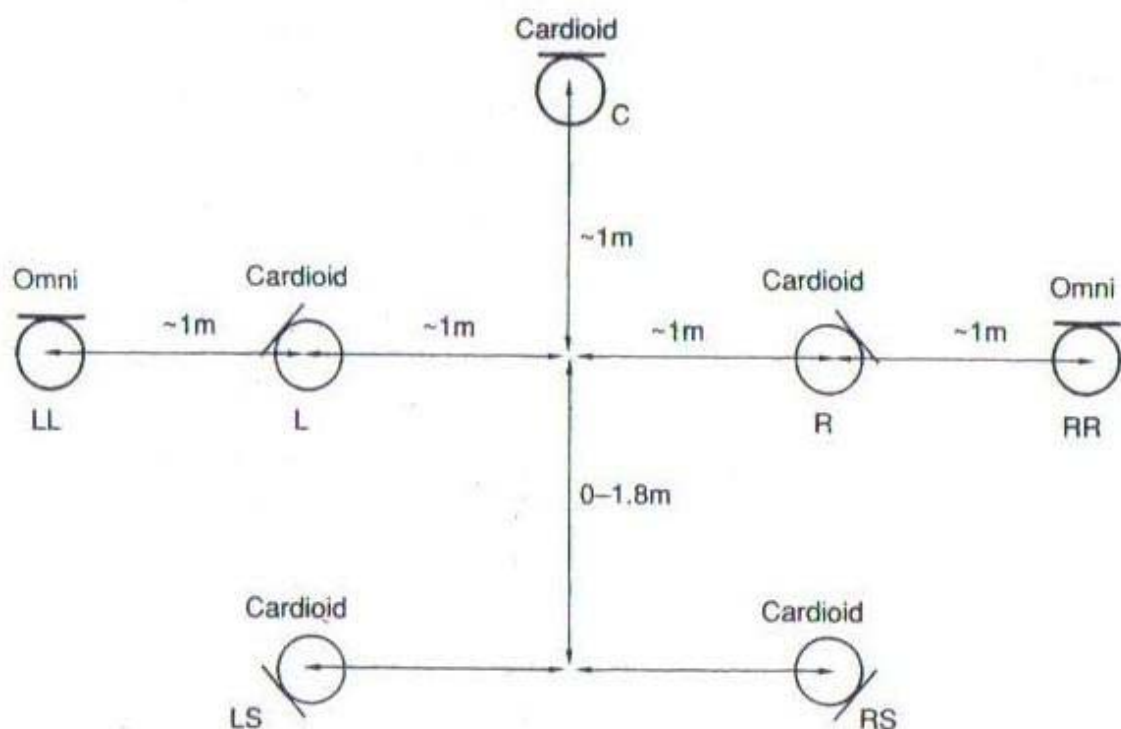


Configuração de microfones cardioides segundo o formato INA-5. Fonte: RUMSEY, 2001.

Muitas abordagens alternativas implementadas a partir de microfones básicos para tratamento *surround* 5.1 utilizam a imagem estéreo frontal obtida separadamente da captura da reverberação natural do ambiente e suas componentes, e outras abordagens são híbridas sem uma base teórica muito clara. Muitas fazem isso adotando variações das técnicas de microfonação estéreo e de três canais, acompanhadas de uma combinação de microfones em diferentes posições para a captura da ambiência. Algumas vezes os canais frontais também contribuem para a captura da ambiência, no entanto o importante é que essas configurações não estejam diretamente intencionadas a captar uma imagem sonora de 360°.

A denominada “*Fukada Tree*” ou “árvore de Fukada” se fundamenta na “*Decca Tree*”, mas ao invés de utilizar microfones omnidirecionais, utiliza cardioides, pretendendo captar menos reverberação e mais som direto frontal. Microfones omnidirecionais são adicionados e endereçados para as laterais com o propósito principal de integrar os canais frontais com os traseiros. Os

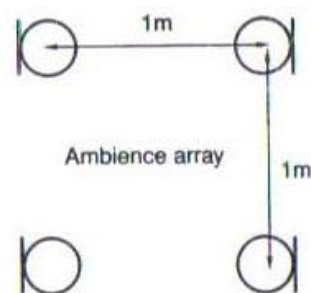
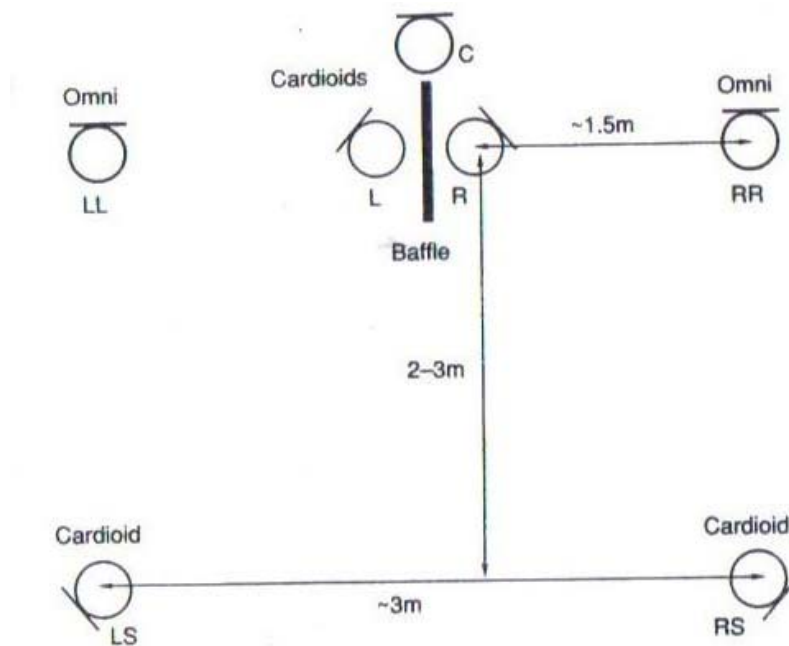
canais traseiros são cardióides e são posicionados além da distância crítica da sala. RUMSEY (2001) menciona que Erdo Groot desenvolveu uma variante desse padrão utilizando microfones omnidirecionais ao invés de cardióides com o propósito de tirar vantagens de sua melhor qualidade e resposta em frequência mais plana. Utilizando o arranjo de omnidirecionais separados 3m entre esquerda e direita assim como frente e trás, ele obteve um resultado espacial onde os canais traseiros e frontais se mostram muito bem integrados, devido ao posicionamento através do controle panorâmico desses últimos entre o plano frontal e o traseiro. O canal central é posicionado um pouco a frente dos canais laterais e é observado que, nesse arranjo, se os canais traseiros forem muito afastados dos frontais têm-se certa separação deles em relação à imagem frontal, como um eco repetindo o som frontal na parte de trás.



A denominada "Fukada Tree" que consiste em cinco microfones espaçados para a gravação em *surround*. Fonte: RUMSEY, 2001.

Um arranjo de microfones proposto por Hamasaki, da NHK, utiliza dois cardióides coincidentes (30cm) separados por um abafador. O microfone cardióide central é posicionado um pouco à frente dos outros microfones e os das laterais extremas espaçados aproximadamente 3m do par principal. Esses

últimos são omnidirecionais e filtrados acima de 250Hz e, ao serem adicionados ao sinal dos canais principais, reforçam a qualidade dos graves. Os canais traseiros são cardioides e espaçados entre 2 a 3m do par principal e 3m entre si. Um outro arranjo de microfones destinados à captação da ambiência consiste em quatro unidades em padrão polar figura oito e direcionados para as laterais espaçados de 1m entre si. Este arranjo pretende capturar reflexões laterais e são posicionados a uma boa altura no espaço de gravação.

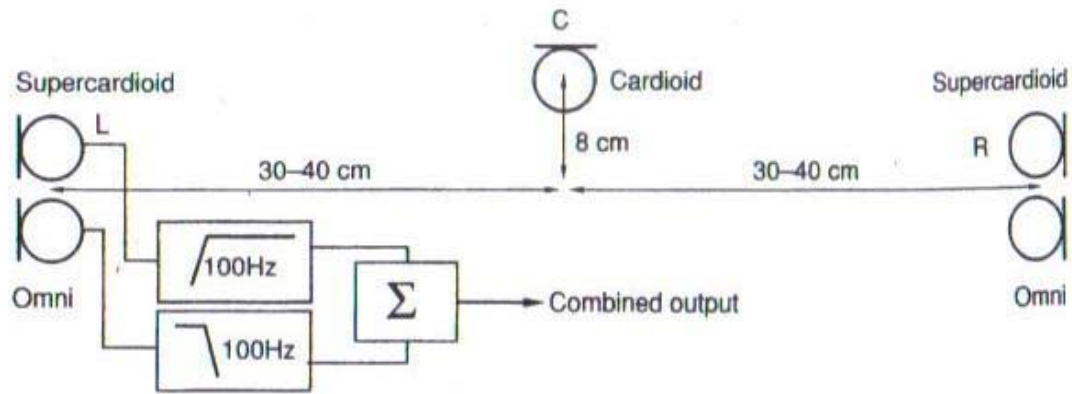


Técnica de microfonação *surround* proposta por HAMASAKI da NHK, que consiste em um arranjo frontal de microfones cardioides, um outro par espaçado de omnidirecionais e um arranjo separado para captação do ambiente. Fonte:

RUMSEY, 2001.

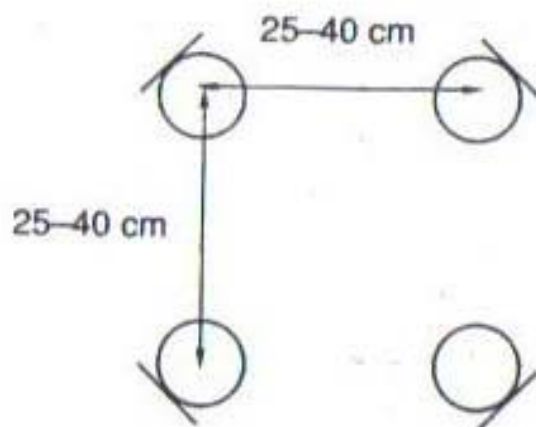
Uma comparação de MASON e RUMSEY (1999) entre diferentes técnicas de captação do ambiente em *surround*, com relação ao resultado sonoro de arranjos baseados em microfones cardióides direcionados para as laterais ou com microfones omnidirecionais afastados do arranjo central, concluíram que a primeira abordagem, que utiliza os microfones cardióides processados por um *delay*, é preferível em relação à outra, quando diz respeito à impressão espacial e definição de imagens sonoras frontais. O atraso aplicado é da ordem de 30ms para evitar que o som direto se sobreponha ao som *surround*.

THEILE (2000) é um outro autor que propõe um arranjo de microfones frontais onde ele reduz os cancelamentos de fase através do uso de microfones supercardioides direcionados para as laterais e um microfone cardióide central. A utilização dos microfones supercardioides é devida ao fato de que eles são mais direcionais do que os cardioides e possuem a maior razão entre som direto e reverberante entre todos os microfones de primeira ordem, além disso possuem o lóbulo traseiro menor do que os hipercardioides. A adoção desse arranjo de microfones tem como objetivo evitar cancelamento de fases entre os segmentos. Ele propõe também um ganho na resposta a frequências graves ao adotar microfones híbridos para os canais esquerdo e direito provenientes da combinação de microfones omnidirecionais filtrados acima de 100Hz com os microfones supercardioides filtrados abaixo dessa mesma frequência. O canal central também é filtrado abaixo dessa frequência (talvez para compatibilidade de coloração tonal ou pela dedicação desse canal, em muitas produções, para diálogos) e, dessa forma, os supercardioides podem ser equalizados para uma resposta plana para sinais frontais num âmbito de 30°. A fabricante Schoeps desenvolveu um protótipo dessa proposta e a denominou OCT "*Optimum Cardioid Triangle*" ou "Triângulo Ótimo de Cardioides".



Arranjo proposto por THEILE e comercializado pela Schoeps com o nome de OCT. Fonte: RUMSEY, 2001.

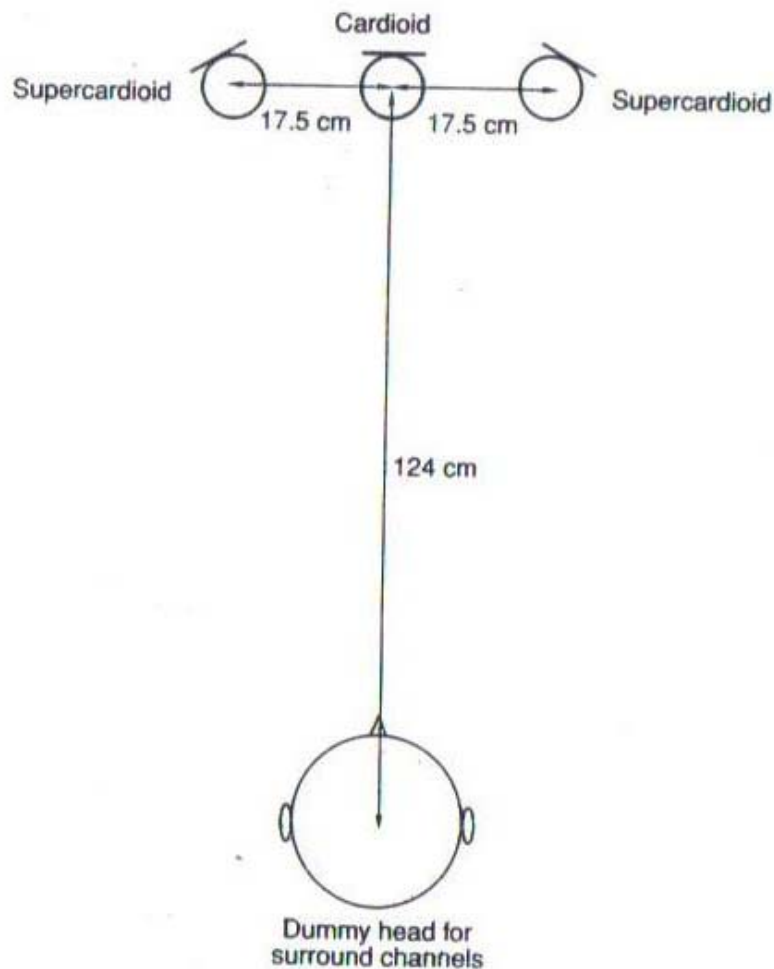
THEILE propõe também um arranjo para a captação da ambiência, com uma configuração cruzada, denominada “*IRT cross*” ou “*atmo cross*”, onde os microfones são cardióides ou omnidirecionais e são afastados entre si de 25cm (cardióide) a 40cm (omnidirecionais), posicionados como nas quatro arestas de um quadrado. O espaçamento pequeno é apropriado para imagens sonoras mais precisas na posição específica do ouvinte, enquanto o espaçamento maior privilegia uma reverberação difusa em uma área maior. Os sinais desses microfones são mixados aos canais L, R, Ls e Rs, mas não ao central.



Configuração “*IRTatmo-cross*” proposta por THEILE para a gravação da ambiência e direcionados para os quatro canais (L, R, Ls e Rs). Os microfones podem ser cardióides ou omnidirecionais. Fonte: RUMSEY, 2001.

KLEPKO (1997) propõe um arranjo frontal composto de um microfone cardióide central e dois supercardioides para os canais L e R formando um ângulo menor do que o proposto por THEILE. Todos os microfones são

alinhados, ao contrário da maioria dos outros formatos que propõe o canal central mais à frente, e os referentes aos canais L e R são afastados 17,5cm do central. A teoria por trás dessa configuração não é muito clara, mas, segundo RUMSEY (2001), o resultado é a de uma grande sobreposição dos ângulos de gravação dos pares L-C e R-C no centro. KLEPKO propõe também a utilização de técnicas pseudo-binaurais (tais como a utilização de *Dummy Heads*) para os canais traseiros.

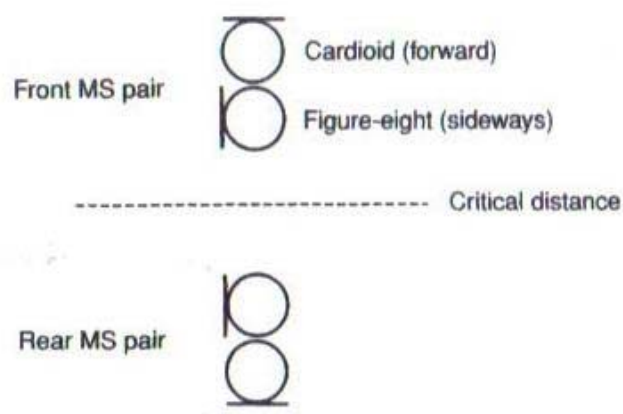


Proposta de KLEPKO para a gravação *surround*. O microfone central é um cardióide, enquanto os laterais supercárdioides. Uma *Dummy Head* é utilizada para a captação da ambiência e direcionada aos canais *surround*.

Fonte: RUMSEY, 2001.

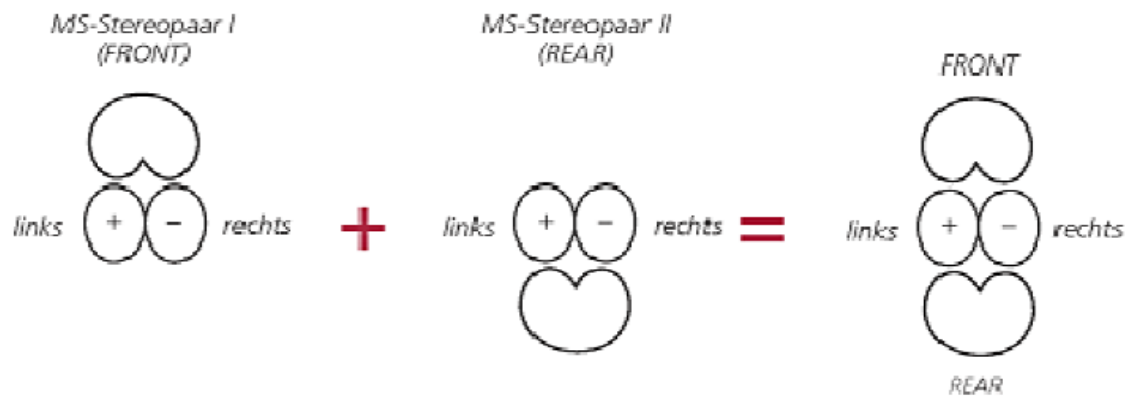
Uma técnica denominada “*Doble MS*”, ou MS duplo, proposto por Curt Wittig e outros, e mencionado por RUMSEY (2001), consiste em dois pares de microfones configurados em padrão estéreo MS, um destinado à captação frontal e outro à traseira. O canal central é alimentado com o sinal do microfone M central. O par traseiro é posicionado além da distância crítica da sala e o

sinal de S de ambos os pares são utilizados para regular a largura da imagem sonora reproduzida tanto na frente quanto atrás. O padrão polar dos microfones M costuma ser cardióide, mas podem variar de acordo com a necessidade de resposta direcional e em frequência desejada. Outros autores propõem a utilização de um microfone cardióide à frente do par frontal para alimentação do canal central, atrasado no tempo para alinhamento com o par principal. Pode-se também utilizar algum atraso para integrar melhor os canais frontais e traseiros.

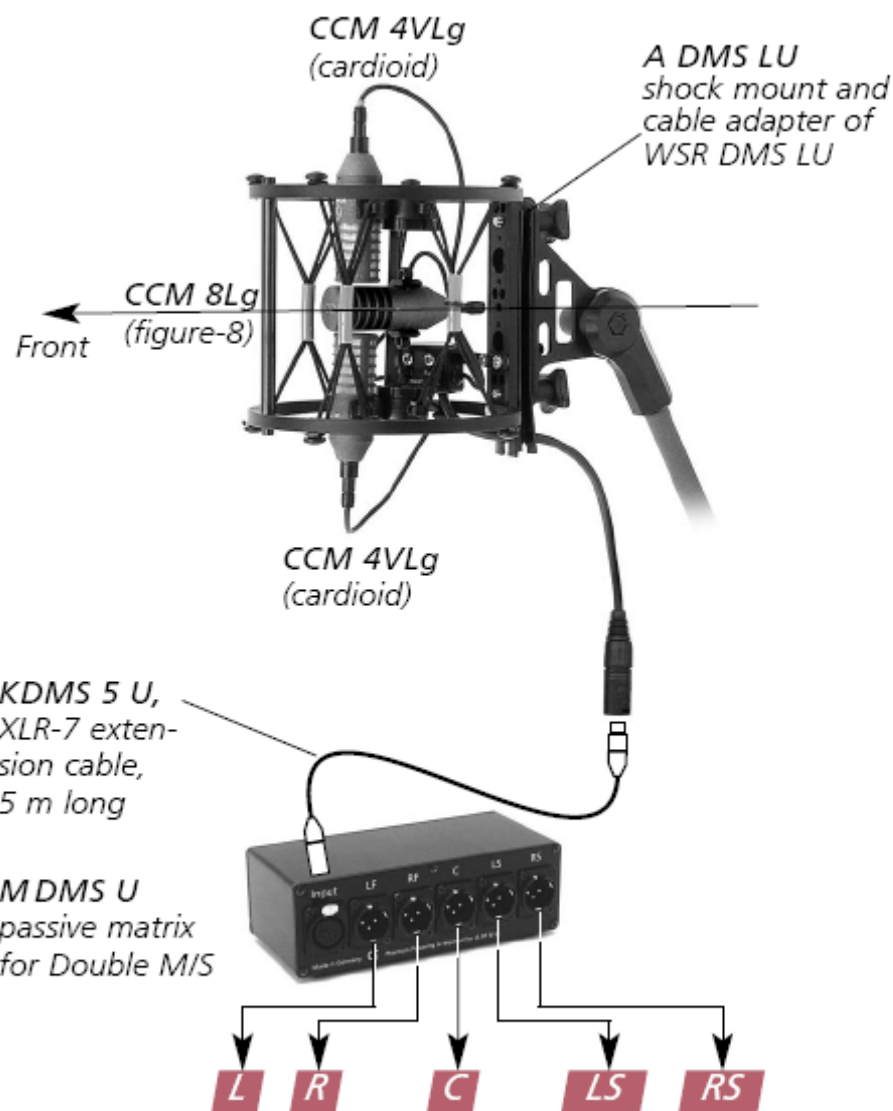


Arranjo de pares "Double MS" com pequeno espaçamento entre os pares frontais e traseiros. Fonte: RUMSEY, 2001.

Alguns autores descrevem uma forma simplificada do arranjo *Double MS*, sendo o microfone figura oito o mesmo para ambos os pares estéreo frontal e traseiro. Com esse arranjo simplificado diversas aplicações são beneficiadas, entre elas a captação sonora de documentários e de rádio novelas. Esse formato foi desenvolvido comercialmente pela Schoeps e descrito por WITTEK, HAUT e KEINATH (2006), onde os autores fazem inclusive uma análise das similaridades desse formato à captação Ambisonics, que pode ser obtido em seu formato B, com exceção da componente z (altura), o que não limita a utilização do formato em questão em sistemas convencionais 2.0, 5.1, entre outros, que não reproduzem tal componente.



Princípio de funcionamento da captação *Double MS*. Somatório de um par estéreo MS dedicado à parte frontal com outro dedicado à parte traseira. Fonte: Schoeps Mikrofone Surround Recording Techniques Brochure.



Arranjo *Surround Doble MS*, comercializado pela fabricante Schoeps. Fonte: Schoeps Mikrofone Surround Recording Techniques Brochure.

Segundo os autores mencionados acima, essa técnica de gravação utilizada tanto no sistema estéreo quanto no multicanal possui vantagens práticas que justificam o crescimento de sua popularidade. Uma dessas vantagens é o baixo nível de cancelamentos que ocorre devido à proximidade das cápsulas dos microfones envolvidos. Eles ressaltam também que a apreciação do arranjo *Double MS* e sua aplicação prática estão diretamente relacionadas à boa decodificação dos sinais provenientes dos microfones envolvidos.

Entre as características mencionadas pelos autores com relação à fidelidade do material gravado com determinado arranjo de microfones e sua reprodução estão a localização, a coloração sonora e a percepção espacial reproduzida. De acordo com suas pesquisas subjetivas, os autores afirmam que as alterações relativas a direcionalidade e tamanho das imagens fantasma criadas entre os alto-falantes a partir de uma gravação com o formato em questão são mais perceptíveis do que as modificações de seu posicionamento e coloração sonora.

