

REGIS ROSSI ALVES FARIA

**AURALIZAÇÃO EM AMBIENTES AUDIOVISUAIS
IMERSIVOS**

Tese apresentada à Escola
Politécnica da Universidade de
São Paulo para obtenção do
Título de Doutor em Engenharia.

São Paulo
2005

REGIS ROSSI ALVES FARIA

**AURALIZAÇÃO EM AMBIENTES AUDIOVISUAIS
IMERSIVOS**

Tese apresentada à Escola
Politécnica da Universidade de
São Paulo para obtenção do
Título de Doutor em Engenharia.

Área de Concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. João Antônio Zuffo

São Paulo
2005

FICHA CATALOGRÁFICA

Faria, Regis Rossi Alves

**Auralização em ambientes audiovisuais imersivos / R. R. A. Faria. – ed. rev. – São Paulo, 2005.
191 p.**

Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

1. Áudio digital 2. Acústica 3. Realidade virtual 4. Multimeios I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II.t.

*As supremas realidades do espírito,
que vos aproximam de Deus,
serão alcançadas pela via que dele vos parecia a mais distante:
a via da ciência objetiva.*

*A Grande Síntese
Pietro Ubaldi*

AGRADECIMENTOS

Agradeço ao prof. Dr. João Antônio Zuffo, meu orientador, pelo seu apoio constante e suporte a essa linha de pesquisa.

Este trabalho pôde convergir para sua forma final graças ao importante suporte dos parceiros institucionais do projeto AUDIENCE, participando da infraestrutura com a cessão de caixas acústicas de alta qualidade (LANDO HiFi), amplificadores (Sankya), cabeamento e conectorização (Cabos Golden). Agradeço o suporte da Intel (*cluster computacional*). Agradeço aos Serviços Gerais da Escola Politécnica da USP, pela fabricação dos pedestais e estruturas de suporte para caixas acústicas.

Meus agradecimentos nas respectivas áreas de apoio a Leandro Ferrari Thomaz (decodificação Ambisonics), a Luiz Gustavo Berro (montagem da infra-estrutura), a Mário Nagamura (artes gráficas do AUDIENCE), a Leonardo Nomura (mundos virtuais), a Breno Teixeira Santos (interfaces com o *PD*), a Hilton Fernandes e Guilherme Lepsch (Fortran), ao prof. Angelo Farina de Parma (Aurora e discussões metodológicas sobre Ambisonics), aos profs. Sylvio Bistafa e José Augusto Mannis (orientações acústicas), a Luciano Soares (*X3D*), a Bárbara Rossi (ilustrações), aos colegas Glauco Todesco e Celso Kurashima (discussões estratégicas e revisões), e ao prof. Marcelo Zuffo (apoio silencioso).

Agradeço ao pessoal da Administração do LSI e a todas as pessoas do LSI que contribuíram no engrandecimento ou construção deste trabalho.

Meus agradecimentos de irmão a todos os meus amigos e a minha família, tias, primas e primos queridos. Agradeço também aos amores e desamores que tumultuaram e enriqueceram minha vida nestes anos.

Finalmente, meus agradecimentos colossais e minha resignada aceitação do poder, apoio e compreensão que emana de pais tão maravilhosos quanto os que eu tenho a oportunidade de ter neste planeta: minha mãe, Aparecida e meu pai Erpino.

RESUMO

Nos últimos anos os avanços em áudio multicanal e sistemas envolventes despertaram um grande interesse pelas técnicas para a auralização de campos sonoros espaciais, capazes de recriar ambientações acústicas com grande realismo, envolvendo o ouvinte completamente. Um estudo sobre as tecnologias e soluções para áudio espacial mostrou que a construção de auralizadores envolve quatro blocos funcionais fundamentais. Mas, embora existam muitas técnicas e ferramentas disponíveis, não há uma arquitetura integradora para produção de áudio espacial que permita ao desenvolvedor selecionar as técnicas que deseja e montar uma solução com um grau de refinamento arbitrário que atenda a seus requisitos. Apresenta-se uma arquitetura modular aberta em quatro camadas para produção de áudio espacial em ambientes audiovisuais imersivos. Como estudo de caso, aborda-se a implementação de um auralizador Ambisonics para a CAVERNA Digital. Verifica-se que a auralização pode fornecer a correta perspectiva acústica do ambiente virtual necessária para uma percepção mais realista do espaço, e que a abordagem aberta para se implementar sistemas de áudio 2D/3D apresenta vantagens. Os resultados da implementação são apresentados e discutidos. Como conclusões, identificam-se desafios na implementação e os trabalhos futuros.

ABSTRACT

Recent advances in multichannel audio and surround systems have contributed to an increasing interest for spatial sound field auralization, capable of delivering acoustic ambience with great realism, and surrounding listener completely. A study on technologies and solutions for spatial sound has shown that building auralizers involve four major functional blocks, but, although there are many techniques and tools available, there is not an integration architecture for spatial audio production which permits developers to select their favorite techniques, and build a solution with an arbitrary refinement level, as wished. It is presented a four level modular open architecture for spatial audio production in immersive audiovisual environments. As a case study, an Ambisonics auralizer is implemented for the CAVERNA Digital CAVE. It is verified that auralization can deliver the correct acoustical perspective for the virtual environment necessary for a more realistic perception of space, and that an open approach presents advantages in 2D/3D audio systems design. Implementation results are presented and discussed. As conclusions, challenges and future works are presented.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

LISTA DE SÍMBOLOS

1	INTRODUÇÃO	1
1.1	RELEVÂNCIA	2
1.2	MOTIVAÇÃO	4
1.3	OBJETIVOS	5
1.4	SUMÁRIO ESTRUTURADO DA TESE	8
2	SISTEMAS DE SONORIZAÇÃO ESPACIAL	10
2.1	SONORIZAÇÃO ESPACIAL (ÁUDIO 2D E 3D).....	10
2.1.1	O espaço sonoro do monaural ao multicanal	11
	O estéreo.....	12
	A quadrifonia.....	15
	Som envolvente	17
2.1.2	Configurações multicanais	19
2.1.3	Espacialização e Auralização	25
	Espacialização	25
	Auralização.....	26
2.2	MÉTODOS DE AURALIZAÇÃO	27
2.2.1	Bi-aural.....	30
2.2.2	Sistema Vetorial de Panorama por Amplitude (SVPA).....	31
2.2.3	Ambisonics	33
2.2.4	Ambiophonics.....	41
2.2.5	<i>Wave Field Synthesis (WFS)</i>	43
2.3	TECNOLOGIAS E SISTEMAS PARA ÁUDIO ESPACIAL (2D/3D).....	45
2.3.1	Plataformas comerciais para áudio espacial	50
	Decodificadores de áudio	50
	Plataformas integradas (<i>hardware e software</i>)	51
	Programas para simulação acústica	54

3	ÁUDIO EM AMBIENTES E SISTEMAS AUDIOVISUAIS IMERSIVOS	57
3.1	NÍVEL DE IMERSÃO SONORA	59
3.2	SISTEMAS DE REALIDADE VIRTUAL IMERSIVA (RVI).....	61
3.2.1	Ambientes <i>CAVE</i> 's: A CAVERNA Digital.....	63
	Aspectos arquitetônicos e estruturais.....	67
	Telas	68
	Sistema computacional.....	69
	Placas Multicanais	71
	Matrizes de alto-falantes (caixas acústicas).....	71
3.2.2	Aplicações audiovisuais de RVI.....	73
3.3	ÁUDIO EM RVI.....	74
3.3.1	Arquiteturas e sistemas para áudio em RV	76
3.3.2	Ferramentas para áudio em ambientes de RVI	80
	<i>VSS</i>	80
	<i>Avango</i>	81
	<i>Bergen Sound Server & Library</i>	81
	<i>FMOD</i>	82
	<i>VRJuggler Sound Library</i>	82
	<i>AuSIM</i>	82
	<i>Lake Huron</i>	83
	<i>Wave Engine</i>	84
3.4	UMA ARQUITETURA MODULAR PARA PRODUÇÃO SONORA ESPACIAL.....	84
	Limitações das arquiteturas fechadas	85
	Vantagens de uma arquitetura aberta.....	86
	Hierarquia de Camadas.....	87
3.4.1	Formatos para descrição e representação de cena.....	91
3.4.2	Modelamento e simulação da cena acústica	96
	Modelamento da fonte.....	97
	Modelamento do meio.....	97
	Modelamento físico versus perceptual	97
	Modelamento baseado em ondas versus baseado em raios.....	100
	Uso da resposta impulsiva	101
	Metodologias de modelamento acústico.....	103
	Métodos analíticos e numéricos (baseado em ondas).....	104
	Métodos geométricos (baseado em raios).....	105
	Modelamento do ouvinte ou receptor	110

3.4.3	Síntese sonora e codificação espacial	111
3.4.4	Decodificação espacial e sonorização da cena.....	112
4	UMA SOLUÇÃO PARA AURALIZAÇÃO EM RVI COM AMBISONICS	114
4.1	INTRODUÇÃO	114
4.2	UM AURALIZADOR AMBISONICS EM CAMADAS	115
4.2.1	Diagrama de Blocos Geral do Sistema: Projeto.....	116
4.2.2	Ambiente acústico virtual de testes	119
4.2.3	Mapeamento dos objetos na cena acústica	119
	Atributos acústicos da cena	120
4.2.4	Simulação acústica	123
4.2.5	Codificação sonora espacial	130
4.2.6	Decodificação do formato espacial e sonorização multicanal	133
	Configurações de alto-falantes	134
	Decodificação para N canais	137
	Decodificação Ambisonics	137
4.2.7	Implantação da infra-estrutura de áudio multicanal.....	138
	Posicionamento dos alto-falantes	140
	Esquema de conexões físicas de áudio	141
4.3	SIMULAÇÕES E TESTES (CASOS DE USO).....	142
4.4	RESULTADOS	146
	Infra-estrutura	146
	Nível de imersão	146
	Codificação espacial Ambisonics	147
	Simulação acústica e auralização.....	147
5	DISCUSSÃO DOS RESULTADOS.....	150
5.1	INTRODUÇÃO	150
5.2	QUALIDADE ESPACIAL PERCEBIDA	150
5.3	ARQUITETURA DE AURALIZAÇÃO.....	151
5.4	LIMITES TÉCNICOS	153
5.4.1	Problemas desafiadores	153
	Acústica local	153
	Modelamento acústico	154
	Eficácia da auralização	154
	Engenharia de <i>software</i>	155

6	CONCLUSÕES	156
6.1	CONCLUSÕES GERAIS	157
6.2	INOVAÇÕES.....	160
6.3	TRABALHOS FUTUROS.....	161
6.3.1	Configurações avançadas para alto-falantes	161
	Anel octogonal virtual	161
	Equalização e calibração	163
	Tratamento da acústica local	164
6.3.2	Auralização integrada com o <i>PD</i>	165
6.3.3	Descrição avançada de cena	166
6.3.4	Modelamento acústico.....	167
6.3.5	Casos de uso progressivos	167
6.3.6	Tendências gerais	168
6.3.7	Ampliação da infra-estrutura e Auralização interativa em tempo real	170
6.3.8	Sistema flexível para auralização multiformato	170
6.3.9	Sistemas para síntese acústica dinâmica.....	171
6.3.10	Protocolos e <i>hardware</i> para distribuição de áudio multicanal	172
7	ANEXOS	174
8	REFERÊNCIAS	181

LISTA DE FIGURAS

Figura 1 – Cenário de tarefas de um sistema para auralização em RVI.....	7
Figura 2 – Efeito estéreo	13
Figura 3 – Microfonação estéreo XY.....	14
Figura 4 – Configuração quadrifônica (alto-falantes nos vértices de um quadrado)	16
Figura 5 – Configurações para 5.1 e 7.1 canais (padrão ITU-R BS.775-1).....	20
Figura 6 – Configurações experimentais (protótipos) de testes de sistemas para som 3D	23
Figura 7 – Incidência de onda plana em uma cabeça indicando diferença em extensão no percurso sonoro entre os ouvidos	31
Figura 8 – Exemplo de configuração de caixas acústicas para SVPA.....	33
Figura 9 – Técnica de Blumlein para captação estereofônica (MS)	35
Figura 10 – Extensão da técnica de Blumlein, permitindo o registro sonoro de um campo 3D (eixo z central apontando para cima)	35
Figura 11 – Perspectiva do padrão de captura Ambisonics (1 ^a ordem).....	36
Figura 12 – Sistemas de coordenadas angulares e cartesiana no Ambisonics	38
Figura 13 - Simulação de posicionamento de fontes sonoras com <i>WFS</i> fora (esquerda) e dentro da sala sonorizada (direita)	45
Figura 14 – Exemplo de painel traseiro de decodificador comercial (<i>receiver</i>)	51
Figura 15 – Máquina de convolução comercial <i>Convolver</i>	53
Figura 16 – Escala para níveis de imersão sonora	61
Figura 17 – Modelo tridimensional da CAVERNA Digital.....	65
Figura 18 – Exemplo de navegação em ambiente virtual na CAVERNA Digital	66
Figura 19 – Diagrama geral de um sistema aglomerado de computadores.....	70
Figura 20 –Aplicativo de RVI audiovisual (<i>Asa Delta</i>).....	74
Figura 21 – Arquitetura modular para produção de áudio espacial em 4 camadas.....	88
Figura 22 – Modelo de referência para interfaceamento entre camadas.....	89
Figura 23 – Nós de áudio na estrutura de árvore do <i>VRML97</i>	92

Figura 24 – Cena audiovisual gerada pelo código <i>X3D</i> da Tabela 7	94
Figura 25 – Geometria de espacialização usada no <i>X3D</i> , <i>VRML</i> e <i>MPEG-4 AudioBIFS</i>	95
Figura 26 – Janelas de tempo típicas de uma IR (reflectograma) e impressões espaciais associadas.....	102
Figura 27 – Principais metodologias para modelamento acústico	104
Figura 28 – Simulação de propagação das ondas sonoras por meio de raios de som (modelamento de reflexão especulares)	105
Figura 29 – Método da fonte-imagem para modelamento acústico discreto e especular.....	107
Figura 30 – Resumo de tecnologias disponíveis por camada da arquitetura	116
Figura 31 – Diagrama de blocos geral de sistema para produção sonora espacial	117
Figura 32 – Sala acústica virtual para testes de auralização (planta baixa)	119
Figura 33 – Cena testes em 3D	123
Figura 34 – Disposição das fontes imagens no espaço acústico	126
Figura 35 – Respostas impulsivas h_w , h_x , h_y , e h_z da cena de testes para 200ms	129
Figura 36 – Vetor de posição de um objeto sonoro O_j e do falante R_j	130
Figura 37 – Esquema de convolução do som O_j pelas <i>IR</i> 's em <i>B-Format</i>	132
Figura 38 – Foto de uma configuração octogonal planar externa à CAVERNA.....	135
Figura 39 –Montagem de alto-falantes em anel octogonal (interno)	135
Figura 40 – Matriz de alto-falantes em configuração cúbica típica (3D).....	136
Figura 41 – Foto ilustrativa da montagem e cabeamento de alto-falante nos cantos inferiores da CAVERNA.....	136
Figura 42 – Decodificação de som 3D (em <i>B-Format</i>) para um canal de saída	138
Figura 43 – Pedestais de madeira reconfiguráveis para posicionamento dos alto-falantes..	141
Figura 44 – Diagrama geral de conexões e distribuição do sistema de áudio.....	142
Figura 45 – Metodologia de testes de auralização	143
Figura 46 – Matriz de alto-falantes em configuração anel octogonal virtual (2D).....	162
Figura 47 – Configuração não-regular prevista para avaliação futura	163

LISTA DE TABELAS

Tabela 1 – Sistemas quadrifônicos.....	15
Tabela 2 – Configurações multicanais segundo o arranjo dos alto-falantes	24
Tabela 3 – Comparando os métodos de auralização bi-aural e multicanal	28
Tabela 4 – Número de canais necessários por ordem da implementação Ambisonics	36
Tabela 5 – Tecnologias para áudio espacial (2D/3D).....	47
Tabela 6 – Níveis de Imersão Sonora	61
Tabela 7 – Exemplo de descrição de cena audiovisual utilizando <i>X3D</i>	93
Tabela 8 – Exemplos de sintaxe descritora de propriedades acústicas para objetos no <i>X3D</i>	122
Tabela 9 – Infra-estrutura geral do sistema de produção sonora espacial.....	138
Tabela 10 – Resultados do enquete subjetivo de qualidade espacial percebida.....	148
Tabela 11 – Ampliação prevista na infra-estrutura para auralização	170

LISTA DE ABREVIATURAS E SIGLAS

ADAT	ALESIS DIGITAL AUDIO TAPE
AES	AUDIO ENGINEERING SOCIETY
AES/EBU	PROTOCOLO DE TRANSMISSÃO DE ÁUDIO DIGITAL AES3
API	APPLICATION PROGRAM INTERFACE
ATSC	AMERICAN TELEVISION SYSTEMS COMMITTEE
B-FORMAT	FORMATO DE ÁUDIO ESPACIAL NATIVO DO AMBISONICS
BIFS	BINARY FORMAT FOR SCENE DESCRIPTION
CAVE	CAVE AUTOMATIC VIRTUAL ENVIRONMENT
CODEC	CODIFICADOR/DECODIFICADOR
DHT	DISTORÇÃO HARMÔNICA TOTAL
DSP	DIGITAL SIGNAL PROCESSING/PROCESSOR
FIR/IIR	FINITE IMPULSE RESPONSE / INFINITE IMPULSE RESPONSE
HDTV	HIGH DEFINITION TELEVISION
HMD	HEAD MOUNTED DISPLAYS
HRTF	HEAD RELATED TRANSFER FUNCTION
HTML	HYPER TEXT MARKUP LANGUAGE
IDE	INTEGRATED DEVELOPMENT ENVIRONMENT
ILD	INTER-AURAL LEVEL DIFFERENCE
IR	IMPULSE RESPONSE (RESPOSTA AO IMPULSO)
ITD	INTER-AURAL TIME DIFFERENCE
ITU	INTERNATIONAL TELECOMMUNICATION UNION
MADI	MULTICHANNEL AUDIO DIGITAL INTERFACE
MIDI	MUSICAL INSTRUMENT DIGITAL INTERFACE
MPEG	MOTION PICTURES EXPERT GROUP
MPI	MESSAGE PASSING INTERFACE
PCI	PERIPHERAL COMPONENT INTERFACE/INTERCONNECT
RMS	ROOT MEAN SQUARE (RAIZ MÉDIA QUADRÁTICA)
RPM	ROTAÇÕES POR MINUTO
RVI	REALIDADE VIRTUAL IMERSIVA
S/PDIF	SONY/PHILIPS DIGITAL INTERFACE FORMAT
SVPA	SISTEMA VETORIAL DE PANORAMA POR AMPLITUDE
TDT	TV DIGITAL TERRESTRE
UHJ	UNIVERSAL HJ
USB	UNIVERSAL SERIAL BUS
VBAP	VECTOR BASED AMPLITUDE PANNING
VRML	VIRTUAL REALITY MODELLING LANGUAGE
VSS	VIRTUAL SOUND SERVER OU VANILLA SOUND SERVER
WFS	WAVE FIELD SYNTHESIS
X3D	EXTENSIBLE 3D
XML	EXTENSIBLE MARKUP LANGUAGE

LISTA DE SÍMBOLOS

O_j	Objeto ou fonte sonora j (som anecóico do objeto/fonte sonora j)
IR_j	Resposta impulsiva onidirecional do ambiente para a fonte sonora j
B_{ji}	Matriz de direção dos raios de para O_j em <i>B-Format</i>
IR_j^B	Resposta impulsiva do ambiente para O_j em <i>B-Format</i>
h_x, h_y, h_z, h_w	Resposta impulsiva do ambiente para O_j em cada eixo <i>B-Format</i>
H^{-1}	Filtro inverso (de dereverberação)
$O_{jx}, O_{jy}, O_{jz}, O_{jw}$	Componentes em <i>B-Format</i> do som do objeto O_j (1ª ordem)
P_ψ	Pressão sonora produzida pela frente de onda da fonte
P_n	Pressão sonora produzida pelo sinal do alto-falante n
W, X, Y, Z	Sinais/canais componentes do formato Ambisonics <i>B-Format</i> para sistemas de 1ª ordem
λ	Comprimento de onda
θ	Ângulo de azimute (plano horizontal)
ϕ	Ângulo de elevação (vertical)
Ω	Impedância (unidade Ohm)
W	Watt (unidade de potência)

1 INTRODUÇÃO

A síntese de som espacial em duas e três dimensões tem sido explorada por décadas, e há uma variedade de técnicas e sistemas disponíveis, atendendo a diversas aplicações e níveis de qualidade. O uso de áudio espacial em ambientes audiovisuais, entretanto, é relativamente novo, especialmente em se tratando da produção de campos sonoros 2D/3D ou auralização, e falta ainda, do ponto de vista sistêmico, uma estrutura de referência que permita a desenvolvedores de sistemas combinarem ou integrar as ferramentas, técnicas ou métodos de áudio espacial que desejem, de forma a produzir efeitos perceptuais de imersão em diversos graus.

Neste trabalho serão apresentadas as principais técnicas de produção sonora espacial e auralização disponíveis para ambientes audiovisuais que envolvam seus usuários, com destaque para os sistemas de realidade virtual imersivos. Segundo sua adequação aos propósitos dos sistemas audiovisuais imersivos, algumas técnicas foram priorizadas, e uma delas, alvo de uma implementação teste (estudo de caso) na CAVERNA Digital da EP-USP¹, como parte integrante e essencial de um projeto mais extenso proposto para sonorização de ambientes audiovisuais imersivos, o projeto AUDIENCE (FARIA, 2004).

A idéia central do trabalho é investigar a viabilidade, a complexidade e a validade do uso da auralização em ambientes audiovisuais imersivos – como os sistemas de Realidade Virtual Imersiva (RVI) – e propor uma arquitetura e implementação de referência para auralização que possa ser inspiradora para a concepção e a industrialização de novas gerações de sistemas para auralização.

Na consecução do trabalho deparamo-nos com dificuldades intrínsecas em se implementar uma infra-estrutura inédita e com recursos escassos, e a definição e implantação de uma estratégia de execução mostraram-se tão importantes para chegarmos aos nossos objetivos quanto os próprios.

¹ URL: <http://www.lsi.usp.br/interativos/nrv/caverna.html>

O projeto e a construção de um sistema para auralização em *CAVE's* constitui a principal contribuição deste trabalho, envolvendo a montagem de uma infra-estrutura para sonorização espacial flexível e escalável, capaz de abrigar e integrar diversas técnicas e sistemas para síntese e produção de áudio espacial, disponibilizando-as para as aplicações visuais. Serão apresentados um sistema de auralização baseado na técnica Ambisonics (GERZON, 1973), a implantação da infra-estrutura de áudio, os testes e resultados.

1.1 Relevância

É interessante notar que os conceitos de panorama sonoro e áudio envolvente não são novos, estando presentes desde os primórdios da história da gravação. Os primeiros sistemas de registro sonoro já tencionavam registrar um campo sonoro tridimensional correspondente às audições musicais, mas a tecnologia disponível só permitia o registro monofônico ou no máximo estereofônico, restringindo-se a dois canais somente.

Até pouco tempo atrás o padrão estéreo de dois canais reinava absoluto no mercado fonográfico e na maioria de aplicações, equipamentos e serviços contendo áudio. Dois dos sistemas mais populares de gravação do século XX facilitaram a disseminação do paradigma de dois canais: os discos de vinil² e as fitas magnéticas. Entretanto, estes mesmos meios, graças a limitações tecnológicas e ao custo elevado em se trabalhar com mais de dois canais, foram em grande parte responsáveis pela acomodação da indústria fonográfica à produção em dois canais, fato que enterrou precocemente o interesse e o desenvolvimento de sistemas multicanais capazes de registrar campos sonoros envolventes, e levou à estagnação e saturação da produção sonora em dois canais por décadas³.

A recente popularização dos sistemas multicanais conhecidos por *surround 5.1* mudou drasticamente este cenário estagnado. Hoje observamos uma incrível

² também conhecidos por “LP’s” (*Long Play’s*)

³ com exceção da indústria cinematográfica, onde os filmes tinham multitrilhas.

tendência à disseminação em larga escala de sistemas baseados em múltiplos canais de áudio, influenciando toda a indústria de consumo e profissional multimídia. Há um enorme interesse na concepção de novos serviços multimídia, com maior grau de realismo sonoro, nos sistemas de armazenamento em estado sólido, codificação e transmissão multicanal (compressão de áudio e distribuição de áudio digital), e na produção de conteúdo e reprodução multicanal espacial, suportando a capacidade de sintetizar ambientes acústicos imersivos dinamicamente intercambiáveis nas plataformas de acesso de usuário final.

Com os progressivos avanços da computação, da microeletrônica, e do processamento de sinais digitais, acrescido de novos meios de armazenamento como os meios ópticos (CD's), as idéias iniciais referentes ao registro de campos sonoros retornaram à cena, tornando possível ao mercado retomar as iniciativas em desenvolver tecnologias para áudio envolvente e multicanal. O mercado consumidor, por outro lado, também exerce pressão pela evolução dos sistemas, para uma oferta de melhor qualidade na experiência audiovisual.

No momento verificamos que muitas correntes tecnológicas se fundem ou convergem para gerar novos serviços multimídia, e que há uma forte tendência à produção de experiências audiovisuais o mais realista possível. O futuro da televisão aponta para a exploração da estereoscopia, da multiprojeção e do áudio envolvente.

Há um enorme interesse em sistemas de telecomunicações capazes de integrar audiovisualmente pessoas remotamente localizadas como se estivessem lado a lado (KURASHIMA, 2005). A demanda por qualidade e realismo na percepção tanto de imagens quanto de sons passa obrigatoriamente pela viabilização de soluções tecnológicas eficazes e de baixo custo capazes de induzir à percepção humana a sensação de imersão. As pesquisas em psicoacústica e realidade virtual que investigam mecanismos e modelos de produção holográfica, estereoscópica e tridimensionais colaboram significativamente para o desenvolvimento e seleção de novas tecnologias que possam amadurecer e uma vez viáveis economicamente, passar por uma engenharia de produto e serem levadas ao mercado de consumo.

Neste contexto evolucionário, a presente pesquisa está inserida como parte deste processo natural de evolução dos paradigmas de sonorização, tão saudáveis e necessários para o estabelecimento de novas tecnologias multimídia. É relevante o estudo e a compreensão das tendências tecnológicas para áudio 3D à luz das descobertas e da investigação das técnicas de sonorização tridimensional, para melhor inferirmos quais tecnologias e formatos têm maior chance de sucesso e quais melhor atendem às demandas em qualidade e realismo desejadas. As pesquisas e experimentos com aplicações imersivas detêm um potencial enorme para prover os avanços tecnológicos necessários para o próximo *breakpoint* na curva de evolução dos sistemas de áudio multicanal.

Por exemplo, o formato Ambisonics é conhecido desde a década de 1970 (GERZON, 1973). O formato *Wave Field Synthesis (WFS)* é relativamente novo (BERKHOUT, 1993). Ambos são formatos de áudio 3D promissores, mas que merecem investigações mais aprofundadas no sentido de se verificar sua real adequação como soluções em larga escala voltadas para a produção de campos sonoros em sistemas audiovisuais.

É relevante prototipá-los no âmbito da CAVERNA Digital, principalmente por dois motivos: i) como nascedouro de aplicações de realidade virtual, a CAVERNA Digital permite a investigação de serviços multimídia imersivos avançados como, por exemplo, a navegação por ambientes virtuais tridimensionais, a telepresença, bem como aplicações audiovisuais para engenharia e medicina onde a percepção da diretividade do som é de extrema importância para simular ou reproduzir situações do mundo real em um ambiente controlado, e ii) a CAVERNA Digital possui um sistema de aglomerado de computadores que torna possível a exploração e a investigação de processos computacionalmente pesados, que exigem grande capacidade de computação, como, por exemplo, modelos acústicos complexos e algoritmos de codificação audiovisual em tempo real.

1.2 Motivação

A inexistência ou carência de soluções eficazes e integráveis para a reprodução de áudio 3D nos ambientes de realidade virtual imersivos é uma das maiores motivações

em se estudar soluções de espacialização sonora e auralização nesta área. Há neste trabalho o desejo implícito em se pesquisar soluções para incrementar o grau de realismo das aplicações de realidade virtual com um áudio de melhor qualidade, de forma a conceber aplicações completamente imersivas.

A maior motivação, entretanto, está no desejo de se implementar e testar formatos de codificação de áudio 3D sofisticados e capazes de alto grau de realismo, e buscar viabilizar soluções de baixa relação custo/benefício, inclusive para o mercado doméstico, estimulando a pesquisa aplicada com fins à engenharia de novos produtos.

1.3 Objetivos

Este trabalho tem os seguintes objetivos fundamentais:

- a) avaliar o estado da arte em tecnologias e soluções tecnológicas para sonorização espacial e auralização voltadas para ambientes audiovisuais imersivos, com especial deferência aos sistemas de realidade virtual imersiva completa do tipo *CAVE's*,
- b) propor uma arquitetura modular, flexível e escalável para produção sonora espacial nestes meios eletrônicos interativos,
- c) projetar, implementar e investigar uma solução para auralização multicanal com Ambisonics para a CAVERNA Digital, utilizando-se matrizes de alto-falantes dispostas ao redor da CAVERNA.