Para gravações de concertos musicais, os autores afirmam que o formato *Double MS* funciona bem em ambientes menores, sendo outros formatos tais como OCT mais indicados para grandes salas de concerto. A combinação do formato *Double MS* com um par espaçado *AB* também deu bons resultados aos autores supracitados. Testes em gravações de pequenos grupos também trouxeram bons resultados com a utilização do formato *Double MS*, assim como a sua combinação com a microfonação próxima. Aplicações em gravações ao vivo de teatros, transmissão ao vivo para televisão e transmissão de radio novelas são muito beneficiadas por esse formato devido à falta de ênfase do mesmo na captação da sala (que muitas vezes não é ideal) e à sua boa compatibilidade com o sistema monofônico.

Em aplicações cinematográficas, a substituição do microfone cardióide central por um mais direcional, supercardiíde ou *shotgun*, favorece a captação de diálogos centrais em produções onde um sinal de som direto multicanal é bem vindo.

Alguns engenheiros experimentaram combinar técnicas binaurais para reprodução *surround*, e como exemplo temos JERRY BRUCK (1997), que adaptou dois microfones com padrão polar figura oito ao microfone esférico da

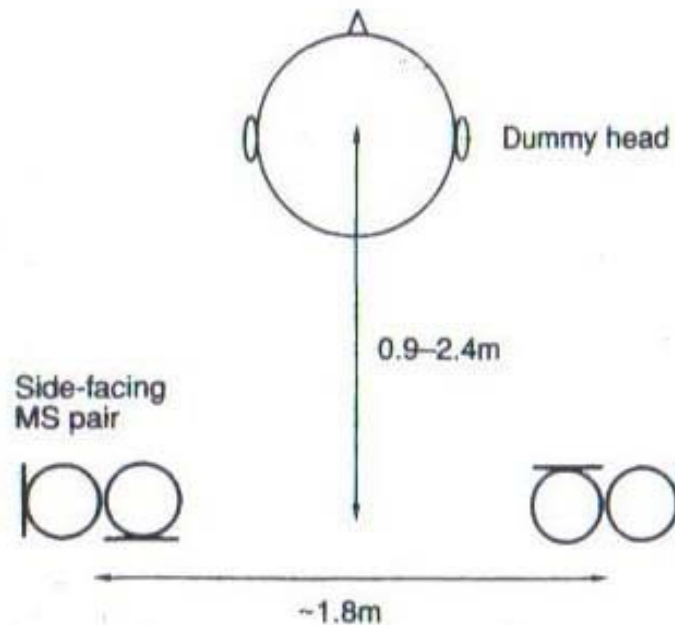
Schoeps, posicionados próximos aos dois microfones omnidirecionais laterais com o eixo direcionado para frente. Atualmente este arranjo é produzido pela própria Schoeps que oferece também um processador cujas saídas são os tradicionais L, R, C, Ls e Rs.



Microfones e processador comercializado pela Schoeps baseado na proposta de microfonação de JERRY BRUCK.

Fonte: Schoeps Surround Brochure, 2006.

Michael Bishop, da Telarc, também mencionado por RUMSEY (2001), foi outro engenheiro que adaptou uma “*Dummy Head*”, destinada à captação frontal, à dois pares estéreo MS, ambos direcionados às laterais, que alimentam os pares laterais do sistema *surround* 5.1 (L-Ls e R-Rs).



Pares "Double MS" direcionados para as laterais e utilizados para alimentação dos pares de alto-falantes laterais, combinados com uma "Dummy Head" cujo sinal é utilizado para a formação da imagem frontal. Fonte: RUMSEY, 2001.

2.2.2 – O controle panorâmico no estéreo e no *surround*

Segundo MALHAM (1998), existem duas formas básicas de se reproduzir uma imagem sonora através de um sistema estéreo: através da diferença de amplitude do sinal entre as caixas ou através da diferença de tempo. A primeira delas é a mais comum e é muito conhecida pela denominação de função *pan* enquanto a outra é mais utilizada para gerar formatos estéreo sinteticamente, principalmente direcionados a reproduções que utilizam técnicas binaurais. Essas duas formas de se criar imagens sonoras também são aplicáveis a sistemas multicanais, tais como o *surround* 5.1.

Apesar disso, RUMSEY (2001) afirma que o panorâmico de sinais entre mais do que dois alto-falantes apresenta inúmeros problemas psico-acústicos, principalmente com relação à distribuição correta da energia sonora entre os canais, a precisão da localização de imagens "fantasma", a escuta fora do centro do sistema e o timbre do som. Diversas soluções têm sido propostas entre elas podemos citar as observações de GERZON, M. (1992) a respeito:

“O objetivo de uma boa lei de panorâmico é pegar cada sinal monofônico e dar a ele níveis de amplitude correspondentes a cada alto-falante, dependendo da ilusão de localização que se pretende, de forma que o som resultante reproduzido proporcione uma imagem “fantasma” convincente. Tal lei de panorâmico deve oferecer uma faixa contínua de direções da imagem “fantasma” para qualquer posição entre os dois alto-falantes mais próximos envolvidos, sem o amontoamento dessas imagens em nenhuma posição específica ou criação de buracos acústicos nos quais a imagem seja muito fraca.” (RUMSEY, 2001, p.209).

O panorama baseado em amplitude para pares de alto-falantes é o tipo de controle com o qual os engenheiros estão mais acostumados por ser a abordagem mais utilizada na maioria dos sistemas estéreo. Ele envolve o ajuste de amplitude de um sinal entre alto-falantes adjacentes para a criação de imagens “fantasma” em qualquer ponto entre eles. Essa técnica é estendida para o panorâmico de três canais assim como para canais laterais, L e Ls, por exemplo, e para os canais traseiros, Ls e Rs. Como discutido anteriormente essa técnica, bem como a baseada em diferenças de tempo, não é muito bem sucedida quando utilizada para alto-falantes muito distantes, resultando em imagens que não se movem linearmente e que tendem a pular de um plano para outro.

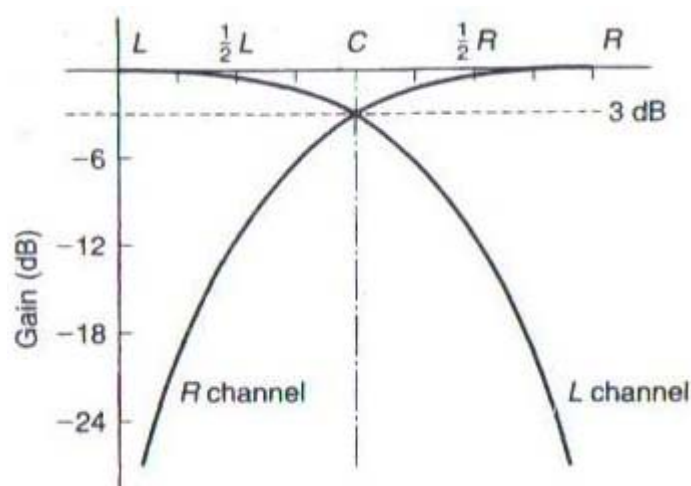
Em decorrência disso alguns sistemas de panorâmico direcionados para a mixagem em *surround* possuem dois sistemas separados, um dedicado ao panorâmico frontal (L-C-R) e outro para o *surround* (Ls-Rs). Ainda assim esses sistemas conseguem bons resultados apenas em algumas regiões sendo importante a experiência e o cuidado para o manuseio dos mesmos, principalmente dos que se utilizam de *joysticks* para realizá-lo.

Essa questão do panorâmico é uma das razões principais pela qual se opta por direcionar um material sonoro mais difuso, muitas vezes correspondente com à ambiência da sala, para os canais traseiros, uma vez que a distância entre eles não favorece uma boa localização das fontes sonoras.

Em gravações com muitos microfones, cada fonte sonora (microfone) alimenta um dos canais de um mixer, onde os níveis são controlados individualmente e o sinal é “paneado” para uma posição virtual entre os canais

esquerdo e direito da saída estéreo principal. O controle de panorâmico (*pan*) transforma um sinal monofônico em dois, controlando a proporção deles direcionada para cada um dos dois alto-falantes envolvidos. Normalmente, leis de controle panorâmico seguem uma curva que subtrai 3dB do sinal quando este se posiciona ao centro, resultando num mesmo nível sonoro quando do movimento da fonte sonora entre os canais. Isso se dá uma vez que os canais dos dois alto-falantes se somam na posição de escuta, fazendo necessária essa atenuação.

Essa lei de panorâmico não se mostra correta se os dois sinais resultantes forem combinados eletronicamente em um sinal monofônico, uma vez que a soma dos dois sinais resultaria em um ganho de 6dB para sinais “paneados” exatamente no centro, fazendo com que uma curva que siga uma lei de -6dB seja mais apropriada para mixers com saídas monofônicas (rádio e tv). No entanto, tal lei resultaria numa atenuação no sinal central quando mixado em estéreo, o que leva algumas empresas a assumirem um compromisso de seguir uma curva com -4,5dB para sinais posicionados no centro.



Lei de panorâmico típica de mesas de mixagens com saídas estéreo. Fonte: RUMSEY, 2001.

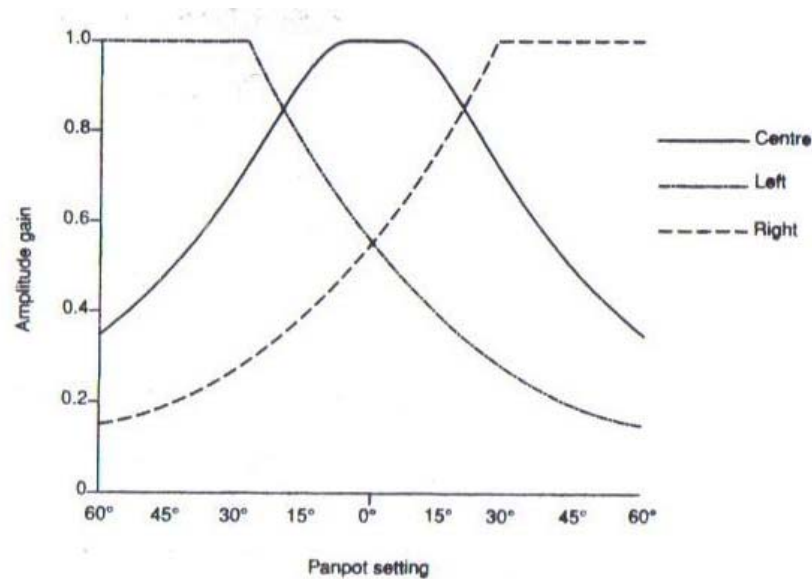
Tais sistemas de panorâmico se fundamentam somente na diferença de amplitude para o posicionamento das fontes sonoras, e o atraso no tempo raramente é utilizado para esses sistemas, o que faz com que a compatibilidade do sinal estéreo com o monofônico seja muito precária, uma vez que a combinação elétrica dos dois sinais resulta em uma atenuação de

6dB com relação ao sinal original estéreo. Reverberação artificial normalmente é utilizada para restaurar o equilíbrio de gravações com múltiplos microfones próximos. A distância de uma fonte também pode ser simulada através da adição de reflexões e reverberação, assim como através de mudanças no espectro sonoro e no nível geral.

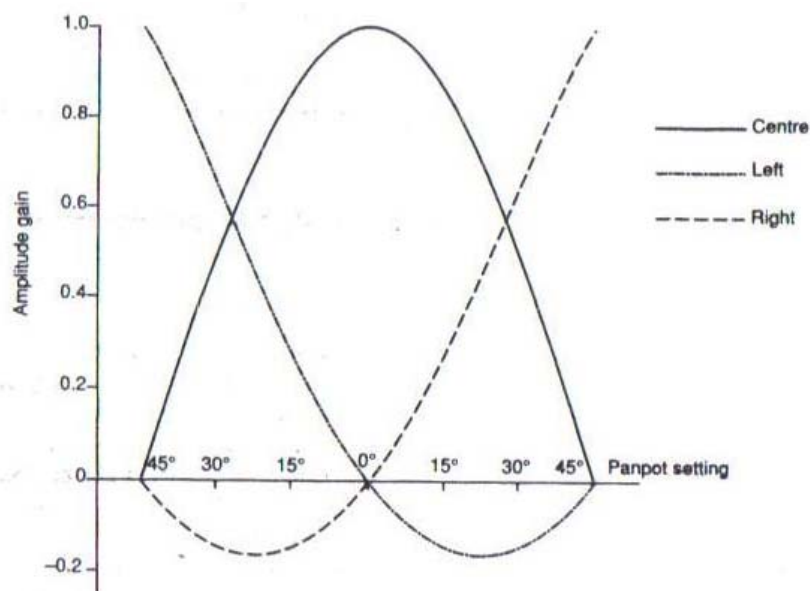
É comum em gravações de música clássica a utilização de microfonação próxima somada a pares de microfones coincidentes ou espaçados, com o propósito de reforçar fontes aparentemente fracas nos microfones principais. Tais microfones são “paneados” o mais próximo possível de suas posições originais percebidas com relação ao par estéreo tido como referência.

A utilização de *delays* para ajustar o tempo de chegada do sinal dos microfones da captação próxima ao do par principal previne a distorção proveniente da distância e acerta o tempo de chegada dos microfones distantes, compensando o efeito de precedência com relação ao par principal. Também é possível o processamento da saída de múltiplos canais monofônicos para a simulação de atrasos binaurais e efeitos relativos à cabeça com o propósito de criar a ilusão do posicionamento dos sons em qualquer posição em torno da cabeça do ouvinte quando o resultado é monitorado através de fones de ouvido.

A expansão das técnicas de gravação em dois canais para três canais frontais é um passo importante nas considerações a respeito da mixagem *surround*. A presença do canal central modifica as leis de panorâmico e as relações dos sinais entre os canais para a precisão das imagens “fantasma” criadas. Uma das propostas para essa nova lei de panorâmico contém componentes anti-fase para que canais posicionados à direita do centro gerem uma componente negativa no canal esquerdo e vice-versa. A lei de panorâmico elaborada pelos Laboratórios Bell, de 1930, utiliza somente componentes positivas, o que gera boa localização de altas frequências, mas não funciona igualmente para frequências graves e gera certa instabilidade, especialmente para ouvintes fora do eixo dos alto-falantes. RUMSEY (2001) menciona o autor GERZON, que acredita que a lei de panorâmico para três canais deve ser dependente da frequência e da aplicação e propõe um panorâmico para três alto-falantes em 90°.

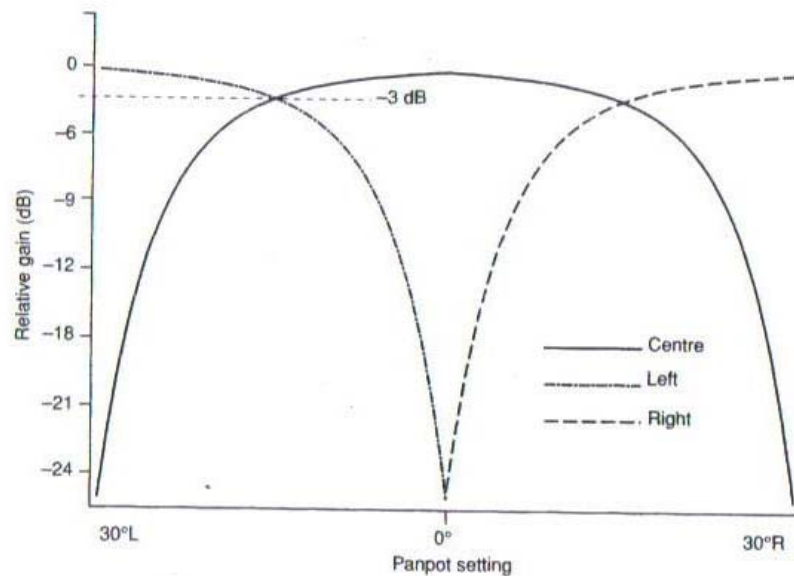


Lei de panorâmico elaborada pelos Laboratórios Bell. Fonte: RUMSEY, 2001.



Lei de panorâmico para três canais frontais proposta por GERZON que prevê certa filtragem de graves e os alto-falantes em um ângulo de 45° com relação ao central. Fonte RUMSEY, 2001.

Panorâmicos dedicados a três alto-falantes tendem a trabalhar com relações positivas de ganho baseados sempre em apenas dois alto-falantes, tratando cada par de uma vez com o panorâmico estéreo simples:



Panorâmico típico para três canais utilizando a relação de nível constante para cada par de canais. Fonte: RUMSEY, 2001.

Muitos *softwares* de gravação multipista possibilitam a mixagem em *surround* 5.1 e outros sistemas ao disponibilizarem interfaces de panorâmico que distribuem cada sinal mono ou pista de áudio em um plano horizontal formado pelo arranjo de alto-falantes definido como saída. Infelizmente são poucos os *softwares* que oferecem informações mais precisas e técnicas a respeito do processamento de áudio realizado por esses *plugins*. A maioria deles parece funcionar a partir do controle de intensidade certamente adicionado a algum outro tipo de processamento, seja ele baseado em *delays* ou simulação acústica de salas.

Entre os *softwares* de gravação multipista está o Digital Performer 7 da fabricante MOTU (*Mark of the Unicorn*), utilizado na mixagem das peças eletroacústicas detalhadas mais a frente neste trabalho. No software em questão são disponibilizados quatro *plugins* com diferentes algoritmos para serem utilizados em mixagens *surround* 5.1 assim como em LCRS (Dolby Surround), Quad, 6.1, 7.1 e 10.2, dependendo da configuração da quantidade de saídas, definida em *bundles*.

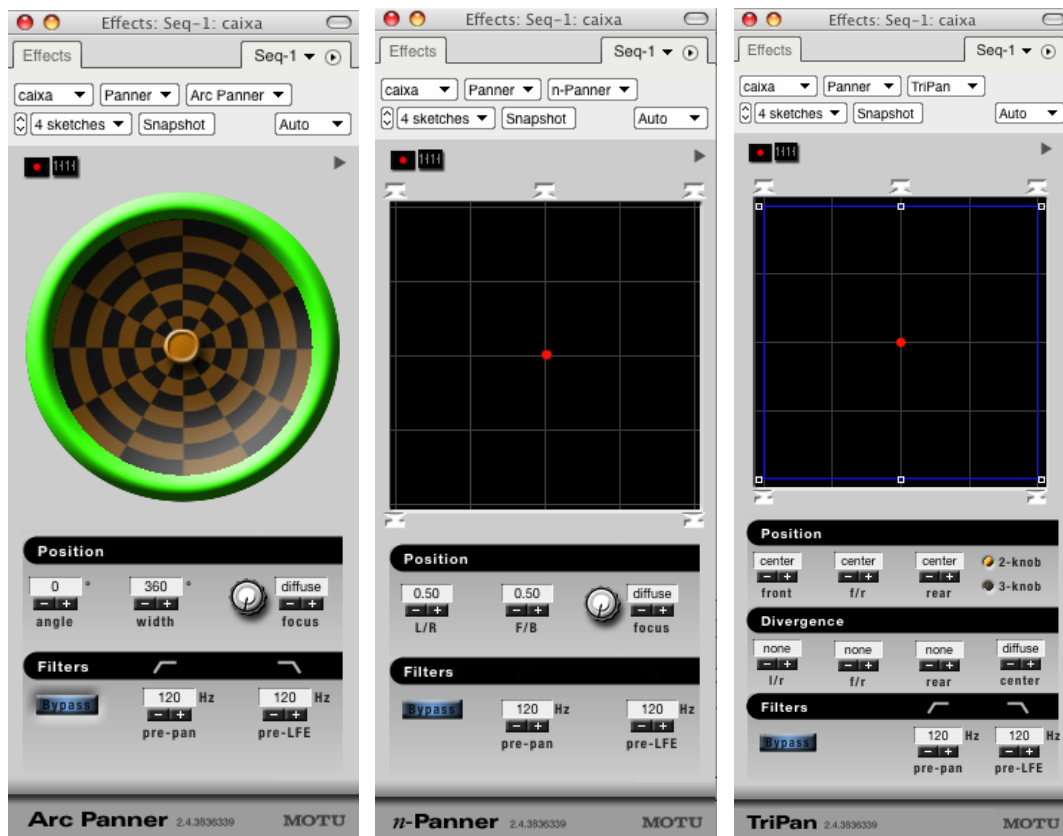
A interface presente no mixer virtual simplesmente dá acesso rápido aos parâmetros posicionais do *plugin* selecionado e, assim como no panorâmico do estéreo tradicional, que possibilita a automação, possibilitando a criação de movimento da fonte sonora, o panorâmico *surround* também pode ser

automatizado através do controle de duas linhas, cada uma delas referente a uma das duas coordenadas cartesianas X e Y. Os *plugins* são: ArcPanner, n-Panner, TriPan e Auralizer, discutidos a seguir.

Arc Panner: nesse *plugin* o campo panorâmico é representado em uma grade radial onde, assim que o ponto correspondente à posição da fonte sonora se movimenta, uma representação da energia daquela fonte sonora se modifica como uma barra verde localizada no arco que circula o campo sonoro. O parâmetro ângulo é definido em graus em relação à posição de escuta e o foco determina a amplitude espacial da fonte sonora no campo representado sendo que, quanto menor o foco mais direcional será a imagem sonora representada pelo sistema de panorâmico e quanto maior, mais difusa e abrangente.

n-Panner: *plugin* de panorâmico mais simples e que pode ser entendido como um par de controles de panorâmico estéreo, sendo que um controla o sinal do par de alto-falantes frontais e o outro o traseiro através de coordenadas cartesianas X/Y. O controle *focus* define a direcionalidade da fonte sonora em questão.

TriPan: versão mais elaborada do n-Panner, pode ser utilizado da mesma forma que o descrito acima ou utilizar o método de panorâmico com três pontos onde além do controle entre L/R no plano frontal e o Ls/Rs no plano traseiro, pode-se controlar o plano frontal/traseiro. Essa técnica permite a transição diagonal de forma suave entre qualquer dois pontos. O controle de divergência pode ser utilizado com valores independentes para cada um dos três pontos descritos acima.



Plugins Arc Panner, n-Panner e TriPan da Motu. Fonte: Digital Performer 7.

Auralizer: simulador de sala que utiliza *matrix surround* para posicionamento preciso de cada fonte sonora em pontos específicos do campo sonoro. A grande diferença entre ele e os descritos acima é que nele a distância do ponto que representa a fonte sonora do centro define a sua distância percebida, não apenas a divergência ou qual alto-falante está ativo. A combinação entre os parâmetros ângulo e distância determina a relação entre a posição da fonte sonora e do ponto de escuta. O parâmetro largura (*width*) cria uma imagem estéreo para fontes monofônicas que corresponda ao seu tamanho. Parâmetros referentes à sala também são configuráveis tais como o seu volume, seu tamanho e seu nível de absorção sonora. Outros parâmetros também podem ser configurados para intensificar a espacialização do áudio no sistema *surround*, tais como o efeito *Doppler* (modificação no domínio da frequência devido ao movimento da fonte sonora em relação ao ouvinte), o *Head delay* (atraso provocado pela distância entre as duas orelhas), *Head shadow* (alteração na resposta em frequência devido ao efeito de sombra provocado pela cabeça), *Distance attenuation* (queda no nível do sinal devido à

distância), controle das primeiras reflexões, *reverb* difuso e nível do sinal enviado para o canal LFE.



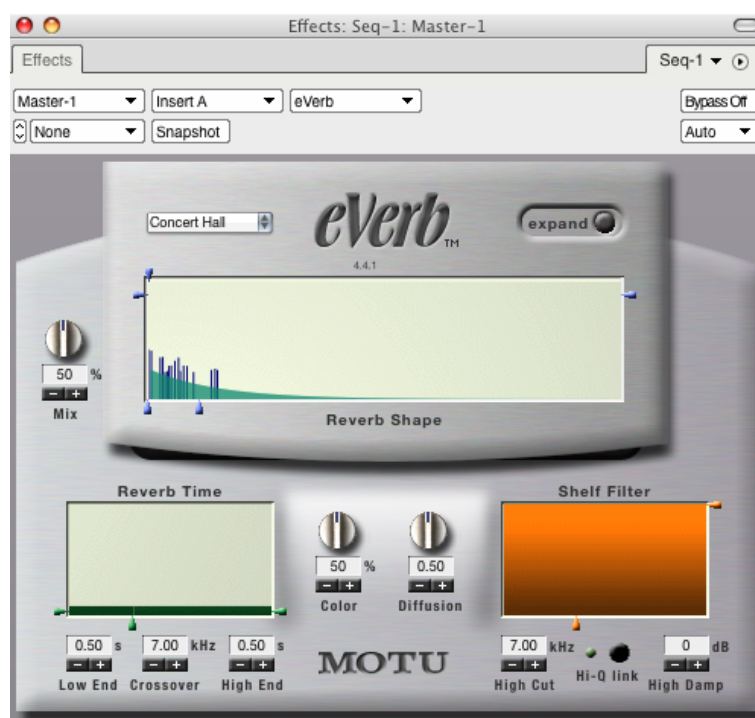
Plugin Auralizer da Motu. Fonte: Digital Performer 7.

No Digital Performer 7 são disponibilizados também outros *plugins* muito importantes no processo de mixagem no sistema multicanal e um deles é o *Bass Manager*, necessário para o controle do sinal grave direcionado ao canal LFE e na simulação de sistemas caseiros de *home-theaters*. Outro *plugin* que pode ser muito útil, também oferecido no software em questão, é o *Calibration*, capaz de auxiliar na regulação de nível entre os alto-falantes do sistema através do auxílio de um microfone omnidirecional posicionado no centro do mesmo.



Plugins Bass Manager e Calibration da Motu. Fonte: Digital Performer 7.

Reverberadores necessários em qualquer mixagem também são disponibilizados em versões com saídas *surround* 5.1, simplificando o processo de conversão de uma mixagem estéreo tradicional ao se adicionar reverberação nos alto-falantes traseiros. No Digital Performer 7 a opção oferecida para esse processamento é o *eVerb*.



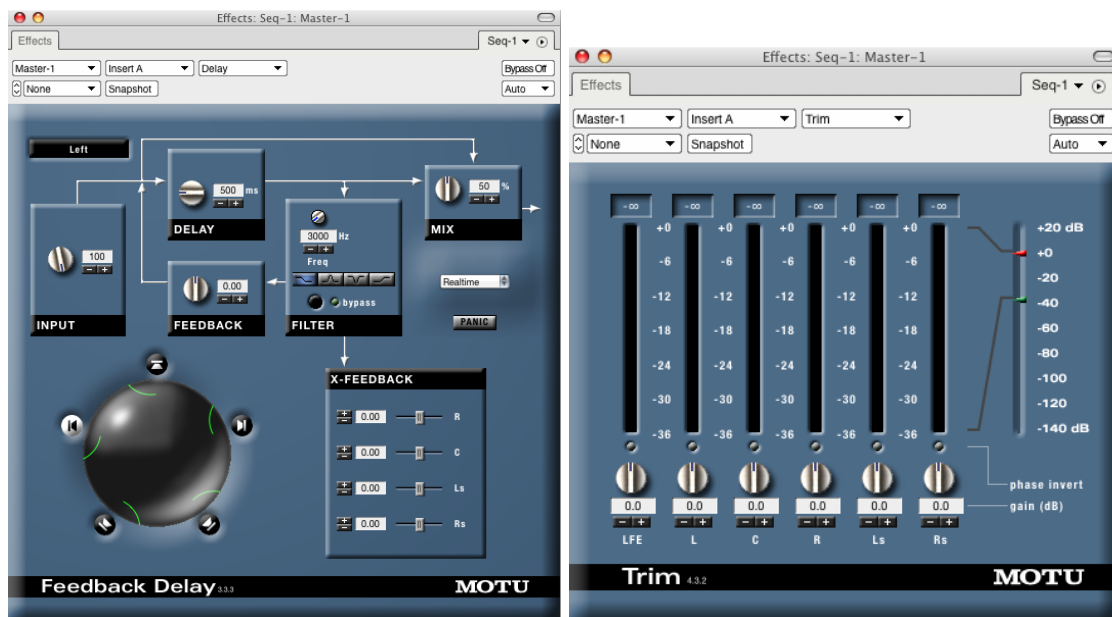
Plugin eVerb da Motu. Fonte: Digital Performer 7.

O *Autopan*, assim como o *Feedback Delay* (este com auxílio também de filtros), possuem saídas para *surround* 5.1, e possibilitam a manipulação de

delays em todo o campo sonoro representado pelo sistema. O *plugin Trim* auxilia na modificação do ganho geral de cada saída de áudio individualmente e conta também com o ajuste de inversão de fase de cada canal, que pode ser utilizado na finalização das mixagens.



Plugin AutoPan da Motu. Fonte: Digital Performer 7.



Plugins Feedback Delay e Trim da Motu. Fonte: Digital Performer 7.

Outros softwares de gravação multipista também oferecem opções de panorâmico em *surround* 5.1. O Sonar 7 Producer Edition, por exemplo, apesar de apresentar apenas uma opção, oferece o controle de diversos parâmetros fundamentais para o posicionamento de uma fonte sonora monofônica no espaço representado pelo sistema *surround* 5.1.

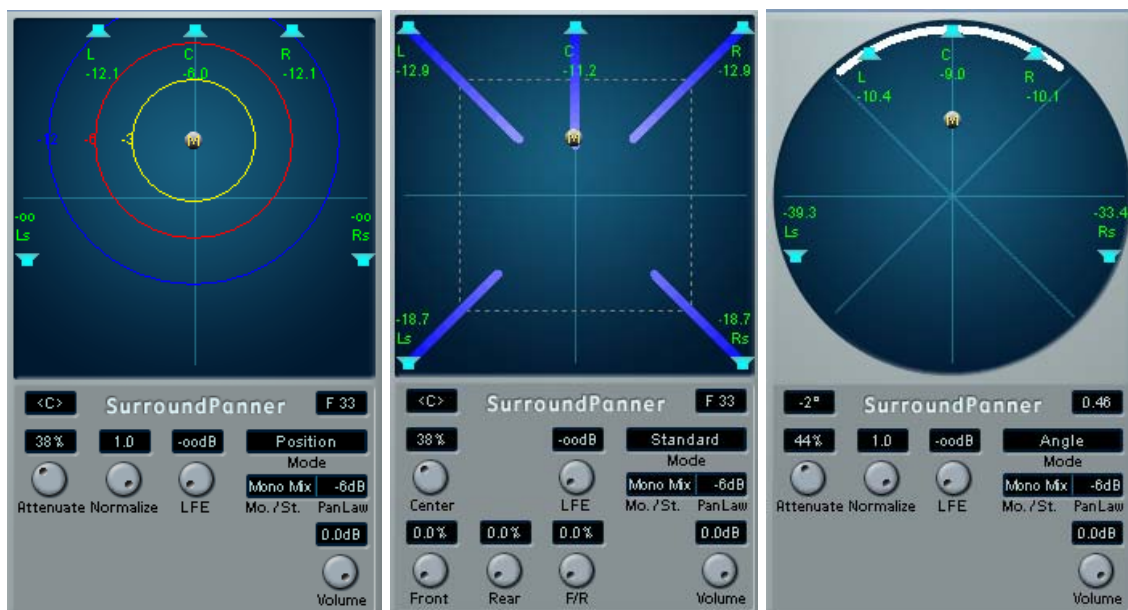


Plugin Surround Panner do Sonar 7 Producer Edition. Fonte: Sonar 7 Producer Edition Help.

Segundo o próprio fabricante, o parâmetro *Angle* (ângulo) define a posição da fonte sonora em relação ao ouvinte, sendo que em 0° a mesma se apresenta exatamente em frente e, em 180° ou -180° , ela se posiciona atrás do mesmo. O parâmetro *Focus* (foco) define a distância da fonte em relação ao ouvinte em uma escala de 0 à 100 onde 0 é o centro do círculo e 100 o perímetro. *Width* (largura) é o parâmetro controlado pelas duas esferas equidistantes da fonte que se refere à abrangência ou dispersão sonora da fonte, podendo variar entre 0° e 360° .

O controle *FrntRrBl* (abreviação de *Front and rear balance* – equilíbrio entre frente e trás) controla o nível do sinal enviado para os planos frontal e traseiro. O *LFE Send* controla o nível do sinal enviado para o canal de Efeitos graves, enquanto o botão *LFE Solo* define o canal em questão exclusivo para efeitos sonoros de frequências graves.

No software Cubase SX 3, as possibilidades de panorâmico são três: *Position*, *Standard* e *Angle*.



Plugins Surround Panner Position, Standard e Angle do Cubase SX 3. Fonte: Cubase SX 3 Help.

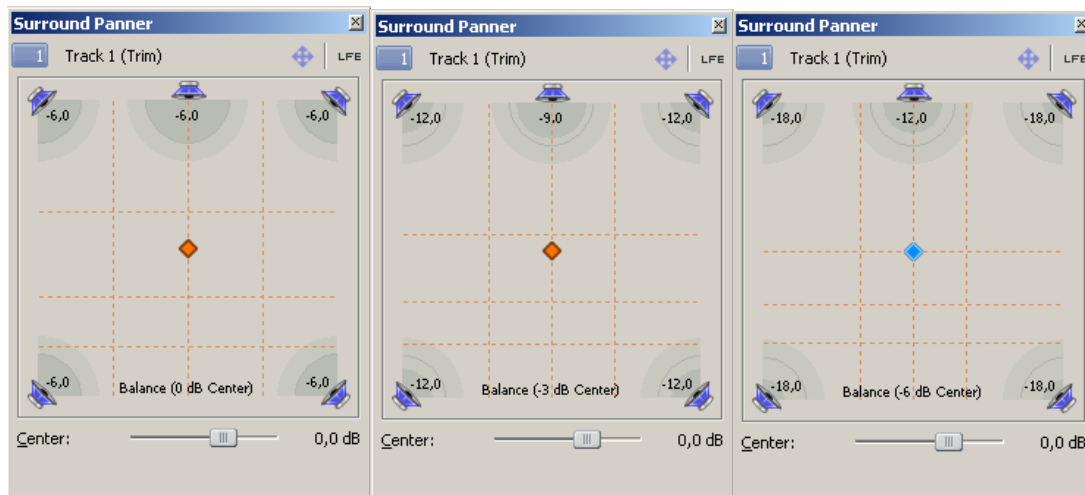
No formato *Position*, assim como no *Standard*, o *plugin* oferecido no Cubase SX 3 nos permite trabalhar com os alto-falantes alinhados no plano frontal, assim como em aplicações cinematográficas, o que significa que os alto-falantes estão posicionados a distâncias diferentes com relação ao centro (posição do ouvinte). O formato *Angle* é o melhor para mixagem de música em *surround 5.1*, uma vez que os alto-falantes estão posicionados equidistantes ao centro e apesar desta não ser a situação real do cinema, é a que se aproxima mais do conceito do formato. Em todos os formatos o controle do sinal a ser enviado para o canal LFE é feito através de um controle dedicado.

O software Sony Vegas 8 apresenta a maior diversidade com relação aos formatos de panorâmico – em número de seis, sendo cinco dedicados ao posicionamento dos alto-falantes de forma linear à frente do ouvinte e um deles com o arranjo como na norma ITU, com os alto-falantes equidistantes da posição central.

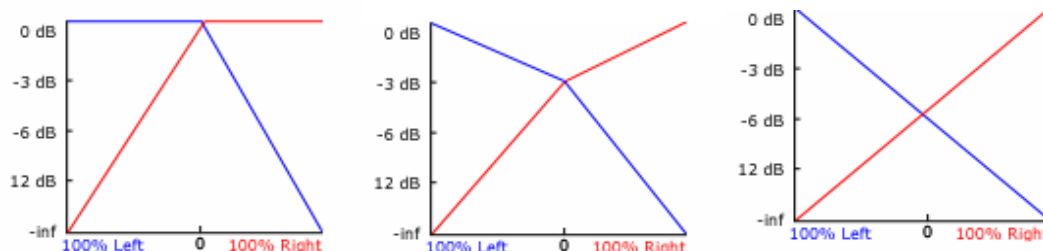
O *plugin Add Channels* funciona a partir de uma curva⁵⁵ que determina a amplitude dos sinais enviados para os alto-falantes frontais de forma que quando uma fonte sonora é posicionada exatamente na frente ela possui sinal saindo igualmente das três caixas frontais, e quando posicionada somente na esquerda ou na direita, a caixa do lado oposto não recebe sinal algum, no

⁵⁵ Neste caso uma reta. Ver gráfico abaixo.

entanto a central recebe sinal com a mesma intensidade da posição em questão. Os *plugins Balance* possuem funcionamento semelhante, no entanto obedecendo a curvas diferentes e não lineares com atenuações referentes à 0dB, -3dB e -6dB nas caixas laterais quando o sinal se encontra no centro frontal exato. Esses quatro primeiros formatos de *plugin* são mais indicados para a espacialização de fontes sonoras estéreo no espaço do *surround* 5.1.



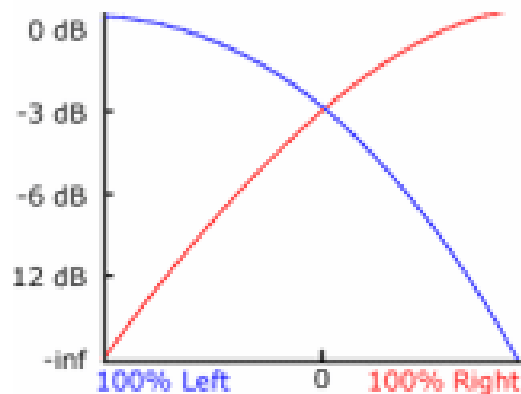
Plugins Surround Panner Balance (0dB Center, -3dB Center e -6dB Center) do Sony Vegas Pro 8.0. Fonte: Sony Vegas Pro 8.0 Help.



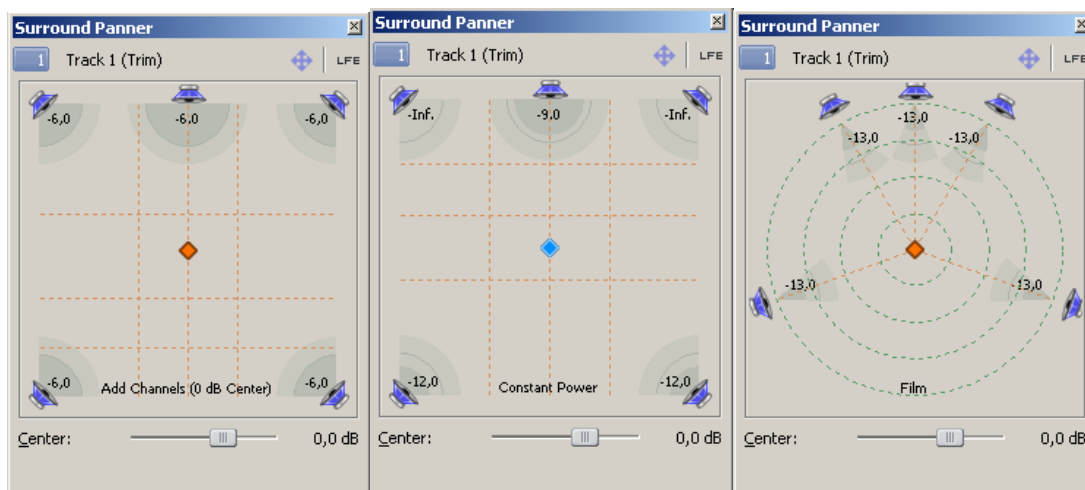
Curvas referentes ao endereçamento frontal do sinal nos plugins Surround Panner Balance (0dB Center, -3dB Center e -6dB Center) do Sony Vegas Pro 8.0. Fonte: Sony Vegas Pro 8.0 Help.

No *plugin Constante power*, mais indicado para espacialização de fontes sonoras monofônicas, a mesma mantém seu volume ao ser posicionada nas diferentes posições do ambiente *surround*, seguindo a curva logarítmica mencionada no início desta seção e estabelecida como padrão para sistemas estéreo. Já o funcionamento do *plugin* no formato *Film* é bem semelhante ao *Constante Power*, no entanto, direcionado para arranjos circulares que seguem

a norma ITU do posicionamento dos alto-falantes de forma eqüidistantes do ouvinte e seguindo a curva já mencionada aplicada a cada par de alto-falantes.



Curva referente ao endereçamento frontal do sinal no *plugin Surround Panner Constante Power* do Sony Vegas Pro 8.0. Fonte: Sony Vegas Pro 8.0 Help.



Plugins Surround Panner *Add Channels (0dB Center)*, *Constante Power* e *Film* do Sony Vegas Pro 8.0. Fonte: Sony Vegas Pro 8.0 Help.

2.2.3 – A questão do monitoramento

RUMSEY (2001) afirma ser de comum acordo que as salas de monitoração não devam interferir no material sonoro da gravação, embora a sala de monitoração não deva ser completamente isenta de reverberação, ou anecóica. Apesar de alguns pensarem que o engenheiro de som deve ouvir apenas o som proveniente dos alto-falantes e nada mais, na realidade isso não é realmente desejável nem muito prático.

Apesar das câmaras anecóicas serem muito utilizadas para testes e como laboratórios experimentais, elas não representam de forma alguma a situação em que uma pessoa ouvirá uma gravação. Uma outra observação diz respeito ao fato de que, devido à grande e rápida queda de pressão sonora em campos abertos, os alto-falantes utilizados em uma sala anecóica teriam de ser extremamente potentes e com muito pouca distorção para funcionarem corretamente em uma sala desse tipo. Outro fator que torna esse tipo de sala inconveniente para a monitoração é que trabalhar nessas salas é extremamente desconfortável durante qualquer período de tempo, sendo cansativo e artificial. Conseqüentemente, uma sala para a monitoração do áudio deve possuir certa reverberação, no entanto esta deve ser extremamente controlada.

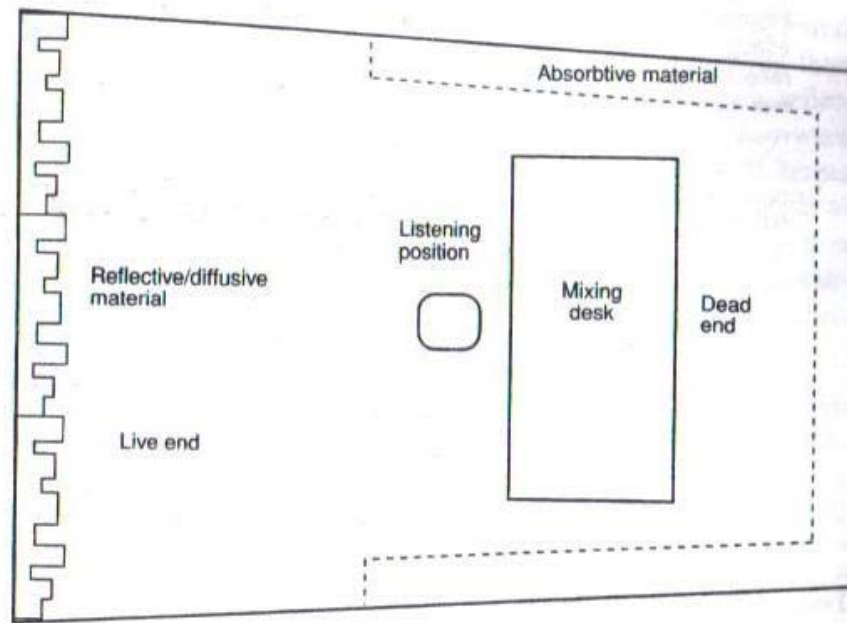
Muitos concordam com o fato de que as primeiras reflexões que chegam ao ouvinte entre 15 e 20 ms após o som direto devem ser minimizadas para evitar modificações no timbre e na imagem espacial gerada pelos alto-falantes, que podem surgir a partir da interação dessas reflexões com o som direto dos mesmos. O efeito dessa interação depende da natureza do sinal e raramente a percepção das primeiras reflexões acontece como pequenos ecos, mas sim como uma modificação no timbre.

Segundo RUMSEY (2001), os resultados de numerosos estudos declaram que as reflexões provenientes das laterais da sala têm maior efeito na percepção de espacialidade e na formação da imagem estéreo; elas são fracas o suficiente em salas de controle para não produzirem um efeito notável, no entanto o devido cuidado deve ser tomado ao se posicionar tanto o equipamento da sala quanto as superfícies duras. Para assegurar isso, diversos padrões internacionais especificam que, nas salas de monitoração, as primeiras reflexões que ocorrem em até 15ms devem ser pelo menos 10dB abaixo do som direto (entre 1 e 8KHz), especificação difícil de ser seguida em estúdios onde temos mesas de mixagem posicionadas entre os alto-falantes e o engenheiro de som, o que favorece muito as reflexões na mesma. Para isso, a especificação considera apenas salas de escuta onde temos somente o ouvinte e os alto-falantes, onde as reflexões do chão e do teto podem ser realmente problemáticas, mais do que a presença de uma mesa de mixagem.

Na região grave do espectro sonoro as salas podem apresentar regiões de ressonância, denominadas “modos da sala”, que são padrões estáticos máximos e mínimos de pressão sonora diretamente relacionados ao comprimento de onda de determinadas frequências. Para solucionar isso se deve tomar cuidado com as medidas da própria sala assim como com o posicionamento dos alto-falantes, que não deve estar em nenhum nó ou antinó desses modos, o que alteraria consideravelmente a percepção da sua resposta em frequência.

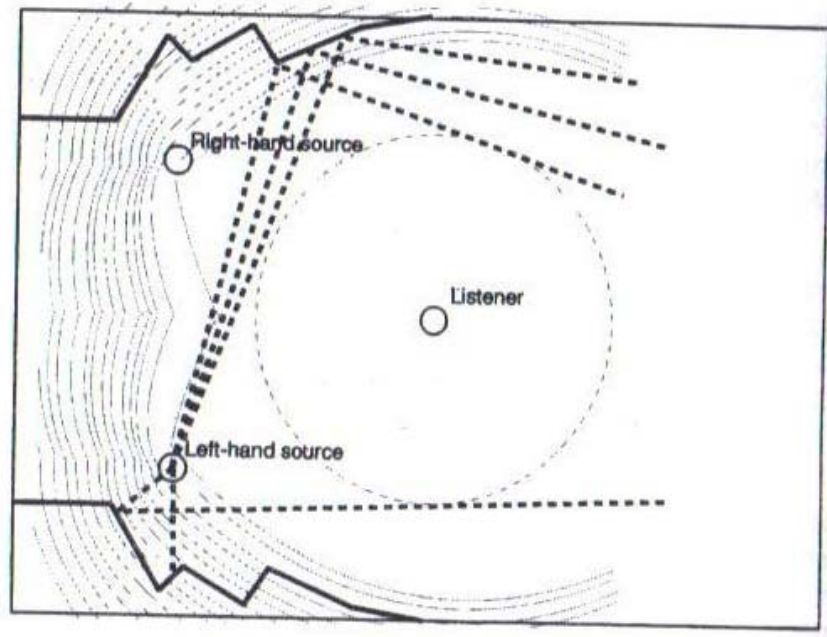
Sistemas de monitoração de dois canais em estúdios são geralmente instalados nos limites da sala (montados nos cantos) em cada um dos dois lados da janela para a sala de gravação, ou simplesmente atrás da mesa de mixagem. Monitores *near-fields*, ou de campo próximo, normalmente são mais utilizados como uma segunda opção de monitoração, menos afetada pelas reverberações da sala e mais próxima da escuta em ambientes domésticos.

O design denominado LEDE (*live-end-dead-end*) foi utilizado largamente na construção de estúdios, onde uma área frontal da sala era tratada com material absorvedor para minimizar as reflexões frontais, do teto e laterais, enquanto o fundo da sala era mais reflexivo, recebendo difusores para criar uma reflexão atrasada e difusa em torno da posição de escuta. Esse método dava à sala um som natural com primeiras reflexões controladas, apesar das provenientes da mesa de mixagem e da janela serem muito dificilmente evitadas.



Sala construída a partir do princípio LEDE. Fonte: RUMSEY, 2001.