Este trabalho cobre parte dos desenvolvimentos e metas previstos para a primeira fase do projeto AUDIENCE (*Audio Immersion Experience by Computer Emulation*) (FARIA, 2004), em curso na CAVERNA Digital. O projeto tem como objetivo dotar a CAVERNA de um conjunto de soluções flexíveis e escaláveis para sonorização, inclusive e principalmente soluções para áudio 2D/3D baseadas em simulações de campos sonoros espaciais.

Uma das maiores barreiras em realidade virtual é a demanda computacional necessária para a produção de experiências realistas, usualmente vinculadas à

emulação ou simulação de processos físicos reais. Para o domínio visual, a carga computacional usualmente é alta para o cálculo envolvido nas renderizações de objetos e nas atualizações das projeções. SOARES (2002) discute com maiores detalhes os problemas e soluções relacionadas à demanda computacional no caso visual.

Para o domínio auditivo, o modelo acústico e o processo de codificação e decodificação de áudio 3D implicam numa alta carga computacional devido principalmente à elevada complexidade dos algoritmos de mapeamento do espaço acústico, bem como devido ao uso extensivo de funções de DSP, operações trigonométricas, e aos requisitos de operação em tempo real. Estes se tornam mais críticos na medida em que o número de objetos sonoros em simulação cresce e também na medida em que a resolução e precisão do modelo acústico aumentam.

Na execução do estudo de caso, concebemos um auralizador Ambisonics e testamo-lo no ambiente da CAVERNA. Concebemos um sistema a ser implementado no decorrer do projeto AUDIENCE que tem como *software* navegador para RVI o Jinx⁴, um navegador de ambientes virtuais 3D desenvolvido na CAVERNA Digital (SOARES, 2004a). Um meio para descrição de cena acústica, seus atributos e parâmetros, foi adicionado ao padrão de descrição de cena adotado no Jinx: o *X3D* (WEB3D CONSORTIUM, 2005) para se permitir o controle da síntese de campos sonoros.

A Figura 1 ilustra o cenário de tarefas envolvidas no desenvolvimento de auralizadores. O auralizador em si tem dois componentes principais: i) um modelador acústico (ou *render* acústico), que será responsável por calcular no ambiente virtual quais são os sons ouvidos no ponto do usuário, e ii) um codificador de áudio 2D/3D, que será responsável por processar os sons individuais de cada objeto sonoro (a princípio armazenados em formato de arquivos *Wave*) e gerar uma representação global contendo tanto a informação temporal quanto espacial dos sons, isto é, sua qualidade acústica e sua diretividade.

⁴ URL: <http://www.lsi.usp.br/interativos/nrv/jinx.html>

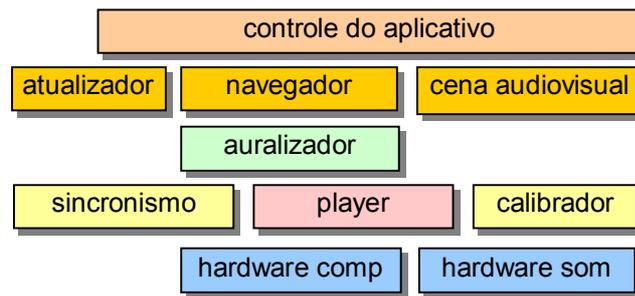


Figura 1 – Cenário de tarefas de um sistema para auralização em RVI

O sistema de áudio espacial em desenvolvimento no AUDIENCE está sendo construído a partir da integração dos diversos módulos constituintes sobre uma interface de programação sonora gráfica, o *PD*⁵. Numa sessão do *PD* podem-se usar os blocos ferramentas que se deseja e montar uma rede e sequência de interligações entre eles, para processar a auralização desejada.

A arquitetura proposta prevê uma infra-estrutura computacional distribuída que será empregada futuramente para alocar os recursos computacionais e hospedar os programas, algoritmos, e ferramentas de áudio e vídeo que se desejar usar, como o sistema navegador 3D (aplicação), os programas e algoritmos do auralizador, programas para reprodução (*API's* de áudio 2D/3D) e *drivers* do *hardware* de áudio específicos.

A CAVERNA utiliza um aglomerado de realidade virtual (*VRCluster*) baseado em computadores convencionais. Cada estação ou nó do aglomerado é responsável por diversas funcionalidades, que podem variar dinamicamente, abrangendo tarefas como geração de vídeo para as telas (projeções visuais), auralização e codificação 3D de sons, entrada de dados (leitura de sensores), dentre outros cálculos necessários para ambas as simulações visual e auditiva.

O sistema deve ainda incorporar uma hierarquia de recursos de sincronização de vídeo, áudio, imagens e dados, para apresentar um ambiente virtual distribuído com coerência visual e auditiva. Uma idéia é utilizar vários nós para rodarem instâncias

⁵ *PD (Pure Data)* é uma ambiente de programação gráfico em tempo real para música, áudio e multimídia. Mais detalhes em <http://www.puredata.info/> (acesso em 30 maio 2005).

do auralizador (para cada objeto sonoro em cena), e um ou dois nós para a mixagem dos vetores sonoros, proceder à decodificação e geração final de até 16 canais de saída de som. Este nó de sonorização final hospedará uma ou mais placas multicanais, capazes de comandar as saídas/alto-falantes. Futuramente este processo deverá ser integrado como subsistema de áudio avançado ao sistema de gerenciamento da CAVERNA Digital.

É importante notar que, no ramo de áudio 3D, o acesso a documentos e arquivos sonoros multicanais não é suficiente para que se possa reproduzi-los adequadamente, dada a necessidade de o decodificador estar disponível e, mais importante, de haver um sistema de reprodução e matrizes de alto-falantes adequadas, nas configurações necessárias, o que na maioria das vezes implica num alto custo para montagem de um *setup* experimental.

O levantamento e análise de tecnologias correlacionadas propiciou o conhecimento teórico e prático sobre diversas técnicas necessárias para implementação de sistemas de áudio imersivo (realidade virtual, *X3D*, sistemas *CAVE*, processamento distribuído, *cluster computing*, codificação *surround*, modelamento acústico, auralização, Ambisonics, etc.) e permitiu a convergência para uma proposta de implementação (experimento) inédita, relevante e viável do ponto de vista tecnológico.

1.4 Sumário estruturado da Tese

Esta tese está estruturada em 6 capítulos a saber:

Capítulo 1 – Introdução: Introdução ao tema abordado na tese, descrevendo sua relevância, motivações, e os objetivos do trabalho em linhas gerais.

Capítulo 2 – Sistemas de Sonorização Espacial: Apresentação dos sistemas e tecnologias de sonorização bi e tridimensionais mais relevantes.

Capítulo 3 – Áudio em Ambientes e Sistemas Audiovisuais Imersivos: Introdução ao cenário de aplicações alvo do trabalho, os ambientes de realidade virtual imersiva do tipo *CAVE*. Apresenta soluções e sistemas de áudio em uso nestes ambientes, e a

proposição de uma arquitetura para produção de áudio espacial baseada no conceito de auralização, mais adequada para uma experiência de imersão audiovisual completa em *CAVE's*.

Capítulo 4 – Uma Solução para Auralização em RVI com Ambisonics : Estudo de caso baseado na implementação de uma solução para auralização com Ambisonics na CAVERNA Digital. Apresenta detalhes de implementação, testes e simulações realizadas, e resultados.

Capítulo 5 – Discussão dos Resultados: Apresenta uma análise e discussão dos resultados à luz dos sucessos e limitações verificadas.

Capítulo 6 – Conclusões: Apresenta uma conclusão geral do trabalho, identificando inovações e trabalhos futuros.

2 SISTEMAS DE SONORIZAÇÃO ESPACIAL

A história dos sistemas de sonorização inicia-se no século XIX, com os primeiros dispositivos para gravação de som inventados na década de 1880⁶. Entre aproximadamente 1877 e 1925 a captura e reprodução sonora foram baseados em princípios de transdução acústico-mecânico, e possibilitaram o surgimento dos primeiros sistemas para registro e reprodução sonora. Thomas Edison, Chichester Bell, Charles Tainter e Emile Berliner encontram-se no prestigiado *hall* de inventores e detentores de patentes cobrindo uma variedade de invenções e meios de armazenamento (entre cilindros e discos) para registro sonoro.

A história da sonorização espacial inicia-se também nesta fase a partir do momento em que a percepção ou sugestão de um ambiente acústico surge através da audição dos registros sonoros, que além das fontes de som, instrumentos e vozes, trazem ecos e reverberações originais do ambiente da gravação.

Este capítulo introduz o conceito de espacialização sonora e sua natural evolução: a auralização. As principais técnicas e/ou métodos de produção de som bi- e tri-dimensional existentes são apresentadas, com enfoque nas multicanais, que empregam conjuntos de alto-falantes.

2.1 Sonorização espacial (áudio 2D e 3D)

O que torna a sonorização espacial diferente da sonorização convencional? Seria a capacidade de envolvimento ou imersão do ouvinte? Seria a reprodução realista e natural do som? Seria a percepção dos sons vindo dos mais diferentes pontos, ou a impressão de estar em um ambiente específico? Ou mesmo a condição de haver dois ou múltiplos canais independentes e várias caixas de som?

Todas as respostas acima estão corretas no sentido de que exploram uma ou outra característica importante relacionada à qualidade espacial percebida do som. Sabe-se que a sonorização espacial está intimamente ligada aos aspectos psicoacústicos de

⁶ <http://memory.loc.gov/ammem/berlhtml/berlgramo.html>

como percebemos o som com qualidade “espacial”, bem como ligada aos recursos tecnológicos, como número de canais e alto-falantes, que são empregados para produzir um campo sonoro com esta qualidade. O suporte tecnológico adequado e a correta manipulação e condicionamento do material sonoro vão permitir a geração e a projeção de campos sonoros bi (2D) ou tridimensionais (3D).

Abordamos aqui a influência do suporte tecnológico na capacidade de se produzir efeitos espaciais na experiência auditiva. A trajetória da produção da percepção espacial sonora dos sistemas monaurais aos atuais multicanais será apresentada. As tecnologias mais relevantes para gravação (registro), geração (síntese), codificação, distribuição, manipulação e/ou processamento, e finalmente para reprodução de campos sonoros espaciais serão abordadas, bem como exemplos de arquiteturas, sistemas comerciais e meios eletrônicos existentes para áudio espacial.

As questões relacionadas à psicoacústica e os conceitos importantes envolvidos na geração e na percepção da qualidade espacial do som serão vistos mais adiante, quando abordarmos a questão da imersão sonora.

2.1.1 O espaço sonoro do monaural ao multicanal

A produção de som por meio de um transdutor eletroacústico produz uma fonte sonora pontual. O transdutor mais popular é o baseado num cone de alto-falante. Similarmente aos microfones, os alto-falantes possuem um padrão de irradiação, que confere certa diretividade à frente de onda dentro de um ângulo de dispersão.

Um só alto-falante pode fornecer alguma informação espacial sobre o ambiente original da gravação, sugerindo que seja um teatro, um túnel, uma igreja ou uma sala anecóica, porque carrega informação escalar sobre a amplitude do som direto e das reflexões e reverberações do som neste ambiente. Todavia não traz informações vetoriais que permitam mapear a diretividade e posicionamento das fontes sonoras no ambiente original de gravação.

Adicionalmente, ao tocar um programa musical como por exemplo o som de um instrumento de sopro, ainda que possa reproduzir com grande fidelidade o espectro de frequências e amplitudes do instrumento natural, o alto-falante não será capaz de

reproduzir o padrão de irradiação natural deste, tornando a experiência de ouvir um trombone ou uma harpa não natural. Claramente há a necessidade de um sistema que possa acondicionar estes indicadores de diretividade⁷ (*cues*) à reprodução sonora artificial, de forma a aproximá-la do que a audição natural permitiria.

Com dois canais já se pode introduzir mais informação espacial quanto à posição do objeto sonoro no espaço de gravação, podem-se introduzir efeitos de movimento e sugerir um palco sonoro. Entretanto, com dois transdutores discretos e pontuais separados de uma distância de poucos metros, os objetos sonoros ainda vão parecer emergir dos cones dos alto-falantes, isto é, do ponto de vista acústico ainda não seremos capazes de gerar uma frente de onda plana que possua o padrão de irradiação, localização no espaço e diretividade do programa natural. Haverá fenômenos de interferência e construção distribuídos em pontos de uma grade virtual no espaço de audição.

O estéreo

As origens dos experimentos com sistemas bi-aurais – isto é, baseados em 2 canais – retrocedem até 1881. Entretanto, a captação, gravação e reprodução de 2 canais tornaram-se mais populares e conhecidas através do advento do sistema estéreo, cujo principal objeto era permitir a formação de uma “imagem sonora estável” entre as caixas acústicas.

O efeito estéreo foi descoberto por volta de 1930 quando se utilizando dois canais de gravação (dois microfones) descobriu-se poder adicionar um efeito “tridimensional” à gravação, agradável à sensação de espacialidade. O efeito claramente explorava as diferentes indicações temporais (atrasos) entre a recepção de um ouvido em relação ao outro, e a partir disso diversas técnicas de gravação estéreo foram propostas que permitem a captura do som por cápsulas espaçadas entre si e apontadas para planos vizinhos no espaço de gravação. O intuito era poder direcionar os microfones

⁷ em Inglês, usa-se o termo *cues*, como em: *auditory cues*, *directional cues*, etc.

imitando o padrão de diretividade dos ouvidos humanos numa audição de frente para o palco sonoro.

Conforme a técnica de microfonação e estando o ouvinte exatamente posicionado na linha de audição ou no *hot-spot* uma imagem sonora estável poderia ser percebida vindo de uma região entre as caixas. Qualquer desvio entretanto da área de audição correta, e as fontes sonoras passavam a emergir das caixas, não mais de posições virtuais.

No decorrer de décadas, entretanto, o uso do efeito estéreo não se limitou a promover a ilusão do correto posicionamento de fontes sonoras, mas em verdade mostrou-se até mais útil para a sugestão de “efeitos tridimensionais” e exploração de movimentos. Memoráveis álbuns produzidos, por exemplo, na década de 1970 exploraram com maestria este recurso e até hoje são considerados precursores neste sentido, como por exemplo o álbum “The Wall” do grupo britânico Pink Floyd.

Entretanto, já por volta de 1950, as falácias deste sistema em registrar com precisão um campo sonoro 3D foram reconhecidas, embora até pelo menos a metade da década de 1970 houvesse ainda pouca concordância quanto a questões psicoacústicas associadas à localização estéreo (GERZON, 1974).

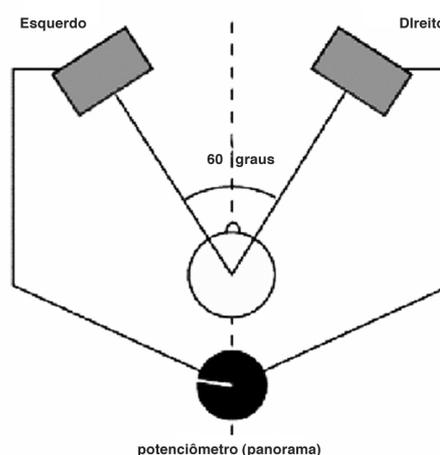


Figura 2 – Efeito estéreo

A Figura 2 ilustra o mecanismo conceitual do efeito estereofônico. Ponderando-se a intensidade de sinal relativa entre as caixas pode-ser sugerir um efeito de “imagem

sonora” entre as caixas. Este efeito é perdido sensivelmente quando o ângulo de abertura entre as caixas excede 60° , e também quando o ouvinte não mais se encontra equidistante de ambas as caixas. A correta região de audição deve estar em um ponto do eixo entre as caixas que atravessa o nariz do ouvinte (BAMFORD, 1995).

Existem várias técnicas de microfonação propostas e utilizadas para gravação estéreo. (VAN ASSELT, 2002) e (MARTIN, 2004) abordam estas técnicas para gravação estéreo, capazes de registrar indicadores *ITD* e *ILD*, e também técnicas de gravação para *surround 3/2* (i.e., 3 canais frontais e 2 de envolvimento, como no arranjo 5.1). As configurações dos microfones sugerem o espaço que eles são capazes de registrar, cuja forma e extensão variam em função dos padrões de captura dos mesmos.

A Figura 3 mostra por exemplo uma configuração comum em “XY”, muito utilizada em gravações clássicas, onde dois microfones cardióides são arranjados num ângulo de 90° . Uma configuração assim pode capturar leituras de variação de intensidade (*ILD – inter-aural level difference*) entre as cápsulas, permitindo a síntese de estéreo de amplitude. Mas não é útil para capturar indicadores de atraso (*ITD – inter-aural time difference*), porque os microfones estão muito próximos no mesmo *hot-spot*.

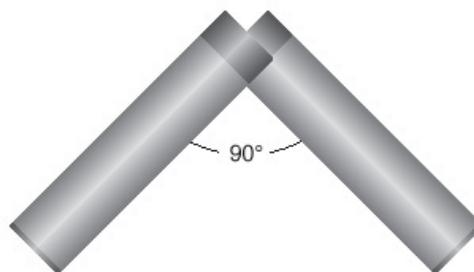


Figura 3 – Microfonação estéreo XY

O website “Stereo” (*Sterophonic Sound*) fornece um histórico interessante sobre o som estéreo em <http://history.acusd.edu/gen/recording/stereo.html>. (BAMFORD, 1995) cobre em sua dissertação maiores detalhes do sistema estéreo. Um retrospecto amplo sobre a história técnica da gravação sonora pode ser consultado em <http://www.recording-history.org/>.

A quadrifonia

A adição de mais alto-falantes permite um aumento nas possibilidades de reprodução “fiel” de um espaço acústico e dos campos sonoros criados por objetos produzindo sons ali dentro. Quatro canais dispostos num plano permitem uma ampliação na capacidade de localização, e melhor reprodução de reverberação. A adição de canais dispostos em cubos ou outras configurações tridimensionais podem ampliar mais este efeito, permitindo a exploração do sentido de elevação e aumentando a qualidade subjetiva de envolvimento do ouvinte (*listener envelopment*) e a percepção da difusão, entre outras propriedades importantes inerentes às salas acústicas, como a impressão subjetiva de “calor”, a clareza (C80), a razão de baixos (*BR – bass ratio*), o índice de qualidade bi-aural (*BQI – binaural quality index*), entre outros (FIGUEIREDO, 2004).

Uma variedade de sistemas e formatos quadrifônicos distintos foram concebidos e introduzidos, inclusive comercialmente na década de 1970. Alguns formatos eram discretos, baseando-se na utilização de canais independentes para obtenção do efeito, o que certamente levava a resultados superiores. Outros, utilizavam uma codificação do tipo matricial, onde por exemplo 3 ou 4 canais eram codificados por meios de operações matriciais em 2 canais (WHITE, 1976), ou mesmo utilizando técnicas de modulação, com a inserção de uma sub-portadora no sinal (GERZON, 1985). Embora tenham sido utilizados sobre vários meios de armazenamento, inclusive o disco de vinil, estes formatos apresentavam resultados melhores quando utilizadas fitas magnéticas. Resultados virtualmente superiores eram obtidos quando da disponibilidade de canais discretos, evitando-se as operações de codificação matricial (ELEN, 2001).

A Tabela 1 abaixo resume uma amostragem dos formatos quadrifônicos desta época.

Tabela 1 – Sistemas quadrifônicos

Formato	Descrição
SQ (<i>Surround Quadraphonic</i> , CBS, 1972)	Matricial, para vinil. Codificava 4 canais em 2 (codificação “4-2-4”)
QS (Sansui)	Matricial. “4-2-4”

CD-4 (<i>Compatible Discrete 4</i> ou <i>Quadradisc</i> , JVC/RCA)	Matricial: sinal quadrifônico codificado com os sinais soma e diferença em disco vinil. Restrições no <i>setup</i> : necessária agulha e cápsula especial, CD-4 demodulador e receptor quadrifônico. Menos popular.
Quad-8 (<i>Quadraphonic 8-Track</i> , RCA)	Discreto, para fita. Baseado num player de 8 pistas quadrifônico, pistas pares combinadas no programa de áudio 1, pistas ímpares no programa de áudio 2. Incompatível com tocadores estéreo.
UHJ (<i>Ambisonic C-Format</i>) ⁸	Matricial. “4-2-4”, “4-2-3”, “4-2-2”. Compatível com mono/estéreo. Para disco, fita, e FM.
<i>Electro-Voice EV Stereo-4</i>	Matricial.

A maioria destes sistemas quadrifônicos analógicos caíram em desuso no decorrer do década de 1980, tendo virtualmente se extinguido após a introdução das mídias contendo áudio digital, como o *CD*, hoje sendo somente encontradas em mãos de colecionadores e audiófilos.

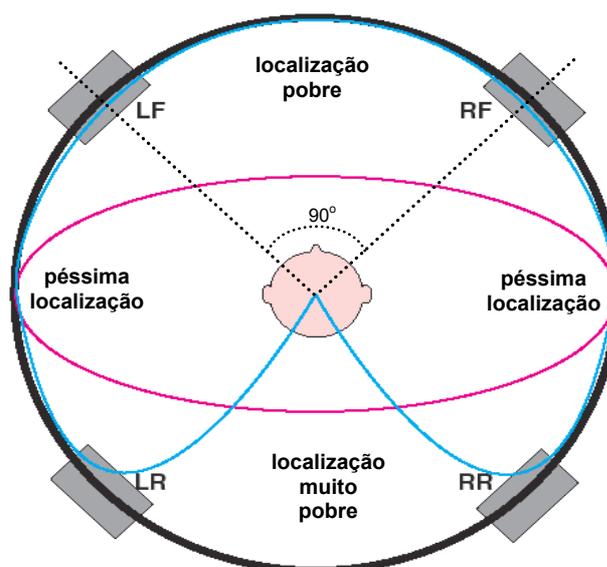


Figura 4 – Configuração quadrifônica (alto-falantes nos vértices de um quadrado)⁹

⁸ Maiores detalhes sobre o formato *UHJ*, muito utilizado para transmissão de gravações Ambisonics comerciais disponíveis, veja o *FAQ* em *Ambisonic.Net* em <http://www.ambisonic.net>. Importante ressaltar que o *UHJ* admite vários arranjos de decodificação e não somente uma configuração quadrangular, rígida nos outros formatos quadrifônicos (RUMSEY, 2003)

⁹ Imagem adaptada de (ELEN, 2001).

Os sistemas quadrifônicos com configuração dos falantes em quadrado (Figura 4) pecam por utilizar ângulos de abertura entre as caixas de 90°, o que prejudica a formação de imagens estáveis na linha entre as caixas duas-a-duas. A localização de fontes sonoras nesta configuração é considerada pobre, e piora quando as imagens são projetadas intencionalmente nas laterais ou, pior ainda, atrás do ouvinte (ELEN, 2001).

Ao se fazer uma imagem circular entre as caixas (círculo preto na Figura 4) o que se ouvia na verdade era um percurso elipsoidal como indicado em vermelho na maioria dos sistemas, exceto para o sistema QS (precursor do Dolby Surround) (ELEN, 2001) que apresentava uma curva percebida no formato cardióide (linha em azul na mesma figura, entrando por trás na cabeça).

Similarmente à incapacidade dos sistemas estéreo em reproduzir campos sonoros 3D, os sistemas quadrifônicos mostraram-se falhos na criação de imagens sonoras estáveis (localização de objetos) e também na codificação de canais de *surround* adicionados aos tradicionais 2 canais. Frequentemente as configurações quadradas para os alto-falantes produziam regiões vazias ou “buracos” no campo sonoro reproduzido. As tentativas de codificação 4-2-4, isto é, codificar-se 4 canais em 2 e na reprodução recuperar-se 4 canais, mostraram-se infrutíferas, com resultados medíocres na localização das fontes sonoras, especialmente porque o campo sonoro reproduzido apresentava sempre um padrão de cobertura diferente daquele apresentado pelo campo sonoro gravado ou intencionalmente projetado (ELEN, 2001).

Entretanto, as fundações dos sistemas quadrifônicos e os conhecimentos adquiridos através dos anos com as diversas gerações de sistemas comerciais e sucessivas tentativas e erros para produzir com sucesso uma codificação *surround* conduziram naturalmente à evolução destes formatos até os atuais sistemas 5.1 (3/2/1).

Som envolvente

Som envolvente ou circundante (*surround sound*) é o conceito de se expandir a imagem espacial do áudio de 1-dimensão (monaural/esquerdo-direito) para 2- e 3-

dimensões. O termo sugere que o som envolve o ouvinte, sendo portanto formado ao seu redor ou sugerindo que o usuário esteja imerso no campo sonoro. Isso é frequentemente exigido para um ambiente de áudio mais realista, ativamente empregado nos sistemas de som para cinema, *technical theatre*, entretenimento caseiro (os “teatros domésticos” ou *home-theater*), *video arcades*, jogos e um grande número de aplicações.

Sucedendo o aprendizado com a fase quadrifônica, novos formatos e configurações foram propostas adicionando-se mais canais. Um sistema de 3 canais contendo o palco estéreo tradicional nos 2 canais frontais e um canal especialmente contendo programação sonora envolvente (canal *surround*) foi proposto.

O grande mercado consumidor que estimulava estes desenvolvimentos era o cinema. Grandes explosões e demais sons de efeitos especiais usualmente são bastante energéticos nas baixas frequências. Um canal de baixas frequências para este material era um bom candidato para canal adicional. Assim, foi também adicionado um canal contendo basicamente efeitos em frequências baixas, resultando no total de 3 canais mais um de efeitos (3.1). O canal de efeitos, por ter largura de banda menor (usualmente de 20 a 120Hz) passou a ser referido como “.1”, e ficou conhecido como o canal dedicado a efeitos de baixas frequências (*LFE – low frequency effects*), normalmente mapeados em um falante do tipo *sub-woofer* (RUMSEY, 2003).

O *Dolby Surround* (analógico) também foi introduzido nesta fase, explorando a codificação de um canal *surround* nos 2 canais principais (estéreo, esquerdo e direito). Nos sistemas que tinham um decodificador *Dolby Surround*, era possível extrair este canal adicional dos canais principais, sem entretanto adicionar artefatos à reprodução estéreo tradicional. Até hoje muitos álbuns musicais são gravados com *Dolby Surround*, que permite a reprodução de efeitos circundantes interessantes, como por exemplo a “separação” de alguns sons do corpo misturado no conjunto estéreo. É muito comum encontrar-se gravações comerciais neste formato que mapeiam, por exemplo, *back vocals* para o canal *surround*.

Após o formato 3.1, a próxima geração buscou meios de se corrigir o “buraco” que frequentemente forma-se no centro do palco sonoro frontal quando a direção do som

codificado em estéreo não corresponde nem à direção de uma das caixas nem se encontra no espaço subentendido entre as caixas. Um falante cobrindo a região frontal (central) foi então introduzido, resultando num sistema com 3 canais para serem apresentados através de alto-falantes frontais, um canal *surround* e um *LF* (arranjo 4.1). Esta configuração atendia novamente outra demanda da indústria cinematográfica: a de garantir que os sons nos cinemas vindo da direção da tela parecessem realmente emergir por detrás da mesma, o que frequentemente não era conseguido pelo posicionamento das caixas nas laterais da tela, daí a necessidade de um reforço frontal, especialmente para as falas.

Ainda com 4 canais, é importante comentar o famoso formato LCRS, adotado como padrão de áudio no cinema. Os quatro canais são: L e R são os canais frontais laterais, C é o canal central que fica por detrás da tela, e S, o canal *surround*, que usualmente é replicado em diversos pontos na parte traseira e laterais da sala (e adequadamente atrasado em relação aos canais frontais dependendo do tamanho da sala). O sistema *Dolby Surround* introduzido por volta de 1982 e o *Dolby Pro Logic*, introduzido por volta de 1987 são sistemas compatíveis LCRS. Entretanto, é importante notar que ambos não são baseados em 4 canais discretos, e utilizam para transmissão somente 2 canais, consistindo em formatos do tipo “4-2-4” (BAMFORD, 1995). A Tabela 5, adiante, lista os produtos multicanais da Dolby e outros.

2.1.2 Configurações multicanais

Incrementando o efeito envolvente, finalmente foi proposta uma configuração com 3 canais frontais e 2 canais de *surround* discretos (3/2), bastante satisfatória para permitir um maior nível de imersão, e que posteriormente inspirou em 1992 a padronização constante da norma ITU-R BS.775-1, que ficou conhecida como a norma do sistema 5.1 (INTERNATIONAL TELECOMMUNICATIONS UNION, 2002).

Configurações de 5, 6 ou mesmo 7 e 8 canais/alto-falantes tornaram-se mais populares após a publicação deste padrão. Estes sistemas são úteis para se criar efeitos complexos de envolvimento, reverberância, e criar imagens sonoras no plano

(2D) ou até no espaço (3D). Podem ainda sugerir a posição de objetos sonoros numa cena audiovisual, e simular movimentos com mais realismo, graças à disponibilidade de múltiplos canais independentes (discretos) no meio digital e à disponibilidade de técnicas digitais para o processamento de sinais.