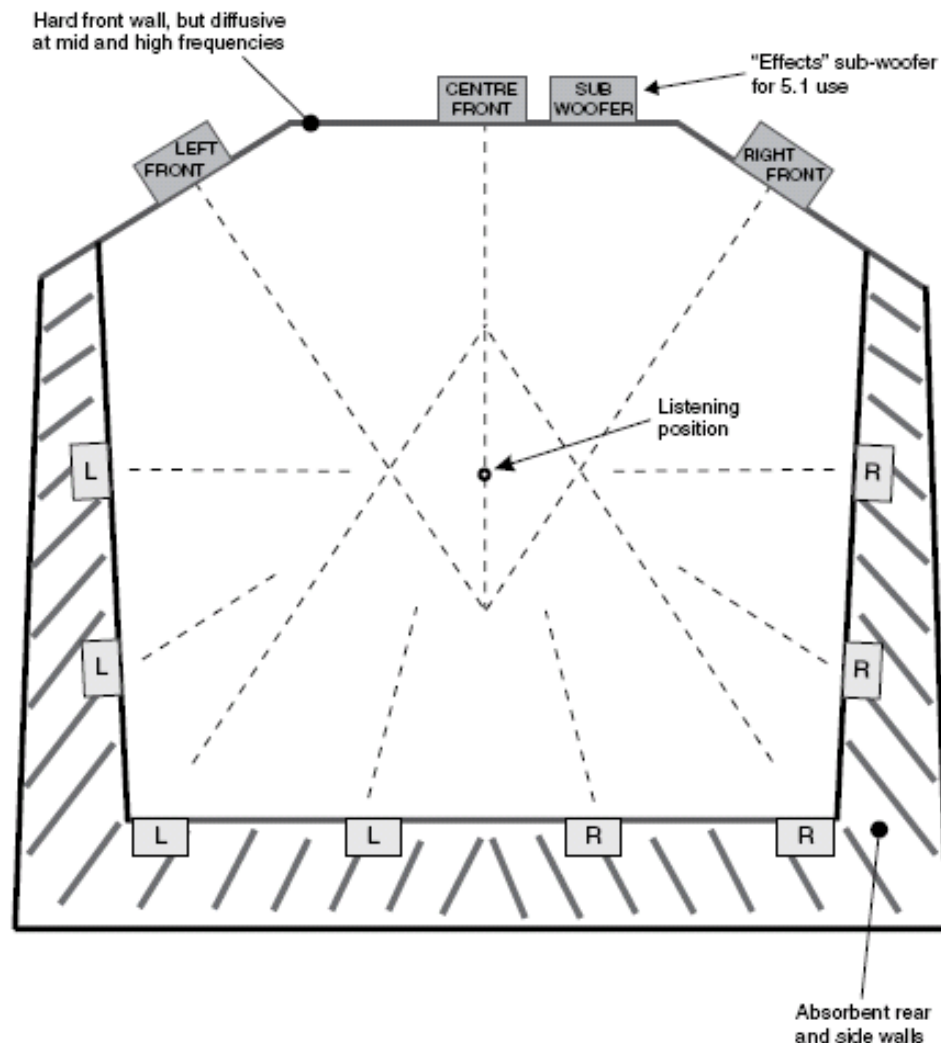
Outro método muito utilizado e denominado de “zonas livres de reflexões” consiste na colocação de material reflexivo próximo aos alto-falantes de forma que as ondas sonoras se dirigissem para longe do ouvinte e para o fundo da sala, criando um período de tempo em que a resposta próxima ao ouvinte era quase que a de uma sala anecóica. Uma outra abordagem possível consiste na colocação dos alto-falantes nas arestas de uma parede frontal muito sólida que funciona como um abafador, e a parede traseira funciona como um atenuador de graves, o que acabou criando o conceito de “sala sem ambientação”. Tal sala tende a se comportar como uma sala anecóica uma vez que todas as paredes, exceto as frontais, são fortemente tratadas com material absorvedor; apesar disso, as reflexões próximas à região dos alto-falantes proporcionam uma sensação natural de espaço, resultando em imagens sonoras “fantasma” altamente precisas.



Sala construída a partir do princípio das zonas livres de reflexões. Fonte: RUMSEY, 2001.

Um grande número de salas desenvolvidas para monitoração estéreo são essencialmente bidirecionais, ou seja, as laterais são simétricas enquanto as pontas não o são, o que representa uma grande dificuldade na utilização dessas salas para a monitoração em *surround* 5.1.

Há em curso a formação gradual de um consenso, segundo o qual salas desenvolvidas para monitoração multicanal devem possuir uma distribuição uniforme de material absorvedor e difusor, de forma que os alto-falantes traseiros funcionem de forma semelhante aos frontais. Tal observação traz certos conflitos, uma vez que sugere que mais da metade das salas de mixagem do mundo são impróprias para a mixagem em sistema *surround* 5.1 e que uma sala de monitoração não consegue ser ótima para mixagem *surround* e para estéreo ao mesmo tempo.



Sala construída especificamente para monitoração em *surround* 5.1 a partir do conceito de se utilizar múltiplos alto-falantes *surround* sem a reverberação da sala. Fonte: Newell, 2003.

Outra discordância surge com relação à função dos canais traseiros, uma vez que, segundo o proposto pelo seu posicionamento, as reflexões das paredes laterais se tornam mais importantes do que da parede frontal. Como já mencionado, existem diferentes opiniões a respeito do material sonoro direcionado para esses canais, embora a sua intenção inicial sempre tenha sido a de acrescentar ambiência não direcional e informações de efeitos que adicionam impressão espacial. Segundo o padrão ITU "(...) não é necessário que os alto-falantes traseiros/ laterais sejam capazes de recriar imagens sonoras fora da região dos alto-falantes frontais". Se tal determinação for seguida, se torna menos arriscado a questão da influência das diferentes acústicas onde o som for reproduzido, uma vez que o sinal proveniente dos

canais frontais irá mascarar as reflexões indesejadas geradas a partir dos canais traseiros. Normalmente os sinais dos canais traseiros possuem o nível mais baixo em relação aos frontais, o que minimiza mais ainda as reflexões em ambientes impróprios para a monitoração de sistemas *surround*.

Outra questão referente ao posicionamento dos alto-falantes de um sistema *surround* 5.1 é que ele exige que a sala seja mais larga do que longa, o que se torna muito difícil em salas previamente projetadas para sistemas estéreo, inclusive pelo fato de que, com a rotação do eixo da sala, o tratamento acústico e a posição de portas e janelas ficam comprometidos. Em todo caso, o ideal é que a sala seja projetada desde o início para a reprodução em *surround* 5.1, o que implica inclusive na distribuição mais uniforme de materiais absorventes e difusores.

Segundo RUMSEY (2001), diversos guias para a monitoração de som *surround* 5.1 têm sido desenvolvidos por diversos grupos especializados em todo o mundo, entre eles podemos citar o fórum de som *surround* alemão, o fórum japonês HDTV e o comitê técnico da AES em áudio multicanal e binaural. De forma semelhante, sistemas específicos, como o THX, também disponibilizam guias com recomendações para que os estúdios entrem em conformidade com os requisitos acústicos para monitoração em seus sistemas. Nesses guias estão estabelecidos limites de tempo de reverberação, níveis de ruído, níveis de reflexões e razões entre as dimensões da sala, assim como especificações referentes aos monitores de áudio. Existem também diferenças nas especificações com relação a salas de monitoração e de mixagem, onde existe equipamento presente, principalmente nas salas onde existe também a mixagem de material para cinema ou TV, nas quais monitores de vídeo fazem parte do equipamento.

Dentre os diversos detalhes a respeito do posicionamento dos alto-falantes em um sistema *surround* 5.1, encontra-se o posicionamento do canal central no mesmo arco criado pelos outros alto-falantes, o que pode interferir no posicionamento do monitor de vídeo, problema que pode ser contornado se alterando-se o posicionamento do monitor de vídeo ou, como sugere a Dolby, posicionar o alto-falante central um pouco abaixo dos laterais, mas posicionando os alto-falantes agudos (*tweeters*) dos monitores laterais para baixo e o do monitor central para cima, mantendo certo alinhamento do

espectro sonoro relativo às informações direcionais, encontradas na região mais aguda. O *subwoofer*, por sua vez deve ser posicionado em algum canto da sala, onde sua resposta em frequência seja favorecida na região grave. Se utilizados dois *subwoofers*, estes devem ser posicionados simetricamente na sala, e em relação ao ponto de escuta do engenheiro de som.

Com relação ao nível de monitoração, em geral é recomendável que se meça, com ruído rosa, um sinal médio de -18dBFS RMS mas, segundo a Dolby, o nível de monitoração deve estar situado entre 85dBC e 79dBC (que corresponde à aproximadamente -20dBFS), quando existe material *surround* com diálogos centrais. O canal LFE deve apresentar 10dB acima dos outros canais.

RUMSEY (2001) discorre também brevemente sobre os sistemas de *home-theater* virtuais, uma aplicação de sistemas binaurais através de fones de ouvido para a criação de salas de monitoração multicanal virtuais em ambientes onde a monitoração através de alto-falantes não é prática ou possível, como mencionado na seção 1.3.3.1 do presente trabalho.

Capítulo 3

Estudo de casos: análises, práticas de mixagem em 5.1 e transcrições para o sistema binaural

3.1 – Recriação de espaços e movimentos em gravações em 5.1: apreciação de trechos de filmes e produções musicais

Aqui realizamos uma breve análise de obras cinematográficas e musicais, focando na utilização do espaço sonoro pelas mesmas, descrevendo e ilustrando algumas particularidades. De uma forma geral, na maioria das produções cinematográficas, os diálogos se apresentam no canal central e a ambientação nos canais traseiros (Ls e Rs). No entanto, como já mencionado anteriormente, alguns efeitos são explorados, aproveita-se o espaço sonoro representado pelo sistema *surround* 5.1, em sua grande maioria com o intuito de intensificar uma experiência observada visualmente ou de dar mais dramaticidade a eventos marcantes. Nesse contexto é comum notarmos objetos ou personagens que produzem sons específicos e característicos, tendo sua imagem seguida pelo seu som. Como exemplo disso observamos a apresentação do esporte Quadribol, no filme “Harry Potter e a Pedra Filosofal” (2001).

Nesse trecho, onde o jogo dos bruxos é apresentado ao personagem principal, podemos observar a presença de sons ambientes compostos principalmente por sons de pássaros, no início da cena, predominantemente nas partes de trás e laterais da paisagem sonora. Segundo a classificação sonora elaborada por CHION (1994), esses sons se enquadrariam em sons ambientes ou sons *offscreen* passivos, que refletem uma extensão da cena ao ilustrar um território cercado de árvores. O ouvinte que assiste o filme sabe que atrás do lugar onde os jovens se encontram há uma floresta, que permanece em cena sonoramente. Assim que os personagens entram em cena, suas

vozes, como utilizado na maioria das produções, são posicionadas na região frontal do espaço representado pelos alto-falantes.



Na apresentação das bolas que compõem o jogo temos um bom exemplo do som característico de cada objeto o acompanhando-o pelo espaço sonoro. O “balaço”, ao ser arremessado para além dos limites visuais, é acompanhado pelo seu som, inicialmente para a direita, rebatida para a esquerda e, ao retornar, se movimentando da esquerda para a direita até ser apanhada por um dos personagens.



Na apresentação do “pomo de ouro” temos outro exemplo de um objeto visual sendo acompanhado pelo seu som. Inicialmente, ainda na palma da mão do personagem principal, nos é apresentado o ruído decorrente do bater das asas do objeto. Logo em seguida quando ele levanta vôo, seu som o acompanha se posicionando inicialmente ao centro, esquerda superior, direita superior, esquerda superior novamente e segue para o que seria uma extensão da imagem passando ao lado e se dirigindo para trás do espectador, se tornando um som acusmático, segundo nomenclatura adotada por CHION

(1994), ou seja, sem representação visual, mas ainda assim diegético⁵⁶, dentro do contexto da narrativa.



Um outro trecho extraído do mesmo filme é a cena onde a personagem Hermione se encontra com um Ogro e seus amigos vão socorrê-la. Nesse exemplo observamos uma abordagem de ambientação não estritamente conectada à imagem apresentada, assim como uma exploração pontual dos efeitos de baixas frequências. Nessa cena, os sons referentes aos trovões, assim como a música de fundo são direcionados principalmente aos canais *surround* (Ls e Rs), de forma semelhante e também estática é posicionada a sonoplastia referente à água das torneiras após o Ogro quebrar as pias. Nessa ocasião, apesar da fonte sonora (vazamento das pias) se apresentar ora na frente do ouvinte, ora atrás do mesmo, o som referente à mesma permanece fixo nos canais frontais, principalmente L e R, assim como o som referente aos trovões permanecem sempre atrás do ouvinte.



⁵⁶ Dentro do contexto da narrativa.

Outro efeito extremamente interessante nessa cena se dá quando da queda do Ogro no chão, devido ao seu enorme peso e tamanho, causando um tremor que repercute no sistema *surround* através do *subwoofer* com material do canal LFE.



Depois disso o som da água vinda dos canos estourados da pia continua na frente, independente da tomada de cena da câmera e ponto de vista do ouvinte.



No trecho seguinte, extraído do primeiro filme da trilogia “The Matrix”, Neo e Trinity, protagonistas do filme, estão no terraço de um prédio onde enfrentam policiais até a chegada de Smith, o antagonista. Neo se vê frente a frente com ele e descarrega sua munição, não obtendo sucesso em nenhum tiro devido à rapidez dos movimentos de Smith. Nesse momento, observamos um efeito semelhante ao que CHION (1994) descreve como voz de frente e voz de costas, no entanto relativo ao timbre dos tiros da arma de Neo. Sempre que a imagem apresenta o rosto do personagem principal e de sua arma, o timbre se mantém claro, mais presente e com o timbre bem realístico. Quando vemos os movimentos de Smith ao desviar das balas, o timbre se modifica, imitando o timbre de como se ele viesse das nossas costas, sem clareza na região aguda do espectro sonoro.



Depois do insucesso em acertar o agente Smith, cabe a Neo tentar desviar dos tiros realizados pelo antagonista. Nesse momento, temos basicamente quatro pontos de vista e de escuta. Com a câmera posicionada atrás do personagem principal ouvimos o som das balas do agente Smith vindo ao nosso encontro, no entanto, como Neo desvia delas, os sons passam ao nosso lado. No segundo ponto de vista, vemos Neo de lado e o som das balas faz o percurso da direita para a esquerda, acompanhando a imagem em câmera lenta.

No terceiro ponto de vista, nos encontramos de frente para Neo, que está se abaixando, e o timbre do som das balas é modificado e passa a se originar de trás do ouvinte. No quarto ponto de vista estamos novamente atrás de Neo e a bala o acerta na perna. Seu som vem da frente do ouvinte na tentativa de simular a sensação de que a bala também acerta o ouvinte.



Um outro efeito interessante e definido por CHION (1994) como *offscreen trash*, ou sobras fora da tela, acontece nessa mesma cena quando Trinity atira à queima roupa no personagem Smith. O som se inicia no centro

da tela e se propaga por todo o campo sonoro representado pelo sistema *surround* 5.1.



Um outro trecho de filme onde o espaço sonoro representado pelo sistema *surround* 5.1 é abordado de forma muito interessante é o trecho do discurso do rei em “A Múmia – Tumba do Imperador Dragão”. Inicialmente, observamos nesse trecho, tomado pelo sentimento predominante de início de guerra, o papel da música de fundo, que é o de intensificar esse sentimento, inclusive com pontos de sincronização envolvendo os passos do imperador e as batidas fortes da música. No entanto, o foco principal desse trecho é ilustrar o que CHION (1994) define como voz de frente e voz de costas, seguida de uma intensa expansão do campo sonoro – *superfield* – onde até a própria voz do antagonista é reposicionada, quebrando completamente com os parâmetros tradicionais do cinema de manter as vozes no centro.

Quando o imperador se dirige ao seu exército, temos nove momentos de diferentes sonorizações para diversos pontos de vista e de escuta espaciais de seu discurso. Na primeira cena o imperador fala a seu exército e o ponto de vista que temos é de suas costas. Ouvimos sua voz timbricamente modificada trazendo realismo à cena, como se estivéssemos realmente às suas costas. Logo em seguida vemos e ouvimos seu discurso frente a frente. Nesse momento visualizamos seu rosto e sua voz é clara e posicionada bem próxima, no centro da imagem sonora à nossa frente.



Num terceiro momento o nosso ponto de vista e de escuta se localiza em frente ao exército e ouvimos a voz do imperador como vindo de longe, no entanto ainda não posicionada às nossas costas, mesmo que o imperador o esteja lá, sem o vermos. Logo em seguida temos novamente o rosto do imperador exibido na tela e sua voz volta a ser clara e posicionada bem próxima ao centro e à nossa frente.

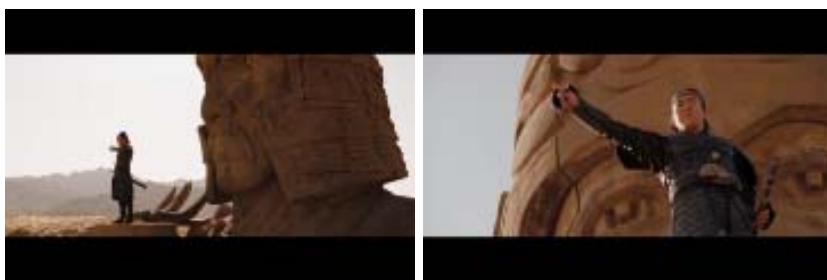


Nas quinta e sexta posições da câmera e pontos de escuta temos a voz do imperador à nossa frente, no entanto não tão próxima quanto nas cenas frontais anteriores. Na sétima cena o ponto de vista se volta para o exército e dessa vez, uma vez que o ouvinte possui o timbre da voz do imperador em mente e bem reconhecível, a voz do imperador se mostra longe e, no sistema *surround* 5.1 localizada bem atrás do espectador, representando o *superfield* de CHION, sem apresentar o estranhamento decorrente de um som *offscreen* que, por não estar sendo representado por uma imagem, segundo o autor, pode se tornar algo independente e não-diegético⁵⁷. No caso apresentado, o som é absolutamente diegético mesmo sendo apresentado completamente *offscreen*, ou fora do cenário apresentado pela imagem.



⁵⁷ Fora do contexto da narrativa.

Nos dois últimos planos da imagem que temos é um som posicionado à frente do espectador com dois níveis de distanciamento, recurso já explorado no sistema monofônico e ainda utilizado como recurso extremamente expressivo com relação ao posicionamento de sons *onscreen* (dentro da cena) em relação às suas respectivas fontes sonoras.



Quando é apresentada a resposta do exército ao discurso do imperador, o mesmo artifício utilizado nas duas cenas anteriores é repetido, agora com um outro material sonoro (vozes do exército) e outras fontes sonoras.



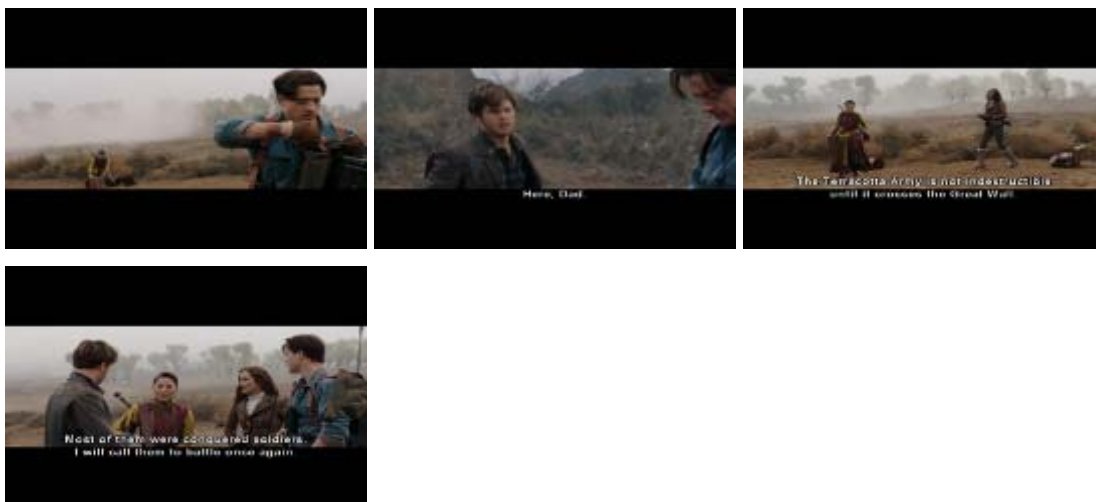
Nesse próximo trecho, também selecionado do mesmo filme “A Múmia – Tumba do Imperador Dragão”, quando os protagonistas vão resgatar uma das personagens, indo de encontro ao exército do imperador, temos um claro exemplo de expansão do espaço representado na tela através do som em sistema *surround* 5.1.

Inicialmente observamos aquele artifício de distanciamento aplicado ao som do avião que acabou de deixá-los em terra e sai de cena pela esquerda da imagem. Logo em seguida temos o que CHION denomina de som etéreo – *on-the-air* – representado pelo som vindo do rádio, não visualizado e posicionado à direita do ouvinte. Logo em seguida ouvimos a voz do personagem principal, facilmente reconhecível e posicionada completamente fora do enquadramento à direita, enquanto no centro da tela temos a visualização das duas mulheres.



Esse efeito já nos revela a posição do personagem ao responder o rádio mesmo que não o vejamos. É um exemplo claro também do que CHION chama de som “acusmático”, sem dúvida alguma inserido na narrativa, que é apresentado apenas sonoramente para depois ser visualizado. Com relação à sua denominação de som fora do enquadramento, esse trecho mostra o que CHION chama de um som *offscreen* relativo, pois mesmo que o personagem não esteja na tela, ele se apresenta na cena através da expansão do campo sonoro – *Superfield* – ampliado pelos efeitos laterais.

Após a revelação de sua imagem, sua voz caminha junto à mesma para o centro da paisagem sonora, onde todas as outras vozes, dos outros personagens, se encontrarão durante o resto do diálogo.



A observação desses trechos de filmes contemporâneos nos revela que, embora ainda timidamente explorado em aplicações cinematográficas, o sistema *surround* 5.1 possibilita a expansão da paisagem sonora de forma bastante realística, enriquecendo a produção para além da tela do cinema,

principalmente quando essas exploram a movimentação do som nos 360° oferecidos pelo sistema de reprodução sonora.

O último trecho sobre o qual faremos algumas observações foi extraído do filme “O Senhor dos Anéis – O Retorno do Rei”, da cena da batalha em frente à Minas Tirith (cidade branca). É interessante observar nessa cena a exploração do canal LFE para sonoplastia dos passos dos Olifantes, assim como a exploração da movimentação das flechas, principalmente entre os canais frontais e, de forma semelhante, o grito dos soldados ao cair do Olifante, ora à direita, ora à esquerda, e novamente a utilização do canal LFE na queda do Olifante quando abatido.



Podemos observar também, na cena da luta entre um dos personagens principais e o vilão, a sonoplastia da movimentação da luta entre os canais frontais, a corrente da massa carregada pelo vilão posicionada como na imagem à esquerda, assim como o deslocamento da mesma (da esquerda para a direita) quando acontece um golpe, e a utilização do canal LFE quando ela acerta o chão.



Nas abordagens adotadas pelas produções cinematográficas vemos, em geral, a exploração de efeitos sonoros no espaço sonoro representado pelo sistema *surround* de forma bastante intensa em momentos extremamente importantes, dramáticos ou no clímax do enredo, intensificando tais momentos. Podemos fazer uma crítica com relação a isso no que se refere à possibilidade de utilização desses efeitos, movimentação e envolvimento sonoro, durante toda a produção, o que não acontece.

Já as produções musicais que se utilizam do sistema *surround* foram classificadas em três categorias, para facilitar a análise:

- > Produção Cinematográfica Musical, ou simplesmente conhecida como Musical;
- > Produção musical de shows ou eventos musicais acompanhados de imagens;
- > Produção puramente musical.

Na categoria produção cinematográfica musical, tomamos como exemplo “O Fantasma da Ópera”, onde observamos a mixagem da música sempre independente da imagem, consistindo em uma paisagem sonora estática. Os diálogos, como nas produções cinematográficas tradicionais, são centrados na tela e as vozes secundárias posicionadas tanto no canal central como no estéreo frontal. Nas árias principais onde temos um solista, o mesmo se posiciona na frente do ouvinte e a orquestra ou grupo instrumental espalhado nos cinco canais do sistema *surround*, sempre com o objetivo de proporcionar uma sensação de imersão sonora mais intensa. O mesmo pode ser observado com a reverberação da voz ou instrumento solista. Quando um coro se apresenta, ele é posicionado nos canais frontais e sua reverberação nos canais traseiros.



Os aplausos, de uma forma geral, aparecem mais intensos na frente, com reverberação nos canais traseiros, o que, assim como no observado anteriormente, pode ser reflexo da necessidade de uma compatibilidade com o sistema estéreo. O canal LFE é utilizado basicamente para a região grave do espectro sonoro assim como para efeitos especiais utilizados no decorrer do enredo.

Efeitos especiais, tais como o das espadas, nas lutas entre os personagens, e gritos, são posicionados no canal central e explorados de forma muito sutil com relação à movimentação, como na cena da capa do fantasma se movimentando da esquerda para a direita. Algumas vezes esse efeito acompanha a cena como no galope de cavalos e os passos dos personagens à esquerda, sempre com o objetivo de dar mais dramaticidade ao evento visual, aproximando-se dos exemplos das produções cinematográficas tradicionais.



As produções de shows observadas, tais como o DVD “Metallica The Vídeos 1989-2004” lançado pelo selo Vertigo, adotam a estética de posicionamento das fontes sonoras principais, localizadas visualmente no palco, no conjunto de alto-falantes frontais, exatamente como sempre foi adotado pelo sistema estéreo tradicional. O material direcionado aos alto-falantes traseiros é limitado a alguns efeitos e, principalmente, à reverberação do ambiente, responsável pela sensação de envolvimento dessas produções. Apesar do ponto de vista se modificar constantemente, o ponto de escuta

permanece sempre o mesmo, e o mesmo das produções estéreo, apenas aproveitando o campo sonoro *surround* 5.1 para uma pequena parcela do conteúdo musical. A mixagem sonora, seja ela estéreo ou *surround* 5.1, é, na maioria das vezes, estática e está desprendida da imagem visual apresentada, muitas vezes cheia de movimentação.



Uma abordagem um pouco fora desse padrão pode ser observada no DVD do Pink Floyd gravado em um show ao vivo do disco “Pulse”, onde foram incluídas algumas faixas extras de músicas do disco “The Dark Side of the Moon” com a mixagem *surround* 5.1 realizada durante o show, também com difusão *surround* 5.1. Nessas faixas, alguns efeitos que foram espacializados ao redor do público durante o show e também são reproduzidos no DVD.

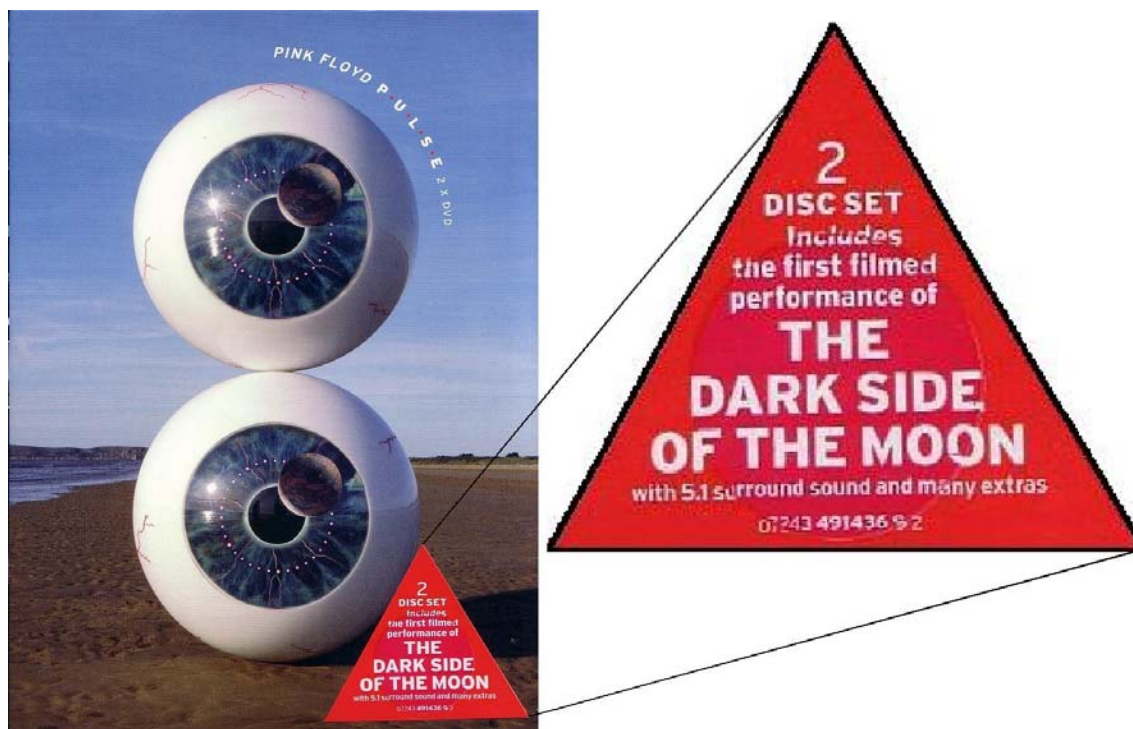
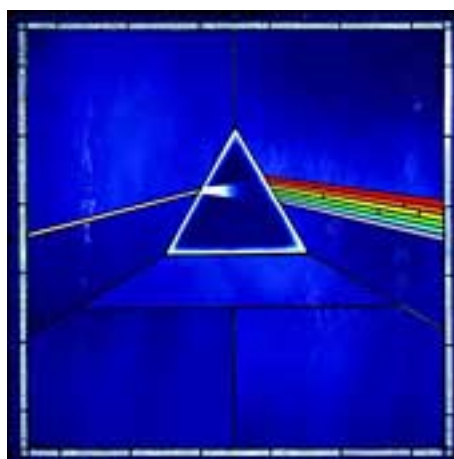


Foto da capa do DVD Pulse do Pink Floyd com detalhe para a nova produção *surround* 5.1. Fonte: DVD Sony Music Pulse do Pink Floyd.

Além dessa versão, baseada na performance realizada ao vivo, o Pink Floyd possui um lançamento em SACD⁵⁸ de 2003 que apresenta uma outra mixagem do álbum “The Dark Side of the Moon” utilizando o sistema *surround* 5.1 para reproduzir uma abordagem sonora estética espacial diferente da tradicional estéreo (mais conhecida) e que só havia sido lançada em um LP quadrifônico na década de 1970, não tendo atingido muitos usuários. Tal mixagem nos apresenta a concepção musical original da época e se mostra completamente diferente da que conhecemos, incluindo o espaço como elemento musical fundamental. Em “Money”, por exemplo, observamos cada um daqueles sons de caixas e moedas em uma posição específica no espaço sonoro, representado agora em 360º horizontais ao redor do ouvinte. Efeitos de movimento também são observados nas faixas “Speak to me” e “Us and Them”.



Capa do SACD “Dark Side of the Moon” do grupo Pink Floyd. Fonte:

<http://www.cduniverse.com/search/xx/music/pid/5690739/a/Dark+Side+Of+The+Moon.htm>

Outro tipo de produção que se comporta de forma muito semelhante às de shows e eventos são as de óperas e orquestras, que também apresentam uma mixagem estática e baseada na estética tradicional de posicionamento das fontes sonoras no plano frontal e da reverberação do ambiente no plano traseiro. Apesar da riqueza dos componentes visuais, principalmente das óperas, assim como da sua movimentação, a mixagem dos eventos sonoros não a acompanha. Como exemplo desse tipo de produção podemos citar o

⁵⁸ Super Áudio CD.

DVD da ópera “Götterdämmerung” (Crepúsculo dos Deuses) de Richard Wagner, gravado pelo selo da Deutsche Grammophon, na interpretação da The Metropolitan Opera Orchestra and Chorus, e regência de James Levine. Nessa ópera, apesar da riqueza de figurino, cenário e movimentação da imagem, a mixagem sonora permanece estática e basicamente frontal.



Foto da capa do DVD da ópera Götterdämmerung de Richard Wagner. Fonte: DVD Universal Music Götterdämmerung.

Nesse contexto de abordagem estética sonora predominantemente estática e frontal encontramos uma produção que nos chamou muito a atenção, uma vez que foge completamente desses padrões.

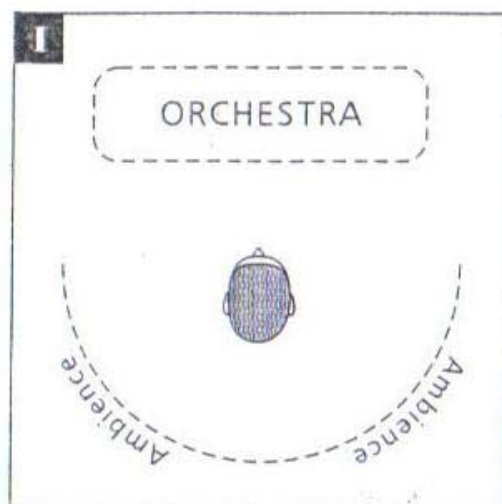
Nos concertos de Brandenburgo de Johann Sebastian Bach, lançada pelo selo TACET Real Surround Sound, que podemos classificar como uma produção puramente musical, diferentemente das diversas outras produções dessa categoria, a mixagem sonora em *surround* 5.1 aborda a idéia de posicionar o ouvinte no centro dos instrumentos musicais, com um posicionamento diferente para cada um dos concertos, dependendo da instrumentação e número de solistas. Tal procedimento se justifica pelo contexto histórico em que essas obras eram interpretadas, de forma bastante

livre com relação ao posicionamento dos instrumentistas entre si e em relação ao ouvinte, sempre favorecendo a apreciação da corte, na época financiadora e maior consumidora das atividades artísticas.



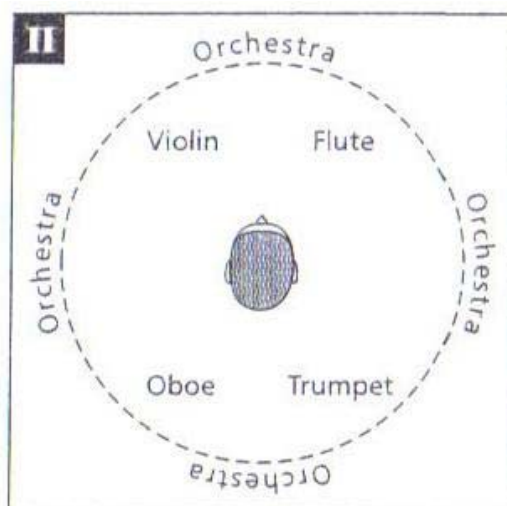
Foto da capa do Super Áudio CD Tacet Real Surround Sound com os seis concertos de Brandenburgo de Johann Sebastian Bach. Fonte: SACD Tacet.

No primeiro concerto, o que temos é a abordagem tradicional já observada anteriormente, com toda a orquestra, composta de cordas, um violino piccolo como solista, três oboés, fagote, duas trompas e um contínuo, tocando à frente do ouvinte, como se ele estivesse sentado na poltrona central de um teatro, sala de concerto ou igreja, e onde ele é envolvido por toda a reverberação do ambiente natural, material direcionado aos canais traseiros.



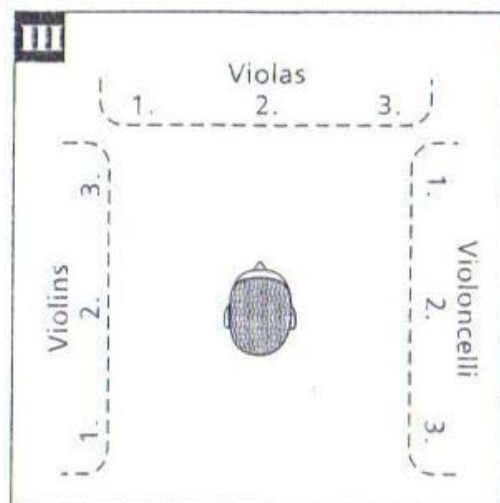
Abordagem estética de posicionamento dos instrumentos da produção musical do primeiro concerto de Brandenburgo de Bach, da TACET Real Surround Sound. Fonte: DVD-Audio TACET, 2000.

No segundo concerto, o violino solista se posiciona à frente e à esquerda do ouvinte, a flauta à frente e à direita, o oboé atrás e à esquerda e o trompete atrás e à direita, todos eles eqüidistantes do ouvinte. Entre esses instrumentos solistas estão posicionados os outros instrumentos da orquestra.



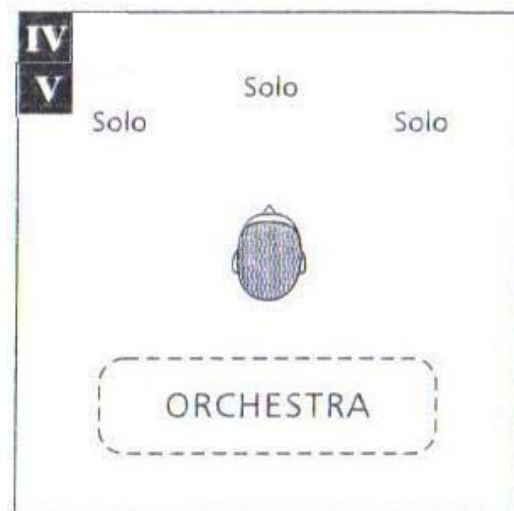
Abordagem estética de posicionamento dos instrumentos da produção musical do segundo concerto de Brandenburgo de Bach, da TACET Real Surround Sound. Fonte: DVD-Audio TACET, 2000.

No terceiro concerto temos três violinos, três violas e três violoncelos. O grupo de violinos está distribuídos por toda a esquerda, as violas preenchem a frente e os violoncelos estão posicionados ao lado direito do ouvinte, ao todo os nove músicos estão ao redor do mesmo, exceto no plano traseiro.



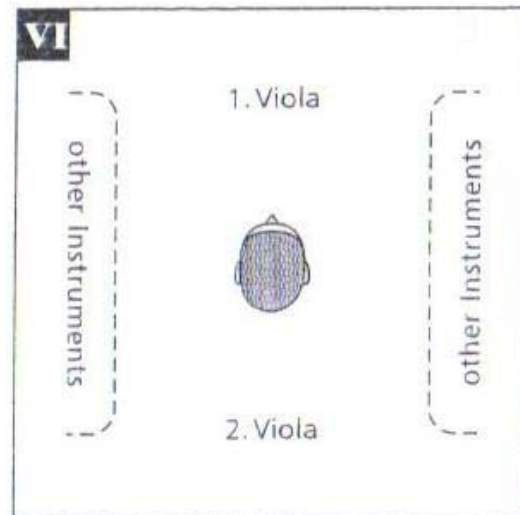
Abordagem estética de posicionamento dos instrumentos da produção musical do terceiro concerto de Brandenburgo de Bach, da TACET Real Surround Sound. Fonte: DVD-Audio TACET, 2000.

Nos quarto e quinto concertos, os solistas são ouvidos à frente. No quarto concerto existem duas flautas, uma à direita e a outra à esquerda, e um violino ao centro. No concerto nº 5 o violino está à esquerda, o cravo no centro e a flauta à direita. Em ambos os concertos a orquestra é percebida atrás, posicionando o ouvinte entre os solistas e a orquestra.



Abordagem estética de posicionamento dos instrumentos da produção musical do quarto e quinto concertos de Brandenburgo de Bach, da TACET Real Surround Sound. Fonte: DVD-Audio TACET, 2000.

No sexto concerto, as duas violas solistas que tocam com as violas da gamba e o continuo estão dispostas uma na frente e outra atrás do ouvinte. Os demais instrumentos são percebidos em torno do ouvinte.



Abordagem estética de posicionamento dos instrumentos da produção musical do sexto concerto de Brandeburgo de Bach, da TACET Real Surround Sound. Fonte: DVD-Audio TACET, 2000.

Já no âmbito de produções com o propósito de divulgação e distribuição de música contemporânea, o cenário é bastante diferente e como exemplo disso encontramos o DVD “Immersion”, que trás obras com eletrônica quadrifônica, onde o estéreo frontal e traseiro se apresentam mais equilibrados, com pouca utilização do canal central e de LFE, refletindo a concepção dessas obras num sistema quadrifônico, que aborda os alto-falantes de forma mais independente.

Essa produção traz também obras concebidas no formato *surround* 5.1 com o propósito de explorar todas as possibilidades desse novo formato com resolução de áudio superior a qualquer outra.



Capa do DVD Immersion, que utiliza o sistema *surround* 5.1 para a difusão da música eletroacústica espacializada. Fonte: DVD Immersion.

Em todas essas produções cinematográficas ou musicais, o que observamos é uma utilização interessante, embora ainda muito tímida, de todo o campo sonoro reproduzido pelo sistema *surround* 5.1 na distribuição e posicionamento das fontes sonoras, atitude justificável pela necessidade de compatibilidade com o sistema estéreo tradicional ainda dominante. No caso de produções com o intuito de divulgação e distribuição de música contemporânea espacializada, percebemos uma utilização mais equilibrada desse campo sonoro, no entanto as mesmas ainda se mostram em pequeno número, refletindo a emancipação de um sistema ainda em processo de adaptação, tanto dos profissionais que o manipulam quanto do público ouvinte.

3.2 – Mixando em 5.1: a teoria na prática

Durante os dois anos do meu curso de mestrado fiz a gravação de alguns eventos nos quais realizei mixagens em sistema *surround* 5.1 utilizando

diversos métodos com o propósito de adquirir alguma experiência no formato em questão. Dentre essas gravações selecionei algumas para serem descritas aqui e que apresentam certa diversidade em seu processo de produção.

Em conversa do presente autor, durante o Festival de Inverno de 2009 em Campos de Jordão, com o engenheiro de gravação e produtor Jakob Handel, o mesmo mencionou o fato de que, em produções de áudio em *surround* 5.1 para DVDs, devido à necessidade de compatibilidade com o sistema estéreo, ele prefere sempre realizar a gravação no sistema estéreo tradicional com microfonação próxima e microfonação com pares de microfones ambientes e, a partir dessa mixagem e através de processamento digital, obter o sinal correspondente aos seis alto-falantes do sistema *surround* 5.1. Para isso ele utiliza um processador de efeitos da Lexicon e o que se obtém é uma maior abertura do espaço representado frontalmente nas três caixas do sistema, uma reverberação direcionada aos canais traseiros e o sinal do canal LFE a partir de uma filtragem do espectro sonoro total emitido pelas outras cinco caixas⁵⁹.

Diante disso e observando que diversas outras produções também utilizam esse tipo de artifício para a espacialização no formato *surround* 5.1 (o que pode ser comprovado em algumas produções que declaram já em sua contracapa "5.1 - simulado"), a primeira mixagem descrita aqui, resultado de uma gravação estéreo tradicional de um concerto de música antiga que ocorreu na 2ª Semana de Música Antiga entre os dias 27 de outubro e 2 de novembro de 2009, parte dessa idéia e possui uma abordagem mais simplificada baseada na ambientação *surround*.

⁵⁹ Diversos *plugins*, como o Auralizer do Digital Performer, já mencionado anteriormente, também funcionam a partir desse mesmo princípio.



Foto da ópera gravada na II Semana de Música Antiga. Fonte: site oficial do evento.

Nessa gravação, o material inicial foram os sinais da microfonação próxima de cada um dos instrumentos da orquestra de câmara assim como os dois microfones direcionados ao palco, que foram processados no Digital Performer 7 com os *plugins* panorâmicos Auralizer e ArcPanner, já descritos no capítulo anterior. A estética da mixagem se manteve estática e com o ponto de escuta centrado na posição de um ouvinte situado no centro da platéia. Vale observar que não foi aproveitado o som proveniente da própria platéia ou mesmo da sala em questão na mixagem *surround* 5.1, uma vez que estes não foram gravados para essa finalidade.

Na reprodução através do sistema *surround* 5.1, observamos a reverberação adicionada através do processamento do Auralizer (que também é um simulador de salas), elemento não presente na mixagem realizada utilizando-se o Arc Panner e que pôde ser obtida através de outro *plugin*, o eVerb. Este efeito foi inserido em cada canal individualmente com o propósito de processar o sinal mono do mesmo em um correspondente aos cinco alto-falantes do sistema. A possibilidade de se utilizar um canal auxiliar também é possível se o mesmo for configurado para receber um sinal já processado para o sistema 5.1, ou seja, possuir uma entrada mono e uma saída 5.1, através de *buses surround*. Caso contrário, se a entrada ou a saída não for compatível

com a configuração do sistema, a imagem sonora estabelecida pelo som direto em 5.1 pode se comprometer.

A orquestra, posicionada originalmente em um fosso abaixo e à frente do palco era composta por dois violoncelos, dois violinos, duas flautas, duas espinetas e duas teorbais. O palco, onde se encontravam os cantores e o coro foi captado com dois microfones cardioides dedicados, cada um, para uma metade do mesmo. Com relação ao posicionamento em ambiente virtual das fontes sonoras no plano de 360° oferecido pelo sistema *surround* 5.1, realizamos a distribuição do sinal dos microfones do palco de forma proporcional ao eixo do sistema na sua parte frontal, e da mesma forma realizamos a distribuição dos outros pares de instrumentos de forma não correspondente à situação real, no entanto mais equilibrada e possibilitada somente devido à captação próxima e individual dos instrumentos. Nessa gravação foi possível também trazer o coro e cantores solistas para a frente da orquestra.

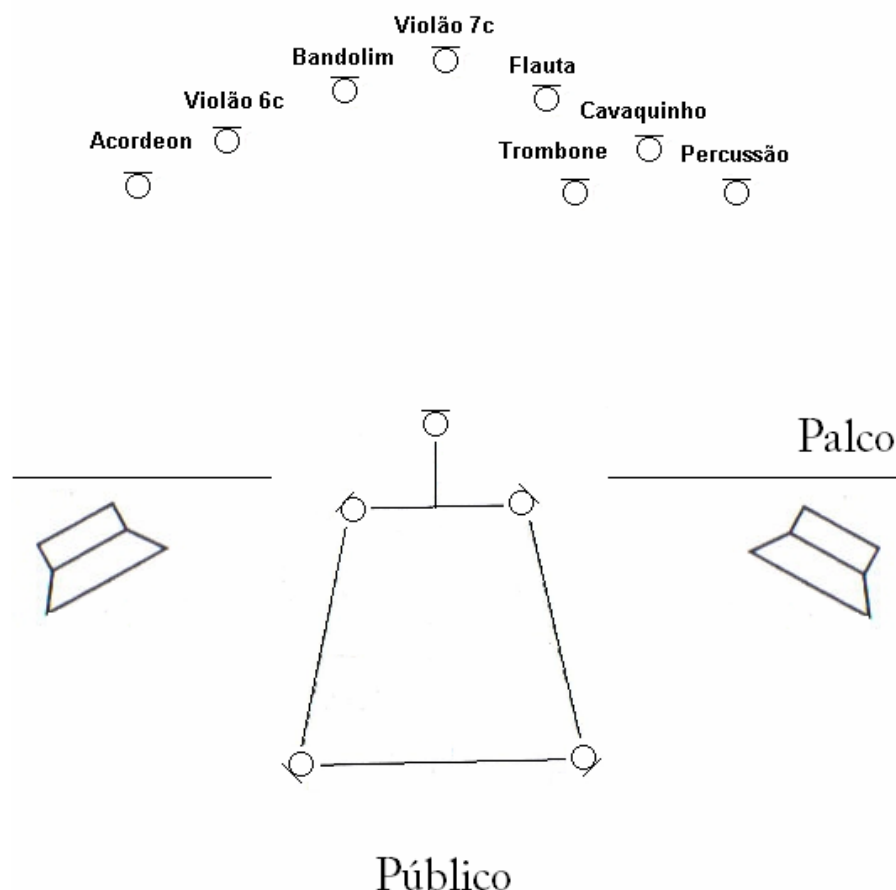


Representação da redistribuição simétrica frontal dos instrumentos da orquestra de câmara e dos cantores, adotada na mixagem da ópera reproduzida pelo sistema *surround* 5.1.