Recomendações quanto a práticas de desenvolvimento e configurações para sistemas multicanais *surround* são abordadas no documento informativo AESTD1001.1.01-10 (AUDIO ENGINEERING SOCIETY, 2001). Os sistemas multicanais podem ser designados por “*n/m/s*”, indicando o uso de “*n*” canais frontais e “*m*” traseiros/laterais, mais “*s*” canais de baixas frequências (*LF* ou *LFE*). Alternativamente, a nomenclatura mais popular “*n.m*” indica o uso de “*n*” canais frontais e *surround* mais “*m*” canais de baixas frequências, como 5.1, 6.1, 7.1, e 10.2.

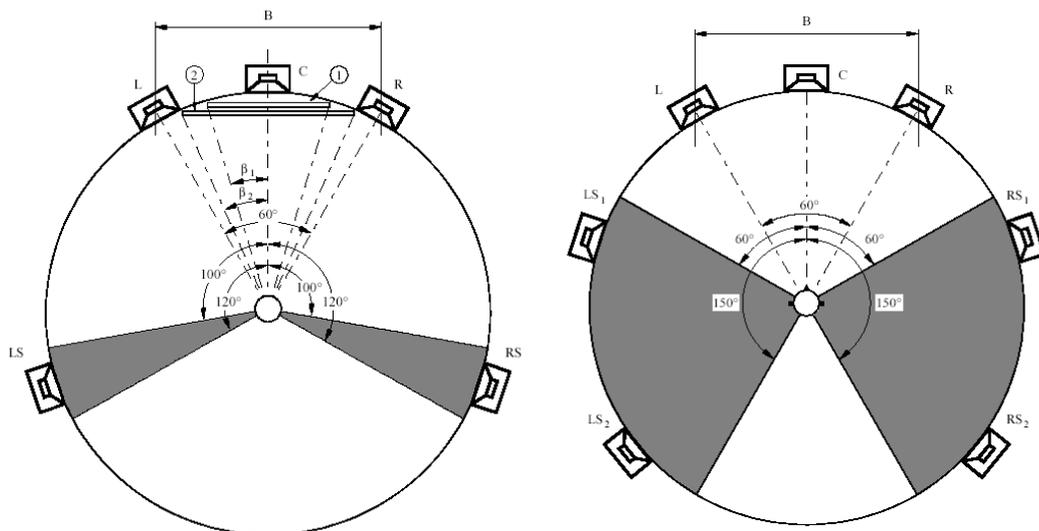


Figura 5 – Configurações para 5.1 e 7.1 canais (padrão ITU-R BS.775-1)

Campos sonoros complexos usualmente requerem programas (canais) distintos enviados para cada alto-falante, não sendo portanto adequados em contextos onde há limitações no número de canais, como nos antigos sistemas analógicos. A representação do áudio digitalmente é praticamente um pré-requisito para a implementação de bons sistemas para auralização tridimensional, haja visto que é possível com facilidade associar (multiplexar) programas distintos de áudio no mesmo feixe de bits. Alguns trabalhos mais atuais têm ainda investigado e explorado a capacidade de configurações *n.m* em permitir a síntese de campos 2D, planos onde

posição e imagens estáveis podem ser geradas, e até a síntese de campos 3D (MILLER, 2003).

No limite, podemos conceber um campo sonoro criado por alto-falantes densamente distribuídos numa configuração específica e que reproduzam com muita resolução e qualidade do campo sonoro produzido por diversas fontes sonora irradiando imersas num espaço acústico como por exemplo uma orquestra tocando numa sala de espetáculos, com suficiente informação quanto a diretividade, posicionamento e acústica do ambiente (DEVRIES, 1999).

A reprodução de uma audição 3D deve considerar os seguintes casos:

- i) o material sonoro foi gravado adequadamente com o auxílio de microfones registrando o campo sonoro real, ou
- ii) o campo sonoro 3D reproduzido foi construído de forma artificial, através da aplicação de um modelo de propagação acústico satisfatório aplicado ao material sonoro “seco” (anecóico).

As técnicas que conseguirem capturar todos os parâmetros acústicos do meio e modelar todos os fenômenos acústicos que ocorrem numa extensa faixa de frequências e possam se comportar com resultados semelhantes nos dois casos de uso acima serão provavelmente as mais corretas e completas do ponto de vista matemático-físico.

Várias configurações têm sido objeto de testes e teorias para reconstrução de campos sonoros nos últimos anos. Citamos os sistemas perifônicos Ambiphonic, Ambisonics e *Wave Field Synthesis (WFS)* como exemplos relevantes abordados neste trabalho mais adiante. (GERZON, 1974, 1976, 1985, 1992); (GLASGAL, 2004); (INFORMATION SOCIETY TECHNOLOGIES, 2001); (BRIX, 2001)

Os sistemas Ambiphonic e Ambisonics usualmente registram campos sonoros 2D e 3D reais ou calculam campos artificiais, reproduzindo através de matrizes de alto-falantes dispostos em configurações ao redor do(s) ouvinte(s), requerendo um mínimo de 4 falantes e um máximo indeterminado, mas que usualmente gira na faixa

de 8 a 32 falantes, por questões de custo-benefício, ou porque o acréscimo além de certo número não implica em ganho de qualidade proporcional.

Outras técnicas, como por exemplo *WFS*, fundamentadas na teoria da propagação de frentes de onda de Huygens (VOGEL, 1993), requerem uma discretização do espaço de audição e o emprego de um número elevado de alto-falantes¹⁰, densamente distribuídos ao redor do espaço de audição. Estes alto-falantes assim dispostos agem como fontes sonoras “secundárias” que têm como função reproduzir as frentes de onda de fontes primárias reais localizadas em outro lugar do espaço.

Estas técnicas podem eventualmente necessitar de dezenas ou até mesmo centenas de alto-falantes para uma simulação de alta qualidade, e também que a distância entre eles (espaçamento) seja o menor possível, para minimizar a formação de artefatos nas frequências mais altas.

As soluções para áudio espacial não se esgotam com os formatos de auralização mais sofisticados. No meio científico e industrial a todo o momento há proposições tanto de configurações experimentais de alto-falantes, quanto de modelos acústicos e formatos de codificação sonora 3D, muitas vezes voltadas para aplicações específicas. É o caso, por exemplo, da configuração em anéis ortogonais ilustrada na Figura 6 (à esquerda), pesquisa do laboratório japonês Suzuki (*Advanced Acoustics Information Systems*) da Universidade de Tohoku¹¹ para avaliar a percepção sonora 3D, bem como da incomum (e no mínimo curiosa) cadeira *surround*¹².

¹⁰ de 32 até 256, ou mais.

¹¹ <http://www.ais.riec.tohoku.ac.jp/> (acesso em 09 set. 2004).

¹² “*The Surround Chair*” (*SS Chair*) é comercializado pela empresa D+S Sound Labs (Flórida, EUA). Trata-se de uma cadeira dotada de 5 alto-falantes montados numa configuração envolvente. O produto aponta como principal desvantagem as reflexões existentes nas salas de audição, que alteram a experiência direcional do som até os ouvidos. Custa US\$3000,00.

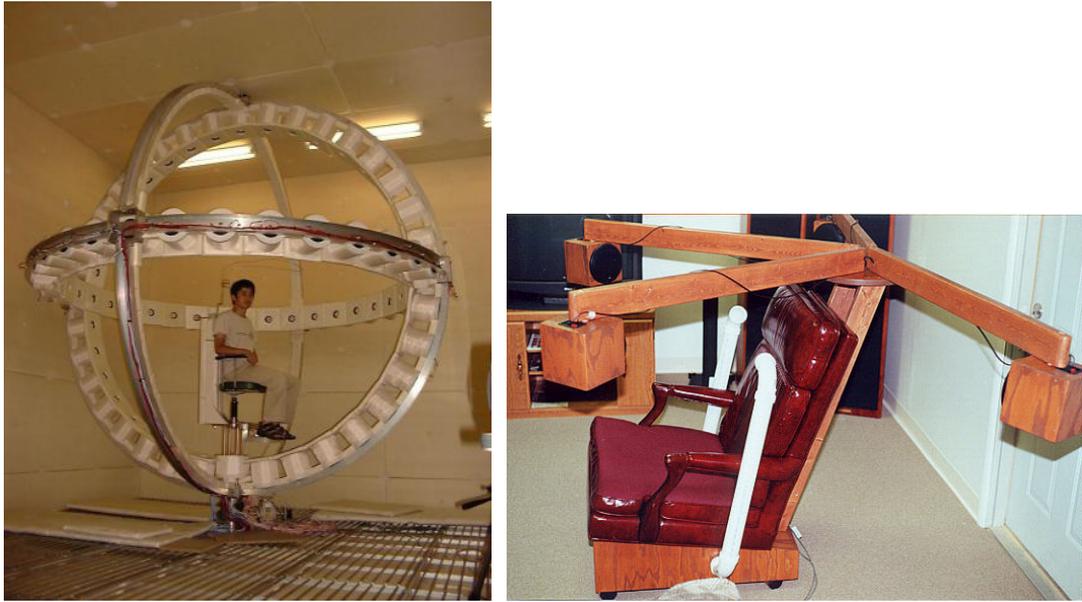


Figura 6 – Configurações experimentais (protótipos) de testes de sistemas para som 3D

É claro que a tarefa de recriar ou sintetizar campos sonoros com realismo não é trivial, pois requer um nível de precisão, sincronização, e correção de fase e amplitude muito crítico entre os sinais sendo reproduzidos por diferentes transdutores distribuídos no espaço de audição, tal que as ondas interferentes possam produzir um campo coerente, sem batimentos, pontos nulos ou cancelamentos que produzam artefatos, e se obtenha uma cobertura homogênea pelo espaço de audição.

Sem falar que ainda existem outros fatores de difícil controle que afetam o resultado como a acústica local, os fenômenos de difração, absorção e reflexão causados por objetos no meio de audição (inclusive os próprios gabinetes dos falantes e os ouvintes).

Alternativamente, existem os sistemas baseados em fones de ouvido, que podem produzir, num ambiente mais controlado e com muito realismo, a percepção de um ambiente acústico tridimensional.

As configurações multicanais podem ser descritas em função do número de canais finais de reprodução, isto é, em função da configuração e do número de alto-falantes empregados na matriz reprodutora. A Tabela 2 traz uma breve listagem dos arranjos típicos existentes, em função do número de alto-falantes empregados. Nos cinemas,

auditórios e espaços amplos pode-se (e até é comum) adotar uma configuração ou arranjo de canais da tabela acima, mas mapeada num número maior de alto-falantes, de forma a cobrir regularmente vastas extensões onde o número de canais típico da configuração escolhida não atenderia.

Tabela 2 – Configurações multicanais segundo o arranjo dos alto-falantes

Número de canais na reprodução	Configuração
Dois canais (2/0/0)	estéreo típico
Três canais (3/0/0)	estéreo mais canal de reforço frontal
Quatro canais (3/1/0) (LCRS)	estéreo mais canal central e canal de efeitos (envolvente)
4, 6, 8, 12, 16 canais (Ambisonics)	caixas formando geometrias regulares (polígonos 2D e 3D)
5.1 canais (3/2/1) (ITU-R.BS 775-1)	estéreo frontal mais canal central e estéreo envolvente
6.1 canais (3/3/1)	5.1 mais canal central traseiro, efeito envolvente melhorado
7.1 canais (3/4/1) (ITU-R.BS 775-1)	5.1 mais dois canais envolvente adicionais
7.1 canais (5/2/1) (SDDS)	5 canais frontais mais dois canais envolventes, para cinemas
10.2 (7/3/2)	5.1 mais dois canais laterais frontais, mais dois canais frontais em elevação, mais um canal central envolvente traseiro, mais um canal de baixas frequências.
16, 32, 64, 128, ... canais (WFS)	painéis ou matrizes de alto-falantes densamente distribuidos

Em termos de geometrias de posicionamento dos alto-falantes, a maioria dos sistemas e configurações multicanais (principalmente as comerciais e domésticas) emprega geometrias planas, isto é, os alto-falantes são dispostos num plano, criando um campo sonoro de referência planar.

Geometrias tridimensionais são também empregadas em muitos sistemas para o posicionamento dos alto-falantes, em especial nos formatos SVPA, Ambisonics e *WFS* (que serão vistos adiante). Estes exploram arranjos octaédricos, cúbicos, dodecaédricos e esféricos, além dos arranjos poligonais planares e regulares típicos, como os hexagonais, octogonais, circulares ou matrizes lineares.

2.1.3 Espacialização e Auralização

Neste trabalho procuramos diferenciar os sistemas de áudio 2D/3D entre aqueles que *sugerem* campos sonoros envolventes, ambiências e uma sensação de espacialidade (sistemas para espacialização) e aqueles que foram concebidos com o objetivo primário de *simularem* ou reproduzirem com o máximo de fidelidade possível campos sonoros 2D/3D legítimos (sistemas de auralização), estes usualmente mais precisos na capacidade de prover a percepção da diretividade e distância (posição) dos objetos sonoros dispostos na cena, bem como reproduzir com mais realismo as características acústicas de uma cena. É indicado neste ponto estabelecer uma separação entre os conceitos de espacialização sonora e auralização, especialmente para melhor caracterizar este último.

Espacialização

Espacialização se refere à produção de um efeito aural tridimensional perceptível, que pode explorar diversos mecanismos psicoacústicos inerentes à percepção da localização, do envolvimento por um ambiente acústico, ou de movimento. Várias técnicas de processamento de sinais podem ser empregadas para forjar efeitos desta natureza, sem necessariamente utilizar um modelo formal de propagação de frentes de onda.

A maior parte dos sistemas comerciais em utilização no mercado presta-se mais à criação de ambiências e efeitos de envolvimento e espacialização para filmes e para entretenimento de modo geral (por exemplo jogos) do que propriamente à reprodução de campos sonoros 2D/3D coerentes. Neste mercado, até onde a cultura musical estabelecida alcance, não é importante a localização precisa dos objetos sonoros ou mesmo a percepção de sua formação (imagem) próxima ou muito além do ouvinte, em pontos do espaço não correlacionados com a posição dos alto-falantes.

O sistema surround 5.1 que se popularizou nas salas de cinema e posteriormente nos equipamentos de *DVD* para *home theater* é um exemplo de sistema multicanal efetivamente padronizado e estabelecido no mercado. Como ocorreu com o MP3, nascido da padrão MPEG-1 da ISO/IEC, o sistema 5.1 ganhou notoriedade a partir de

sua padronização pela norma ITU-R BS.775-1, publicada pela União Internacional de Telecomunicações entre 1992-1994¹³, e foi virtualmente adotado pela indústria cinematográfica graças às possibilidades enormes para permitir a reprodução de efeitos de baixa frequência, como explosões, a lateralização de movimentos sonoros, o reforço na percepção de movimentos e translações de objetos do vídeo, e para a centralização da voz dos interlocutores.

Auralização

Auralização se refere mais especificamente à produção de campos sonoros tridimensionais matemática e/ou fisicamente corretos e coerentes com o esperado na realidade da propagação dos sons em um ambiente acústico. Frequentemente as técnicas ou métodos de auralização envolvem modelos de propagação sonora formais e simulações para obter o efeito desejado.

Neste sentido, a auralização é uma versão mais sofisticada de espacialização, onde, por exemplo, as diferenças de fase entre os sons emergindo de diferentes alto-falantes é crucial para se obter a fusão dos mesmos numa percepção integrada do meio acústico e seus objetos sonoros. Os sistemas genéricos de espacialização não têm restrições tão rígidas, podendo produzir efeitos tridimensionais artificiais ainda muito convincentes ou satisfatórios com combinações mais livres.

Auralização é o método pelo qual se reproduz um campo de som (2D e/ou 3D) criado por uma fonte situada no espaço, onde também se situam o(s) ouvinte(s) e eventuais objetos adicionais componentes da cena audiovisual. Algumas definições referem-se à “técnica para a reprodução do campo sonoro de uma sala de concertos” (<http://ymec.com/hp/pref2/theme02.htm>).

Auralização é um termo introduzido em analogia à visualização, para descrever a produção audível (e imaginária) de campos sonoros. KLEINER et al. (1993) define auralização como “... o processo de tornar audível, por modelamento físico ou matemático, o campo sonoro de uma fonte no espaço, de maneira a simular a

¹³ e em 2002, com adições (emendas).

experiência bi-aural de audição numa dada posição dentro do espaço modelado”. Segundo esta definição, não é nem possível produzi-la com técnicas de panorama (como SVPA) porque não há em si um modelamento físico ou matemático do campo sonoro produzido no espaço pela fonte embutido nestes métodos.

Do ponto de vista de construção ou representação de ambientes e cenas acústicas, há vários métodos em acústica arquitetônica disponíveis para modelamento, os mais importantes e usuais sendo o de fonte-imagem (*Image-source*, (ALLEN, 1979) e (BORISH, 1984)), traçado de raios (*Ray tracing*¹⁴) e traçado de feixes (*Beam tracing*, (FUNKHOUSER, 1999)).

Segundo KUTTRUFF (1993) “o termo ‘auralização’ denota um novo procedimento que nos permite criar impressões auditivas a partir de respostas (impulsivas) de salas simuladas por computador”. Ele cita que a idéia de auralização é tão antiga quanto de 1934, quando Spandöck propôs tocar sinais musicais com a faixa de frequência alterada dentro de modelos (físicos) em escala reduzida, gravá-los novamente de dentro da maquete, e apresentá-los a um ouvinte em fone de ouvido após retransformá-los de volta à faixa de frequência original. Schroeder e outros em 1961 propuseram procedimentos similares mas usando-se modelos computacionais, primeiro aplicados a caixas simples retangulares por ALLEN e BERKLEY (1979). A primeira auralização bi-aural (baseada em salas retangulares) é devida a Pösselt, em 1987.

2.2 Métodos de auralização

Existem diversos métodos para permitir a síntese de áudio espacial ou tridimensional (3D), que podem ser classificados segundo a técnica de renderização ou reprodução final utilizada (meio), basicamente bi-aural ou multicanal. Ou também segundo o objetivo final (fim), que pode ser o de emular a reverberância e envolvimento de um ambiente físico específico ou de permitir a localização de objetos sonoros no espaço.

¹⁴ Referências para esta técnica incluem Schroeder, M.R. 1970, Wayman, J.L. 1980 e outros (não consultadas).

As técnicas de auralização bi-aurais utilizam-se das teorias levantadas a mais de um século por Lorde Rayleigh, sobre os indicadores auditivos que influenciam na percepção da diretividade do som. Seja qual for a mistura de fontes sonoras realizada, o mapeamento final sempre convergirá para dois canais, o direito e o esquerdo, direcionados a cada um dos ouvidos.

As técnicas baseadas em reprodução multicanal se caracterizam pela renderização para matrizes de alto-falantes, isto é, os sons tocados são mapeados para múltiplos canais. Estas são mais adequadas para emular ambientes acústicos reais para mais de uma pessoa ao mesmo tempo, sem a necessidade de utilização de fones de ouvido, projetando campos sonoros no ambiente tridimensional ao redor dos ouvintes, como ocorre na natureza.

A Tabela 3 compara características de ambos os métodos de auralização, identificando-se algumas vantagens e desvantagens de cada abordagem.

Tabela 3 – Comparando os métodos de auralização bi-aural e multicanal

bi-aural	multicanal
vantagens 👍 e desvantagens 👎	
👍 estabilidade do ouvinte e da audição indiferente à posição do fone no ambiente	👎 estabilidade da audição em região limitada no ambiente (“ <i>hot spot</i> ” versus “ <i>sweet spot</i> ”)
👎 criticidade da posição da cabeça no campo sonoro (para dipólo estéreo, com cancelamento de <i>cross-talk</i>)	👍 robusteza em relação a movimentos da cabeça na área de audição, sem perda da estabilidade da imagem (para várias técnicas)
👎 liberdade de movimento restrita para fones com fio	👍 liberdade de movimento assegurada dentro do volume de audição
👍 baixo custo do ambiente (não necessário câmeras anecoicas ou especialmente tratadas)	👎 custo elevado, proporcional ao número canais (caixas acústicas)
👎 uso individual	👍 alcance simultâneo a múltiplos ouvintes
👎 requer uso de fone de ouvido (exceto para sistema dipólo estéreo utilizando-se par de alto-falantes)	👍 sem necessidade de usar fones de ouvido
👎 campo sonoro ideal para uma pessoa	👎 campo sonoro ideal para uma região no ambiente

☞ instabilidade da medida HRFT	☑ não restrito ao uso de <i>HRTF's</i>
☑ formato tradicional para codificação espacial utiliza 2 sinais (<i>HRTFs</i>)	☞ formato tradicional de codificação espacial utiliza mais de 2 sinais
☑ requer 2 canais para sonorização final	☞ requer n canais para sonorização final ($2 < n < \infty$)
☞ necessária equalização do fone para excelentes resultados	☞ necessária equalização do ambiente (quase anecóico) para excelentes resultados
☞ problemas de localização espacial com fones de ouvido, principalmente confusão frente-trás	☑ confusão frente-trás reduzida ou eliminada com configurações envolventes (panto ou perifônicas)
☞ formação da imagem dentro da cabeça	☑ formação da imagem externa (no ambiente)
☞ diferenças sensíveis entre ouvintes para mesmas <i>HRTFs</i>	☑ experiência semelhante do campo entre ouvintes diferentes
☑ menor complexidade na infra-estrutura	☞ infra-estrutura complexa

Os efeitos da padronização e alcance dos sistemas surround 5.1 são sentidos também junto à comunidade acadêmica, que voltou interesse na matriz multicanal 5.1 como uma possibilidade para se auralizar objetos sonoros codificados noutros formatos menos populares ou em fase de desenvolvimento, propondo ferramentas transcodificadoras (MILLER, 2003). Sistemas para emulação da acústica de salas de concerto para uma audição mais realista de uma peça musical podem hoje ser encontrados transcodificados para 5.1.

Para os sistemas de realidade virtual do futuro somente a espacialização de eventos sonoros não será suficiente para impor o realismo necessário. Faltarão a percepção da diretividade, da localização dos objetos no espaço. Por exemplo, através da estereoscopia hoje é possível localizar um objeto visual numa *CAVE* formando-se bem próximo ao usuário, como também é possível navegar no ambiente ao redor de um objeto projetado. O mesmo espera-se para os atributos sonoros deste objeto e dos atributos acústicos do ambiente: a capacidade de percebê-lo em posições tanto além do ambiente físico da *CAVERNA*, quanto em posições dentro dela, e navegar ao redor dos objetos, inclusive levando-se em consideração as diferentes propriedades e padrões de irradiação sonora que cada objeto possui no espaço.

Os seguintes métodos de auralização serão abordados nesta seção, em função de sua relevância e validade para as aplicações em vista: bi-aural, panorama de amplitude vetorial, Ambisonics, Ambiophonics, e *Wave Field Synthesis*.

2.2.1 Bi-aural

Os métodos bi-aurais de produção de campos sonoros 3D baseiam-se na geração de 2 canais de som, um para cada ouvido, contendo todas as informações temporais e espaciais necessárias para que o sistema auditivo perceba ou possa reconstruir um espaço acústico tridimensional. Para tanto, estes métodos fazem uso dos principais indicadores ou leituras que o mecanismo de audição utiliza para inferir a diretividade dos sons.

Indicadores espaciais auditivos são produzidos na natureza por reflexões e absorções ao redor da cabeça, torso e ouvido externo, e têm sido extensivamente catalogados como função da posição 3D dos objetos sonoros em respeito à cabeça. Os principais indicadores espaciais auditivos são conhecidos por ITD (*inter-aural time difference* ou diferença temporal interauditiva) e ILD (*inter-aural level difference* ou diferença de intensidade interauditiva) e fornecem respectivamente a diferenças de tempo e intensidade em que os sons chegam a um ouvido em respeito ao outro.

Em outras palavras, o *ITD* explora as diferenças de fase (tempos de chegada) percebida entre um ouvido e o outro, e o *ILD* explora as diferenças na intensidade percebida de um som entre os ouvidos (veja Figura 7). Usualmente o primeiro indicador funciona melhor para os sons de baixas frequências, enquanto o segundo, para os sons de altas frequências. Com bases nestas diferenças, o sistema auditivo pode inferir a diretividade dos sons no meio.

Para se proceder a auralização, o percurso dos sons entre a fonte e cada ouvido são adequadamente modelados por meio de funções de transferência relativas à cabeça (*HRTF* ou *Head Related Transfer Function*), e os dois canais de som são apresentados ao ouvinte preferencialmente via fones de ouvido. Os canais também podem ser enviados aos ouvidos através de alto-falantes adequadamente posicionados em relação à cabeça do ouvinte, utilizando-se técnicas de dipólo estéreo (com cancelamento da conversa cruzada entre os alto-falantes). A síntese de som

espacial usa então técnicas de filtragem lineares para impor pares (estéreo) de funções de transferência (uma para o falante esquerdo do fone, outra para o direito) cujas respostas em amplitude e atraso reproduzem aquelas dos indicadores auditivos conhecidos.

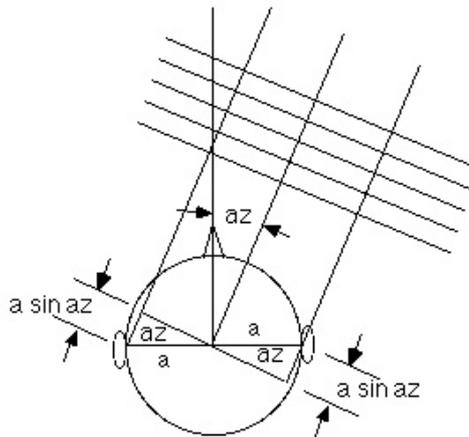


Figura 7 – Incidência de onda plana em uma cabeça indicando diferença em extensão no percurso sonoro entre os ouvidos

Diversos trabalhos abordam as técnicas e sistemas para reprodução bi-aural, modelamento do ouvinte, modelamento, cálculo e processamento de *HRTF's*. Citamos, por exemplo, (SAVIOJA, 1999), (SUSNIK, 2003), (MOUCHTARIS, 1998), (BEGAULT, 1998), (CHEUNG, 1998).

2.2.2 Sistema Vetorial de Panorama por Amplitude (SVPA)

O Sistema Vetorial de Panorama por Amplitude (SVPA) – ou, originalmente do Inglês, *Vector Based Amplitude Panning (VBAP)* – é um método que permite o deslocamento das fontes virtuais e seu posicionamento no espaço com precisão. A técnica não limita o número de alto-falantes empregáveis, e utiliza-se num dado instante o menor número de alto-falantes necessário: um, dois ou três. Nesta técnica o conceito de panorama por amplitude foi estendido a um modelo mais genérico que permite combinações de alto-falantes em localizações arbitrárias (RUMSEY, 2003).

SVPA é mais eficiente computacionalmente que os métodos bi-aurais, que utilizam filtros *HRTF*. Alguns autores consideram SVPA melhor que Ambisonics. PULKKI (1997) considera SVPA mais flexível que Ambisonics porque o posicionamento dos

falantes em SVPA é mais livre. Esta afirmação no entanto é questionável, dado que se pode escolher previamente uma configuração e distribuição desejada para os alto-falantes, e então obter-se as coordenadas destes para serem utilizadas nas equações de decodificação. A cada nova configuração, uma nova situação de coordenadas deve ser usada.

Diferentemente do Ambisonics, o método de SVPA em muitos casos, principalmente em três dimensões, necessita de mais de 4 canais de som (PULKKI, 1997). Como em todos os métodos de panoramas por intensidade (amplitude), a fonte virtual não pode ser posicionada fora do arco ou região ativa, enquanto que com Ambisonics, uma fonte virtual pode ser posicionada em qualquer ponto dentro da esfera de referência (MALHAM, 1987).

O erro máximo na localização da fonte virtual com SVPA é proporcional às dimensões da região ativa. Se uma precisão maior é necessária numa área de audição maior, então as dimensões das áreas ativas devem ser reduzidas, e isso é feito aplicando-se mais alto-falantes à região pretendida. A Figura 8 mostra um exemplo de posicionamento de três alto-falantes numa configuração esférica, cobrindo um segmento triangular onde fontes sonoras nos vértices são precisamente localizadas (fontes reais) e imagens virtuais são formáveis na região de cobertura entre as caixas.

Comparativamente, para que se tenha melhor precisão na localização de fontes virtuais com Ambisonics pode-se aplicar ordens mais altas, como por exemplo Ambisonics de 2^a ou 3^a ordem, que naturalmente implicam num número mínimo de alto-falantes também maior, respectivamente 8 ou 16 alto-falantes.

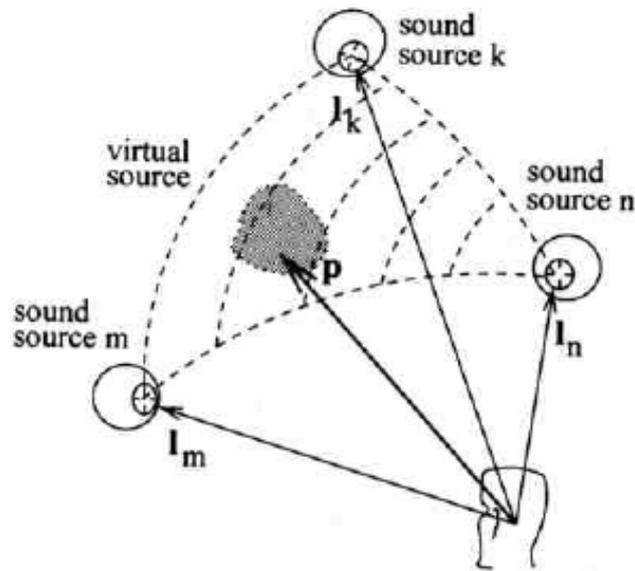


Figura 8 – Exemplo de configuração de caixas acústicas para SVPA¹⁵

Vários sistemas de áudio 2D e 3D para realidade virtual utilizam técnicas de panorama por amplitude (OGI, 2003), (NAEF, 2002). Esta escolha é comum em função do baixo custo da técnica, simplicidade teórica, e pela flexibilidade no posicionamento dos alto-falantes, que não precisam atender a uma disposição rígida ou geometricamente restritiva ou regular.