Uma outra gravação com a mesma abordagem estética, mas obtida através de outro processo, foi realizada no projeto Viva Música no dia 31 de março de 2010, na apresentação do grupo de choro “Chora o sete”. Nessa mixagem *surround* 5.1 foi feita a captação próxima de cada instrumento e a

captação da sala onde aconteceu o concerto com um arranjo de microfones no padrão INA, do qual cada sinal foi direcionado para um dos alto-falantes do sistema. Somado à imagem *surround* estabelecida pelo padrão de microfonação INA foi adicionado um pouco dos sinais da microfonação próxima para uma maior clareza e definição espacial dos instrumentos do grupo, com o panorâmico processado através do *plugin* Arc Panner. Não foi utilizado nenhum outro tipo de processamento para o resultado final. A estética adotada foi a do posicionamento estático do ponto de escuta no centro da sala de concerto, no centro do padrão de microfonação INA, com os instrumentos posicionados no palco em frente ao ouvinte e o público ao seu redor e mais acentuado na parte de trás, facilmente perceptível no momento dos aplausos.

Uma vez que alguns dos instrumentos estavam sendo amplificados por duas caixas direcionadas para o público, a captação do ambiente a partir do arranjo de microfones INA ficou um pouco prejudicada, o que pode ser observado claramente na audição. É interessante observar também o posicionamento dos instrumentistas em meia lua e como isso é refletido na reprodução a partir do sistema *surround* 5.1. Outra observação deve ser feita com relação ao posicionamento da platéia em relação ao ouvinte que, captada pelos microfones traseiros e reproduzida pelos alto-falantes *surround*, reproduz a sensação ao ouvinte de estar entre a platéia e o palco.



Situação do concerto e gravação do grupo de choro.

A gravação da peça “4 sketches em movimento” de Sérgio Freire, escrita para instrumentos de percussão, eletrônica pré-gravada e espacialização ao vivo de alguns deles, é um exemplo atípico e um dos principais alvos dessa pesquisa, uma vez que consiste em uma obra contemporânea onde o espaço é de extrema importância na sua performance ao vivo e na qual tentamos manter esse elemento no processo de gravação ao utilizar uma mídia de maior capacidade que o cd tradicional, capaz de armazenar o conteúdo dos diversos canais, e preparar o material para a reprodução em um sistema específico, no caso, o *surround* 5.1. Normalmente a obra é apresentada em concerto com quatro alto-falantes posicionados no formato do sistema quadrifônico, distribuídos nos 360° ao redor do público com separação de 90° entre eles, e

os instrumentos acústicos posicionados no palco em frente aos ouvintes e ao redor do instrumentista, atrás dos alto-falantes frontais.

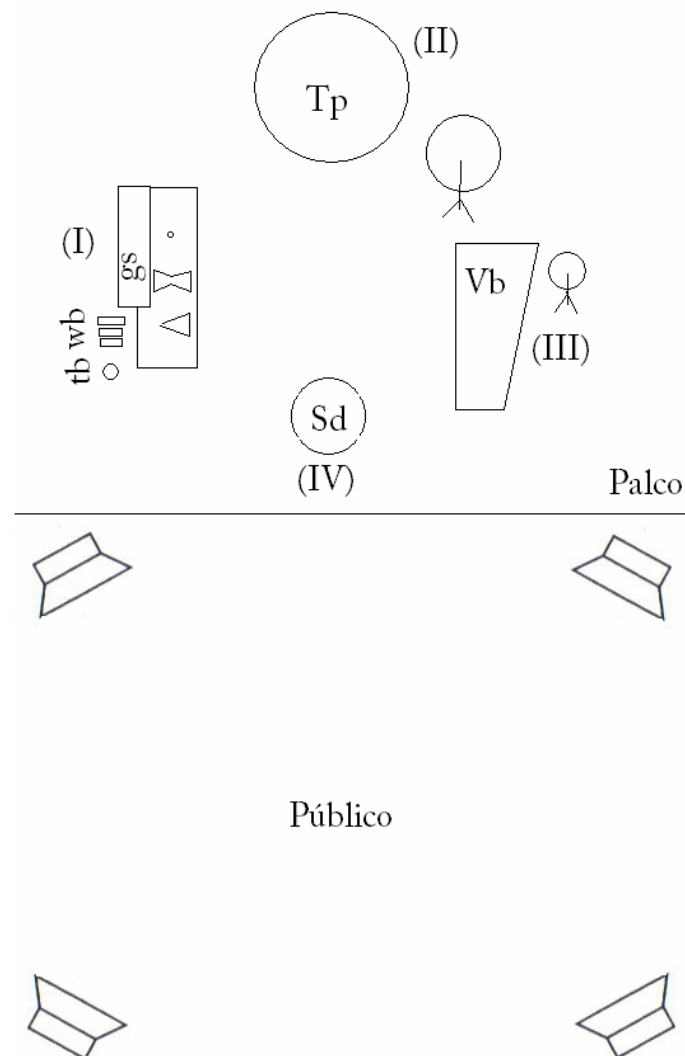
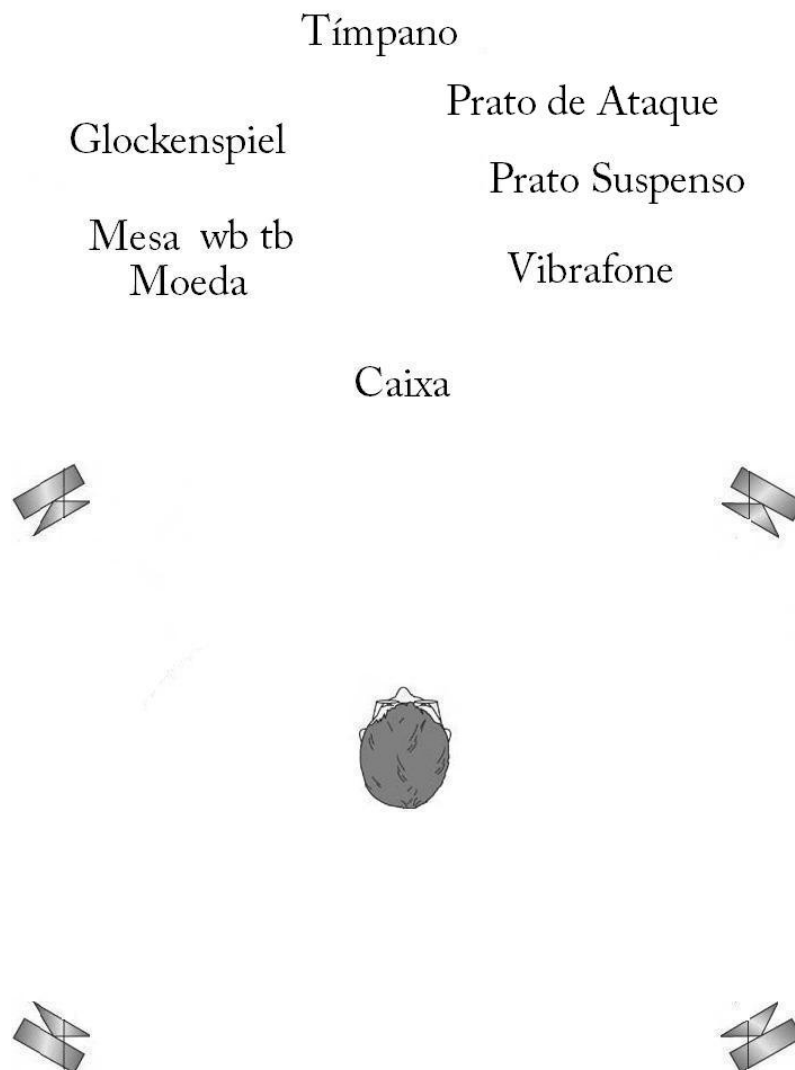


Ilustração da situação de concerto da peça "4 sketches em movimento". Fonte: SOUSA, 2010.

A peça foi gravada ao vivo em um concerto no dia 25-05-2009, realizado no auditório da Escola de Música da UFMG. O material utilizado na mixagem *surround* 5.1 foi o obtido da microfonação próxima de cada instrumento assim como do sinal processado e especializado ao vivo. A mixagem, assim como nas peças mencionadas anteriormente, foi realizada no software Digital Performer 7, com uma interface de áudio MOTU 828mk3 e caixas Yamaha MSP-5, respeitando-se a geometria da disposição padrão das caixas em estúdios 5.1. Nesse caso em especial optamos por estabelecer duas mixagens com estéticas diferentes, apesar de ambas se apresentarem estáticas.

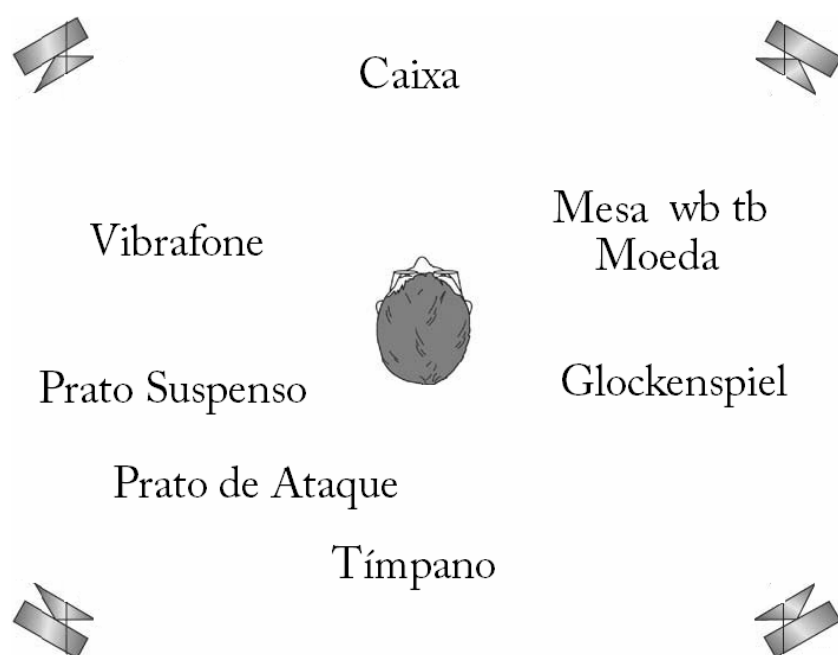
Na primeira delas os sinais de áudio gravados foram posicionados de acordo com o ponto de escuta do ouvinte com os quatro alto-falantes posicionados ao seu redor como no sistema quadrifônico, e o instrumentista posicionado em um palco a sua frente.



Estética adotada na primeira mixagem da peça. Fonte: SOUSA, 2010.

Na segunda mixagem, um pouco fora dos padrões tradicionais, adotamos o ponto de escuta do ouvinte no centro dos instrumentos e, simultaneamente, no centro dos quatro alto-falantes (nas mesmas posições do sistema quadrifônico). Essa estética, impossível em situações reais, se apresenta bastante interessante do ponto de vista da produção musical em estúdio. Segundo o próprio compositor, tal abordagem estética proporciona uma boa experiência, não mais a de uma sala de concerto que, apesar de

perder um pouco o caráter instrumental e criar uma certa confusão entre o material musical pré-gravado e o realizado ao vivo, apresenta um nível maior de imersão sonora. Para o instrumentista, essa situação se apresenta como um ideal, onde ele se encontra rodeado pelos sons gerados nos seus próprios instrumentos e, ao mesmo tempo, pelos sons da parte pré-gravada que, em situação de concerto, é direcionada ao público.



Estética adotada na segunda mixagem da peça. Fonte: SOUSA, 2010.

É importante observar nessa peça, por exemplo, o som do relógio que nos segundos iniciais da peça se movimenta ao redor do ouvinte, elemento importante a ser preservado tanto nas duas mixagens em *surround* 5.1 com estéticas diferentes como na sua transcrição para o sistema binaural, visto mais à frente.

Outra observação deve ser feita com relação ao som do tímpano no caso da segunda mixagem com o ponto de escuta centrado no instrumentista, que fica claramente prejudicado uma vez que suas componentes mais graves são direcionadas, a partir do sistema de gerenciamento dessa região espectral, para o canal de LFE, posicionado fisicamente na parte frontal do sistema *surround* 5.1 doméstico.

Outro exemplo dessa abordagem de mixagem de peças contemporâneas foi utilizada na peça “O Templo dos mil Sinos” composta pelo professor Fernando Rocha para uma instalação sonora, a ser apresentada por Laura Belém em Liverpool, na Inglaterra, ainda neste ano de 2010. Nessa obra, diversos sons pré-espacializados através de processamento no MAX/MSP, inicialmente para um sistema quadrifônico, foram mixados em sistema *surround* 5.1 obedecendo seu endereçamento para os alto-falantes esquerdo e direito frontais e traseiros, posicionando a voz isoladamente no alto-falante central e realizando uma filtragem em todos esses cinco canais para a obtenção do sinal endereçado para o canal LFE.

Nessa peça também observamos os sinais graves prejudicados pelo posicionamento físico do *subwoofer*, e como exemplo disso podemos mencionar o som do navio que inicialmente deveria ser apresentado se movimentando ao redor do ouvinte, mas fica limitado à parte frontal do sistema, devido à sua composição espectral grave.

3.3 – Transcrevendo para o sistema binaural

Uma vez que temos o formato 5.1 como um formato já estabelecido e estruturado em diversos estúdios pela indústria cinematográfica, tomamos uma mixagem realizada para o mesmo como ponto de partida. Na transcrição desse material mixado em um sistema multicanal para um sistema binaural, tomamos como fontes primárias o sinal proveniente dos próprios alto-falantes, originalmente fontes secundárias, segundo FARIA (2005), para a realização do processamento de transcrição para sistema binaural. Essa adoção das fontes originalmente secundárias (alto-falantes) como fontes sonoras primárias se torna necessário e extremamente viável, uma vez que as ferramentas para a mixagem em sistemas *surround* 5.1 já estão mais desenvolvidas, disseminadas e dominadas pelos profissionais da área.

Esse processo se assemelha muito com o que RUMSEY (2001) define como *Virtual Home-Theaters*, onde os alto-falantes se tornam fontes sonoras virtuais reproduzidas através do fone de ouvido. Tal processo pode ser reproduzido também através de apenas dois alto-falantes, no entanto a proposta de fazê-lo direcionado para fones de ouvido segue observações

realizadas em KALLINEN e RAVAJA (2004) a respeito da maior imersão sonora e envolvimento provocados pelos mesmos.

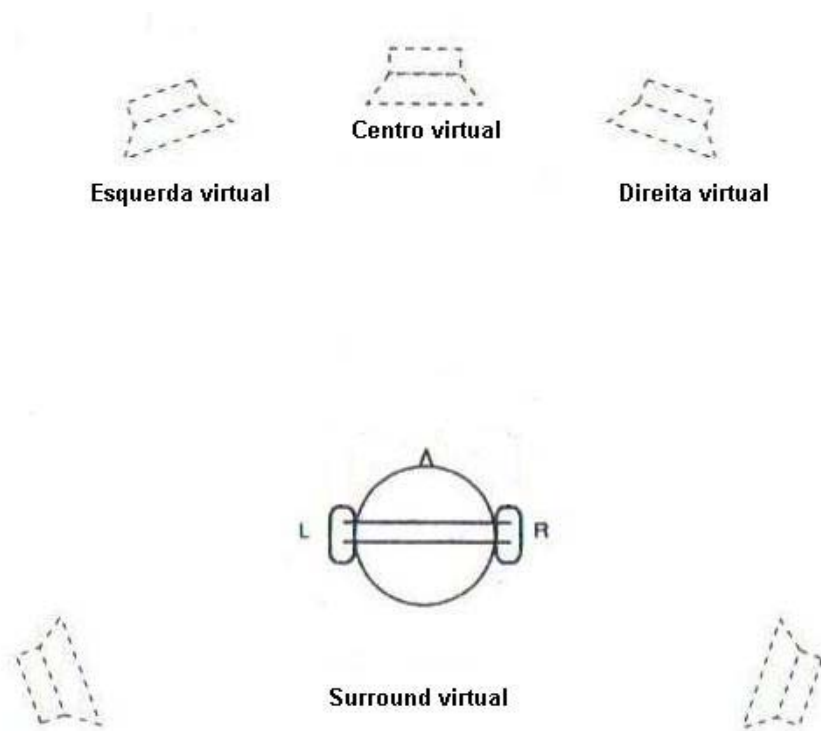


Ilustração representando o Home-Theater Virtual reproduzido por fones de ouvido, possibilidade do processamento digital de áudio a partir de HRTFs.

KALLINEN e RAVAJA (2004) afirmam que, cada vez mais, a escuta através de fones de ouvido se torna mais popular, seja ela através de tocadores de CD, telefones portáteis ou computadores de bolso. Uma observação interessante realizada pelos autores supracitados diz respeito ao fato de que na audição através de fones de ouvido o objeto, ou seja, a música ou áudio ouvido, se torna mais próximo do ouvinte e os ruídos externos a essa integração são muito atenuados. Essas condições que isolam o ouvinte do mundo exterior, possibilitadas pelos fones de ouvido, segundo os autores, já foram mencionados em outros trabalhos e descritos como um facilitador para a imersão sonora e o envolvimento, o que leva também a uma maior atenção do ouvinte à mídia e a uma experiência de escuta mais positiva.

As pesquisas de KALLINEN e RAVAJA (2004), apesar de estarem inicialmente direcionadas à escuta de notícias e a absorção de conteúdo informativo pelos ouvintes através dos dois meios de reprodução (alto-falantes

e fones de ouvido) podem ser aplicadas ao presente trabalho, uma vez que possui conclusões importantes tanto a respeito das características de cada um desses meios de escuta quanto ao que diz respeito à preferência dos ouvintes.

Segundo eles existe uma preferência de 60% dos ouvintes à escuta através de fones de ouvido, apesar da maioria considerar a escuta através de alto-falantes mais realística. A maioria dos ouvintes envolvidos na pesquisa dos autores supracitados também afirma a experiência auditiva proveniente de fones de ouvido ser mais positiva, mais atenciosa e, por isolar o ouvinte dos ruídos externos, mais prazerosa. Apesar dessas conclusões os autores afirmam que a preferência do ouvinte com relação aos alto-falantes ou fones de ouvido é dependente da aplicação e do ambiente onde os mesmos se situam.

A opção de se criar um tipo de *Virtual Home-Theater* através de fones de ouvido, especificamente, parte também da necessidade de um sistema com baixo consumo de processamento, que seria maior no caso de uma reprodução através de alto-falantes (ambiosonics ou transaural). Outra possibilidade dessa transcrição para sistemas binaurais é a monitoração de uma produção para *surround* 5.1 através de fones de ouvido em estúdios onde os meios físicos, relacionados à acústica ou aos equipamentos, não a possibilitam. Segundo RUMSEY (2001), o desenvolvimento de um *Virtual Home-Theater* se dá, normalmente, através de métodos tradicionais a partir da diferença de intensidades no caso dos canais central, esquerdo e direito, e os canais traseiros são processados por HRTFs correspondentes aos ângulos de 110° e -110° a partir do eixo central. Segundo ele, o maior desafio de um sistema desse tipo é a externalização das fontes sonoras, ou seja, fazer com que elas sejam percebidas externamente à cabeça do ouvinte nas laterais e à frente, principalmente.

SO, R. H. Y. et al (2006) mencionam que uma das primeiras aplicações das HRTFs (funções de transferência relativas à cabeça) a alcançar o consumidor de eletrodomésticos foi realizada a partir da implementação do que eles chamam de sistema *surround* virtual baseado em fones de ouvido (VHSS - *Virtual Headphone-based Surround-system*). Segundo os autores supracitados, devido ao fato de o *mixdown* (processo de redução do áudio multicanal *surround* 5.1 para apenas dois canais – Dolby stereo) remover consideravelmente as informações referentes ao posicionamento de fontes

sonoras frontal-traseiras e degradar muito os efeitos *surround*, uma possibilidade surge a partir do desenvolvimento de um VHSS. Eles afirmam que para criar um sistema *surround* baseado em fones de ouvido:

“(...) cada um dos cinco canais do som *surround* é processado por um par de HRTFs apropriadas de forma que eles são transformados em cinco canais de som binaural que carregam conteúdo sonoro com suas respectivas informações direcionais. Quando os cinco canais de som binaural e o canal de efeitos de frequências graves (VLF) são reproduzidos através de um fone de ouvido, as informações direcionais embutidas no sinal binaural podem simular um efeito de som *surround*.” (SO, R. H. Y.; et al. 2006. p.697)

Diversas questões surgem no momento dessa transcrição do sistema *surround* 5.1 para um sistema binaural. Entre elas estão a definição de qual biblioteca de HRTFs utilizar, uma vez que temos a disposição ou podemos desenvolver bibliotecas de HRTFs gravadas individualmente para cada indivíduo, generalizadas – média entre diversos indivíduos, gravada a partir de um manequim, ou até mesmo obtidos a partir de modelos gráficos virtuais. Com essa diversidade, uma forma mais simples, inclusive sugerida por SO, R. H. Y. et al (2006), é a de se utilizar o banco de HRTFs gravado a partir do manequim KEMAR (*Knowles electronics manikin for acoustic research*) de Gardner e Martin, em 1995, no Instituto de Tecnologia de Massachusetts (MIT), mais acessível e disponível no objeto spat~ desenvolvido para a plataforma MAX, como veremos a seguir.

Com relação aos fones de ouvido tanto para a monitoração quanto para a escuta do ouvinte, podemos dividir em dois tipos básicos de fones de ouvido: os extra-auriculares (mais comuns) e os intra-auriculares (mais conhecidos como *in-ears*). A diferença entre eles está na sua proximidade ao ouvido interno, sendo que o som reproduzido pelos fones extra-auriculares ainda estão sujeitos a uma certa filtragem realizada pela orelha do ouvinte, enquanto os intra-auriculares reproduzem o som quase que diretamente no ouvido interno.

Esse ponto pode ter implicações na escuta no que diz respeito à melhor percepção direcional, no entanto, segundo experiência própria, está mais relacionado à preferência de cada ouvinte, podendo acontecer de alguns se adaptarem bem a fones intra-auriculares ou se sentiram mais confortáveis com

os extra-auriculares, assim como ocorre a adaptação às HRTFs. Com relação à resposta em frequência dos fones de ouvido, esta deve ser o mais linear possível e completamente ausente de equalização, uma vez que a percepção de direção espacial está diretamente ligada ao domínio da frequência.

Essa questão relativa aos fones de ouvido nos leva também à reflexão a respeito da nossa própria percepção auditiva que, como mencionada anteriormente, não é focada apenas nos ouvidos, mas percebida pelo corpo como um todo e principalmente também através da cabeça e caixa torácica⁶⁰. Um problema que surge com relação ao espectro sonoro e observado no processo de transcrição do sistema *surround* 5.1 para o sistema binaural diz respeito ao posicionamento espacial do canal destinado aos efeitos de baixa frequência que possuem natureza omnidirecional e não costumam ser bem reproduzidos através de fones de ouvido, principalmente por estes não atuarem sobre o corpo do ouvinte como um todo e sim somente sobre o ouvido, eliminando a possibilidade de percepção sonora pelo torso, por exemplo. SO, R. H. Y. et al (2006) sugerem que esse canal seja somado na sua forma original – sem processamento – aos outros cinco canais binaurais já processados, uma vez que a região grave do espectro sonoro não possui informações referentes a direcionalidade⁶¹.

Uma outra questão muito importante diz respeito à forma como o movimento do ouvinte é tratado na reprodução por um sistema binaural, uma vez que esse artifício é de extrema importância para a nossa percepção espacial⁶². Alguns autores, a exemplo de HORBACH (1999), apresentam tal preocupação e propõem formas de rastreamento do movimento da cabeça do ouvinte para que as HRTFs se modifiquem em tempo real e recriem essa nossa ferramenta natural de percepção espacial.

Um exemplo de sistema comercialmente desenvolvido com implementação de rastreamento de movimento que merece ser citado é o *Headzone* da *Beyerdynamics* que, segundo o próprio fabricante, é o primeiro sistema móvel de fones de ouvido que permite a monitoração de um sistema *surround* 5.1 virtual em qualquer lugar. O sistema se apresenta como um

⁶⁰ HAMIL e PRICE, 2008.

⁶¹ BLAUERT, 1997.

⁶² KEDALL, 1995.

sistema de monitoração portátil com reprodução 5.1 virtual que cria uma sala de controle com ambiente simulado através de *software* e possui um sistema de rastreamento de cabeça que fixa a localização do ouvinte na sala e compensa qualquer modificação relacionada ao seu movimento.



Sistema *Headzone* da *Beyerdynamics*, primeiro sistema de monitoração *surround* 5.1 através de fones de ouvido e com rastreamento de movimento distribuído no mercado. Fonte: *Beyerdynamics*.