Nos sistemas de espacialização sonora que exploram as indicações ou variações de intensidade de sinal entre as caixas, como são os baseados em panorama, na apresentação dos sons há uma maior tolerância às incoerências sônicas devido ao posicionamento fora do *hot-spot*. Nos sistemas onde as indicações temporais ou de fase são exploradas, há um maior desconforto à medida que se distancia do *hot-spot*, com a introdução ou formação de artefatos e distorções.

2.2.3 Ambisonics

Ambisonics foi proposto por Gerzon na década de 1970 (GERZON, 1973, 1974, 1976). Trata-se de um método de perifonia¹⁶.

¹⁵ imagem de (PULKKI, 2001c).

Temos duas formas para introduzir esta técnica. Na primeira forma, mais formal, recorreremos às suas fundações teóricas matemáticas, embutidas nas equações de representação de frentes de onda, utilizando-se a teoria das harmônicas esféricas e conceitos oriundos do teorema de Green. O Anexo III traz uma introdução simplificada às fórmulas de codificação e decodificação Ambisonics. BAMFORD (1995) aborda esse desenvolvimento de forma detalhada em sua dissertação.

A segunda forma, de fato mais fisicamente sugestiva, permite compreender o conceito de registro tridimensional no sistema de som Ambisonics de uma forma prática, utilizando-se para isso de uma extensão da técnica de Blumlein para gravação estéreo. Em verdade, na prática Ambisonics teve sua origem muito ligada à procura de uma técnica que permitisse gravar um campo sonoro 3D para posterior reprodução análoga.

Blumlein introduziu uma técnica para se gravar estéreo conhecida por “MS”, na qual um par de microfones (*mid* e *side*) era utilizado: um *onidirecional*, que captava toda a região esférica a seu redor, i.e. capturava L+R, e um segundo microfone com padrão do tipo *figura-de-8*, apontando para a esquerda, num eixo horizontal. Um diagrama polar deste tipo ao captar um sinal vindo de um lado registra um deslocamento positivo em relação ao outro lado, registrando de fato no eixo do segundo microfone uma diferença de pressão entre os lados, i.e, L-R.

Uma simples operação algébrica permite recuperar os canais direito e esquerdo através da soma e da diferença entre estes sinais. Similar efeito de captura estéreo pode se obter com cápsulas cardióides dispostas horizontalmente a 90° sem necessidade de algebrismo.

¹⁶ Perifonia refere-se à cobertura esférica total do campo sonoro, levando-se em conta tanto o campo horizontal (2D) quanto a elevação, resultando numa cobertura 3D (GERZON, 1973). Já pantofonia refere-se somente ao campo sonoro 2D, no plano horizontal.

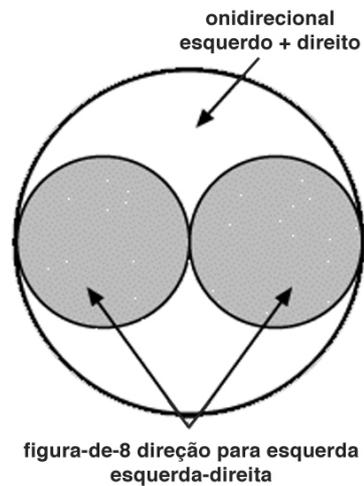


Figura 9 – Técnica de Blumlein para captação estereofônica (MS)

Se estendermos esta técnica para 3 dimensões, teremos uma implementação Ambisonics, capaz de codificar fontes sonoras de qualquer direção no espaço 3D. A técnica equivalente empregaria uma cápsula onidirecional mais três cápsulas em figura-de-8, totalizando uma captura com 4 diagramas polares complementares.

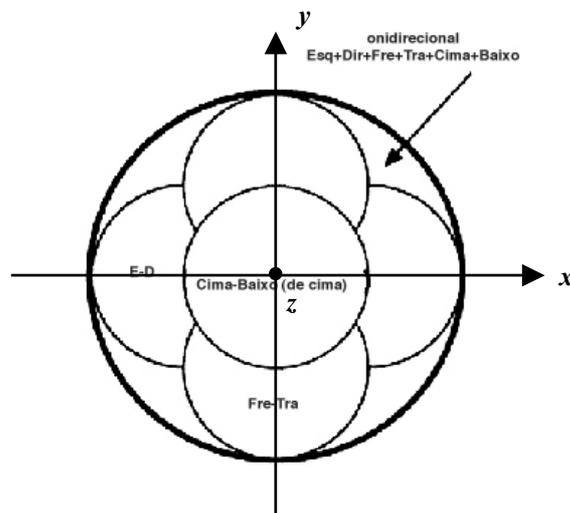


Figura 10 – Extensão da técnica de Blumlein, permitindo o registro sonoro de um campo 3D (eixo z central apontando para cima)

A Figura 11 mostra numa perspectiva tridimensional os lóbulos de captura em cada eixo. Quatro canais são necessários para se registrar completamente um campo

sonoro 3D, um arranjo multicanal (formato de codificação multicanal) que ficou conhecido por “*B-Format*”¹⁷.

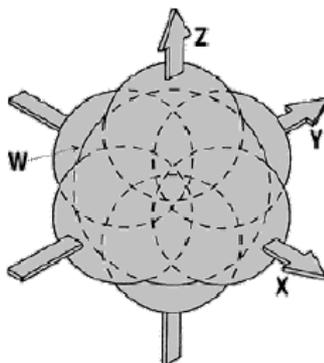


Figura 11 – Perspectiva do padrão de captura Ambisonics (1ª ordem)

Dependendo da ordem considerada do sistema Ambisonics e também se estamos considerando ou não a elevação (se 2D ou 3D) teremos um número diferente de equações, isto é, um número diferente de sinais (canais) que serão necessários para registrar a informação sonora mais a informação espacial. A Tabela 4 abaixo mostra para cada circunstância quantos canais (mínimos, mas não máximos) serão necessários empregar.

Tabela 4 – Número de canais necessários por ordem da implementação Ambisonics

no. de canais	sinais (canais)	ordem da implementação Ambisonics
3	<i>WXY</i>	primeira ordem sem elevação
4	<i>WXYZ</i>	primeira ordem com elevação
5	<i>WXYUV</i>	segunda ordem sem elevação
6	<i>WXYUVW'</i>	segunda ordem sem elevação, dados para ordem mista 1ª/2ª
9	<i>WXYZRSTUV</i>	segunda ordem com elevação
10	<i>WXYZRSTUVW'</i>	segunda ordem com elevação, dados para ordem mista 1ª/2ª
16	<i>WXYZRSTUVKLMNOPQ</i>	terceira ordem com elevação

Para Ambisonics de 3ª ordem, que teoricamente apresenta melhor descrição e resolução espacial, podendo produzir campos sonoros de maior qualidade em *sweet-*

¹⁷ RUMSEY (2003) informa que no caso de gravações o formato é conhecido por “A-Format”.

spots ampliados (comunicação pessoal com Angelo Farina) o requisito em número de alto-falantes necessários é de 16.

Para Ambisonics de 1ª ordem temos as seguintes equações para registrar o campo sonoro, isto é, as seguintes equações codificadoras, em coordenadas esféricas (MALHAM, 1993):

$$X_j = O_j * \cos \theta * \cos \varphi$$

$$Y_j = O_j * \sin \theta * \cos \varphi$$

$$Z_j = O_j * \sin \varphi$$

$$W_j = O_j * 0,707$$

onde X_j , Y_j e Z_j equivalem às diferenças de velocidade de partícula registradas no ponto de audição devido ao campo sonoro produzido pelo objeto j (O_j), e W_j corresponde a uma medida onidirecional de pressão sonora registrada em relação à origem devido ao mesmo campo. Uma normalização prática exige que $(x^2 + y^2 + z^2) \leq 1$ para posicionar sons na superfície ou dentro da esfera unitária tomada como referência na codificação. Se o raio de O_j é maior que “1” o objeto não irá decodificar corretamente, e tenderá a “pular” para a caixa mais próxima (MALHAM, 1995). A Figura 12 mostra o sistema de coordenadas esférica e cartesiana de referência.

Alguns autores propõem fórmulas alternativas, considerando-se que a codificação basicamente registra a diretividade do som, e que a percepção da distância é forjada manipulando-se as amplitudes. Segundo Farina, a esfera de referência ou nocional admitida na codificação não guarda relação alguma com a esfera nocional da decodificação, esta última associada à matriz de posição dos alto-falantes¹⁸. Isto implica que as equações de decodificação não fazem referência alguma à esfera de referência usada na codificação.

¹⁸ informação pessoal obtida com o prof. Angelo Farina, da Universidade de Parma, Itália.

Também destaca que a esfera nocional de referência faz sentido no caso de gravações reais utilizando-se um microfone do tipo *soundfield*, que na verdade não registra o campo sonoro em um ponto idealizado, mas na prática numa pequena esfera de cerca de 12mm. Como todos os sons de importância que chegam a esta esfera estão longe, ela age como uma lente sonora, focada no infinito. Todos os sons gravados são considerados pois virtualmente emanando a partir desta superfície, sendo possível eventualmente extrapolar o objeto sonoro para dentro, mas a um custo de distorção.

Já na codificação virtual não temos estas limitações, uma vez que podemos tomar os cálculos aplicados a um ponto (microfone virtual pontual). Uma esfera de referência para codificação entretanto é sugerida a uma distância de, por exemplo, 10m, para que se tenha uma referência de amplitude unitária de um objeto situado nesta distância, e também para que possamos trabalhar com distâncias humanamente úteis, considerando nosso microfone virtual a distâncias realistas das fontes sendo “gravadas”. Outra vantagem muito importante na codificação Ambisonics virtual é que não temos os inconvenientes da distorção quando os objetos caminham ao centro, isto quer dizer que podemos ter focos mais próximos, aquém da esfera nocional. A distorção, é claro, ocorre em distâncias muito próximas de zero.

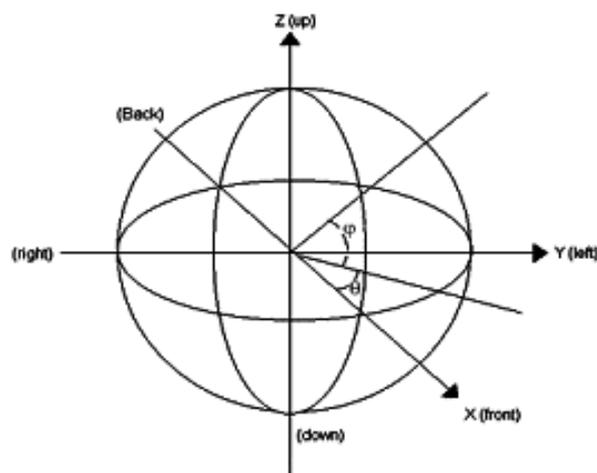


Figura 12 – Sistemas de coordenadas angulares e cartesiana no Ambisonics

O leitor interessado em pormenores quanto às equações de propagação das frentes de onda e sua representação no espaço por meio de séries de cossenos e funções de Bessel cilíndricas é estimulado a recorrer a BAMFORD (1995), inclusive como

referência para uma introdução matemática às equações de codificação (baseada na representação da diretividade de frentes de onda varrendo o meio) e decodificação (baseada na reconstrução da mesma frente de onda em alguns pontos do espaço por meio de alto-falantes), as quais não são abordadas em profundidade neste trabalho.

É interessante notar que, embora as equações acima remetam a uma simples e direta decomposição no sistema esférico ou cartesiano de contribuições parciais, elas guardam íntima consonância com os resultados obtidos matematicamente pela manipulação das equações das harmônicas esféricas, simplesmente vindo a confirmar a teoria com uma lógica elegante. As limitações e erros cometidos pela codificação e decodificação Ambisonics no que se refere à exatidão da representação e da reconstrução dos campos sonoros estarão associados com as margens de erros devido a truncagens das séries infinitas, e também a erros na decodificação devido ao crítico posicionamento dos falantes em configurações poligonais regulares.

As equações para decodificação dependem a) do número de alto-falantes que será empregado, e b) na quantidade de esféricas harmônicas utilizadas (i.e., no número de canais do sinal codificado). O sinal enviado a cada falante é uma somatória da contribuição de todas as harmônicas. Esta formulação matemática transcende o escopo deste trabalho e portanto não será detalhada aqui, mas é abordada por BAMFORD (1995). O *website* “Ambisonics.Net” em <http://www.ambisonic.net> contém uma enorme quantidade de informações sobre esta técnica, e referências adicionais (ELEN, 2004).

TORGER (2004) cita bem que Ambisonics não teve um sucesso comercial apreciável não por causa da tecnologia, mas devido a sua chegada ao mundo, numa época (década de 1970) em que não havia um suporte para gravação multicanal adequado e também o insucesso do sistema quadrifônico ecoava ainda fresco na memória dos ouvintes. Com a recente disponibilização de mídia multicanal digital o sistema está sujeito a uma renascença.

Para gravação dos 4 canais do *B-Format* de 1ª ordem existem há muito tempo microfones direcionais para captura de campo sonoro. Melhor desempenho na precisão da localização dos objetos sonoros e melhor projeção do palco sonoro são

esperados com o uso de 2ª ordem, que requer 9 canais. Entretanto, não há ainda microfones para registro em *B-Format* de 2ª ordem, com 9 canais, restando o caminho da síntese virtual (artificial) para a mídia de transporte em 2ª ordem (TORGER, 2004).

Quando foi ao mercado nas décadas de 1970/80, as gravações Ambisonics eram codificadas num formato conhecido por *UHJ*, uma codificação matricial analógica do tipo 3-2-3 transmitida sobre dois canais baseado num arranjo matricial de 3 componentes: os componentes W, X e Y da matriz *B-Format* de 4 canais (desprezando-se a elevação Z). Segundo a terminologia *UHJ*, os sinais transmitidos são o Σ (L+R) e Δ (L-R), mais um terceiro T (que poderia ser limitado em frequência ou de largura de banda plena) transmitido codificado. GERZON (1985) apresenta as equações para isso.

Há alguns *CD's* lançados trazendo este formato, sem a capacidade de reproduzir a elevação (Z). No caso em que a elevação era desejada para uma reprodução completa 3D, um quarto sinal era adicionado ao *UHJ* (sinal Q). Este novo formato foi proposto por GERZON (1985) para transmissão multicanal em *FM* de som espacialmente codificado com *UHJ* e/ou Ambisonics. Gerzon também chegou a propor codificadores e decodificadores avançados para 5 ou 6 canais para uso em *HDTV* (GERZON, 1992). Nestes últimos artigos e também em FARINA (1998) muito sobre a questão da psicoacústica envolvida na reprodução utilizando-se filtros do tipo “*shelf*” é esclarecido, e diversas equações são apresentadas.

Hoje, aparentemente, os direitos das patentes do Ambisonics estão em poder da *Nimbus Records*. Vários sítios na Internet e alguns artigos abordam esta tecnologia, mas os detalhes de implementações (principalmente sobre decodificadores) são abordados com níveis diferentes de profundidade. FARINA (1998) faz uma prática abordagem a implementação de codificação e decodificação Ambisonics. Vários artigos, todavia, são omissos quanto à apresentação plena das equações. Completos detalhes da implementação podem ser encontrados na patente de Gerzon, disponível no site de patentes do governo norte-americano (GERZON, 1976).

FURSE (1999) aborda as equações de decodificação para diversas configurações regulares de alto-falantes em <http://www.muse.demon.co.uk/ref/speakers.html>. MALHAM (2005) mantém um sítio Internet sobre Ambisonics em http://www.york.ac.uk/inst/mustech/3d_audio/, com referências a instituições, pesquisadores, literatura geral, e sistemas comerciais. *Plugins VST* para Ambisonics (codificadores e decodificadores) foram desenvolvidos por Malham sob o suporte do *Swiss Center for Computer Music*, e estão disponíveis em http://www.dmalham.freemove.co.uk/vst_ambisonics.html. Finalmente, algumas empresas possuem equipamentos decodificadores Ambisonics lançados comercialmente, como o SSP-1 da Cantares (<http://www.cantares.on.ca/decoder.htm>).

2.2.4 Ambiophonics

Ambiophonics é uma técnica mista para gravação e reprodução de som envolvente que combina mecanismos de sugestão tridimensional utilizados na tradicional estereofonia, utilizados em Ambisonics e nas técnicas bi-aurais (GLASGAL, 2004). Alguns autores julgam que esta técnica apresenta um elevado grau de imersão¹⁹ semelhante ao obtido com *WFS* (GLASGAL, 2003).

Ambiophonics foi introduzido por Ralph Glasgal e consiste num método para reprodução multicanal empregando princípios psicoacústicos para condicionar os sinais para alguns alto-falantes (dependendo de sua posição em relação aos ouvintes) e criar (recriar) um elevado nível de realismo na reprodução de salas de concerto e posicionamento das fontes sonoras virtuais no ambiente de projeção.

É mais indicado para um numero reduzido de usuários (um ou dois) numa sala de audição dedicada e para reproduzir programas de áudio musicais (TORGER, 2004), com destaque à projeção correta e realista do palco sonoro musical. Uma característica interessante é sua capacidade de reproduzir em formato Ambiophonics gravações estereofônicas originais, *DVD's* ou mesmo *SACD's* utilizando seus princípios de condicionamento de sinal para uma ótima recriação acústica.

¹⁹ vide item 3.1 (Nível de Imersão Sonora), mais adiante neste texto.

Um sistema completo consiste numa matriz reverberante de 8 ou mais alto-falantes orientada à reprodução do campo reverberante da sala de concerto, e de um “ambiopólo” (sinal em dois canais livres de *crosstalk*) para reprodução do som referente à área frontal central. Comparado ao sistema estéreo tradicional, o desempenho na reprodução do palco sonoro e da localização precisa é superior. Existem varias implementações alternativas de configurações Ambiphonics, algumas ganhando nomenclaturas também alternativas, mas apoiando-se sobre variações da mesma técnica, atingindo-se vantagens como menor numero de canais necessário ou focando-se na melhor reprodução do campo reverberante.

Usualmente, uma implementação de Ambiphonics pode ser conseguida aliando-se uma matriz Ambisonics com um dipolo estéreo²⁰ (ambiopólo) frontal (TORGER, 2004). Ambiphonics combina assim no mesmo *setup* o uso de ambiopólos e o uso de uma configuração *surround* de alto-falantes, provendo a superimposição de ambiência através de convolução de sons “secos” com respostas impulsivas de salas de concerto, bem como provendo a correção para os efeitos da função de transferência alto-falante/ouvinte (GLASGAL, 2003).

Ambisonics e Ambiphonics guardam diversas diferenças explicitadas por TORGER (2004), uma delas é que o embasamento do Ambiphonics é mais *psicoacústico* e o do Ambisonics é mais *matemático* e generalista (em termos de posicionamento em qualquer lugar e em relação ao programa sonoro). Outra é que o *sweet spot* é mais suave degradado com Ambisonics do que com Ambiphonics, que é extremamente exigente na posição do ambiopólo para render uma projeção de áudio correta. A combinação de dipolos (frontal e traseiro) e uma matriz de 4 canais para Ambisonics são citadas como forma de gerar uma configuração Ambiphonics.

O *website* <http://www.ambiphonics.org/> contém uma excelente introdução à técnica e apresenta uma literatura apreciável e completa sobre ela.

²⁰ pares de falantes em configuração estereofônica mas eliminando o efeito de correlação cruzada

2.2.5 *Wave Field Synthesis (WFS)*

A soma da contribuição de diversos alto-falantes para a reprodução de um campo sonoro usualmente pode garantir a reconstrução das propriedades temporais corretas para o campo original em qualquer ponto da área de audição. Todavia, na prática, os padrões de diretividade dos alto-falantes provocam fenômenos de interferência, e a percepção da diretividade só fica correta numa região restrita (DEVRIES, 1999). Tanto uma redução do erro na percepção da diretividade quanto a ampliação da região de audição podem ser conseguidas empregando-se matrizes de alto-falantes para a síntese de campos de onda.

O conceito de *wave field synthesis (WFS)* foi introduzido por Berkhout por volta de 1988. *WFS* permite a geração de campos sonoros 2D e/ou 3D preservando suas propriedades temporal e espacial dentro de volume ou área cercada por uma matriz de alto-falantes. As bases teóricas remontam ao princípio de Huygens, e aos teoremas de representação de Kirchhoff e Rayleigh conforme detalhados em (BERKHOUT, 1993), (BERKHOUT, 1999) e (VOGEL, 1993).

O princípio de Huygens estabelece que qualquer ponto de uma frente de onda pode ser considerado uma fonte secundária. Segundo a formulação integral de Kirchhoff-Helmholtz um campo sonoro dentro de um volume V provocado por fontes primárias fora de V é totalmente definido pelo campo sonoro na superfície (ou contorno) S deste volume, isto é, o campo sonoro num ponto interno de V pode ser determinado calculando-se o campo devido a fontes secundárias ao longo da superfície S (VOGEL, 1993), (BERKHOUT, 1993).

WFS está baseado numa formulação computacionalmente mais complexa, envolvendo o modelamento físico da propagação de ondas sonoras no meio, e que leva em consideração uma discretização do espaço físico da área de audição, além é claro da discretização do tempo, para se trabalhar no domínio digital. Uma outra discretização é feita no número de fontes secundárias, para que se possam utilizar alto-falantes como fontes secundárias.

Um efeito da discretização do espaço está na limitação de uma frequência máxima de trabalho, além da qual observamos efeitos de *aliasing* espacial. Isto requer que as

fontes secundárias reais (alto-falantes) tenham uma separação mínima entre seus cones. Por exemplo, para sintetizar um campo sonoro com frequências até 5 kHz uma matriz de alto-falantes com uma amostragem espacial (separação) de pelo menos 34mm é necessária (VOGEL, 1993). Uma consequência é que o método requer alto-falantes pequenos e simplificações quanto aos ângulos e área de cobertura, de maneira a também reduzir o número de alto-falantes necessários.

Nesta técnica, o objetivo principal é sintetizar uma frente de onda que seria produzida por fontes primárias reais em um ambiente acústico específico, através de matrizes de alto-falantes densamente distribuídas no contorno da área de audição, que agem como fontes secundárias. Uma vantagem desta técnica é sua maior tolerância a variações na carga acústica presente na área de audição (múltiplos ouvintes). E também sua capacidade em induzir a percepção de objetos sonoros localizados numa área de audição maior, inclusive havendo a possibilidade de se posicionar objetos formados no meio do ambiente de audição, posicionados ao redor do(s) ouvinte(s) e formados na frente dos alto-falantes (veja Figura 13).

A síntese da frente de onda é uma tarefa crítica, e pode se implicar num custo computacional elevado. No projeto AUDIENCE em curso na CAVERNA Digital prevê-se abordar este formato utilizando-se um aglomerado computacional para distribuir as tarefas e cálculos associados à auralização pelos nós de um aglomerado computacional.

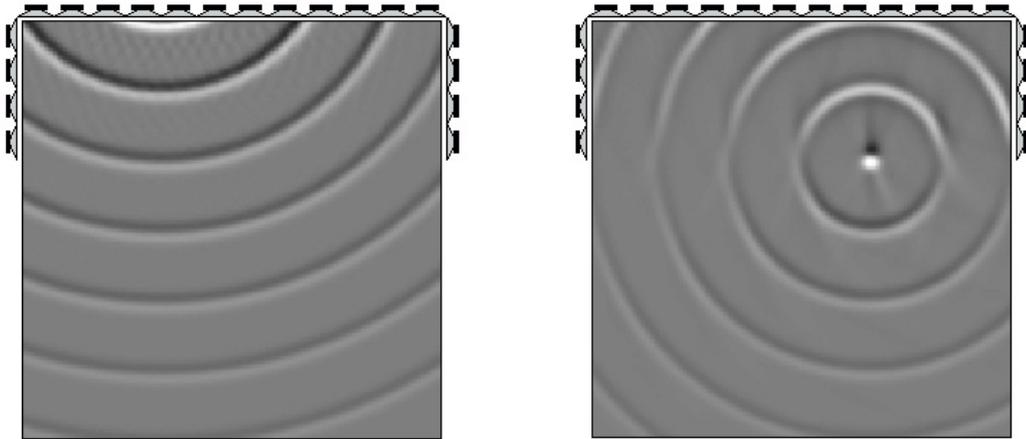


Figura 13 - Simulação de posicionamento de fontes sonoras com *WFS* fora (esquerda) e dentro da sala sonorizada (direita)²¹

WFS oferece a chance de um controle maior e mais preciso sobre as ondas propagando no meio e sobre os fenômenos de interferência entre si e entre os objetos obstáculos. Porém, devido a sua maior complexidade não somente computacional, mas também do ponto de vista de instalação, sua implementação não foi selecionada como alvo neste trabalho. cremos que os resultados do trabalho de investigação e implementação de Ambisonics na CAVERNA proverão subsídios relevantes para orientar os melhores caminhos para então uma futura implementação de *WFS*.

2.3 Tecnologias e sistemas para áudio espacial (2D/3D)

Existe um número grande de tecnologias comerciais e abertas, formatos proprietários e livres, e padrões para áudio estéreo, multicanal, áudio envolvente (*surround*), áudio 2D e áudio 3D. A maioria se manifesta na forma de componentes de *software*, bibliotecas de programação (*API's*) programas e *plugins* portáteis e/ou instaláveis em uma variedade de plataformas computacionais. Muitos são embarcados em equipamentos decodificadores (*hardware*) e dispositivos portáteis (celulares, *PALM*, gravadores e reprodutores de mídia em estado sólido – como *MP3 player/recorders*, entre outros).

²¹ imagem de (DEVRIES, 1999).

Os sistemas disponíveis que suportam funções para auralização em sua maior parte são programas ou módulos de *software* que implementam funcionalidades limitadas, frequentemente não cobrindo todos os aspectos da cadeia completa, da produção à reprodução (seja bi-aural ou multicanal). No caso de auralização, a maior parte das implementações é orientada aos métodos bi-aurais, envolvendo a manipulação de *HRTF's*.

Alguns poucos sistemas para auralização são disponibilizados em plataformas dedicadas de circuitaria (*hardware*), sendo usualmente aqueles onde a complexidade computacional dos algoritmos e os requisitos em capacidade de processamento são elevados. Isso se dá porque o uso da convolução nestes modelos é intenso. Frequentemente estes sistemas objetivam alta resolução, operação em tempo real e elevada taxa de bits ou de geração/processamento simultâneo de muitos canais ou fluxos independentes, também resultando em sistemas cuja portabilidade é restrita e os custos, elevados.

Já os sistemas e formatos populares de som envolvente têm evoluído durante décadas, e incluem os quadrifônicos, *Dolby Surround 5.1* (e sucessores), *DTS*, *SDDS*, e outros formatos multicanais (mais ou menos sofisticados), e que podem eventualmente reclamar a capacidade de também reproduzir frentes de onda ou campos sonoros 2D/3D com mais precisão, suportando, por exemplo, Ambisonics ou dipólos estéreo com *HRTF's* e cancelamento de *cross-talk*. Há também diversas tecnologias para som *surround* que utilizam uma saída de dois canais (estéreo, dois falantes) para simular ambientes envolventes²².

Há uma tendência da indústria de meios eletrônicos e multimídia em incorporar soluções para espacialização de som em dispositivos eletrônicos portáteis, *set-top boxes* e decodificadores multi-formatos, utilizando *DSP's* e/ou dispositivos *FPGA* embarcados. O aumento na capacidade de processamento (*throughput*), melhorias na miniaturização e redução do consumo de potência (além da redução do custo desses circuitos em larga escala) têm permitido ao mercado profissional e de consumo a

²² como uma referência recente, citamos o sistema *SRS® TruSurround*, em <http://www.srslabs.com>

integração de tecnologias para processamento e reprodução de áudio antes disponíveis somente em equipamentos *desktops*.

Na Tabela 5 a seguir compilamos, de forma bem resumida, os padrões, formatos, tecnologias e ferramentas de áudio mais relevantes aos propósitos de aplicações em áudio espacial, que cobrem uma extensa gama de funcionalidades para geração de áudio 1D/2D/3D, simulação ou produção de campos sonoros 2D/3D, codificação e transmissão, processamento espacial e interatividade, e reprodução bi-aural ou multicanal. RUMSEY (2003) aborda diversas delas com mais detalhes.

Estas tecnologias são amplamente utilizadas em jogos, para áudio de computadores pessoais, em receptores de som de consumo e profissionais, na indústria fonográfica e cinematográfica, na televisão e rádio, em plataformas de telecomunicações (por exemplo, celulares), em mídias comuns (por exemplo, *CD*, *DVD*) e em codificação e decodificação de áudio.

Tabela 5 – Tecnologias para áudio espacial (2D/3D)

Tecnologia	Descrição
<i>A3D (Aureal 3-Dimensional) v.3.0</i>	<i>API</i> de áudio espacial da Aureal, para placa de som com aceleração de áudio por <i>hardware</i> , similar a <i>DS3D</i> , inclui tecnologia de oclusão e reflexão “ <i>Wavetracing</i> ”, compatível <i>EAX</i> e <i>I3DL2</i> , simulação de efeito 3D através de 2 alto-falantes, para jogos.
<i>ASIO (Audio Stream I/O; Steinberg) ASIO 2.0</i>	Arquitetura/protocolo multiplataforma para transferência de áudio multicanal digital, suportado por vários fabricantes de equipamentos e placas de som.
<i>Circle Surround</i>	Sistema matricial “5-2-5” e “4-2-4”, baseado em Lt/Rt, (soma/diferença, como outros), foco em reprodução estéreo e música.
<i>DirectSound / DirectSound3D (DS3D) (Microsoft)</i>	<i>API</i> de áudio e bibliotecas para áudio espacial, explora aceleração de áudio por <i>hardware</i> , arquitetura da Microsoft.
<i>DirectX/MME/WDM (Microsoft, v.8.0, v.9.0)</i>	<i>API</i> , bibliotecas e drivers para áudio <i>hardware</i> , <i>MIDI</i> e multimídia, arquitetura da Microsoft.
<i>Dolby® Dolby Stereo Dolby Surround Dolby Surround EX Dolby Prologic</i>	Sistemas e tecnologias para redução de ruído, codificação “4-2-4”, codificação <i>surround</i> discreta (multicanal) e processamento multicanal, referência da indústria cinematográfica, <i>DVD</i> , e padrão de áudio do sistema <i>ATSC</i> (TDT nos EUA).

<i>Digital® Pro Logic II®</i> <i>Dolby Digital</i> ou <i>AC-3</i> (multicanal)	
<i>DTS (Digital Theater Systems)</i> <i>DTS ES® (Discrete 6.1, Matrix 6.1, Neo:6)</i>	Sistema para áudio <i>surround</i> multicanal, referência para cinema e <i>home-theaters</i> , usa taxa de bits superior à do Dolby e do <i>SDDS</i> (menor compressão para potencial maior qualidade de áudio).
<i>DVD-A (DVD-Audio)</i>	Sistema multiformato multicanal de alta resolução e capacidade em mídia óptica.
<i>EAX 1.0/2.0/3.0/4.0 (Creative Labs)</i> <i>(Environmental Audio Effects)</i>	Tecnologia de espacialização sonora, bibliotecas e <i>API</i> para som suportada por diversas placas de som para <i>PC</i> (com aceleração de <i>hardware</i> para processamento de áudio); padrão industrial (não oficial), particularmente muito adotado pelo segmento de jogos.
<i>HDCD (High Definition CD, Microsoft)</i>	Tecnologia provê maior faixa dinâmica codificando áudio qualidade 20-bits em 16-bits
<i>I3DL2 (IASIG 3D level 2)</i> <i>(Interactive Audio Special Interest Group)</i> <i>MIDI Association)</i>	especificação do <i>IASIG</i> de <i>API</i> para reverberação, aplicação em jogos, suportado consórcio empresas; versão 2 (1999) modela posição de objetos e ouvinte, distância, (atenuação e reverberação), obstrução, oclusão, velocidade, efeitos, radiação, ambiência.
<i>Logic 7 (Lexicon)</i>	Processo de codificação/decodificação <i>surround</i> matricial, alternativo ao <i>Dolby Surround</i> .
<i>MLP</i> <i>(Meridian Lossless Packing)</i>	sistema para codificação e empacotamento de áudio digital sem perdas, permite preservação integral do conteúdo de áudio e registro multicanal em alta-definição, não utiliza compressão perceptual, não alcança taxas de compressão superior a 3:1.
<i>MPEG (Motion Pictures Expert Group)</i>	família de tecnologias da ISO/IEC para codificação de áudio, vídeo e metadados.
<i>MPEG-1 parte 2 (2 canais)</i> <i>MPEG-2 parte 2 (multicanal: BC)</i> <i>MPEG-2 parte 7 e MPEG-4 (AAC)</i>	família de codecs de áudio da ISO/IEC para 2 canais, 6 canais (5.1) e até 48 canais (<i>Advanced Audio Coding – AAC</i>).
<i>MPEG-4 (Structured Audio)</i>	integração de ferramentas avançadas para codificação, descrição, composição, transmissão, síntese e reprodução de diversos tipos de áudio.
<i>MPEG-4 BIFS (Binary Format for Scenes)</i> <i>Advanced BIFS</i>	formato de descrição de cenas audiovisuais; parte de áudio inclui modelamento da posição e diretividade da fonte, da propagação e atenuação no meio, e metadados para auralização.
<i>OpenAL (Open Audio Library)</i>	interface aberta de <i>software-hardware</i> ; <i>API</i> para áudio 3D, bibliotecas (e primitivas de conexão com sistema operacional),

	independente de plataforma, sintaxe semelhante ao estilo OpenGL, compatível com <i>IASIG</i> 3D (nível 1 e 2), ferramentas para modelar diretividade e distância, efeitos, reflexões, obstruções, reverberação, transmissão, permite ao programador especificar objetos e operações para saída multicanal e arranjos de fontes em volta do ouvinte.
<i>SACD (Super Audio CD, Philips/Sony)</i>	sistema multiformato multicanal de alta resolução e capacidade em mídia óptica, opera com conversão/codificação do tipo <i>DSD (Direct Stream Digital, de 1bit)</i> , alta qualidade.
<i>SDDS (Sony Dynamic Digital Sound)</i>	sistema de codificação multicanal (7.1) de áudio envolvente da Sony usado em cinema, usa codificação <i>ATRAC (Sony)</i> .
<i>SPAT (IRCAM Spatialisateur)</i>	programa (bibliotecas para MAX/MSP) espacializador e auralizador do <i>IRCAM</i> , para som 3D bi-aural, transaural ²³ e multicanal (4 a 8 alto-falantes), compatível com o formalismo do <i>MPEG-4</i> , suporta parâmetros perceptuais (presença, brilho, calor, envolvimento, etc.) e parâmetros de baixo nível (físicos).
<i>SRS WOW</i> <i>SRS TruSurround / TruSurround XT</i> <i>SRS Circle Surround® 5.1</i> <i>SRS Circle Surround II (6.1)</i> <i>SRS Headphone, e SRS VIP</i>	tecnologia proprietária para simulação de áudio espacial (“ <i>virtual surround</i> ”) para fone de ouvido, para pares de alto-falantes (estéreo melhorado), e para decodificadores de 5.1 e 6.1 canais.
<i>THX</i> <i>THX Select®, THX Surround EX®</i>	sistema proprietário de referência e certificação de qualidade para áudio envolvente em cinemas e teatros.
<i>VST (Virtual Sound Technology, Steinberg, 1996)</i>	padrão para implementação de <i>softwares</i> de áudio (geradores, processadores, efeitos) e integração transparente com outros sistemas por meio de “ <i>plugins</i> ”; utiliza arquitetura <i>ASIO</i> .
<i>Waves 360 Surround Tools</i> <i>Waves M 360 Surround Manager</i> <i>Waves R 360 Surround Reverb</i> <i>Waves S 360 Surround Imager</i> <i>Waves L 360 Surround Limiter</i> <i>Waves P 360 Surround Panner</i> <i>Waves C 360 Surround Compressor</i> <i>Waves IDR 360 bit requantizer</i> <i>Waves LFE 360 low pass filter</i> <i>Waves M 360 Surround Mixdown</i>	módulos de <i>software</i> para produção de áudio espacial em formato <i>surround</i> multicanal (5.1) voltados para edição de parâmetros <i>surround</i> , masterização e produção de trilhas 5.1.

²³ o termo “transaural” refere-se ao sistema dipólo estéreo, onde o sinal para cada ouvido é enviado por 2 alto-falantes frontais em configuração estéreo, com cancelamento de *cross-talk* entre os falantes.

2.3.1 Plataformas comerciais para áudio espacial

Existem hoje muitos sistemas comerciais para áudio 2D e 3D tendo em vista objetivos tão diversos como simulação acústica, espacialização, ambientação de espaços 3D, auralização, localização precisa de objetos, etc. Para simulação acústica vários métodos e abordagens diferentes são empregados nos sistemas comerciais, assim como para a auralização e sonorização final, utilizando-se técnicas tão diversas como panoramas, matrizes (ex: codificação de 3 canais em 2 ou “3-2-3”), convoluções com respostas impulsivas (simulação de acústica de salas), auralização bi-aural com *HRTF's*, e auralização multicanal, como Ambisonics, Ambiophonics, *Wave Field Synthesis*, e técnicas mistas, entre outras.

Citamos alguns sistemas comerciais (programas, equipamentos e plataformas integradas de “*hardware/software*”) mais importantes, seja pelo destaque, popularidade ou desempenho.

Decodificadores de áudio

Plataformas decodificadores ou receptores/sintonizadores profissionais (e cada vez também os domésticos) usualmente contam com processadores de sinal digital (*DSP's*) de 32 bits (ponto flutuante), alta resolução de áudio *PCM* (96KHZ de taxa de amostragem, alguns já com conversores *AD/DA* de 192 kHz) suporte para diversos formatos de codificação de áudio e número de saídas (*MP3*, *AAC*, 5.1, *Dolby*, *THX*, *DTS*, *SACD*, *DVD-A*, entre outros formatos e codecs).

Várias plataformas suportam funcionalidades para áudio espacial, como generosa memória e programação para diversas configurações de salas acústicas (funções para simulação acústica de salas, usualmente onidirecional), saídas componentes e multicanais (ópticas e/ou elétrica) e frequentemente têm utilizado circuitos reconfiguráveis (*FPGA*) como núcleo do sistema. A título de ilustração, a figura abaixo mostra a traseira de um receptor de áudio/vídeo²⁴ cheio de conectores para

²⁴ equipamento que integra funções de coletor de sinal de áudio/vídeo de aparelhos reprodutores (*CDs*, *DVDs*, etc.), função de processador de sinal, decodificador de mídia digital, amplificador e distribuidor de sinais para monitores e alto-falantes.

áudio e vídeo, capaz de atender a uma diversidade de configurações e saídas, entre analógicas e digitais, para vídeo e áudio multicanal.



entradas: "phono", nível de linha (CD, DVD, TV, DBS, VCR, VDP), 8 canais áudio (7.1), digital I/O (6 ópticas e 2 coaxiais), vídeo componente, composto, e saídas para VCR e monitor (conectores RCA e S-Video)

saídas: vídeo componente (Y-Pb/Cb-Pr/Cr), 7.1 (4 canais surround, 2 frontais, 1 central e sub-woofer)

Figura 14 – Exemplo de painel traseiro de decodificador comercial (receiver)

Até recentemente os resultados de pesquisas puras e proposições elegantes como o Ambisonics não lograram penetração e alcance suficiente para firmar tendências no mercado consumidor. Até existem decodificadores comerciais para Ambisonics disponíveis no mercado, mas não populares, em função do custo final para o usuário e a baixa disseminação da tecnologia.

O equipamento SSP-1 da Cantares é um decodificador de Ambisonics, *UHF* (2 canais), "enhanced stereo", bi-aural, quadrifônicos *SQ*, *QS*, *Circle Surround* e formatos de espacialização. Entretanto, sua produção está num "hiato" e as informações do site são de 2001 (<http://www.cantares.on.ca/decoder.htm>).

Plataformas integradas (hardware e software)

*IOSONO wave-field-synthesis System*²⁵. *IOSONO* é uma primeira experiência comercial resultante do emprego da técnica de *wave-field synthesis (WFS)* para a sonorização espacial de amplos espaços, como cinemas e teatros. Segundo divulgam

²⁵ Sítio Internet: <http://www.iosono-sound.com/> (acesso em 10 abr. 2005).

os autores, o sistema é o resultado de mais de 15 anos de pesquisa e desenvolvimento de diversas universidades europeias, e foi aperfeiçoado pelo instituto Fraunhofer, da Alemanha, estando pronto para introdução comercial, havendo já um cinema de 100 lugares em Ilmenau (Alemanha) equipado com 198 alto-falantes operando com o sistema desde 2003. O sistema tem sido também pesquisado para aplicações domésticas (*home-theater*).

O sistema *IOSONO* consiste num anel contínuo de caixas de som configuradas em painéis de alto-falantes (duas vias) espaçados de 15cm entre si, montados em paredes na periferia do ambiente de audição, e conectados via cabo óptico à plataforma central *IOSONO* responsável pela “renderização” sonora. Assume-se que fontes sonoras podem ser “colocadas” em qualquer lugar no espaço e que a imagem é estável, isto é, um ouvinte sentado na frente do auditório terá a mesma experiência auditiva de uma pessoa sentada no centro da sala, ainda que esteja em movimento poderá perceber a fonte sonora numa posição fixa, como na vida real. Ainda, baseando-se nas vantagens da técnica *WFS*, os sons podem se mover pelo ambiente, atravessando-o, e fontes sonoras são percebidas como que emanando do ambiente, não a partir das caixas acústicas individuais.

A ferramenta utilizada para produzir as cenas sonoras espaciais no sistema *IOSONO* é a *IOSONO SAW (Spatial Audio Workstation)*, uma plataforma *stand-alone* integrável a equipamentos de estúdio e gravações existentes, compatível com formatos *surround* existentes (e.g. estéreo, 5.1, *DTS*, e *SDDS*) contando com um módulo de processamento, um módulo de renderização e um painel do tipo *touch-screen* como interface de usuário. Objetos sonoros podem ser organizados em camadas, permitindo agrupamentos como diálogos, efeitos sonoros e música de fundo.

No processo de produção, o registro da cena *master* (orientada a objetos) é feito com uma descrição de cena e dados de áudio *PCM* não comprimidos. Para distribuição, o formato é baseado no *MPEG-4* combinado com áudio comprimido em alta-qualidade, ambos os formatos exportáveis a partir da plataforma central.

Convolvotron. Um dos sistemas comerciais mais creditados é o *Convolvotron*²⁶, que é uma máquina de convolução massiva integrando *DSP's Texas Instruments* da família TMS320C25, cada um contendo 128 16x16 ULA's (unidade lógico-aritmética) capaz de 320 milhões de operações de multiplicação/adição/deslocamento por segundo, acoplada a uma memória *RAM* estática (*dual port*) compartilhada pelo sistema *DSP*. Até 8 placas podem ser associadas em paralelo, para espacializar 32 sons anecóicos ou 8 fontes sonoras com reflexões de primeira ordem simulando um ambiente acústico.

A tecnologia de áudio CRE²⁷ 3D foi originalmente implementada no *Convolvotron* (veja Figura 15). A *CRE Acoustetron* é um subsistema de áudio 3D, “*hardware*” e “*software*”, que provê uma interface de controle comum para qualquer sistema hospedeiro. Internamente, uma *Acoustetron* pode ser configurada para qualquer combinação de *hardware* de espacialização CRE.

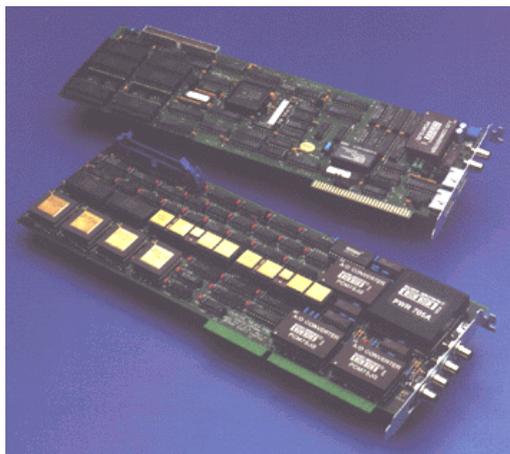


Figura 15 – Máquina de convolução comercial *Convolvotron*

Um *hardware* CRE de entrada inclui um *Convolvotron* e uma placa de baixo custo *Beachtron*, com síntese *MIDI* e funcionalidades para tocar arquivos de som (formas de onda). Uma configuração *Acoustetron* típica de “topo de linha” combinaria 4

²⁶ informações no sítio: <http://www-cdr.stanford.edu/DesignSpace/sponsors/Convolvotron.html> (última atualização era de 9 fev. 1994).

²⁷ *Crystal River Engineering*.

Convolvotrons e 4 *Beachtrons*, com estas últimas suprindo efeito Doppler, reverberação, síntese, reprodução de amostras (*samples*), e efeitos especiais.

Contudo, este *hardware* tem hoje pelo menos 11 anos, e os recursos lógico-aritméticos e de *DSP* disponíveis atualmente, por exemplo, em um dispositivo reconfigurável *FPGA* de última geração, ultrapassam em muito estas figuras de desempenho e capacidade. Uma *FPGA* de entrada da família Stratix II (*Altera*) possui 6 estruturas *MAC's* completas, cada uma delas capaz de 370 milhões de multiplicações/adições/deslocamento por segundo, excetuando as *MACs* que poderiam ser implementadas em lógica reconfigurável no dispositivo. No total, poder-se-ia alcançar com um destes chips o montante de 2,2 bilhões de operações *mult/add/shift*. A *FPGA* mais sofisticada da linha tem 96 *MAC's* nativas, podendo alcançar em capacidade de processamento até 35,52 bilhões de operações *mult/add/shift* por segundo (excetuando-se as que poderiam ainda ser implementadas em lógica)²⁸.

RX6 Piranha Multifunction Processor. Esta é uma das plataformas multi-*DSP* da empresa *Tucker-Davis Technology*, orientada ao mercado de pesquisa em som 3D. Para isso inclui funcionalidades como geração de ondas, estímulos auditivos, aquisição e análise de dados, geração de áudio espacializado (3D) usando *HRTF's*, etc. Acoplada a *softwares* da empresa (como *OpenEX* e *BioSigRP*) a plataforma permite implementar várias funcionalidades de áudio 3D, como suporte a múltiplas fontes sonoras, processamento em tempo real de *HRTFs*, etc. O sítio da empresa é <http://www.tdt.com/Workstations/3D.htm>.

Programas para simulação acústica

Muitos programas para simulação acústica de salas, previsão e projeto de comportamento acústico foram desenvolvidos desde o artigo inicial proposto por Krokstad, Strom e Sorsdal em 1968 sobre simulações em computador de acústica de salas (VORLÄNDER, 1995). Os programas comerciais são encontrados para várias

²⁸ informações pessoais obtidas com representante da *Altera* no Brasil.

plataformas (com preponderância para Windows[®] da Microsoft[®]) e frequentemente importam/exportam arquivos em diversos formatos, particularmente os formatos compatíveis com *AUTOCAD*.

Em 1994 o instituto nacional alemão de metrologia²⁹ propôs uma competição entre programas de simulação acústica, com o intuito de dar uma chance a desenvolvedores de comparar os resultados de seus programas com medidas confiáveis, e ao fim comparar o desempenho com outros programas. O projeto, conhecido como “*round robins on acoustical simulation*” teve outras versões, a última finalizada em 2002 (terceira versão) e contou com a participação de 21 programas, entre eles o RAIOS (v.3.0), desenvolvido no LAVI-UFRJ³⁰ (CAMILO, 2002), o *RAMSETE*, de Angelo Farina, e famosos e prestigiados programas para simulação acústica e auralização, como o *CATT Acoustics*, o *ODEON Room Acoustics Software*. O procedimento do projeto/competição é detalhado por (VORLÄNDER, 1995).

O projeto *ODEON* foi fundado em 1984 (pela empresa dinamarquesa *Orsted DTU* e consultores especializados) com o objetivo de desenvolver um programa de predição acústica de salas confiável e fácil de se usar. O programa resultante (*ODEON*), hoje na sua versão 7.0, permite prever o comportamento acústico de espaços públicos, auditórios e ambientes industriais, e cobre também o controle de ruído³¹.

O programa *EASE (Enhanced Acoustic Simulator for Engineers)* é um programa para predição acústica e modelamento de reforço sonoro com resultados muito precisos para visualização e auralização de espaços criados no computador. Integra funcionalidade de auralização que permite renderizar o som ouvido numa dada posição da maquete ou modelo arquitetônica em computador. O módulo *EARS*

²⁹ PTB – Physikalisch-Technische Bundesanstalt: <http://www.ptb.de/en/suche/suche.html>

³⁰ Laboratório de Acústica e Vibrações, Universidade Federal do Rio de Janeiro, <http://www.lavi.coppe.ufrj.br/>

³¹ Maiores detalhes na página do programa na Internet: <http://www.at.oersted.dtu.dk/~odeon/>

licencia um algoritmo da *LAKE DSP* para gerar auralização (bi-aural e dipolo estéreo) em tempo real de streams de áudio, com latência inferior a 300ms³².

O *Aurora*, de Angelo Farina, é um módulo (extensão) para o programa de áudio multicanal *CoolEditPro (Adobe Audition)* com funções para análise e simulação acústica voltados para auralização e acústica de salas.

³² segundo informações em http://svconline.com/mag/avinstall_ease_windows/index.html e <http://www.auralisation.de/index.php>.

3 ÁUDIO EM AMBIENTES E SISTEMAS AUDIOVISUAIS IMERSIVOS

A imersão em ambientes audiovisuais está diretamente relacionada com a quantidade e qualidade visual e auditiva da experiência da espacialidade e da fidelidade em reproduzir um ambiente ou situação.

Existem diversos sistemas audiovisuais que produzem graus de percepção imersiva variados. O cinema é o sistema audiovisual mais popular que oferece experimentações imersivas. A tela grande à frente frequentemente projeta um espaço de dimensões largo o suficiente para que possamos avaliar partes do ambiente da cena, e até focalizarmos nossa atenção para objetos que estão praticamente distribuídos à nossa frente, produzindo uma impressão de envolvimento. Fenômenos ilusórios frequentemente podem ser atingidos numa dada cena: a ilusão é o objeto de desejo do cinema, e a arte, seu meio.

O ganho de espacialidade e a imersão na cena também podem ser alcançados graças à sugestão psicológica da trama e pela qualidade da montagem audiovisual, forjando a forma que prende e adiciona o espectador, este fortemente tomado pela sinestésica e temporária euforia provocada pela sequência visual em suas retinas, pela trilha sonora e o envolvimento produzido pelo *surround*, e pela forma em que o roteiro explora as sutilezas de suas preferências, pela exotividade dos pontos de vistas, das perspectivas e situações, pelo seu casamento com uma cena audiovisual eloquente. É a realização de uma experiência singular (e segura) de uma realidade artificial além das possibilidades reais. Sem dúvida, a produção sonora envolvente e a sonorização multicanal respondem hoje neste meio pela maior parte da percepção ou experiência de imersão.

Os sistemas de realidade virtual são os principais ambientes audiovisuais que objetivam especificamente a questão da experiência envolvente, a criação e a reprodução da realidade e seus atributos físicos. Entretanto, menos atrelados às questões psicológicas e artísticas que modulam a experiência imersiva no cinema, estes sistemas estão mais voltados para os aspectos técnicos, cognitivos e perceptuais

dos sistemas neurofisiológicos humanos. Dentre eles, os ambientes imersivos do tipo “CAVE” são os que alcançam os resultados mais excepcionais, e servem como incubadores de novas gerações de sistemas audiovisuais como nunca antes vistos. Uma diversidade de aplicações audiovisuais possíveis e integrações de diversos dispositivos a estes sistemas estimulam atividades práticas de pesquisa e desenvolvimento que apontam uma direção de investigações neste campo que nos leva, progressivamente, a conceber sistemas eletrônicos mais robustos, compactos, fáceis de instalar e configurar, e com melhor capacidade de projeção visual e auditiva.

Podemos definir um ambiente de realidade virtual como aquele que reúne um conjunto de técnicas, metodologias e algoritmos capazes de permitir ao ser humano interagir e navegar em mundos sintetizados por computador. Existem inúmeros sistemas e recursos tecnológicos disponíveis capazes de propiciar esta experiência.

Uma dificuldade intrínseca a estes sistemas está na capacidade de visualização e sonorização com qualidade e em tempo real. Estes requisitos têm estimulado no decorrer dos anos intensa pesquisa em aplicação de sistemas de alto desempenho nesta área.

A televisão dotada de sistemas de áudio *surround* permite criar sugestões de espacialidade e pode assim incrementar substancialmente a experiência de imersão, em especial quando uma configuração adequada de alto-falantes e um decodificador *surround* estejam disponíveis. Para uso doméstico, entretanto, não somente os custos são variáveis inibidoras da popularização destes sistemas em larga escala, mas também a impossibilidade de se destinar uma sala adequadamente tratada e configurada para este fim.

Diversos esforços em pesquisa e desenvolvimento são orientados para buscar soluções tecnológicas viáveis, que libertem usuários comuns e cenários domésticos de requisitos restritos em demasia, e que possam mesmo reconfigurar uma sala da casa do usuário para acomodar um sistema de projeção audiovisual sem maiores intervenções físicas, contando-se principalmente com recursos eletrônicos como técnicas de sensoriamento remoto e processamento de sinal para inferir sobre o

estado e a configuração do ambiente real existente, e, com base na quantidade e qualidade de sistemas de projeção e sonorização disponíveis, computar todas as correções, dereverberações, e compensações necessárias no sinal visual e auditivo para que produzam o campo audiovisual envolvente (especialmente correto) desejado.

Neste capítulo abordamos os ambientes audiovisuais imersivos e a produção sonora nestes meios. Introduzimos o importante conceito de imersão sonora e abordamos os sistemas de realidade virtual imersiva, em especial a CAVERNA Digital, como ambiente de RVI de referência neste trabalho. Discute-se a metodologia de construção de cenas sonoras nestes ambientes, e finalmente apresenta-se uma proposta de implementação de soluções de áudio espacial para RVI, com potencial para estimular a compatibilização e integração de ferramentas voltadas para produção de áudio espacial.

3.1 Nível de imersão sonora

O mais importante é saber quão imersiva é a experiência sonora, o grau de imersão que o ouvinte experimenta num sistema de sonorização, e não quantas caixas de som ou qual sistema ou marca comercial são usados.

O grau de imersão percebida pelo ouvinte pode variar sensivelmente entre a aplicação de uma técnica e outra e também em função do número de alto-falantes utilizado. Mesmo utilizando-se somente um canal (monofônico) pode-se transmitir uma sensação de espacialidade, por exemplo, convoluindo um sinal sonoro com a resposta ao impulso (onidirecional) de uma sala. Com dois canais, pode-se explorar o efeito estéreo. Utilizando simplesmente uma técnica de gravação estereofônica e estando o ouvinte adequadamente posicionado no campo de formação do efeito estéreo entre 2 alto-falantes, é possível explorar a formação de imagens no campo entre os falantes, e simular movimentos. A percepção de bidimensionalidade e tridimensionalidade é maior na medida em que se usam técnicas de auralização bi-aurais ou multicanais, e no caso desta última, também depende do número de alto-falantes utilizado. As técnicas de auralização permitem simular com mais eficácia a sensação de imersão em campos sonoros 3D.

Existe mais de uma centena de programas, *plugins* ou codificadores/decodificadores de *surround* no mercado. Frequentemente sistemas comerciais combinam técnicas diferentes num mesmo produto para cobrir uma faixa maior de efeitos espaciais. Como então podemos, antemão, classificar a eficiência ou eficácia de uma técnica de espacialização, *surround* ou de auralização, ou mesmo traçar comparações?

Muitas formas de classificação ou critérios de comparação podem ser usadas, como por exemplo, pela técnica utilizada, pela quantidade de efeitos de espacialização ou acústicas simuladas (de auditórios, cavernas, debaixo d'água, etc.), pela implementação do efeito via processamento espectral ou temporal, pela implementação via *software* ou via *hardware*, pelo número de canais que utiliza, pelo modelo acústico utilizado, pela duração ou cobertura temporal que o efeito cobre, pela quantidade de técnicas diferentes que combina, etc. Cada qual pode ser útil para um levantamento ou estudo específico.

Entretanto, de uma maneira geral no contexto de áudio espacial, e em especial para os propósitos deste trabalho, julgamos necessário e oportuno propor uma escala para mensurar a qualidade e quantidade percebida de imersão. Uma escala de graus de imersão pode ser útil como referência comparativa e instrumento de classificação aplicáveis a qualquer método, sistema ou técnica de sonorização, mesmo nos diversos cenários de comparação citados acima.

Propomos uma escala de níveis de imersão gradual, como na Figura 16, partindo do nível onde nenhuma sensação ou grau de imersão existe, passando por graus de imersão progressivamente maiores, até o nível mais alto, onde a percepção da imersão é mais forte, estável e bem condicionada. Propomos 5 níveis de referência nesta escala, mas é admissível numa avaliação prática que a sensação ou percepção do grau de imersão esteja em qualquer ponto da escala, não exatamente nos níveis discretos.

A **Tabela 6** associa os níveis de imersão à sensação espacial, e também exemplifica quais técnicas ou efeitos estão relacionados. Um maior grau de imersão também sugere um maior grau de liberdade e cobertura espacial, na medida em que uma percepção no espaço 3D é mais completa que a percepção em um plano (2D).

Também, na medida em que o grau de precisão na localização de objetos no plano ou espaço é maior, também a correção do efeito de síntese imersiva é maior. A percepção de campos sonoros (2D e 3D) usualmente está associada a uma melhor precisão na localização dos objetos sonoros, mais realismo e correção na simulação do ambiente acústico e reprodução de campos sonoros.