Sistema *Headzone Pro XT* da *Beyerdynamics*, sistema de monitoração *surround* 5.1 através de fones de ouvido com rastreamento de movimento. Fonte: *Beyerdynamics*.

O presente trabalho registra uma tentativa de rastreamento de movimento a partir da utilização do *Wii Remote Motion Plus*, da Nintendo, que envia sinais correspondentes à sua rotação nos três eixos para o MAX/MSP através de *Bluetooth* e atualiza as HRTFs utilizadas no processamento binaural, mantendo o sistema 5.1 estático enquanto o ouvinte movimenta sua cabeça. Tal etapa do processamento pretende dar maior realismo à experiência binaural, uma vez que a movimentação se mostra indispensável na nossa percepção de espaço, principalmente no que diz respeito à resolução de

problemas existentes no denominado “cone de confusão”, onde sons provenientes de trás podem ser confundidos facilmente com sons dianteiros devido aos valores iguais de ITD e ILD⁶³.

Uma outra ferramenta importante na recriação do espaço sonoro que deve ser considerada é a inclusão de conteúdo visual na produção, uma vez que, assim como quando ouvimos determinado som no espaço, tentamos encontrá-lo visualmente, a visualização de um objeto associado a determinado som pode auxiliar na sua percepção sonora. Essa ferramenta pode ser uma facilitadora muito interessante no que diz respeito à identificação de fontes frontais, principalmente.

Depois da mixagem em *surround* 5.1 finalizada, a transcrição desse material espacializado nos 360° ao redor do ouvinte para dois canais no formato binaural, para serem ouvidos através de fones de ouvido, objetiva manter a espacialização da obra, mesmo em dois canais, e foi realizada através de duas ferramentas: a primeira delas foi o objeto spat~ do MAX/MSP para as suas versões 4 e 5, e o plugin VST KLT-3D desenvolvido pelo Grupo de Processamento de Áudio da Universidade Federal do Rio de Janeiro, sob orientação de Luiz Wagner Biscainho.

Nessa etapa da pesquisa nos propusemos também a pegar trechos de obras cinematográficas e musicais que foram lançadas no formato 5.1 e analisadas no segundo capítulo, para comparação com sua transcrição para um sistema binaural, verificando principalmente a compatibilidade do espaço representado por ambas.

Vale mencionar aqui uma necessidade que surgiu desse processo de transcrição. Uma vez que escutamos e realizamos uma mixagem em *surround* 5.1 em uma sala típica doméstica ou em um estúdio que, por menores que sejam, não estão isentos de reflexões, surge a necessidade de se adicionar certa reverberação também no sinal direcionado para os fones de ouvido. Essa reverberação adicional se torna fundamental para a sensação de externalização das fontes sonoras através desse meio de reprodução.

⁶³ KENDALL, 1995.

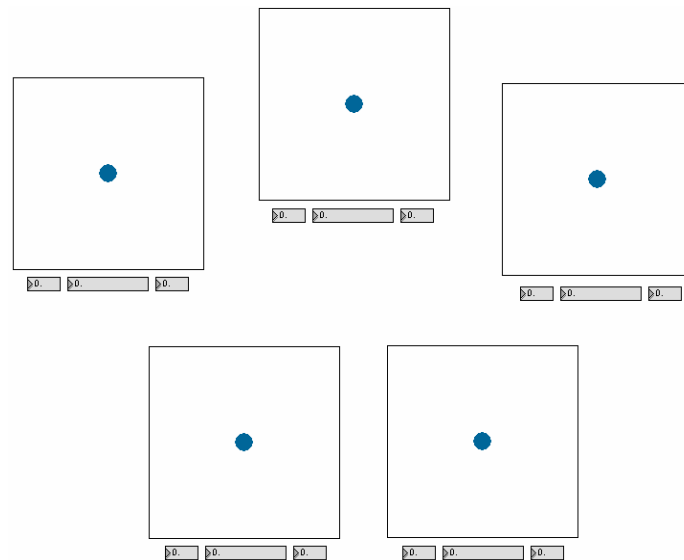
3.3.1 – O objeto spat~ do MAX/MSP e o KLT3D – transcrevendo o áudio 5.1 para outros formatos

Segundo JOT e WARUSFEL (1995), o objeto spat~ é um software de processamento em tempo real desenvolvido pelo IRCAM para funcionamento dentro do ambiente gráfico MAX de processamento de sinais. Ele nos oferece uma biblioteca de módulos elementares que são integráveis e direcionados para a manipulação do posicionamento de fontes sonoras e de caracterização de ambientes acústicos, e pode ser configurável para diversos formatos multicanais de reprodução de áudio através de alto-falantes ou fones de ouvido.

O objetivo central do desenvolvimento do objeto em questão foi proporcionar aos compositores, intérpretes e engenheiros de som o controle da difusão sonora em espaços reais ou virtuais. Ele recebe o sinal proveniente de uma fonte sonora (inicialmente isenta de qualquer tipo de reverberação), adiciona espacialização em tempo real e libera um sinal específico para o meio de reprodução eletroacústico definido a priori, seja ele um arranjo de alto-falantes ou fones de ouvido.

No *patch*⁶⁴ que desenvolvemos inicialmente no MAX/MSP 4.6, a posição de cada fonte sonora (alto-falante) era definida a partir de uma interface gráfica simples, onde os valores definidos em um plano x-y correspondiam ao azimute e distância. Tais parâmetros, quando manipulados, nos davam resposta em tempo real. Utilizando o Digital Performer 7.0 (*software* de gravação multipista) enviando o sinal mixado em *surround* 5.1 para o MAX/MSP através de um driver interno (*Soundflower*) temos a possibilidade de monitorar a mixagem *surround* 5.1, já no formato binaural, através de fones de ouvido.

⁶⁴ *Patch*: Programa do MAX constituído de objetos ligados entre si, na interface gráfica, através de linhas que permitem a troca de mensagens ou dados entre os mesmos, uma alusão às conexões físicas que interligavam os módulos dos sintetizadores analógicos. WINKLER, T. (2001).



Interface gráfica desenvolvida no MAX/MSP. Fonte: Sousa, 2010.

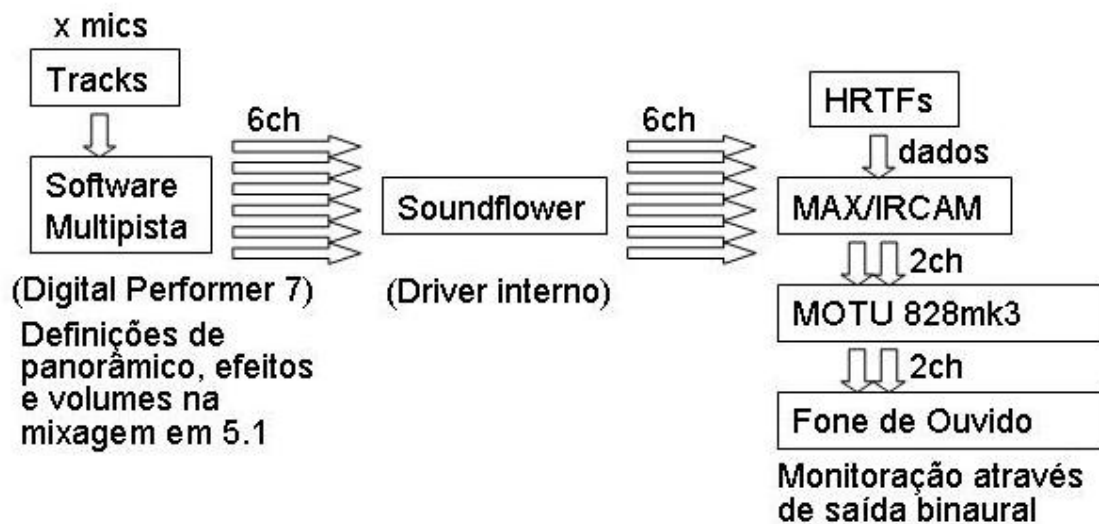


Diagrama de blocos ilustrando o fluxo de sinais estabelecido no sistema descrito acima.

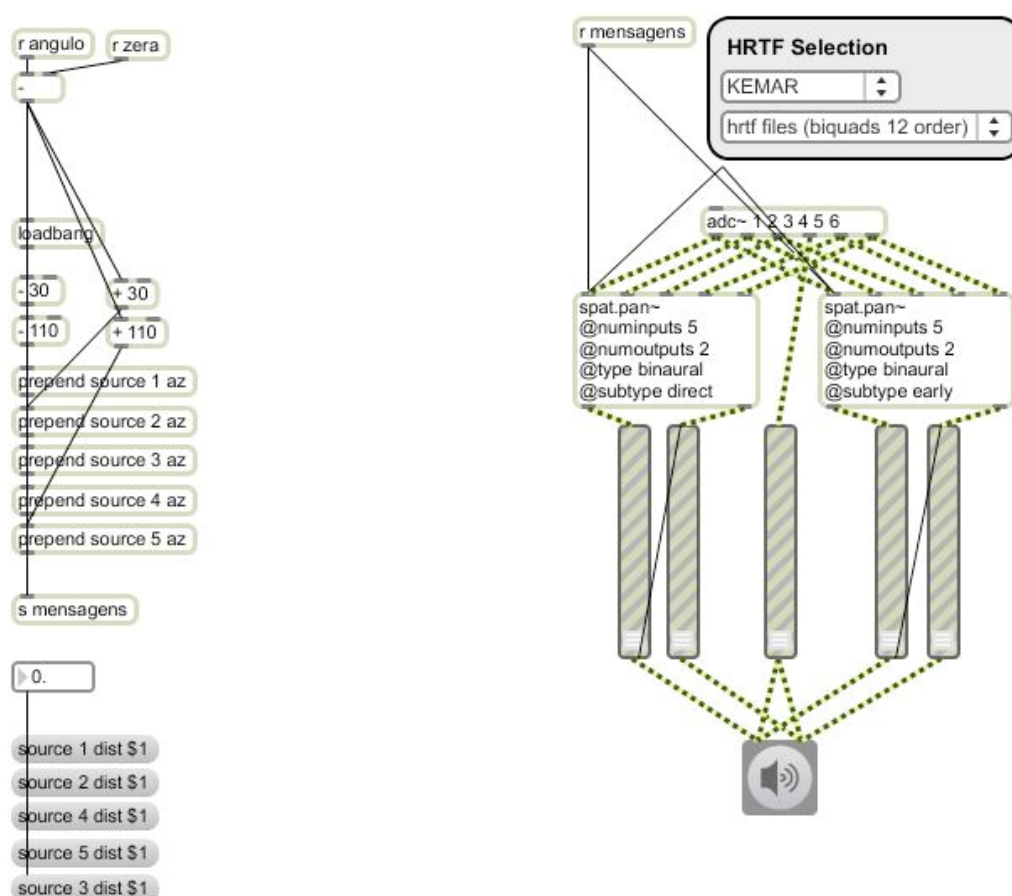
Após o processamento, observamos que a externalização das fontes sonoras é bastante prejudicada na reprodução através de fones de ouvido; por isso uma etapa de adição de um reverberador antes do envio para os fones foi muito bem vinda. Um fator importante observado foi que o exagero e, principalmente a qualidade do reverberador utilizado tem grande influencia nessa percepção de externalização das fontes sonoras. Considerando que, quando ouvimos qualquer reprodução em um sistema *surround* 5.1, seja ele doméstico ou em ambiente de estúdio, o sinal que chega aos nossos ouvidos já

possui certa reverberação do ambiente em que estamos, a adição de uma etapa de reverberação extra se torna bastante coerente.

O nível de reverberação adicionada nessa etapa tem estreita relação também ao tipo de material musical em questão, sendo que quando o material possui ataques rápidos ou caráter transiente a reverberação pode auxiliar bastante; no entanto quando o material apresenta ataques pouco definidos ou caráter mais contínuo o excesso da mesma acaba prejudicado a localização das fontes sonoras.

No MAX/MSP 5 e com disponibilização de uma nova versão do objeto *spat~* pelo IRCAM, o *patch* desenvolvido ficou bem mais simples, uma vez que definimos a distância para todos os alto-falantes com uma única mensagem (aproximadamente entre 1 e 2 metros – correspondente à distância dos alto-falantes nos sistemas domésticos) e o azimuth correspondente à cada um dos alto-falantes com um número, correspondente ao ângulo do círculo do plano horizontal ao redor do ouvinte, e definido para o alto-falante central em 0, para o direito em 30, o esquerdo em -30, o traseiro direito em 110 e o traseiro esquerdo em -110 graus.

O sinal correspondente ao ângulo foi estabelecido inicialmente como sendo zero e, em uma tentativa de rastreamento do movimento do ouvinte, como se verá a seguir, é proveniente de um outro *patch* dedicado a receber o sinal do *wii remote* (da Nintendo) e define o eixo de todo o sistema e, através de operações simples, os ângulos correspondentes aos outros alto-falantes. O sinal recebido em “r zero” corresponde a um controle de calibração que define o valor recebido do *wii* como eixo central (ângulo zero).

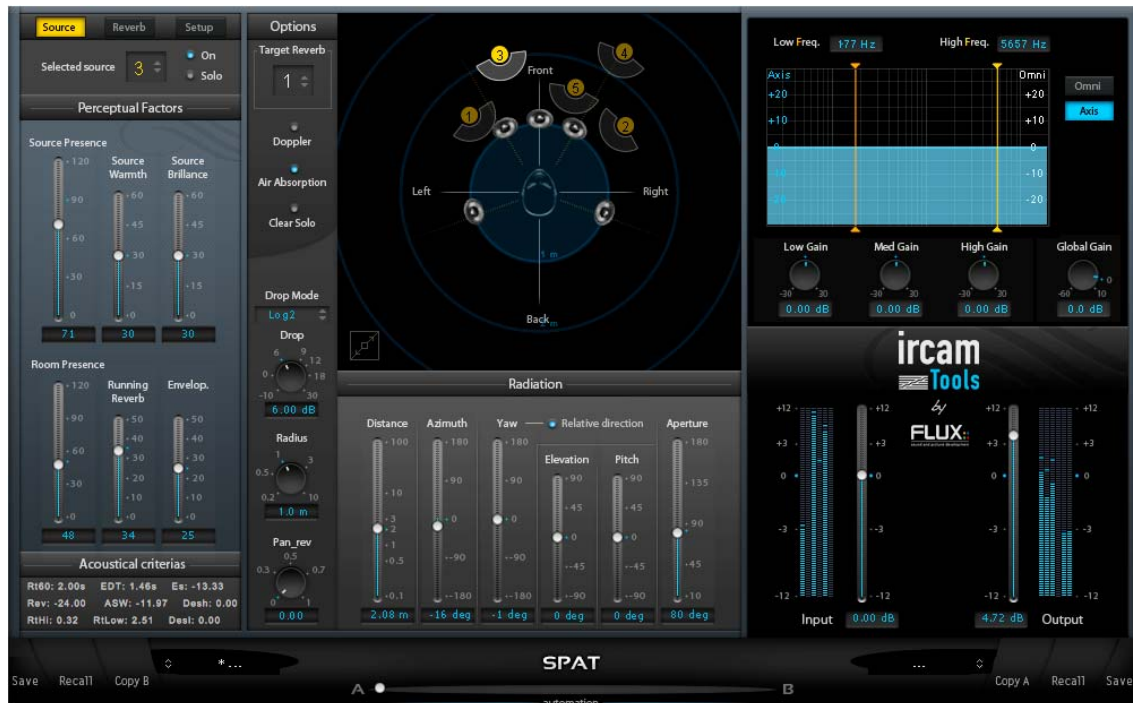


Patch desenvolvido no MAX/MSP para o processamento binaural recebendo os seis sinais de uma mixagem em *surround* 5.1 e já com a implementação do rastreamento de movimento da cabeça.

Na presente pesquisa, utilizamos o objeto descrito acima somente com o propósito de processar o sinal direcionado para cada um dos cinco alto-falantes do sistema *surround* 5.1 para seu correspondente em um sistema binaural para fones de ouvido, processando cada um dos cinco canais com HRTFs correspondentes à posição do alto-falante ao qual era destinado originalmente. O canal de LFE foi reproduzido sem processamento para ambos os lados do fone de ouvido, segundo observações de SO, R. H. Y. et al (2006). Além dessa aplicação, o objeto `spat~` possui diversas outras funcionalidades que não serão abordadas aqui, entre elas a espacialização para sistemas em ambisonics, transaural (ambisonics) e simulação de salas.

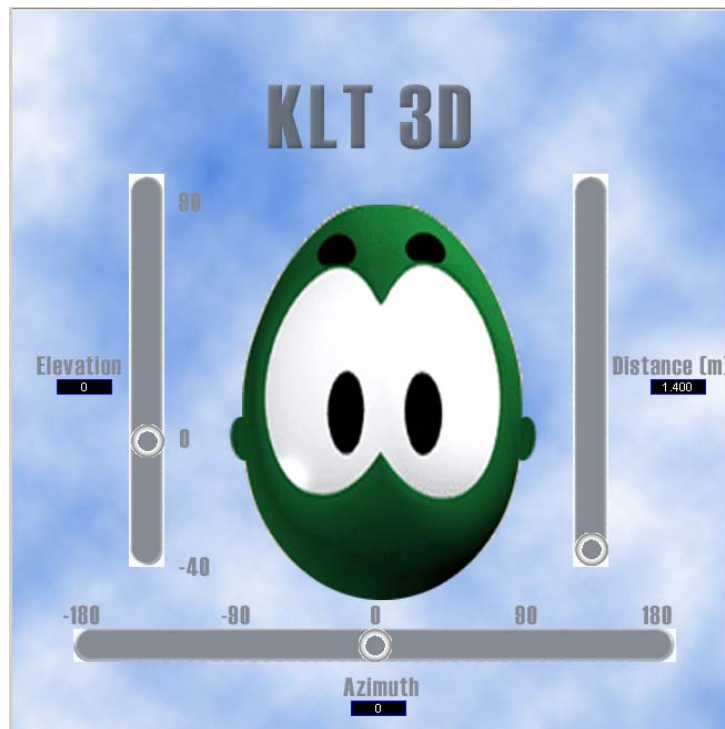
Uma opção mais simples do que a utilização do objeto descrito acima na plataforma MAX/MSP seria um *plugin* VST para utilização em qualquer software (como o proposto pelo FLUX IRCAM Tools). Recentemente, a empresa de *plugins* FLUX parece ter comprado os direitos do objeto

mencionado acima e prevê um lançamento ainda este ano dessa ferramenta, com o nome de IRCAM Tools, em um formato mais amigável para utilização em qualquer *software* multipista.



Plugin IRCAM Tools da fabricante Flux: Fonte:www.ircamtools.com

Nessa linha de desenvolvimento de *plugins* não podemos deixar de mencionar a pesquisa realizada no Grupo de Processamento de Áudio da UFRJ, por Alan Tigel e Luiz Wagner Biscainho, apresentada no 7º Congresso da AES Brasil em 2009, que deu origem ao KLT3D.



Plugin KLT3D, desenvolvido pelo Grupo de Processamento de Áudio da UFRJ. Fonte: *plugin KLT3D*.

Após a mixagem em *surround* 5.1 e obtenção dos cinco canais individuais, correspondentes a cada um dos alto-falantes, criamos no software multipista uma nova sessão e inserimos o *plugin* configurado para o processamento correspondente a cada uma das fontes (alto-falantes) em seu respectivo canal. Esse processo facilita inclusive o tratamento da reverberação mencionado anteriormente, quando se objetiva um produto final em sistema binaural.

Utilizando este *plugin*, o presente autor não pôde deixar de observar uma certa coloração tonal no conteúdo musical processado, sendo a região aguda do espectro sonoro bastante favorecida. Essa observação nos leva também a refletir que grande parte do conteúdo correspondente à direcionalidade de uma fonte sonora se encontra nessa região e que, favorecendo a mesma, o processamento pode realizar sua função com maior êxito, apesar de diminuir a fidelidade da mesma com a mixagem original.

Vale observar aqui que a possibilidade de se processar cada fonte individualmente, independentemente da mixagem *surround* é possível e, além de muito interessante, provavelmente mais bem sucedida quando se objetiva apenas uma mixagem para um sistema binaural. Prova disso é que a proposta

do presente trabalho de utilizar uma mixagem já finalizada em *surround* 5.1 encontra diversas barreiras, entre elas a proporção de reverberação entre os dois sistemas e a correspondência tímbrica entre ambos.

No processo de rastreamento da movimentação do ouvinte que utiliza os fones de ouvido e da modificação das HRTFs aplicadas ao sinal processado, a manipulação ocorre no nível dos alto-falantes tomados inicialmente como fontes sonoras primárias.

Como já mencionado, este trabalho faz uma tentativa de rastreamento de movimento da cabeça do ouvinte através da utilização do *wii remote* (da Nintendo) com o *motion plus*, que envia um sinal correspondente à movimentação angular da cabeça do ouvinte para atualizar o grupo de HRTFs utilizada no processamento. No *patch*, o valor 0 (zero) direcionado à HRTF correspondente ao canal central é atualizada pelo sinal do wii e somado aos valores correspondentes às outras caixas (-30, 30, -110 e 110). Um botão do próprio controlador, quando acionado pelo ouvinte, estabelece a posição 0 (zero), que deve ser atualizada sempre que o ouvinte está olhando diretamente para a frente, necessidade observada a partir da instabilidade do sensor.



Usuário utilizando *Wii Remote* adaptado para o rastreamento dos movimentos de sua cabeça.

3.3.2 – Observações a partir da escuta

Essa parte do presente trabalho descreve algumas observações significativas a respeito do processo de transcrição para o sistema binaural sem a pretensão de ser definitiva e com a consciência de não ter estabelecido uma metodologia mais sistemática para avaliar a percepção dos colegas e professores participantes, o que poderia ser obtido a partir de resultados quantitativos e sistematizados.

Na transcrição do material mixado em *surround* 5.1 para o sistema binaural, como esperado, devido principalmente ao meio de reprodução (fones de ouvido) e às diversas diferenças já mencionadas anteriormente, houveram diferenças significativas; no entanto tal processo não deixa de ser extremamente interessante e de importante utilidade, necessitando de aprimoramentos futuros e algumas adaptações.

Quando ouvimos algumas das transcrições do material obtido a partir dos trechos das obras cinematográficas, observamos algumas características muito bem representadas no sistema binaural. Como exemplo disso podemos citar o movimento das bolas do jogo quadribol no trecho do filme “Harry Potter e a Pedra Filosofal” (2001), mencionado no segundo capítulo. Apesar de não serem exatamente compatíveis com a reprodução do sistema *surround* 5.1, a movimentação das bolas no sistema binaural não deixa de ser interessante. Uma observação feita por alguns colegas que ouviram ambas as reproduções diz respeito à extrema dificuldade de localizar as vozes na parte frontal, assim como qualquer material sonoro posicionado nessa mesma posição. Tal observação pode ser justificada pela não visualização da imagem representada pelo som em questão e pode ser parcialmente solucionada a partir da exibição da mesma.

As peças que possuem a estética de representar um palco em frente ao ouvinte e a reverberação da sala nos alto-falantes traseiros, seja através de processamento ou de uma microfonação específica para o sistema, apresentaram uma certa dificuldade de externalização das fontes frontais, que se misturam um pouco com as traseiras e muitas vezes causam até um certo conflito, resultando em uma sonoridade pouco definida. Nesse ponto, a quantidade de reverberação aplicada na mixagem fica difícil de ser controlada,

pois ao mesmo tempo em que o exagero de reverberação pode ajudar na externalização das fontes sonoras, ela acaba diminuindo a definição das mesmas, fato que pode estar relacionado também ao material musical dessas peças, como mencionado anteriormente. Esse efeito relativamente problemático e de difícil resolução foi observado na abertura da ópera do evento de música antiga e na apresentação do grupo de choro.

Já nas peças “4 sketches em movimento” e “O Templo dos mil Sinos”, provavelmente devido ao material musical adotado por elas, esse efeito do excesso de reverberação resultando em pouca definição direcional das fontes sonoras é bem reduzido e os efeitos de movimentação observados no sistema *surround* 5.1 são reproduzidos muito bem, e também a externalização das fontes sonoras foi melhor resolvida. Nessas peças, mixadas inicialmente em *surround* 5.1, nos momentos em que elas perderam movimentação da região grave do espectro sonoro devido ao sistema de gerenciamento de graves (direcionados ao *subwoofer*), observamos certa atenuação desse problema, na transcrição para o sistema binaural, quando direcionamos o canal LFE para ambos os lados do fone de ouvido sem processamento; no entanto, a sensação corporal que decorre da reprodução dessa região ainda assim não pode ser reproduzida.