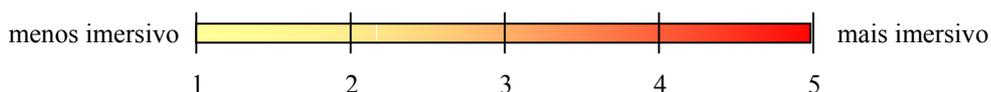


Figura 16 – Escala para níveis de imersão sonora

No contexto proposto, os níveis mais altos da escala englobam os níveis imediatamente anteriores. Uma avaliação referente ao nível M também atende aos requisitos dos níveis anteriores (M-1, ..., 1).

Tabela 6 – Níveis de Imersão Sonora

Nível	Técnica ou métodos	Percepção (resultados)
0	monaural, sinal “seco” (anecóico)	imersão nula
1	reverberação, ecos	espacialidade, ambiência
2	panoramas (entre alto-falantes), estereofonia, multicanais <i>surround</i>	direção do som, movimentos
3	panorama por amplitude, SVPA	localização correta em regiões limitadas
4	<i>HRTF</i> , auralização, técnicas perifônicas (Ambisonics, WFS, Ambiophonics)	campo sonoro 2D estável
5	<i>HRTF</i> , auralização, técnicas perifônicas (Ambisonics, WFS, Ambiophonics)	campo sonoro 3D estável, distância e localização precisas

(FARIA, 2005a) aborda e explica em detalhes a proposta da escala de níveis de imersão para medição da qualidade espacial percebida.

3.2 Sistemas de Realidade Virtual Imersiva (RVI)

A busca da sinestesia e da onipresença parece ser a cola que motiva e, em última instância, até guia a pesquisa atual por sistemas de realidade virtual imersiva. Este fio, muitas vezes não considerado dentro do universo técnico e de engenharia, se faz objeto de preocupação e investigação pelos cientistas das humanidades, filosofia e

psicologia, na busca de compreender as motivações da cultura atual, e onde ela pretende chegar.

No século XIX a humanidade observou o advento da mídia elétrica, o que trouxe modificações estruturais substanciais e avassaladoras na forma de criação, transmissão, distribuição e armazenamento da linguagem, dos símbolos, ícones desta cultura. Os cultos da cultura simbólica a partir de então passaram a ser praticados num nível de abstração superior, onde a semiótica impera, onde as ligações entre símbolos são múltiplas e podem ser instanciadas através de ferramentas como as mídias eletrônicas. A experiência e o consumo das informações sensoriais também ficaram mais transitórios, talvez até mais voláteis na cultura contemporânea, e esta tendência ainda não demonstrou saturação porque novas tecnologias, com base em ferramentas físicas, têm constantemente aberto novas perspectivas para experiências sinestésicas e sensoriais mais elaboradas.

Agora mesmo, provavelmente estamos guiando a cultura simbólica para experimentações que afirmarão algumas tendências no trato dos símbolos, e também extinguirão outras. A segunda oralidade já veio com o advento da mídia elétrica, e há autores que verificam o quanto a experiência auditiva já se mesclou com a táctil. Talvez estejamos presenciando o consumo do som *surround* também como uma “massagem” nos cinemas e *home-theaters* armados com poderosos *sub-woofers*.

Grupos de pesquisa tecnológica no mundo atual tendem fortemente a incluir nas estruturas hierárquicas de suas implementações o conceito amplo de nível de abstração da experiência sensorial. O grupo MPEG da ISO/IEC, por exemplo, desenvolveu o padrão de codificação multimídia MPEG-7, que busca a classificação e indexação de objetos multimídia. A extensão de classificação a que se propõe tal ferramenta é notoriamente uma soma e uma exacerbação desta tendência em abraçar e conectar a informação de baixo nível, como atributos e características únicas e irrefutáveis de objetos, facilmente tratável com ferramentas simples, e a informação de alto nível, pertencente a um nível de abstração semiótico mais alto, e portanto sujeita a múltiplas interpretações ou universo de solução mais amplo (FARIA, 2002b).

Um fator preponderante na capacidade de firmar uma tendência neste contexto evolucionista e nesta era econômica é a garantia da qualidade da emulação de um efeito ou concretização de uma experiência sensorial, e a exequibilidade da sua implementação com recursos convencionais. A CAVERNA Digital na EPUSP tem como uma de suas referências paradigmáticas a busca de soluções nestes patamares de qualidade e factibilidade, utilizando como referência recursos convencionais.

Este capítulo pretende abordar os sistemas integrados que viabilizam a experiência da realidade virtual imersiva, tendo a CAVERNA Digital como sistema de referência, e com especial ênfase na discussão sobre o uso de áudio nestes sistemas.

Até a presente data observa-se poucos trabalhos prévios em sistemas de realidade virtual que abordam o uso de técnicas para a renderização de áudio espacial. A convergência das pesquisas e resultados obtidos nas áreas de acústica de salas e sistemas de auralização comparados com os obtidos nas frentes de pesquisa em aglomerados gráficos e projeção visual estereoscópica estão bem aquém do ponto de amadurecimento tecnológico para permitir uma experiência completa de realidade virtual imersiva.

3.2.1 Ambientes CAVE's: A CAVERNA Digital

A CAVERNA Digital é um sistema de realidade virtual do tipo *CAVE* (*CAVE Automatic Virtual Environment*TM) que consiste de uma sala cúbica na qual cada “parede” é uma tela de projeção. O ambiente virtual (cena audiovisual) de interesse é projetado nas telas criando para o(s) usuário(s) que se encontra(m) dentro do volume das telas uma sensação imediata e intensa de imersão naquele ambiente.

Sistemas de projeções têm sido utilizados há muito tempo, contudo os mais comuns são sistemas de uma única projeção, como no cinema, atrações de parques temáticos, ou mesmo nos sistemas domésticos. O que torna difícil a projeção de mais de uma fonte é como combinar as imagens de forma que se apresentem contínuas, e como fazer para que as fontes de vídeo estejam sincronizadas para apresentar, não só na mesma velocidade as imagens, mas também para que os tempos de apresentação de ambas as visões sejam idênticos, garantindo coerência visual.

Sistemas de projeção cúbica podem ter de 3 a 6 faces e requerem sistemas computacionais capazes de gerar e sincronizar imagens em alta resolução a partir da mesma base de dados. Os primeiros sistemas deste tipo foram implementados na Universidade de Chicago em 1992, e são conhecidos por *CAVE* (CRUZ-NEIRA, 1993).

As *CAVE's* são sistemas de realidade virtual baseada em multiprojeções desenvolvido no Electronic Visualization Lab³³ concebidos por Carolina Cruz-Neira, Dan Sandin e Tom DeFanti. A *CAVE* foi demonstrada em público pela primeira vez na conferência SIGGRAPH'92, e a partir daí foram implementadas em vários locais. Entre alguns sistemas de RV baseados em *CAVE's* estão o *ImmersaDesk*³⁴ e o *Iwall* (uma tela larga de uma projeção mas contando com 4 níveis gráficos para resolução aumentada). A *CAVE* e o *ImmersaDesk* são agora produtos comerciais, vendidos pela empresa *Fakespace Systems* (previamente *Pyramid Systems Inc.*). A biblioteca *CAVE Library* é vendida e distribuída pela VRCO.

O primeiro sistema deste tipo na América Latina – a *CAVERNA Digital*³⁵ – foi construída na Escola Politécnica da USP em 2001. Dentre as várias possibilidades de sistemas de multiprojeção imersivos, decidiu-se implementar para a *CAVERNA Digital* uma *CAVE* de 5 faces: as paredes laterais e o chão, conforme ilustra a Figura 17. Os aspectos do projeto e detalhes de implementação da *CAVERNA Digital* são descritos por (ZUFFO, 2001) e (SOARES, 2002).

³³ <http://www.evl.uic.edu/>

³⁴ O *ImmersaDesk* é um dispositivo semi-imersivo de uma projeção. Detalhes disponíveis em <http://www.evl.uic.edu/pape/CAVE/idesk/>

³⁵ **CAVERNA Digital** é uma marca registrada da Universidade de São Paulo.

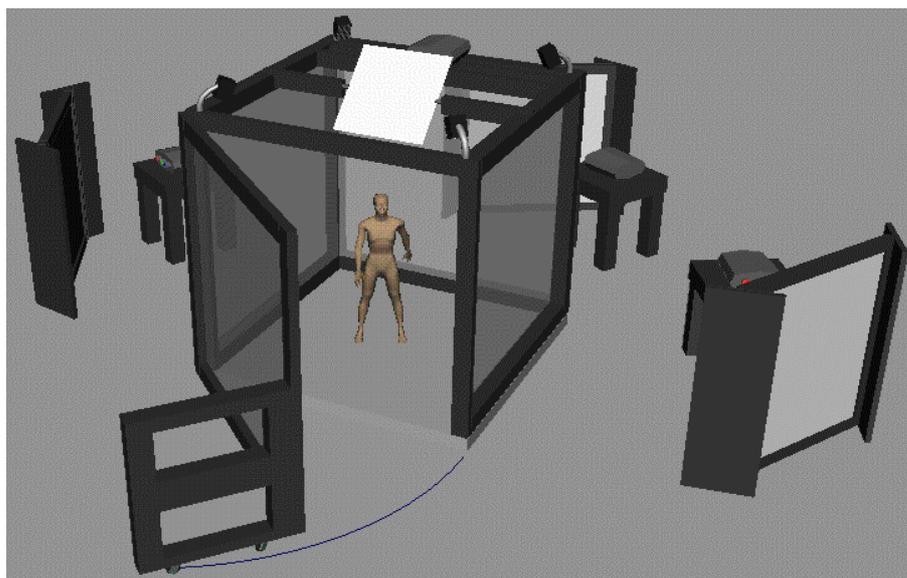


Figura 17 – Modelo tridimensional da CAVERNA Digital

Um dos grandes diferenciais da CAVERNA Digital em relação aos outros sistemas de projeção cúbica ao redor do mundo é devido a esta ser acionada por um aglomerado de computadores convencionais para aplicações gráficas e de realidade virtual (*VRCluster*). Todos as tarefas envolvidas na navegação de um mundo virtual são executadas em nós do aglomerado. Os processos são atribuídos a cada nó de forma a garantir uma distribuição de carga computacional e assim otimização do uso da capacidade integrada. Tais processos incluem por exemplo a renderização do mundo virtual, a geração da imagem para cada tela, o controle e sincronismo da projeção de cada tela, o rastreamento do usuário (e eventualmente seu ponto de vista) no mundo virtual, a captura de comandos do usuário para realizar translações, e a geração de som.

Um sistema de gerenciamento global é utilizado para iniciar as aplicações e os respectivos processos em cada nó do aglomerado, controlar a comunicação e o fluxo de dados entre nós/processos, em outras palavras incumbido da integração e sincronia dos processos. Atualmente o sistema SIRIUS é responsável pela geração, gerenciamento e multiprojeção de imagens na CAVERNA Digital.

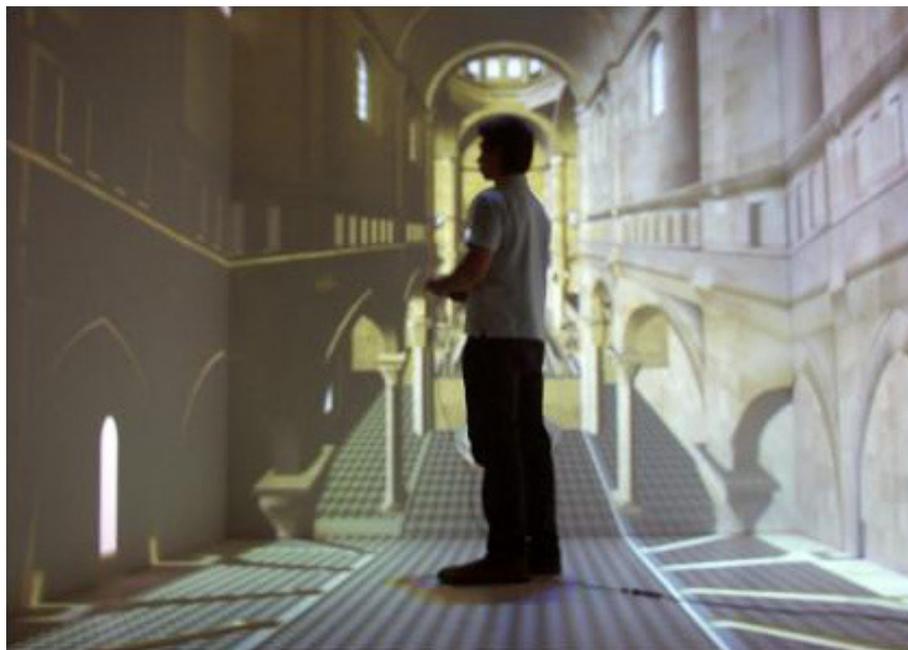


Figura 18 – Exemplo de navegação em ambiente virtual na CAVERNA Digital

Dentre os vários subsistemas de *hardware* destacam-se os projetores de imagens, o comutador de vídeo, o sistema de áudio (ex: placas de som, amplificadores), o sistema de iluminação, e o sistema de rede. Todos podem ser controlados de forma transparente ao usuário. Usualmente uma aplicação de RVI pode ser disparada a partir desta interface de gerenciamento, que, a partir do disparo inicial, aciona e prepara e controla todos os subsistemas necessários e utilizados na aplicação.

Os sistemas cúbicos apresentam atualmente uma qualidade de imersão melhor que qualquer outra forma de sistema de realidade virtual, como por exemplo os capacetes de realidade virtual *HMD (Head Mounted Display)*, que são óculos pesados e apresentam sérios problemas de rastreamento, navegabilidade, campo de visão e podem induzir ao mal estar no usuário. Na CAVERNA Digital os recursos incorporados ao usuário para a imersão são geralmente mínimos, baseando-se apenas em óculos para suportar a estereoscopia e dispositivos de rastreamento.

Uma grande vantagem da CAVERNA Digital é sua flexibilidade, em ter a possibilidade de compartilhamento do mundo virtual por vários usuários simultaneamente, a quantidade de pessoas está limitada diretamente às dimensões físicas da CAVERNA utilizada, como também é possível o recálculo do ponto de

vista do usuário, baseado em um sistema de rastreamento para um usuário e assim se produzir o efeito perfeito de imersão e navegabilidade.

Dentre as possíveis aplicações da CAVERNA Digital, podemos destacar: a visualização de lençóis petrolíferos para a maior facilidade de perfurações, as maquetes digitais na indústria automobilística para solucionar problemas de design e aerodinâmica, a análise meteorológica como a visualização de massas de ar, além de simulações cirúrgicas na medicina, planetário virtual, óperas virtuais, dentre muitas outras.

Aspectos arquitetônicos e estruturais

A CAVERNA Digital é composta por um cubo de dimensão 3m x 3m x 3m com projeção nas 4 “paredes” laterais mais o piso. Para aplicações coletivas, o sistema pode comportar até 6 usuários no seu interior, compartilhando a mesma experiência virtual.

No leiaute arquitetônico da CAVERNA Digital da EPUSP, um dos aspectos levados em conta foi embutir totalmente os projetores e outros dispositivos ópticos eletrônicos, escondendo do usuário final as particularidades técnicas do sistema. Mesma orientação é válida para o posicionamento previsto para os alto-falantes que devem ser camuflados ou posicionados por trás da tela fora das regiões proibidas de projeção. Essa requisito é muito restritivo no caso do *setup* dos alto-falantes, conforme veremos adiante na proposição de configurações para posicionamento destes no experimento.

A escolha das faces da CAVERNA Digital, onde há projeção, foi baseada no potencial de uso de aplicações relevantes para ciências gerais, engenharias e indústria brasileira, na qual há a necessidade de se visualizar objetos, e o maior interesse é a sua visualização por cima, e ambientes para passeios virtuais, onde a referência do piso é importante para a sensação de imersão do usuário.

Para não comprometer as características de isolamento acústico foi necessário optar-se pela utilização de sistemas de ar-condicionado de alto volume (60.000 BTUs) com baixa emissão sonora. Também, objetivando a redução de reflexões e reverberação

no espaço atrás da CAVERNA, onde estão os projetores e espelhos, jateou-se espuma acústica nas paredes. O teto sobre a CAVERNA, embora construído com uma camada de placas que separam o espaço acústico da CAVERNA da laje do prédio, não têm nível de isolamento satisfatório, deixando entrar ruídos externos.

Haja vista a prioridade pelo projeto visual na concepção da CAVERNA, não houve um projeto acústico integrado desde o início, havendo problemas estruturais de ordem acústica que precisarão ser corrigidos nas próximas fases de atualização desta infra-estrutura.

Os detalhes estruturais e de construção da CAVERNA são abordados em (ZUFFO, 2001) e (SOARES, 2002).

Telas

As imagens das quatro telas laterais são projetadas pela parte traseira, em películas de material plástico (polímero branco). O piso recebe uma projeção direta pela parte superior.

A projeção sonora também deve ocultar dos ouvintes a posição dos alto-falantes. A maioria dos falantes deverão ser posicionados por detrás das telas, direcionando a irradiação sonora para o centro da CAVERNA. Não se conhece, a priori, as propriedades acústicas da tela existente. Supõe-se que tenham uma curva de resposta em frequência com atenuação seletiva, e aumentando com a frequência. Também se supõe que apresentem propriedades difusoras, o que pode contribuir para “borrar” a localização da imagem sonora. Estão previstas medidas acústicas para se levantar a função de transferência das telas, e conhecer suas propriedades acústicas, inclusive avaliações em baixas frequências, para se determinar a existência de modos estacionários e interferências construtivas/destrutivas entre as telas, dentro do volume da CAVERNA. Esse conhecimento será fundamental para se projetar filtros inversos adequados para cancelar os efeitos de distorção introduzidos pela função de transferência do sistema transdutor-tela (KIRKEBY, 1999).

Em algumas configurações do decodificador Ambisonics, será possível posicionar algumas caixas sobre a estrutura, irradiando por cima diretamente para o centro,

evitando-se a difusão através das telas e os artefatos sonoros conseqüentes. Isso é preferível do ponto de vista sonoro por evitar a ação das telas sobre o som. Contudo, um dos pontos relevantes neste estudo será avaliar o quanto é comprometedora a ação das telas e o quanto se pode compensar desta interferência, por meio de filtragens inversas, descorrelacionando sua função de transferência do processo global. Se esta tarefa for feita adequadamente, do ponto de vista auditivo será como se as telas e os alto-falantes não existissem, somente os objetos sonoros irradiando no meio. O processo de sintonia da CAVERNA a um sistema de sonorização multicanal poderá ser portanto investigado.

Sistema computacional

Nos últimos anos observamos uma rápida e crescente disseminação de sistemas computacionais com desempenho melhorado baseados em aglomerados de computadores convencionais (*commodity computer cluster*) encontrando aplicações em diversas áreas, desde as ciências e engenharia até a educação e entretenimento (SOARES, 2004a).

Tais sistemas são de interesse especialmente nas aplicações avançadas voltadas para a realidade virtual, onde as operações de computação gráfica e as técnicas mais sofisticadas para a simulação e síntese sonora espacial exigem alta capacidade de processamento.

Os aglomerados de computadores são um conjunto de computadores capazes de compartilhar seus recursos para um objetivo comum. Cada computador é denominado um nó do sistema. Os aglomerados devem ser providos de um mecanismo de gerenciamento. O ganho de desempenho global será proporcional ao número de estações utilizadas, podendo equiparar a capacidade computacional em executar cálculos de grande complexidade numérica, transacional e gráfica de computadores maciçamente paralelos, porém a uma fração reduzida de seu custo.

O diagrama do sistema típico de um aglomerado computacional multiprocessado é mostrado na Figura 19. Cada nó do aglomerado possui recursos próprios de memória e dois processadores.

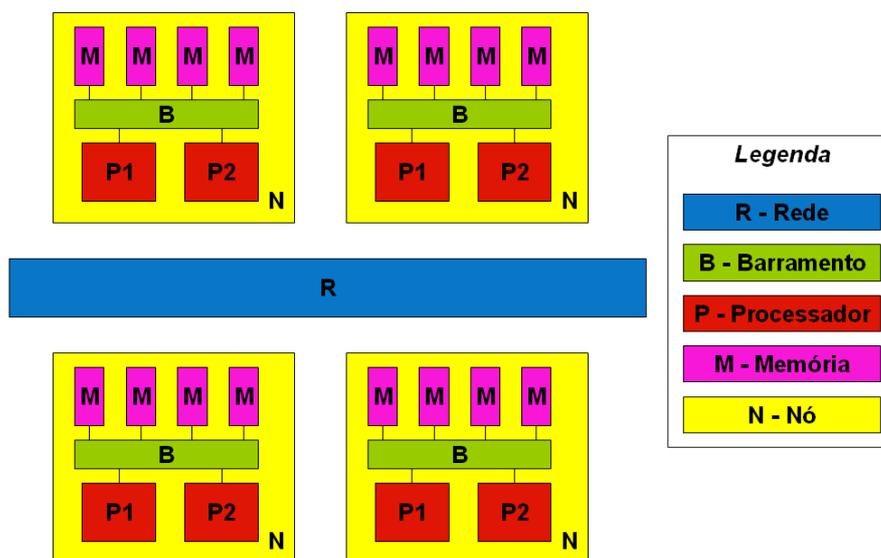


Figura 19 – Diagrama geral de um sistema aglomerado de computadores

Uma rede de alta velocidade permite a comunicação de dados entre nós, incluindo aí não somente o envio/recebimento de mensagens entre processos/nós, mas também mídia codificada/comprimida, metadados e parâmetros para composição, controle e atualização de cenas audiovisuais³⁶.

A cada nó no aglomerado podem ser atribuídas tarefas ou funcionalidades diferentes, dependendo da aplicação desejada, da capacidade do sistema de gerenciamento em administrar a alocação de recursos, das limitações e características do sistema de comunicação entre eles, e também dos dispositivos de *hardware* específicos instalados em cada um. Os nós dedicados ao processamento gráfico e projeção de telas terão instalados placas gráficas. Os nós dedicados ao processamento de áudio e simulação acústica deverão possuir recursos compatíveis de memória, processador e interface de rede. O nó dedicado ao processamento e sonorização final deverá ter instalada(s) a(s) placa(s) de áudio multicanal, ou possuir conectividade (interface) a módulos de áudio multicanal instalados em *racks*.

³⁶ como por exemplo a passagem de mensagens para atualização de movimentação de câmera do servidor de rastreamento para os nós renderizadores, conforme o usuário se desloca no ambiente virtual.

Placas Multicanais

No sistema atual, é possível instalar-se placas de som (barramento *PCI*) em um ou mais nós. Para tarefas de espacialização com nível de imersão até 3, podemos admitir que um sistema de *software* sincronizado com algum mecanismo de *data-lock* seja suficiente. Entretanto, para tarefas mais complexas objetivando níveis de imersão sonora acima de 3, também um mecanismo mais fino de sincronismo entre placas deve ser estabelecido (*wordclock*).

Uma solução mínima que atenderia os sistemas de auralização baseados em Ambisonics até 3^a. ordem, bem como técnicas como 5.1, 6.1, 7.1 seria de 1 ou 2 placas contemplando até 16 canais simultâneos.

As gerações de placas existentes no mercado atualmente não possuem todas as características desejadas, mas permitem soluções escaláveis e combinações de equipamentos para atingir o objetivo primário de prover até 16 saídas de áudio analógicas e/ou digitais³⁷ simultâneas, sincronizadas (*word clock* compartilhado) e com baixa latência (atrasos).

Matrizes de alto-falantes (caixas acústicas)

A técnica de espacialização ou auralização usualmente é determinante na escolha do número e a disposição das caixas acústicas ao redor da área de audição. No caso da CAVERNA, a disponibilidade de área para instalação dos alto-falantes é restrita, a maior parte das opções encontra-se ao redor da estrutura, por baixo dos feixes de projeção de imagem, por cima (irradiando-se para dentro do volume cúbico) e aquém dos espelhos, distante das telas e atrás dos feixes de projeção.

Para sistemas de auralização, a regularidade da geometria escolhida para posicionamento das caixas frequentemente é crítica e deve ser respeitada para permitir os efeitos desejados.

³⁷ dependendo da capacidade das caixas em receber sinal digital direto.

As caixas de som devem obrigatoriamente apresentar blindagem eletromagnética, para não interferir na operação dos rastreadores. Também devem apresentar baixa distorção, curvas de resposta de frequência planas, cobertura de ampla faixa de frequências (20-20 kHz), baixo consumo, alta relação sinal/ruído, baixa distorção, pressão sonora superior a 70dB@1m, e padrão de irradiação não-estrito. Várias caixas ativas e passivas atendendo a esses requisitos estão disponíveis no mercado. As caixas acústicas ideais deveriam atender os seguintes critérios: a) tamanho reduzido (volume ocupado menor), b) peso inferior a 1 kg (menor carga física), c) recebam sinal digital. Não identificamos, porém, nenhuma marca/modelo no mercado que atenda ao conjunto de todos estes requisitos.

As melhores opções são as caixas bi-amplificadas, por poderem receber um sinal balanceado em nível de linha (baixa tensão, alta imunidade a ruído, impedância característica de linha) e por evitarem *crossovers* passivos que introduzem ruídos. É necessário que todas as caixas sejam de marca/modelo idênticos, para se evitar variações grandes entre as curvas de resposta de frequência e coloração espectral distintas, e garantir a produção regular e homogênea do campo sonoro por toda a área de audição pretendida.

Adicionalmente, é recomendável um sistema de alimentação (potência) balanceado e condicionado à prova de variações da rede, e que sejam capazes de filtrar harmônicos espúrios e entregar potência AC de qualidade. O uso de *no-breaks* chaveados (*PWM*), filtros de linha elétrica convencionais e a maioria dos estabilizadores e filtros de linha não são recomendados por introduzir ruídos e harmônicos no sistema que não são totalmente eliminados nas filtragens DC e contaminam a composição espectral dos sinais de áudio. Este fato que pode contribuir para a produção de artefatos em algumas faixas de frequências, e interferir na correta relação de fase esperada para produzir imagens sonoras estáveis.

Estamos utilizando na CAVERNA Digital as caixas acústicas modelo Merlin 100 da LANDO, parceira no projeto AUDIENCE. Estas caixas apresentam uma resposta em frequência relativamente plana, potência de até 100W rms, e sua maior vantagem está na alta qualidade de som que irradiam, permitindo uma excelente reprodução sem mascaramento do conteúdo espectral do programa original, nem artefatos.

3.2.2 Aplicações audiovisuais de RVI

A maior parte das aplicações e aplicativos hoje existentes de realidade virtual imersivos utilizam até 2 canais de áudio (estéreo). Alguns aplicativos exploram aspectos relacionados à espacialidade sonora, seja para sugerir uma ambiência ou condicionar a distância da fonte sonora ao usuário. A maior parte dos sistemas não excede o nível 3 de imersão sonora, ou não apresentam áudio de espécie alguma. Por exemplo, Kauff e Schreer introduziram um sistema de videoconferência 3D imersiva usando o que chamam de ambientes virtuais compartilhados de times de usuários. A idéia é uma abordagem à telepresença imersiva com uma implementação que garanta o contato visual e a reprodução de gestos entre os participantes remotos da videoconferência, como diferenciais para ampliar a sensação da telepresença ou realismo, entretanto sem oferecer uma estrutura de referência de áudio compatível para o sentido auditivo (KAUFF, 2002). Ausência de áudio em aplicações onde este sentido é fundamental é decepcionante.

As aplicações possíveis para ambientes de RVI são infinitamente amplas, e uma compilação compreensiva transcende os objetivos deste texto.

Alguns aplicativos existentes na CAVERNA possuem áudio (estéreo). Citamos por exemplo:

- *Crayonland* (desenvolvido com CAVELIB, roda em plataforma SGI e utiliza o sistema de áudio (BERGEN, 2004))
- OP_ERA (desenvolvido em C++ utilizando a biblioteca DICELIB para implementar sincronismo visual através dos nós (computadores) com *data-lock* e *frame-lock* através do envio de comandos via rede para cada nó, e sonorizando-se também através do envio via rede de instruções MIDI a um computador Macintosh onde um *setup* quadrifônico especial foi construído com Max/MSP, para espacialização quadrifônica) (CYCLING'74, 2005).

- Asa Delta – Passeio Virtual Pelo Rio de Janeiro³⁸ (desenvolvido na CAVERNA Digital, cenário construído em *Maya*, descrito em *X3D*, navegado com o Jinx, sonorizado com o programa FMOD (FIRELIGHT TECHNOLOGIES, 2004) em estéreo.



Figura 20 –Aplicativo de RVI audiovisual (*Asa Delta*)

3.3 Áudio em RVI

Nas décadas de 1980 e 1990 a sonificação de dados foi pioneira no uso de áudio para estender as possibilidades de análise de grandes volumes de dados, como suporte à visualização de dados. Em um mapeamento indireto, certos atributos do volume de dados eram associados a parâmetros sônicos, como frequência, amplitude ou timbres.

Em realidade virtual o áudio está frequentemente associado a um objeto visual e eventos, num mapeamento direto. Na década de 1990 era comum utilizar-se de ícones sonoros disparados a partir de um evento de navegação no mundo virtual, como por exemplo ao atingir um ponto do espaço ou tocar um objeto. O retorno do

³⁸ URL: <http://www.lsi.usp.br/interativos/nrv/hangGlidingRJ.html>

evento era mapeado em um evento sonoro. Muitas implementações e arquiteturas propostas para interação em mundos sintéticos utilizam este paradigma para retorno de áudio, por exemplo (GOBBETTI, 1993).

Em ambos os casos de uso, os recursos eletrônicos e computacionais mais comuns para se realizar o mapeamento de dados ou atributos do mundo virtual em som eram os sintetizadores comandados via *MIDI*, usualmente com saídas estéreo. Parâmetros, variáveis e eventos constantes no mundo virtual são mapeados em comandos *MIDI*, definindo-se timbre (instrumento), tocando-se seqüências específicas de notas e frases musicais, e definindo-se intensidade e modulações, entre outras possibilidades.

Sintetizadores *MIDI* virtuais (em *software*) e seqüenciadores oferecem também muitos recursos para síntese de som, inclusive mapeando-se espacialmente alguns atributos auditivos para sonorização em dois canais, utilizando efeitos como reverberação e panorama por amplitude. O fato de se poder mapear a leitura de dispositivos e sensores de entrada em comandos *MIDI* tornou esta técnica de mapeamento sonoro muito popular para aplicações em realidade virtual interativas e em tempo real na década de 1990 (SAVIOJA, 1999), (RUSCHIONI, 1997).

Neste ponto identificamos uma mudança de interesse no uso de áudio em aplicações virtuais, o que pode ser atestado pela evolução no emprego de uma terminologia comum de *sonificação de dados* para outra mais ampla de *espacialização sonora*. Presencia-se a partir de então um aumento gradual no interesse pela produção sonora multicanal e uma preocupação pela projeção de campos sonoros mais realistas que envolvam o(s) usuário(s). A produção multicanal, entretanto, tardou para atingir um público maior e uma maior quantidade de aplicações populares, e por muitos anos manteve-se mais conhecida nos meios acadêmicos e nos círculos da música contemporânea e das audições eletro-acústicas.

Embora a tecnologia *MIDI* ainda seja uma excelente ferramenta para controle e mapeamento de comandos útil para se implementar interfaces interativas em tempo real, em termos de qualidade sonora espacial temos hoje disponíveis ferramentas de síntese e produção sonora mais adequadas e completas do que sintetizadores

controlados via *MIDI*, principalmente quando o alvo é a simulação de campos sonoros.

3.3.1 Arquiteturas e sistemas para áudio em RV

Ao contrário do que se supunha inicialmente, a simulação acústica em realidade virtual e sistemas para áudio espacial em ambientes audiovisuais imersivos é uma área bem documentada, com várias implementações e sistemas propostos. (KLEINER, 1993), por exemplo, faz uma revisão bibliográfica a respeito, e cita nada menos que 109 artigos relacionados ao tema auralização. Entretanto, como muitas implementações abordam os sistemas de áudio como tecnologia e desenvolvimento suplementar àqueles de realidade virtual e de computação visual, verifica-se a ausência de uma estrutura de referência ou uma linha coerente de adesão conectando as contribuições.

Uma das arquiteturas mais completas e bem documentadas para áudio espacial em realidade virtual é o *DIVA (Digital Interactive Virtual Acoustics)* da Universidade de Tecnologia de Helsinki. SAVIOJA (1999) e outros abordam num artigo extenso o projeto e a implementação de um sistema de acústica virtual para o *DIVA*. Eles abordam os métodos para o modelamento das fontes sonoras, para o modelamento do espaço acústico, e para o modelamento do ouvinte, explicando praticamente todas as técnicas que consideraram e aquelas que preteriram. Detalhes de implementações de filtros e a arquitetura de auralização interativa no *DIVA* são apresentados, com ênfase na parte de modelamento em tempo real. Os autores citam que uma demonstração completa do *DIVA* foi feita no *SIGGRAPH'97*, consistindo de uma orquestra virtual com quatro instrumentistas.

BURGESS (1993a, 1993b) reporta uma primeira experiência com áudio 3D com ambientes virtuais e a organização de uma arquitetura de áudio 3D para aplicações de realidade virtual em três partes: (1) um cliente de espacialização, nível em que se situa a aplicação de RV, (2) um servidor de espacialização sonora, que faz a geração do som e a produção dos parâmetros necessários para comandar os processos de baixo nível nos canais físicos, e (3) um nível de canais físicos, que perfaz os cálculos finais de reverberação, efeitos e auralização com *HRTFs*. O autor investiga a

qualidade dos experimentos e considera fundamental uma *API* para conectar o nível mais alto das aplicações com o gerenciamento dos recursos físicos. Como parte do projeto *MERCATOR*, eles desenvolveram um sistema para áudio espacial de “baixo custo” utilizando uma *SGI Indigo* (para calcular posições), uma *Sun SPARCstation IPX* e uma placa *DSP Ariel S-56x* para realizar a computação de pares de *HRTF's*, auralizando uma fonte sonora. O autor ressalta que por muitos anos a comunidade científica ignorou a parte de áudio 3D devido aos altos custos de *hardware* especializado, mas que com a emergência de dispositivos *DSP's* não mais se admitia desculpas.

NAEF (2002) introduz o sistema *blue-c*, que roda numa *SGI*, e cujo subsistema de áudio tem *API* própria, não usando as comerciais *DS3D*, ou *Avango*.

Embora visualmente imersivos, identificamos poucos ambientes do tipo *CAVE* possuindo sistemas de áudio imersivo com capacidade de localização de fontes sonoras no meio virtual, e nenhum documentado capaz de auralização 3D, permitindo experiências de imersão precisa em campos sonoros 3D, e com a correta experiência acústica do ambiente projetado.

OGI (2003) e outros desenvolveram um processador de áudio espacial para ambientes do tipo *CAVE*, o *Wave Engine*, baseado numa placa *DSP*. No projeto do sistema os autores demonstram especial preocupação em garantir a suavidade nas transições entre um filtro de auralização e outro. Utilizam uma técnica mista para simulação de campo sonoro (solução da equação integral de Kirchhoff através do método de integração de raios sonoros finito) e geração de *IRs*. Cada raio sonoro calculado é posicionado utilizando-se *SVPA*. Integrando-se as respostas ao impulso divididas por todos os alto-falantes e para todos os raios de som uma função de transferência final é construída (*IR*).

A *Wave Engine* é construída sobre um *DSP ADSP-21065L* da *Analog Devices*, e possui 16 canais de processamento de áudio. Em cada canal (mapeando um alto-falante) há no *pipeline* um filtro para convolução de resposta impulsiva e um filtro de compensação do efeito das telas. Eles concluem que um movimento suave durante

uma navegação pelo ambiente virtual é obtida quando os filtros de convolução são mapeados numa grade de metro em metro (OGI, 2003).

Muitos sistemas exploram técnicas de panoramas estéreo, que sugerem (com limitações) imagens sonoras formadas numa região entre os alto-falantes, e movimento no eixo entre as mesmas. A região onde estes efeitos são mais bem percebidos é com frequência muito estreita, restringindo-se a uma área focal (*hot-spot*) longe da qual os efeitos de posição, movimento ou envolvimento são perdidos ou mal formados.

ECKEL (1998) descreve a implementação do sistema de áudio da caverna “*CyberStage*” do GMD³⁹ (Alemanha), uma CAVE de 4 lados de dimensões 3x3x3m, contando com um sistema de projeção de áudio de 8 canais especializado. Um sistema servidor de áudio baseado em *software* foi desenvolvido no contexto do projeto *ISIS* (*Integrated Simulation of Image and Sound*): o *CSS* (*CyberStage Sound Server*). O *CSS* roda numa máquina *SGI* e suporta, segundo os autores, todos os tipos de interfaces de áudio, incluindo *ADAT'S* de 8 canais. Todos os módulos de renderização de áudio utilizam panorama por amplitude para a codificação da diretividade. A escolha para codificação envolvente se deu por esta técnica ao invés de Ambisonics porque, segundo os autores, ela produz melhores resultados em localização sob as condições acústicas particulares de sua caverna (ambiente reflexivo devido às telas e espelhos).

ECKEL (1998) reconhece a importância da questão da localização sonora na navegação e orientação em ambientes imersivos, bem como a necessidade de imprimir a “assinatura acústica” do ambiente, entre outros aspectos importantes (como o padrão de irradiação das fontes sonoras). A primeira versão do sistema (1996) usava 4 alto-falantes provendo localização 2D. A segunda (1997) usava 8 alto-falantes montados numa clássica configuração de cubo, orientada para localização 3D. As 4 caixas superiores foram instaladas por cima das telas. As inferiores foram montadas por trás das telas, nos cantos, tomando-se o cuidado e a

³⁹ German National Research Center for Information Technology, Sankt Augustin, Alemanha.

atenção para a restrição principal que afeta este tipo de ambiente imersivo: a restrição do cone de projeção visual, que não pode sofrer oclusão. Eckel considera este problema uma questão crítica no *design* de sistemas de multiprojeção como *CAVEs*, ainda sem solução. Medidas acústicas foram realizadas para estimar a atenuação dependente da frequência provocada pelas telas, e projetar filtros para compensar este efeito. A acústica da sala por trás das telas foi tratada, para evitar reflexões.

O projeto CARROUSO (*Creating, assessing and rendering in real-time of high-quality audio-visual environments in MPEG-4 context*)⁴⁰ tem o objetivo de desenvolver nova tecnologia para possibilitar a transmissão de um campo sonoro, gerado virtualmente ou gravado, para um ambiente remoto, contando com meios para controlar aspectos temporais e espaciais e propriedades do campo sonoro, e combinado com informação visual. A intenção é romper com as limitações dos sistemas estereofônicos e multicanais tradicionais, baseando-se primariamente em duas tecnologias: MPEG-4 para codificação orientada a objeto e manipulação interativa de áudio 3D, e *WFS* para produzir espaços sônicos verdadeiros, em contraposição à estereofonia (INFORMATION SOCIETY TECHNOLOGIES, 2001).

O projeto MERCATOR tem como objetivo prover acesso transparente a “*X Windows*” (interfaces gráficas para computador) para usuários cegos ou com a visão severamente comprometida. Primeiro o projeto foca numa estrutura de referência para monitorar, modelar e traduzir interfaces gráficas sem modificar as aplicações em si, e, segundo, o projeto visa a traduzir as interfaces gráficas em não-gráficas, e suas interfaces são feitas de componentes auditivos relacionados com os componentes da interface gráfica, como menus, botões, janelas de diálogos, etc. O sistema faz uso de síntese de voz e ícones sonoros para produzir uma versão auditiva de interfaces gráficas⁴¹.

⁴⁰ O sítio do projeto na Internet é: <http://www.idmt.de/projects/carrouso/index.html>

⁴¹ maiores detalhes em <http://www.cc.gatech.edu/gvu/multimedia/mercator/mercator.html>

3.3.2 Ferramentas para áudio em ambientes de RVI

Em princípio qualquer tecnologia ou ferramentas/bibliotecas para áudio constante da Tabela 5 poderia ser portada ou adaptada para uso em ambientes de realidade virtual. Entretanto, o que se verifica é que a maioria das tecnologias baseadas em modelos perceptuais para espacialização (tipicamente voltadas para jogos) não são preferidas nos ambientes audiovisuais imersivos e em plataformas profissionais de realidade virtual, possivelmente devido a uma combinação de fatores, como a imprecisão física destes modelos e à necessidade real de auralização em tempo real.

Ainda, nos ambientes do tipo *CAVE* é comum a adoção de bibliotecas de gerenciamento globais, que incorporam as ferramentas de computação gráfica, de controle do ambiente computacional, e também bibliotecas para áudio, que na maioria das vezes satisfazem requisitos minimalistas de uso de som, principalmente nos ambientes e aplicações onde a visualização é priorizada.

Centramos nossa presente abordagem nas ferramentas utilizadas para programação e geração de áudio nestas plataformas. Nelas, o cenário de uso envolve atividades de pesquisa e desenvolvimento em engenharia e arquitetura, treinamento em funções especializadas (utilizando modelos e maquetes virtuais), em projeto de estruturas arquitetônicas, civis, e grandes espaços físicos, na navegação virtual em ambientes inóspitos, e finalmente em entretenimento avançado. A navegação nestes cenários virtuais requer técnicas imersivas de visualização e auralização avançadas, para garantir o grau de realismo e nível de imersão necessários.

As seguintes ferramentas são usualmente utilizadas em sistemas e ambientes voltados para aplicações de realidade virtual imersivas:

VSS

O *VSS* (*Virtual Sound Server* ou *Vanilla Sound Server*) é um pacote de *software* independente de plataforma, voltado para produção de som orientado a dados, e controlável pelas aplicações interativas⁴². Trata-se de uma biblioteca para

⁴² http://www.isl.uiuc.edu/VSS/vss_reference/vss3.0ref.html

plataformas Irix, Linux, e Windows criada para complementar aplicações interativas desenvolvidas em C ou C++ com recursos sonoros (GOUDESEUNE, 2003).

O programa permite ao usuário controlar, coordenar e sincronizar sons produzidos por síntese direta por *software*, reprodução de amostras/arquivos, e sintetizadores externos como *MIDI*, *MAX* e *OpenSound Control*. Antes de se utilizar o *VSS* é necessário se criar um servidor para sonorização. Após, é estabelecida uma conexão entre o cliente e o servidor, e o cliente pode enviar comandos para a geração de sons. O controle e comunicação são efetuados por intermédio de atores, estes atores podem controlar diversos sons ao mesmo tempo. A comunicação é feita diretamente por *IP*, esta é especificada em linha de comando.

O sistema oferece as seguintes opções para disposição dos alto-falantes: a) 2 canais, b) 4 canais (frontal direito esquerdo e traseiro direito e esquerdo), e c) 8 canais (cubo).

Avango

Avango consiste numa estrutura de referência orientada a objetos para aplicações em realidade virtual⁴³. A biblioteca Avango usa um servidor de áudio baseado no sistema de processamento de som MAX/FTS do *IRCAM*. O controle é feito via rede local, através do protocolo *UDP*. A integração com a cena audiovisual é descrita por meio de nós especiais no gráfico de cena.

Bergen Sound Server & Library

A biblioteca *Bergen* foi desenvolvida para ser uma versão simplificada do *VSS* distribuída gratuitamente (BERGEN, 2004). Assim como na arquitetura do *VSS*, existem os papéis do servidor e o cliente. Uma versão para Linux foi disponibilizada.

⁴³ Maiores informações no sítio Internet da ferramenta, em <http://www.avango.org/>

FMOD

O FMOD é uma biblioteca de desenvolvimento para aplicações sonoras com bons recursos, bem documentada, e gratuita (FIRELIGHT TECHNOLOGIES, 2004). Com ela é possível sintetizar sons com baixo consumo de processamento e memória. Uma de suas principais vantagens é operar em diversas plataformas (Windows, Windows CE, Linux, Macintosh, Play Station 2, Xbox) o que facilita muito a sua portabilidade. Suporta computadores com até 16 placas de som ou o uso de dispositivos como *videogames*. Também permite explorar recursos multimídia específicos da arquitetura, como por exemplo, o sistema *MMX*⁴⁴ que pode aumentar a performance e qualidade da aplicação em até 2 ou 3 vezes.

Um aspecto atraente é seu suporte a diversos formatos de áudio (*WAVE, AIFF, Ogg Vorbis, MP2, MP3, MIDI, AVI, etc.*). Outra vantagem é sua capacidade nativa de executar simulações de espacialização (panoramas) e suporte a efeitos estereofônicos e Doppler.

VRJuggler Sound Library

O *VRJuggler* é uma biblioteca ampla, alternativa ao *CAVELib*, para gerenciamento de *CAVE's* e para aplicações interativas, e possui uma biblioteca de desenvolvimento exclusiva para áudio, a *VRJuggler Sound Library*, além de suportar os recursos de som da biblioteca *Performer*. Sua utilização é, entretanto, de relativa complexidade. O grande inconveniente desta ferramenta é estar presa à biblioteca *VRJuggler*, limitando seu uso.

AuSIM

A *AuSIM Engineering Solutions* desenvolveu o sistema *AuSIM3D*. A empresa informa que a tecnologia “foi projetada para aplicações interativas como *CAVE's* e outros ambientes imersivos, com a habilidade de integrar dispositivos *head-tracking* e fornecer uma perspectiva do mundo auditivo dinâmica, aliviando o problema

⁴⁴ *Multi Media eXtensions*, tecnologia da Intel, consiste num conjunto de instruções para acelerar aplicações multimídia em *PC*.

conhecido por ‘perspectiva congelada’, e permitindo a correta correlação de todos os objetos gráficos móveis mesmo com o usuário movendo-se⁴⁵.

A solução primária da *AuSIM* é a simulação bi-aural, para fones de ouvido. A segunda solução é panorama por amplitude (SVPA), para matrizes de alto-falantes (multicanal). Para aplicações de *CAVE* suportam auralização independentes via fones de ouvido para diferentes usuários, utilizando fones com rastreamento.

A implementação de SVPA prioriza a liberdade do usuário dentro da *CAVE*, com imagem ótima focada no centro da *CAVE*, e sonorização para um mínimo de 14 alto-falantes sugeridos, para melhores resultados.

O sistema bi-aural da *AuSIM* é utilizado em várias *CAVE*'s, por exemplo a da *NASA Langley's International Space Station (EUA)*. O SVPA é usado por exemplo na Universidade de Helsinki, mas a *AuSIM* não tem *CAVE*'s com soluções suas de SVPA instaladas até hoje⁴⁶.

Lake Huron

Lake Huron é uma plataforma proprietária de *hardware/software* baseada em *DSP* para aplicações otimizadas de áudio. Uma arquitetura modular expansível provê a flexibilidade para implementar soluções de complexidade diferentes. O sistema inclui um equipamento de *rack* para processamento *DSP* e placas de áudio para entrada/saída de áudio (analógico e digital) escalável de 2 até 512 canais.

O sistema provê recursos como ferramentas de configuração, *virtual patchbay* (central de conexões e gerenciador de entrada/saída), *mixer*, equalização, e uma suíte de módulos para simulações acústicas com diversas opções de renderização/auralização que incluem: *AniScape* (modelamento em tempo real do movimento da fonte e ouvinte em ambiente 3D), *HeadScape* (aplicação para avaliação de modelos acústicos baseada em fones de ouvido), *MultiScape* (aplicação

⁴⁵ conforme em <http://www.ausim3d.com/FAQ/audio.html>

⁴⁶ Informação pessoal via *e-mail* com representante da *AuSIM*.

de navegação para múltiplos ambientes, fontes e ouvintes), *BinScape* (para renderização bi-aural com fones de ouvido), *SpaceArray* (aplicação multicanal para posicionamento de múltiplas fontes em espaço 3D utilizando até 50 alto-falantes), e *Convolver* (módulo para convolução de baixa latência e alto desempenho proprietário da *Lake*, capaz de 278.244 *taps* de convolução e instantânea troca de filtros). Este último módulo é talvez o mais atraente para desenvolvedores, e esta tecnologia tem sido licenciada para outros sistemas.

McGRATH (1994) reporta num artigo técnico da *Lake* experimentos com codificação e reprodução de campo sonoro com Ambisonics utilizando 12 alto-falantes e também reprodução em fones de ouvido, numa implementação associada ao módulo *AniScape*, que incorpora suporte a *B-Format*. Reporta que o campo sonoro é criado por simulação virtual acústica codificando em *B-Format* o raio direto e os 6 primeiros raios, além de um algoritmo para modelar a reverberação. A simulação acústica no *AniScape* é suprida pelo pacote *CATT-Acoustics*⁴⁷, e o sistema, proprietário da *Lake*, utiliza os recursos de convolução de baixa latência patenteados da mesma. Dos sistemas revisados, este é o o único que combina ferramentas de alto desempenho para simulação acústica, suporte a auralização por Ambisonics e recursos de convolução de alto desempenho. Entretanto não há implementações adaptadas a ambientes audiovisuais imersivos com *CAVE*'s documentadas.

Wave Engine

O *Wave Engine* é o processador e gerador de áudio multicanal baseado em *DSP* desenvolvido por OGI (2003) e apresentado no item anterior.

3.4 Uma arquitetura modular para produção sonora espacial

A construção e correta apresentação de uma cena audiovisual virtual imersiva envolve uma seqüência de três tarefas principais:

⁴⁷ informação disponível no sítio Internet em: <http://www.euphonia.fr/lakelogiciel.htm> (acesso em 30 maio 2005).

- a) a descrição da cena,
- b) o processamento audiovisual (processamento gráfico e sonoro),
- c) a apresentação audiovisual (projeção e sonorização).

A concepção do aplicativo e as relações temporais e espaciais entre os objetos está embutida na descrição da cena audiovisual. Os modelos empregados na geração do material audiovisual estão embutidos no processamento gráfico e sonoro. A interação com o usuário e a geração dos sinais de vídeo e áudio de saída estão embutidos na apresentação audiovisual.

Para construir e apresentar uma cena audiovisual deve-se mapear os objetos presentes na cena, seus atributos visuais e sonoros que vão guiar a simulação do ambiente audiovisual, e o roteiro de apresentação da cena.

Na hierarquia de tarefas, somente na descrição da cena é que ambos os atributos visuais e sonoros dos objetos são considerados em conjunto. Nas tarefas seguintes estes atributos são considerados separadamente e em paralelo: o sistema de vídeo será responsável pela utilização dos atributos visuais da cena para gerar as projeções finais esperadas, enquanto o sistema de áudio será responsável pela utilização dos atributos sonoros da cena para guiar a correta produção e sonorização do campo sonoro desejado. A simulação visual e as tarefas associadas (cálculos, renderização e projeção) são bem abordadas em um grande número de trabalhos abordando a produção e o processamento da parte visual, e não é coberta neste trabalho.

Limitações das arquiteturas fechadas

Com referência a arquiteturas para produção sonora espacial em RV, um trabalho abrangente, e uma das poucas referências mais completas no sentido de se identificar todas as fases associadas à produção de ambientes acústicos virtuais interativos é apresentado por SAVIOJA (1999).

Sistemas comerciais, bibliotecas e ferramentas existentes para mapeamento de cenas acústicas, síntese e sonorização espacial utilizam estruturas, hierarquias ou componentes de programação (métodos ou procedimentos) diversos, não havendo um consenso ou padrão industrial definido que governe ou possibilite a

interoperabilidade entre aplicações programadas com ferramentas diferentes. Isso limita a portabilidade, interoperabilidade e usabilidade de *software*, bem como a integração de ferramentas que utilizem técnicas distintas (e mesmo complementares) de síntese de áudio 3D.

Em algumas soluções o mapeamento da cena acústica e a geração dos sons finais (canais) a serem endereçados aos alto-falantes são calculados a partir de primitivas de código chamadas diretamente no programa fonte principal, que abstraem do usuário a técnica que está sendo usada para modelar a resposta acústica e produzir os sinais que as caixas de som tocarão. Isso pode ser eficiente do ponto de vista de programação direta, enxuta e simplificada da cena acústica, mas vale somente para algumas plataformas de programação e linguagens, e também limita o usuário a utilizar somente as técnicas e métodos de sonorização espacial que aquela ferramenta implementa. Claramente trazem-se situações onde o uso daquela ferramenta não é adequado em vista da qualidade espacial desejada, isto é, em vista do nível de imersão que se deseja. Exemplos de arquiteturas fechadas *top-down* de programação incluem o *DS3D*, e mesmo o *OpenAL* apresenta limitações neste aspecto. As arquiteturas modulares baseada em *plugins* como o *VST*, entretanto, permitem e estimulam a interoperabilidade e integração de *softwares*.

Vantagens de uma arquitetura aberta

Do ponto de vista de uma arquitetura genérica para escrever e produzir uma cena sonora, propomos uma hierarquia de tarefas modularizada, de tal forma a segmentar as principais fases da produção sonora espacial em blocos distintos, e agrupar as tarefas afins em módulos cujas entradas e saídas sejam bem definidas. Desta forma estaremos contribuindo para que as aplicações de áudio 2D/3D possam ser desenvolvidas planejando-se uma interoperabilidade e capacidade de se utilizar algumas funções de um *software* numa camada, e eventualmente outras de outro pacote computacional noutra camada. Por exemplo, pode-se desejar utilizar o *X3D* para descrever a cena acústica, e utilizar o *FMOD* ou *OpenAL* para tocar cada sinal/canal, e deixar a simulação acústica e cálculos para gerar os sons codificados em 3D com um simulador de acústica virtual, como o *RAIOS* ou o *CATT-Acoustics*, onde a localização precisa das fontes sonoras seja fielmente obtida.

Neste esquema, é importante frisar a importância de haver um formato de representação intermediário dos sons, que contenha, além dos próprios sons das fontes sonoras, a informação espacial acerca do ambiente e de sua propagação acústica neste, incluindo a diretividade das fontes sonoras e as transformações ocorridas entre fonte-ouvinte.

Hierarquia de Camadas

Uma proposta para resolver estas questões e oferecer a flexibilidade desejada para a produção da experiência auditiva espacial está na proposição e adoção de uma arquitetura modular orientada a camadas, absorvendo e agrupando ou adensando as principais funções e tarefas envolvidas em quatro grandes blocos, assim resumindo e formatando todo o processo à execução de quatro grandes fases sequenciais de atividades. As distintas tarefas e processos envolvidos serão adequadamente atribuídos a uma camada na hierarquia de camadas, segundo seus objetivos, similaridades e funções.

Quatro camadas são propostas, como ilustrado na Figura 21, cada qual responsável pela execução das seguintes funcionalidades:

- camada de descrição de cena acústica
- camada de simulação acústica
- camada de codificação espacial
- camada de decodificação e sonorização

A Figura 21 ilustra o conceito de camadas numa arquitetura modular para produção sonora, e as funcionalidades associadas a cada uma.

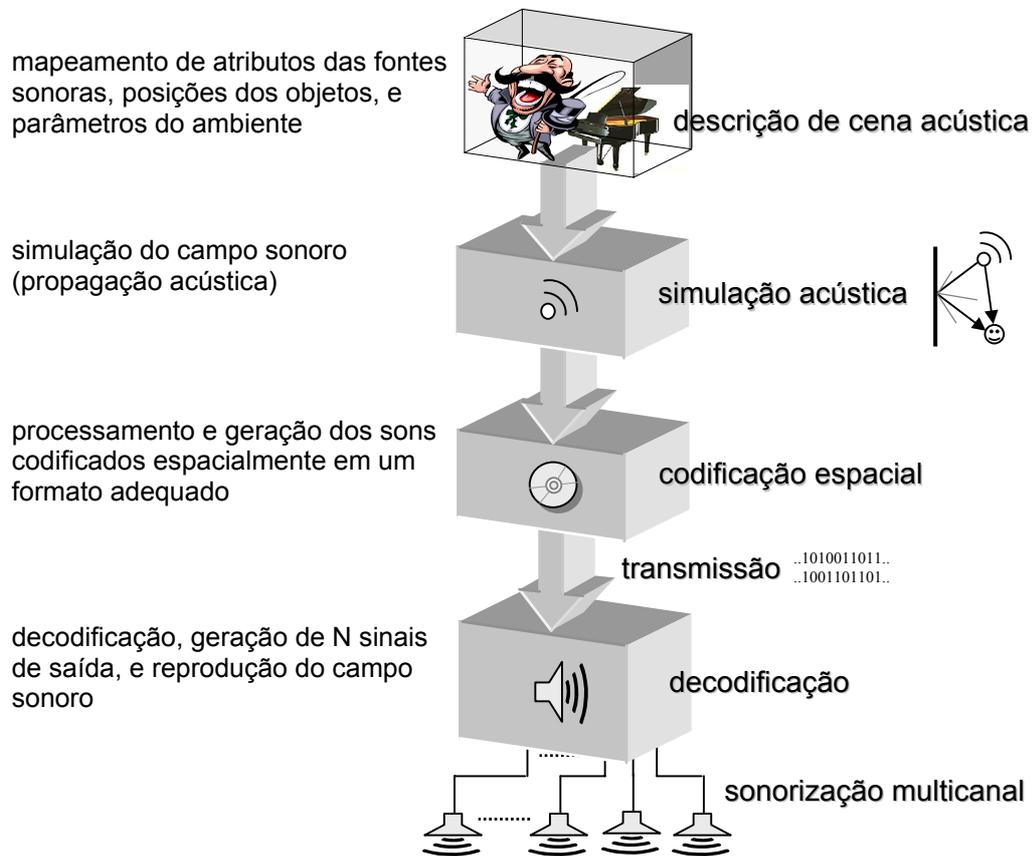


Figura 21 – Arquitetura modular para produção de áudio espacial em 4 camadas

Para que a hierarquia de camadas seja utilizável é preciso especificar a interface entre elas, mais especificamente os possíveis sinais de entrada e saída que poderão estar presentes na interface e vão compatibilizar a saída de um bloco com a entrada do próximo. A Figura 22 ilustra um exemplo de interfaceamento genérico, discriminando um conjunto de possíveis sinais presentes nas interfaces.