Conclusões e futuras pesquisas

Com certeza, ainda existe muito trabalho a se fazer para que uma reprodução completamente tridimensional através de fones de ouvido seja efetivamente realizada; no entanto, a transcrição sugerida, utilizando as ferramentas citadas, já mostra uma certa superioridade desse processamento com relação ao estéreo tradicional, principalmente com relação àqueles obtidos com uma redução de um material originalmente quadrifônico. A possibilidade de distribuição e divulgação de produções pensadas para reprodução em sistemas com mais de dois alto-falantes e que possuem o espaço sonoro como material musical essencial têm, através do processamento proposto, uma importante possibilidade a ser considerada.

Considerando inicialmente a movimentação como elemento essencial para a percepção do espaço sonoro, realizamos uma tentativa de rastreamento de movimento que ainda precisa muito ser aprimorada, talvez em pesquisas futuras, principalmente com relação ao sensor de movimento utilizado que deve ser menor e preferencialmente integrado ao fone de ouvido. Por outro lado já se coloca a questão sua real indispensabilidade nesse processo uma vez que a captação de movimento demanda mais poder de processamento, além de *softwares* e *hardwares* específicos. Um questionamento a respeito do custo / benefício de um sistema desses deve ser feito pela produção, uma vez que o mesmo se afasta do objetivo de utilização de meios mais acessíveis para a divulgação e distribuição de materiais até então praticamente intangíveis.

Com relação a pesquisas futuras também, devemos considerar a possibilidade de se utilizar o mesmo processamento para outros formatos, tais como o ambisonics (em suas diversas configurações de saída), a *wave field synthesis*, o quadrifônico ou até mesmo o estéreo tradicional, neste último caso com a possibilidade de externalização dos dois alto-falantes quando da reprodução através de fones de ouvido. A comparação entre uma mixagem binaural que utilize fontes primárias com outra que utilize fontes secundárias (alto-falantes) também parece uma boa proposta.

A dependência de um reverberador com qualidade direciona a pesquisa para a utilização de softwares baseados em convolução e talvez até para a

criação de respostas a impulso que já considerem o posicionamento da fonte sonora e já carreguem essas informações para serem aplicadas em sistemas binaurais. Fica a questão de que talvez um único processamento - considerando reverberação e HRTFs específicas - seja mais econômico com relação ao processamento e ofereça resultados mais interessantes e realísticos.

Bibliografia

Livros

ARNHEIM, Rudolf. **History of Broadcasting: Radio to Television.** New Hampshire: Ayer Company Publishers Inc., 1986. 296p.

BEGAULT, D. R. **3-D Sound for Virtual Reality and Multimedia.** London: Academic Press. 1994. 293p.

BLAUERT, J. **Spatial Hearing: the Psychophysics of Human Sound Localization.** London, England: MIT Press, 1997. 494p.

BREGMAN, Albert S. **Auditory Scene Analysis: The Perceptual Organization of Sound.** Cambridge: MIT Press, 1990. 773p.

CHION, Michel. **Audio-vision: sound on screen.** Trad. Claudia Gorbman – New York: Columbia University Press, 1994. 239p.

EARGLE, John. **The Microphone Book.** Oxford: Focal Press, 2005. 377p.

HAMILL, Teri A. and PRICE, Lloyd L. **The Hearing Sciences.** United Kingdom: Plural Publishing, 2008. 447p.

HOLMAN, Tomlinson. **Surround Sound Up and Running.** 2nd edition. Oxford: Focal Press, 2008. 253p.

MAXFIELD, J. P. and HARRISSON, H. C. **Methods of High Quality Recording and Reproducing of Music and Speech Based on Telephone Research.** Bell Telephone Laboratories Incorporated, 1926. 32p.

NEWELL, Philip. **Recording Studio Design.** Oxford: Focal Press, 2003. 711p.

RUMSEY, Francis. **Spatial Audio**. Oxford: Focal Press, 2001. 240p.

SCHAFER, R. Murray **A afinação do mundo**. Trad. Marisa Trench Fonterrada. São Paulo: Editora UNESP, 2001. 381p.

WINKLER, T. **Composing Interactive Music. Techniques and Ideas Using Max**. Cambridge, Massachusetts: MIT Press, 2001. 350p.

Capítulo de Livros

STERN, Richard M., BROWN, Guy J. and WANG, Deliang. In: Computational Auditory Scene Analysis – Principles, Algorithms and Applications. **Binaural Sound Localization**. New Jersey: IEEE Press Wiley Interscience, 2006.

PLACK, Christopher J. In: The Sense of Hearing. **Spatial Hearing**. England: Lawrence Erlbaum Associates, 2005.

Teses de Doutorado

FARIA, Regis Rossi Alves **Auralização em ambientes audiovisuais imersivos**. 2005. 204f. Tese (Doutor em Engenharia Elétrica) – Escola Politécnica, Universidade de São Paulo. São Paulo.

Dissertações de Mestrado

THOMAZ, Leandro Ferrari **Aplicação à música de um sistema de espacialização sonora baseada em ambisonics**. 2007. 151f. Dissertação (Mestrado em Engenharia Elétrica) – Escola Politécnica, Universidade de São Paulo, São Paulo.

Artigos

BAUER, B. **Phasor analysis of some stereophonic phenomena**. *Journal of the Audio Engineering Society of America* 33/11, pp. 1536-1539. 1961.

BEGAULT, D. R. **Challenges to the successful implementation of 3D sound.** *Journal of the Audio Engineering Society* 39/11, pp. 864-870. 1991.

BÈKÉSY, G. v. **Zur Theorie des Hörens, über das Richtungshören bei einer Zeitdifferenz oder Lautstärkenungleichheit der beiderseitigen Schalleinwirkungen.** *Physikal. Z.*, vol. 31, pp. 824-835, 857-868. 1930.

BOER, Cf. K. de **Stereophonic Sound Reproduction.** *Philips Technol. Rev.*, vol. 5, pp. 107-114. 1940.

BRANT, H. **Space as an Essential Aspect of Musical Composition.** *Contemporary Composers on Contemporary Music*, pp. 223-242. DaCapo Press, New York. 1978.

CHENG, Corey I.; WAKEFIELD, Gregory H. **Moving Sound Source Synthesis for Binaural Electroacoustic Music Using Interpolated Head-Related Transfer Functions (HRTFs).** *Computer Music Journal*, 25/4, pp. 57-80, MIT Press, 2001.

CLARK, H.; DUTTON, G.; VANDERLYN, P. **The 'stereosonic' recording and reproducing system: a two channel system for domestic tape records.** *Journal of the Audio Engineering Society*, 6/2, pp. 102-117. 1958.

COOPER, D. H.; BAUCK, J. L. **Prospects for Transaural Recordings.** *Journal of the Audio Engineering Society*, 37/1-2, pp. 3-19. 1989.

COOPER, D. H.; SHIGA, T. **Discrete Matrix Multi-Channel Stereo.** *Journal of the Audio Engineering Society*, 20/5, pp. 346-360. 1972.

GRIESINGER, D. **Equalization and Spatial Equalization of Dummy Head Recordings for Loudspeaker Reproduction.** *Journal of the Audio Engineering Society*, 37/1-2, pp. 20-29. 1989.

GERZON, M. A. **Periphony: With-Height Sound Reproduction.** *Journal of the Audio Engineering Society*, 21/5. pp. 2-10. 1972.

KALLINEN, Kari; RAVAJA, Niklas. **Comparing Speakers Versus Headphones in Listening to News from a Computer – Individual Differences and Psychophysiological Responses.** *Science Direct. Computers in human Behavior*, 23. 303-317. Elsevier. 2004.

KENDALL, G. S. **A 3-D Sound Primer: Directional Hearing and Stereo Reproduction.** *Computer Music Journal*, 19/4, pp. 23-46, Massachusetts Institute of Technology. 1995.

KIM, S.; KONG, D.; JANG, S. **Adaptative Virtual Surround Sound Rendering System for an Arbitrary Listening Position.** *Journal of the Audio Engineering Society*, 56/4. pp. 243-254. 2008.

KYRIAKAKIS, C.; TSAKALIDES, P.; HOLMAN, T. **Acquisition and Rendering Methods for Immersive Audio.** *IEEE Signal Processing Magazine*, 99. pp 55-66. January, 1999.

LANGMUIR, I.; et al. **A Study of Binaural Perception of the Direction of a Sound Source.** OSRD Rep. 4079, PB n° 31014 (available through the Office of Technical Services, U.S. Department of Commerce, Washington 25, D.C.). Jun, 1944.

MALHAM, D. G. **3-D Sound Spatialization Using Ambisonic Techniques.** *Computer Music Journal*, 19/4, pp. 58-70, Massachusetts Institute of Technology. 1995.

MALHAM, D. G. **Tutorial Article: Approaches to Spatialisation.** *Organised Sound*, 3/2, pp. 167-77, Cambridge University Press. 1998.

MALHAM, D. G. **Surround Sound, Yesterday, Today and Tomorrow.** *Rycote Microphone Windshields Ltd*, pp 1-4. 2003.

MEYER, E. Brad.; MORAN, David. R. **Audibility of a CD-Standard A/D/A Loop Inserted into High-Resolution Audio Playback.** *Journal of the Audio Engineering Society*, 55/9, pp. 775-779. 2007.

MOLLER, H. **Fundamentals of Binaural Technology.** *Applied Acoustics*, 36. pp. 171-218. 1992.

MOLLER, H. **Head-related transfer functions of human subjects.** *Journal of the Audio Engineering Society*, 43/5, pp. 300-321. 1995a.

MOLLER, H. **Transfer characteristics of headphones.** *Journal of the Audio Engineering Society*, 43/ 4, pp. 203-217. 1995b.

RATLIFFE, P. **Properties of Hearing Related to Quadrafonic Reproduction.** *BBC Research Department Report*, RD 1974/38. 1974.

SO, R. H. Y.; et al. **A low cost, non-individualized surround sound system based upon head related transfer functions: an ergonomics study and prototype development.** *Applied Ergonomics*, 37, pp. 695-707. Elsevier. 2006.

STEINBERG, J. C.; SNOW, W. B. **Physical Factors in Auditory Perspective.** *Bell Syst. Tech. J.*, vol.13, pp. 246-259. 1934.

WALLACH, Hans. **The Precedence Effect in Sound Localization.** *Journal of the Audio Engineering Society*, 21/10, pp 817-826. 1973.

WHITWORTH, R.; JEFFRESS, L. **Time versus intensity in the localization of tones.** *J. Acoust. Soc. Amer.* 33, pp. 925-929. 1961.

YOST, W.; WIGHTMAN, F.; GREEN, D. **Lateralization of Filtered Clicks.** *J. Acoust. Soc. Amer.*, 50, pp. 1526-1531. 1971.

Artigos apresentados em eventos

BRANDENBURG, K.; BRIX, S.; SPORER, T. **Wave Field Synthesis**. In: 3DTV Conference: The True Vision – Capture, Transmission and Display os 3D Video, Postdam, 4-6 Maio. 2009.

BRUCK, J. **The KFM360 Surround – a Purist Approach**. In: 103rd AES Convention, New York, 26-29 Setembro. Preprint 4637. Audio Engineering Society. 1997.

GRIESINGER, D. **Objective Measures of Spaciousness and Envelopment**. In: Procedures 16ª International AES Conference, Rovaniemi, 10-12 Abril, pp.27-41. Audio Engineering Society. 1999.

GRIESINGER, D. **Spatial Impression and Envelopment in Small Rooms**. In: 103rd AES Convention, New York, 26-29 Setembro. Preprint 4638. Audio Engineering Society. 1997.

GERZON, M. A. **Signal Processing for Simulating Realistic Stereo Images**. In: 93rd AES Convention, San Francisco, 1-4 Outubro. Preprint 3424. Audio Engineering Society. 1992.

HORBACH, U.; et al. **Design and applications of a data based auralization system for surround sound**. In: 106ª AES Convention, Munich, 8-11 Maio. Preprint 4976. 1999.

KLEPKO, J. **Five-channel Microphone Array with Binaural Head for Multichannel Reproduction**. In: 103rd AES Convention, New York, 26-29 Setembro. Preprint 4541. Audio Engineering Society. 1997.

MARTIN, G.; et al. **Sound Source Localization in a Five Channel Surround Sound Reproduction System**. In: 107th AES Convention, New York, 24-27 Setembro. Preprint 4994. Audio Engineering Society. 1999.

MASON, R.; RUMSEY, F. **An investigation of microphone techniques for ambient sound in surround sound recordings.** In: 106th AES Convention, Munich, Germany, 8-11 Maio. Audio Engineering Society. 1999.

MILLER, Robert E. Robin. **Compatible PanAmbiophonic 4.1 and PerAmbiophonic 6.1 Surround Sound for Advanced Television – Beyond ITU 5.1.** In: SMPTE 144th Technical Conference and Exhibition, Pasadena, CA. 2002.

SHAXBY, J. H.; GAGE, F. H. In: Medic. Res. Council, Spec. Rep. 166, pp. 1-32 (1932). See WOODWORTH, R. S. **Experimental Psychology.** pp. 529-530. 1938.

SOUSA, Fábio W. J. **Mixagem em Surround 5.1 e transcrição para audição binaural de músicas eletroacústicas especializadas: o caso de “4 sketches em movimento”.** In: 8^o Convenção da AES Brasil, São Paulo, 4-6 Maio. 2010.

THEILE, G. **Multichannel natural music recording based on psychoacoustic principles.** In: 108th AES Convention, Paris, 19-22 Fevereiro. Preprint 5156. Audio Engineering Society. 2000.

WILLIAMS, M.; LE DÛ, G. **Microphone Array Analysis for Multichannel Sound Recording.** In: 107th AES Convention, New York, 24-27 Setembro. Preprint 4997. Audio Engineering Society. 1999.

WILLIAMS, M.; LE DÛ, G. **Multichannel Microphone Array Design.** In: 108th AES Convention, Paris, 19-22 Fevereiro. Preprint 5157. Audio Engineering Society. 2000.

WITTEK, Helmut; HAUT, Christopher; KEINATH, Daniel. **Double M/S – a Surround recording technique put to test.** Tradução de SCHÄNDLES, Kevin. In: Tonmeistertagung. 2006.

Documentos Eletrônicos

cori spezzati. In: _____. Dicionário de Música Oxford Online. Disponível em < www.oxfordmusiconline.com > Acesso em 03 de dezembro de 2009.

Driver de áudio Soundflower. Disponível em < <http://code.google.com/p/soundflower/> > Acesso em 6 de abril de 2010.

Headzone. Disponível em < <http://www.sweetwater.com/store/detail/Headzone/> > Acesso em 05 de junho de 2010.

Headzone Pro XT - Updated version of the unique Headzone® PRO 5.1 monitoring headphone system. Disponível em < <http://archiv.beyerdynamic.de/en/broadcast-studio-video-production/products/headphonesheadsets/headzone-pro1/headzone-pro-xt.html> > Acesso em 05 de junho de 2010.

Holophonics. Disponível em < <http://en.wikipedia.org/wiki/Holophonics> > Acesso em 03 de fevereiro de 2010.

IRCAM Tolls by Flux. Disponível em < www.ircamtools.com > Acesso em 15 de Junho de 2010.

JOT, Jean-Marc; WARUSFEL, Olivier **Spat~: A Spatial Processor for Musicians and Sound Engineers.** IRCAM. CIARM 95, Ferrara (Italy) 1995. Disponível em < <http://articles.ircam.fr/textes/Jot95a/> > Acesso em 15 de junho de 2010.

Performing In Virtual Acoustics - Essay By Tom Beghin. Disponível em < <http://www.music.mcgill.ca/thevirtualhaydn/extras.php?pid=1> > Acesso em 03 de fevereiro de 2010.

QSound. Disponível em < <http://www.qsound.com/index.htm> > Acesso em 03 de fevereiro de 2010.

Roland's RSS. Disponível em <

http://www.soundonsound.com/sos/1996_articles/mar96/rolandrss10.html >

Acesso em 03 de fevereiro de 2010.

The Dark Side of the Moon, Pink Floyd. Disponível em <

<http://www.htforum.com/vb/archive/index.php/t-912.html> > Acesso em 18 de

Julho de 2010.

Pink Floyd Dark Side of the Moon Super-Audio CD. Disponível em <

<http://www.cduniverse.com/search/xx/music/pid/5690739/a/Dark+Side+Of+The+Moon.htm> > Acesso em 18 de Julho de 2010.

The Virtual Haydn Project. Disponível em <

<http://www.music.mcgill.ca/thevirtualhaydn/project.html> > Acesso em 03 de

fevereiro de 2010.

Wave Field Synthesis. Disponível em <

<http://www.holophony.net/Wavefieldsynthesis.htm> > Acesso em 13 de

Setembro de 2010.

Multmídia:

Götterdämmerung. Direção: Brian Large. Produção: Deutsche Gramophon DVD video. Intérpretes: The Metropolitan Opera Orchestra and Chorus; James Levine. Música: Richard Wagner. Hamburg: Universal Music, 2002. 2 DVDs (281 min). 4:3, NTSC, color. German. DTS 5.1, Dolby Digital 5.1, PCM stereo.

Harry Potter and the Sorcerer's Stone. Direção: Chris Columbus. Produção: David Heyman. Intérpretes: Daniel Radcliffe; Richard Harris; Rupert Grint e outros. Roteiro: J. K. Rowling; Steve Kloves. Música: John Williams. USA/ Canadá: Warner Bros Pictures, 2001. 1 DVD (152 min). 2.35:1, color. Inglês. DTS-ES, Dolby Digital EX, SDDS.

Immersion. Música: Merzbow; Meredith Monk; Pamela Z; Paul Dresher e outros. Starkland, 2001. 1 DVD-Audio (65 min). NTSC. Dolby Digital 5.1.

Metallica The Vídeos 1989-2004. Intérpretes: Metallica. Música: Metallica. E/M Ventures, 2006. 1 DVD (141 min). 4:3, Color. Inglês. LPCM 2.0 / Dolby Digital 5.1.

Pulse. Direção: David Mallet. Música: Pink Floyd. UK: Sony Music, 2006. 2 DVDs (240 min). 4:3. NTSC. Inglês. Dolby Digital 5.1, Stereo.

Six Brandenburg Concertos BWV 1046-1051 Complete Edition. Intérpretes: Stuttgarter Kammerorchester. Música: Johann Sebastian Bach. TACET Real Surround Sound, 2000. 1 DVD-Audio (94 min).

The Lord of the Rings: the Return of the King. Direção: Peter Jackson. Produção: Wingnut Films. Intérpretes: Elijah Wood; Ian McKellen; Liv Tyler e outros. Roteiro: Fran Walsh; Philippa Boyens e Peter Jackson. Música: Howard Shore. New Line Cinema, 2003. 2 DVDs (200 min). Widescreen, color. Inglês Dolby Surround 5.1, Inglês Dolby Surround Stereo, português Dolby Stereo e espanhol Dolby Stereo.

The Matrix. Direção e roteiro: Andy and Lana Wachowski. Produção: Joel Silver. Intérpretes: Keanu Reeves; Laurence Fishburne; Carrie-Anne Moss e outros. Música: Don Davis. USA/ Australia: Goucho II Film Partnership, 1999. 1 DVD (136 min). 2.35:1, color. Inglês. DTS, Dolby Digital, SDDS.

The Mummy: Tomb of the Dragon Emperor. Direção: Rob Cohen. Roteiro: Alfred Gough; Miles Millar. Produção: Stephen Sommers. Intérpretes: Brendan Fraser; Jet Li; Maria Bello; John Hannah e outros. Música: Randy Edelman. USA: Universal Pictures, 2008. 1 DVD (112 min). 2.35:1, color. Inglês. Dolby Digital, SDDS, DTS.

The Phantom of the Opera. Direção: Joel Schumacher. Produção: Really Useful Films/ Scion Films. Intérpretes: Gerard Butler; Emmy Rossum; Patrick

Wilson; Miranda Richardson e outros. Roteiro: Andrew Lloyd Webber e Joel Schumacher. Música: Andrew Lloyd Webber. Odyssey Entertainment, 2005. 1 DVD (141 min). Widescreen anamórfico. Inglês Dolby Digital 5.1.

Outros

Cubase SX 3 Help

Digital Performer 7 User Guide. Mark of the Unicorn, Inc. Cambridge. 2009.

Dolby. 5.1 Channel Production Guidelines.

Dolby. 5.1 Channel Musical Production Guidelines.

HAMASAKI, Kimio. **A geração futura – surround 22.2**. AES Latin American Conference 2009 – Áudio para HDTV e sua evolução. Realizado dia 26-05-2009 em São Paulo-SP.

LYSWA, J., BASKIND, A. **Use of Binaural and Transaural Spatialization Techniques in Multichannel 5.1 Production: Examples of Multichannel Music Mixes**. AES Latin American Conference 2009 – Áudio para HDTV e sua evolução. Realizado dia 26-05-2009 em São Paulo-SP.

LYSWA, J., BASKIND, A. **Use of Binaural and Transaural Spatialization Techniques in Multichannel 5.1 Production: Technical and Aesthetic Principles, from Recording to Post-production**. AES Latin American Conference 2009 – Áudio para HDTV e sua evolução. Realizado dia 27-05-2009 em São Paulo-SP.

MAYO, Andrés. **Workshop gravação/mixagem 5.1**. AES Latin American Conference 2009 – Áudio para HDTV e sua evolução. Realizado de 26-05-2009 à 28-05-2009 em São Paulo-SP.

MOTU 828 mk3. Especificações técnicas da interface de áudio. Disponível em < <http://www.motu.com/products/motuaudio/828mk3/summary.html> > Acesso em 13 de Setembro de 2010.

Plugin KLT3D. Grupo de Processamento de Áudio da UFRJ. Disponível em http://gpa.lps.ufrj.br/index.php/pt_BR/KLT3D-VST Acesso em 9 de abril de 2010.

Schoeps Mikrofone Surround Recording Techniques Brochure. 11/2006.

Sonar 7 Producer Edition Help.

Sony Vega Pro 8.0 Help.

Yamaha MSP-5. Especificações técnicas dos monitores ativos de estúdio. Disponível em < <http://www.yamaha.com/yamahavgn/CDA/ContentDetail/ModelSeriesDetail.html?CNTID=543310> > Acesso em 13 de Setembro de 2010.

Anexo: DVD

Exemplos sonoros apresentados com opções de áudio em

- 1) estéreo tradicional (quando existente)
- 2) binaural processado pelo *plugin* KLT3D
- 3) binaural processado através do objeto spat~
- 4) *surround* 5.1 AC3

Obras Cinematográficas

- > Harry Potter and the Sorcerer's Stone – cena da apresentação do quadribol.
- > The Matrix – cena do tiroteio sobre o prédio.
- > The Mummy: Tomb of the Dragon Emperor – cena do discurso do imperador direcionado ao seu exército.
- > The Lord of the Rings: the Return of the King – cena da batalha em frente à cidade branca e da luta entre Éowin e o senhor do escuro

Obras Musicais

- > The Phantom of the Opera – ária de Cristine, ária “Mascarade” e cena da luta entre o fantasma e Raul.
- > Metallica The Vídeos 1989-2004 – clip “Sad but True”
- > Pink Floyd – “Intro” e “Money”
- > Götterdämmerung – árias “Heil dir, Gunther!” e “Brünnhild', die hehrste Frau”
- > Six Brandenburg Concertos Johan Sebastian Bach BWV 1046-1051 Complete Edition - Concerto nº 2, 1º movimento.
- > Immersion – Carl Stone “Luong Hai Ky Mi Gia”

Produções próprias:

- > Ópera *La Descente d'Orphée aux Enfers* – mixagem utilizando o ArcPanner
- > Ópera *La Descente d'Orphée aux Enfers* – mixagem utilizando o Auralizer
- > “Chora não sô” Grupo Chora o sete – mixagem com padrão INA
- > 4 sketches em movimento I – ponto de escuta do instrumentista
- > 4 sketches em movimento II – ponto de escuta do ouvinte
- > “O Templo dos mil Sinos”