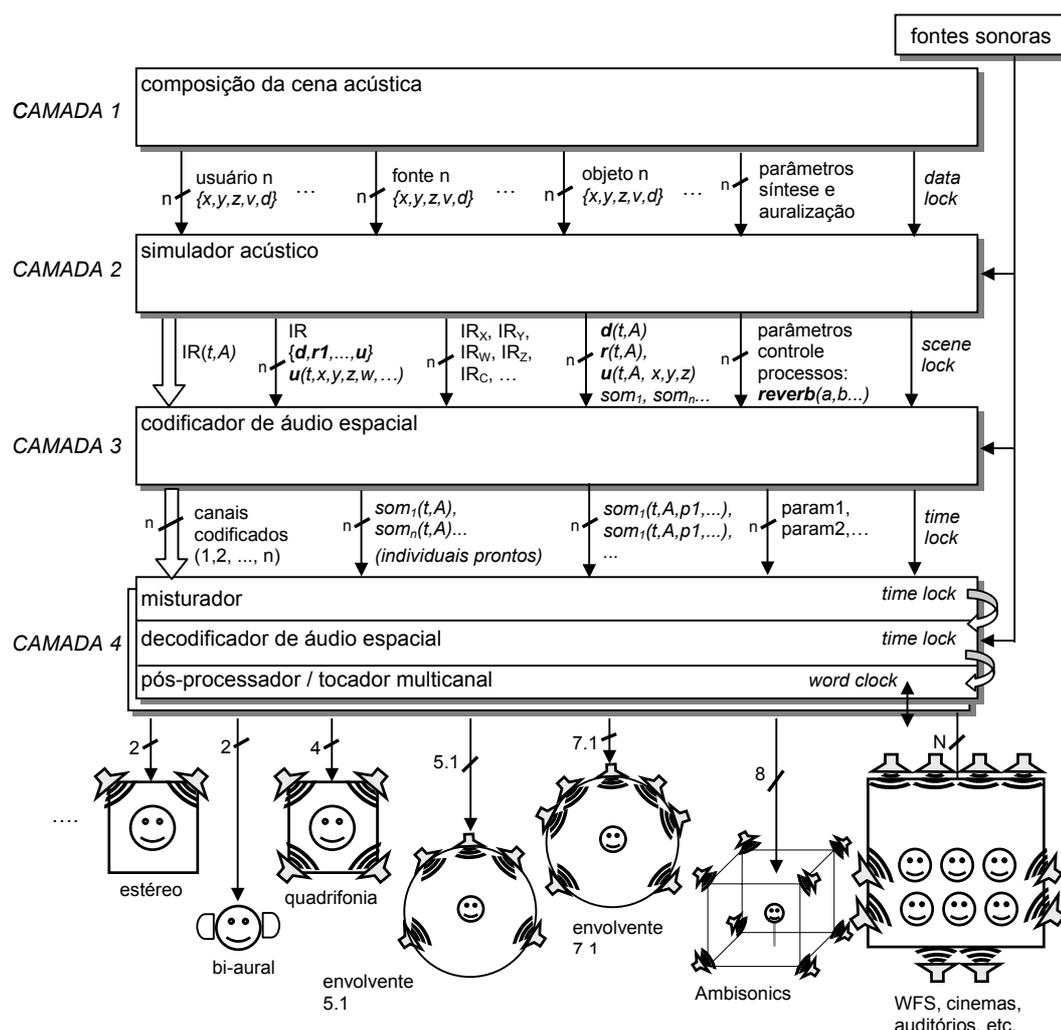


Figura 22 – Modelo de referência para interfaceamento entre camadas

Na interface entre a camada 1 e 2, os parâmetros de síntese e auralização podem determinar o nível de imersão desejado, e transmitir parâmetros de masterização e combinação dos canais (como níveis de edição, multiplexação de canais, coeficientes de matriciação, codificação conjunta, e outros tipos de combinações), comandos para síntese, para auralização e sonorização utilizando-se *API's* e ferramentas para processamento específicas, dados dinâmicos da cena (como oclusão), parâmetros acústicos subjetivos (como calor, brilho, etc.), parâmetros para equalização, etc. O sincronismo de *data lock* permite a coerência de *cache* para evitar que instâncias diferentes dos módulos posteriores tenham acesso a cenas distintas.

Na saída da camada de simulação acústica (camada 2) devemos encontrar já calculados os parâmetros principais da auralização, que refletem as características espaciais e temporais do espaço da gravação ou do espaço virtual (no caso de

acústica virtual). Assim, nesta interface encontramos tipicamente respostas ao impulso do ambiente, que podem estar preparadas e agrupadas segundo a forma de captura (como foram calculadas ou medidas) ou segundo a forma como serão utilizadas, por exemplo para renderizar cada fonte sonora, para conjuntos de fontes, para cada posição medida, etc. Saídas de som já processadas (por exemplo já com algum efeito aplicado) podem ser sinais válidos nesta interface. Dependendo da técnica utilizada na renderização podemos também encontrar nesta interface comandos já renderizados para o controle e a configuração das máquinas de reverberação, circuitos e algoritmos que serão usados na camada de sonorização, ou mesmo parâmetros de baixo nível que criam as linhas de atraso, coeficientes de filtros, etc.

Tipicamente a transmissão do sinal sonoro totalmente codificado se dá entre a camada 3 e 4. Esta transmissão pode se dar utilizando uma camada de transporte e multiplexação, como por exemplo, a utilizada nos padrões de transmissão de TV digital utilizando *MPEG-2* e agora também *MPEG-4*. Mais simples, esta transmissão pode se dar internamente na máquina de auralização por barramentos digitais, *DMA* ou outro mecanismo de compartilhamento de memória, e também por mídias de armazenamento. Os sinais na saída da última camada (camada 4) são os canais de som já preparados para cada alto-falante reproduzir.

Esta arquitetura sugere um modelo de referência genérico para a implementação global ou parcial de várias funções no processo de produção de audio espacial, em especial útil para o universo de aplicações nos sistemas para realidade virtual imersiva, incluindo aí a indústria de jogos e entretenimento audiovisual.

Em seguida apresentamos o escopo, funcionalidades e ferramentas envolvidas em cada camada. Uma proposta de especificação formal da estrutura do protocolo, blocos de dados e sintaxe associada está, entretanto, além do escopo deste trabalho, cabendo a trabalhos futuros orientados a este aspecto específico.

3.4.1 Formatos para descrição e representação de cena

A descrição da cena audiovisual é realizada através da identificação dos objetos presentes na cena, do seu mapeamento na geometria (posição) e da descrição dos seus atributos, visuais e auditivos (acústicos).

A descrição da cena acústica deve prover um conjunto mínimo de parâmetros e características das fontes sonoras para que os processos subjacentes responsáveis pela criação da experiência sonora virtual sejam devidamente alimentados. Sobre a fonte sonora interessa-nos conhecer seu tamanho, origem do som, técnica de síntese, padrão de irradiação, etc. Do ambiente, minimamente interessa-nos conhecer seus contornos (limites), volume, tipos de objetos, as propriedades acústicas dos materiais, suas geometrias, medidas (tamanhos), localização, parâmetros subjetivos de qualidade, etc. Finalmente, acerca do ponto ou área de audição é necessário saber sua geometria, a extensão (tamanho) e posição. O modelamento de cada item constituinte da cena sonora será abordado adiante.

Existem vários formatos para descrição de cenas 2D e 3D. Abaixo relacionamos três mais importantes no caso de aplicações em realidade virtual.

a) VRML

O *VRML* foi concebido com o propósito de descrever uma cena 3D de forma simples e eficiente para que pudesse ser disponibilizado na Internet, através de uma linguagem independente de plataforma. O *VRML* se baseou no *OpenInventor*, e sua primeira apresentação foi em 1994, no *First International Conference on the World Wide Web*. A versão *VRML97* é publicada sob o número 14772-1 pela ISO/IEC em 1998.

A forma mais comum de se acessar um conteúdo *VRML* é através de um navegador de páginas da Internet, que suporte este formato. Usualmente um aplicativo complementar (*plugin*) viabiliza as funcionalidades de visualização e navegação 2D e/ou 3D.

O modelo de áudio do *VRML* oferece meios para composição de cena de áudio orientada a objetos e cobertura de algumas funções de espacialização sonora. Dois

nós na hierarquia são orientados para áudio: `Sound` e `AudioClip` que têm duas tarefas: 1) associar um som a uma cena audiovisual e localizá-lo numa posição específica e 2) apresentar ou reproduzir este som nas coordenadas locais do sistema do terminal de exibição. A Figura 23 mostra os nós de áudio na estrutura de árvore do *VRML*.

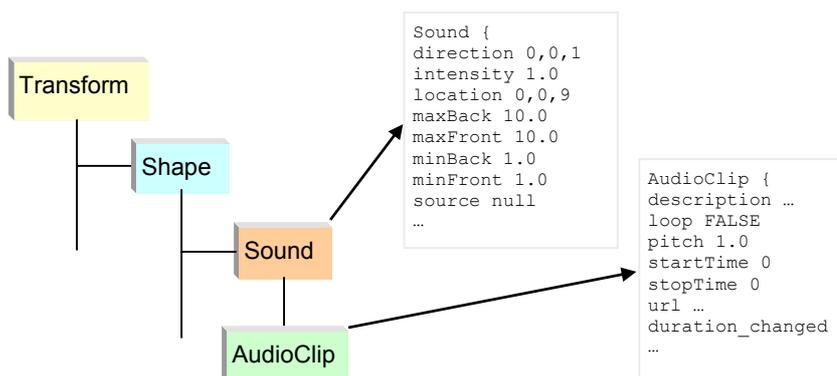


Figura 23 – Nós de áudio na estrutura de árvore do *VRML97*

Numa árvore de descrição de áudio o nó `Sound` ocupa a hierarquia mais alta. Os atributos espaciais são dados pelos campos *location* e *direction*. Os campos *maxBack*, *minBack*, *maxFront* e *minFront* especificam a característica de diretividade e padrão de atenuação dependente da distância à fonte sonora. O nó `AudioClip` permite associar uma fonte sonora real (o som) ao nó `Sound`, e controlar sua reprodução, iniciando, parando ou alterando atributos de *play-back*.

b) X3D

O *X3D* é um padrão aberto desenvolvido e introduzido em Agosto de 2001 pelo *Web3D Consortium* para suceder e expandir as capacidades do *VRML97*, utilizando XML (*Extensible Markup Language*) como linguagem textual base (WEB3D CONSORTIUM, 2005).

O *X3D* foi concebido para suportar diversos tipos de sistemas e dispositivos de RV como *CAVE's*, luvas, panoramas, óculos estereoscópicos, *mouses* 3D, capacetes de realidade virtual *HMD*, e opera em computadores simples com capacidades 3D.

Sua estrutura descritiva busca cobrir as áreas mais importantes para realidade virtual, em especial para aplicações na Internet, como gráficos 3D e 2D, animações diversas, som espacializado, interatividade, programabilidade e simulações físicas. A título de exemplo, a Tabela 7 mostra um trecho de código de uma descrição de cena audiovisual utilizando-se o *X3D*.

Tabela 7 – Exemplo de descrição de cena audiovisual utilizando *X3D*

```

<X3D>
  <Scene>
    <Transform DEF="piso" center="7.5 0.05 10">
      <Shape>
        <Box size="15 0.1 20"/>
        <Appearance>
          <Material diffuseColor="1 1 1"/>
        </Appearance>
      </Shape>
      <Sound/>
    </Transform>
    ...
    <Transform DEF="bola_de_som" translation="-5 1.7 -13">
      <Shape>
        <Sphere/>
        <Appearance>
          <Material diffuseColor="0 0 1"/>
        </Appearance>
      </Shape>
      <Sound>
        <AudioClip url="som.wav"/>
      </Sound>
    </Transform>
    <Transform DEF="ouvinte" translation="0 1.7 0">
      <Shape>
        <Sphere/>
        <Appearance>
          <Material diffuseColor="1 0 0"/>
        </Appearance>
      </Shape>
    </Transform>
  </Scene>
</X3D>

```

A cena acima gera uma cena simples tridimensional, que consiste de um piso de dimensões 15x20x8, um objeto “ouvinte” posicionado quase ao centro (em vermelho), e um objeto sonoro (em azul), próximo a uma parede ao fundo. A Figura 24 mostra a cena produzida por este código.

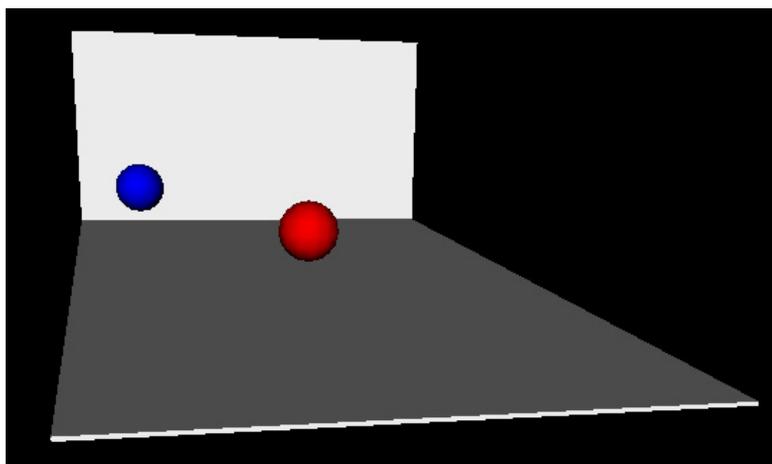


Figura 24 – Cena audiovisual gerada pelo código *X3D* da Tabela 7

O *X3D* permite uma modelagem textual e programável, o que facilita a criação e alteração de cenas, criando bons efeitos gráficos, além de uma interatividade controlada. Foi adotado como ferramenta descritiva de cena no navegador Jinx (SOARES, 2004a), desenvolvido na CAVERNA Digital.

O Jinx embarca recursos para a configuração, disparo e gerenciamento dos processos do subsistema de renderização visual da CAVERNA, sendo executado em diversos nós do aglomerado computacional de forma síncrona, para a geração de cada projeção. Sua estrutura é adequada para embutir também a descrição da cena acústica e gerenciar o disparo e controle do subsistema de auralização, portanto foi adotado como veículo integrador de ambos os subsistemas, visando naturalmente à produção sincronizada e controle unificado da visualização, auralização e interatividade em tempo real.

Assim como no *VRML*, o *X3D* possui um esquema para descrição de cena acústica e também do padrão de diretividade e atenuação do som dependente da distância, mas apresenta uma sintaxe incompleta para cobrir todos os aspectos importantes para as tarefas de produção sonora espacial. A Figura 25 mostra a geometria de referência que define os parâmetros de espacialização do *X3D* e do *VRML* que foram apresentados na Figura 23.

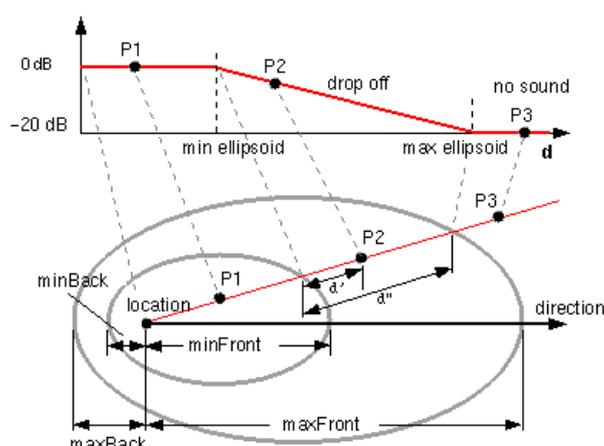


Figura 25 – Geometria de espacialização usada no X3D, VRML e MPEG-4 AudioBIFS⁴⁸

A sua estrutura de formação de nós se assemelha à da hierarquia utilizada no MPEG-4 BIFS⁴⁹, entretanto a parte referente à composição da cena acústica do X3D está muito aquém, requerendo contribuições substanciais para um modelamento e descrição mais completa das fontes sonoras, do meio acústico (mundo virtual do ponto de vista sonoro) e dos pontos de audição.

c) MPEG-4 BIFS

O padrão *MPEG-4 AudioBIFS* e sua extensão, o “*Advanced AudioBIFS*” (segunda versão do *BIFS*) inclui funcionalidades fundamentais para uma descrição e modelamento de cenas acústicas, como o “*advanced sound environment modeling*” para cenas virtuais acústicas.

Dos três padrões, este é o que apresenta uma cobertura mais ampla, especificamente orientada à descrição e composição da cena sonora. São cobertos propriedades do ambiente e atributos, como reflexões, reverberação, efeito Doppler, obstrução/occlusão (devido a objetos entre as fontes e o ouvinte) e modelamento da diretividade da fonte sonora (padrão de irradiação). Tais parâmetros permitem a

⁴⁸ Figura obtida em (WEB3D CONSORTIUM, 2005).

⁴⁹ ambas por sinal derivadas da hierarquia proposta no *VRML97*.

computação dinâmica, realista e em tempo real (para sistemas capazes deste requisito) do ambiente sonoro.

Este padrão de descrição de cenas acústicas foi introduzido no escopo do *MPEG-4 BIFS (Binary Format for Scenes)*. VÄÄNÄNEN (2002) aborda com profundidade o assunto, desde o histórico do formato e heranças comuns com *VRML97* até a implementação de nós avançados no *Advanced AudioBIFS*, que cobrem aspectos de modelamento da propagação acústica, a composição e forma de apresentação de cenas de áudio 3D. Particularmente digno de citação, é a cobertura de descrições de aspectos perceptuais associados à cena acústica, como presença da fonte sonora, calor, envolvimento, peso, leveza, vivacidade, brilho, bem como parâmetros físicos associados aos objetos da cena, como por exemplo parâmetros acústicos dos materiais (coeficientes de absorção, transmissão, etc.).

A cobertura ampla sobre todos os aspectos da codificação multimídia tem permitido eleger o *MPEG-4* como o padrão potencialmente mais promissor para adoção futuramente nos sistemas audiovisuais imersivos profissionais e domésticos, onde seus recursos podem ter impactos na vida cotidiana, pela possibilidade de consumo de mídia de maneira personalizada.

3.4.2 Modelamento e simulação da cena acústica

Este tópico, juntamente com a codificação espacial, é particularmente o mais importante na produção da auralização, e será abordado em mais detalhes.

Em acústica virtual o modelamento de uma cena de som 3D é normalmente dividido em três partes (VÄÄNÄNEN, 2002), (SAVIOJA, 1999):

- modelamento das fontes (os objetos que emitem sons),
- modelamento do meio de transmissão (o ambiente acústico em si, e tudo o que ele encerra),
- modelamento do ouvinte ou receptor (a forma como o som é recebido, percebido, e tratado no destino).

Modelamento da fonte

O conceito de modelar a fonte em acústica virtual refere-se basicamente a atrelar uma fonte sonora (som) a um ambiente e parametrizar seus atributos como diretividade (padrão de irradiação), condição de produção anecóica (som “seco”, nível de imersão zero), tamanho (pontual ou segmentada), qualidade do sinal original (relação sinal/ruído, resolução em número de bits e frequência da amostragem), e origem (natural ou sintético) (SAVIOJA, 1999).

Modelamento do meio

O modelamento do meio refere-se ao modelamento dos espaços acústicos, da absorção (aérea e pelos objetos existentes no percurso) e da propagação sonora pelo meio e reverberação (SAVIOJA, 1999). Incluem-se ainda o modelamento de fenômenos de refração e difusão, dos fenômenos associados a modos ressonantes em baixa frequência, a variação das propriedades acústicas dos materiais em função da frequência do som, e eventualmente, nos modelos mais sofisticados, das variações de temperatura, umidade, e presença de ventos (PATRONIS, 2002), o que podem levar à representação de sistemas extremamente complexos. Felizmente há simplificações admissíveis e abordagens que, ainda que analíticas, produzem resultados numéricos e perceptuais bastante coerentes, e plenamente satisfatórios. A viabilidade de um método ou de uma associação de métodos vai depender da capacidade computacional de que se dispõe e do grau de precisão ou correção final que se deseja simular.

Modelamento físico versus perceptual

Do ponto de vista funcional e do domínio de aplicações, os métodos empregados em acústica virtual conceitualmente se baseiam numa abordagem via modelamento perceptual ou via modelamento físico (VÄÄNÄNEN, 2002).

No primeiro caso – modelamento perceptual – um conjunto de métricas acústicas e parâmetros qualitativos do ambiente são considerados para produzir uma representação auditiva consistente com o meio que está sendo modelado. Podem ser considerados parâmetros subjetivos e perceptuais do som (fontes sonoras) e do ambiente como “calor”, “proximidade”, “presença”, “envolvimento”, e mesmo

parâmetros derivados de medidas objetivas da qualidade percebida, como tempo de reverberação (*RT60*), frações laterais (*LF – lateral fraction*), correlação cruzada inter-auditiva (*IACC – inter aural cross correlation*), clareza (*C80*), tempo central (*CT – central time*), razão de baixos (*BR – bass ratio*), e muitos outros (AHNERT, 2002), que retratem propriedades do som direto e características macroscópicas da cena acústica. Esta abordagem usualmente abrange aplicações onde o controle da acústica virtual pode ser mais intuitivo, onde o usuário poderia eventualmente modificar as condições e situação da cena acústica na reprodução, e requer menor capacidade computacional.

Frequentemente esta abordagem é adotada em processadores de efeitos e máquinas reverberadoras comerciais, onde os efeitos são pré-programados, isto é, mapeados em *presets* como “concert hall”, “slight ambience”, “basement”, “empty garage”, “in the shower”, “warm echo” e dezenas de outros tipos, ou então são programados pelo usuário. Novos efeitos e ambientações sonoras são usualmente programados manipulando-se parâmetros de controle para equalização (seleção de filtros, banda de passagem, fator de qualidade, ressonadores, etc.), para efeitos de *pitch* (velocidade, desvio em semitons, profundidade do efeito, forma de onda, retro-alimentação, etc.), para cadeias de atraso (tempo de atraso, batidas por minuto, realimentação, número de taps, etc.) e para reverberação (níveis e limiares de ataque, sustentação, decaimento, liberação, densidade, difusão, mistura, compressão, reflexão, amortecimento, pré-atraso, etc) (ALESIS, 1995). A seleção de um conjunto de parâmetros controladores permite ao módulo aplicar o processamento de sinal adequado para que se produza a percepção do ambiente acústico desejado. Parâmetros perceptuais e métricas acústicas importantes na percepção e determinação de imersão sonora são abordados por VÄÄNÄNEN (2002) e FARIA (2005a).

No segundo caso – modelamento físico – o objetivo é produzir uma representação mais precisa dos processos acústicos reais que ocorrem no meio, e para isso levam em conta os parâmetros físicos do ambiente, como suas características macroscópicas (tempo de reverberação, volume, medidas, área de absorção, etc.) e microscópicas

(reflexões, coeficientes de absorção, absorção aérea, diretividade, difração, propriedades dos materiais, etc.) (VÄÄNÄNEN, 2002).

Para um grande número de aplicações em realidade virtual em tempo real (jogos, por exemplo) a abordagem física não é essencial para se reproduzir efeitos como oclusão, atenuação com a distância, efeito Doppler, reverberação e ecos, podendo-se valer do modelamento perceptual e métodos de implementações alternativos, eventualmente até propositadamente não correlacionados aos processos físicos reais. Já o modelo físico é fundamental para aplicações de simulações de espaços acústicos, salas de concerto, auditórios e todas aquelas onde o realismo e a precisão devam ser consistentes com os processos naturais que se deseja emular, como por exemplo, nas aplicações de projetos arquitetônicos, projetos industriais (para modelamento do comportamento acústico de cabines e conforto ambiental), etc.

LOKKI (2001) mostra em estudo realizado que, comparando-se sons artificialmente gerados por modelamento físico com sons realmente gravados em uma sala de aula, resultados muito semelhantes podem ser obtidos. Em uma avaliação subjetiva de auralização, os testes de audição mostraram que é possível criar ambientes virtuais auditivamente naturais com modelamento acústico físico e processamento digital de sinais.

O modelamento acústico do meio é tão complexo quanto o grau de realismo final que se deseja. Um fator importante para incrementar a qualidade percebida da auralização em acústica virtual é a modelagem da propagação e representação dos fenômenos acústicos numa extensa faixa de frequências, haja vista que a fenomenologia difere em cada faixa. O uso combinado de técnicas pode levar a uma boa modelagem geral da faixa audível, haja vista que usualmente uma técnica resolve bem a simulação numa faixa, mas não se mostra adequada ou modela satisfatoriamente outras faixas. Nos testes de (LOKKI, 2001), por exemplo, os resultados da auralização artificial foram promissores, mas os testes indicaram que,

para o conjunto de técnicas que ele utilizou⁵⁰, os maiores erros percebidos estavam nas baixas frequências (<400Hz) e nas altas frequências (acima de 8kHz).

LOKKI (2001) atribui a falta nas baixas frequências à ausência de modelamento da difração. Nas altas frequências, entretanto, não identifica uma razão clara, e indica como possível fonte de erro a equalização dos fones de ouvido utilizados em seu experimento.

Modelamento baseado em ondas versus baseado em raios

Do ponto de vista matemático e da implementação (construção de simuladores) existem duas principais escolas para se abordar e equacionar o problema do modelamento acústico: utilizando-se a teoria da **propagação de ondas** (soluções da equação de onda) ou a teoria da **propagação de raios** (modelagem do som por raios). SAVIOJA (1999) e CHÉENNE (2002) abordam os principais métodos computacionais para modelamento acústico, remetendo a outros autores para maiores detalhes quanto a cada técnica.

O som viaja num ambiente radialmente em todas as direções encontrando obstáculos e superfícies que alteram seu trajeto e multiplicam os percursos sonoros. Sons são refletidos por objetos muito mais largos que o comprimento da onda λ incidente (pelo menos 3 vezes) e usualmente passam despercebidos e até sem causar sombras pelo obstáculo que tenha dimensões reduzidas frente à λ . Para efeito de melhor compreensão dos fenômenos de propagação sonora podemos abordar os sons como ondas ou como raios dependendo principalmente da faixa de frequência em questão.

Abaixo de 300-400 Hz os sons são mais bem modelados como ondas. Sons acima de 300-400Hz são mais bem modelados quando considerados como raios (EVEREST, 2001), estando o modelo de sua propagação, portanto, sujeito às regras de espelhamento e leis de reflexão características desta abordagem. Dependendo da superfície refletora, a reflexão pode ser especular (obedecendo à lei de Snell) ou

⁵⁰ método da fonte-imagem para modelar o som direto e primeiras reflexões, e um algoritmo parametrizado para a reverberação posterior.

difusa (produzindo raios emergentes em várias direções como num espalhamento) (CAMILO, 2002). Ondas sonoras refletidas especularmente agem como se originadas de uma fonte imagem, situada atrás do plano da superfície refletora. Para computar a intensidade sonora total num ponto do espaço deve-se considerar a contribuição de todas as fontes sonoras reais e imagens.

Uso da resposta impulsiva

Em simulação acústica virtual a resposta ao impulso do ambiente/sistema (ou *IR* – *impulse response*) é o principal elemento de modelamento e parametrização dos processos de propagação e simulação acústica (VÄÄNÄNEN, 2002). A assinatura acústica do ambiente é registrada ou calculada artificialmente expondo-o a uma excitação pontual e energética, como por exemplo um tiro de pistola ou mesmo até um estouro de balão, e a resposta a este sinal constitui-se no principal produto do simulador acústico. Mesmo os métodos computacionais baseados na teoria das ondas têm em vista a geração de *IR*'s.

Do ponto de vista perceptual ou psicoacústico, a resposta ao impulso de um espaço acústico pode ser dividida em 3 janelas de tempo⁵¹, cada qual associada a um aspecto perceptual específico (VOGEL, 1993). A Figura 26 ilustra as janelas constituintes de uma *IR* ou reflectograma, mostrando também o padrão típico de reflexões num ambiente fechado, e as impressões espaciais associadas a cada janela. A primeira janela de tempo considera o som direto e os primeiros 20ms contados a partir de sua chegada ao ouvinte. A percepção da diretividade do som está principalmente associada à chegada do som direto e ao conteúdo sonoro (reflexões) desta faixa temporal.

⁵¹ Alguns autores preferem dividir a resposta ao impulso em 4 janelas, conforme seus objetivos. (BERKHOUT, 1993). Alguns autores consideram o som direto numa janela distinta, separada das primeiras reflexões (VÄÄNÄNEN, 2002).

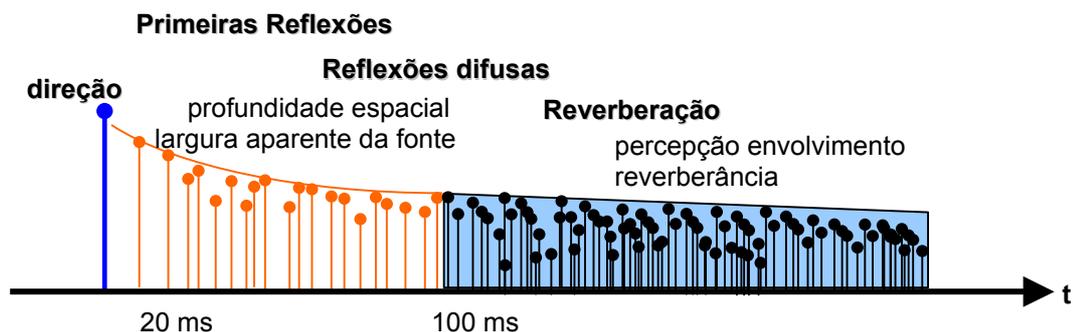


Figura 26 – Janelas de tempo típicas de uma IR (reflectograma) e impressões espaciais associadas

A segunda janela cobre a faixa de aproximadamente 20 a 100ms, onde se manifestam propriamente as primeiras reflexões⁵². As primeiras reflexões contribuem efetivamente para a percepção do espaço, isto é, estão usualmente associadas à sensação de “espacialidade”, à percepção da “presença” do som no espaço (BERANEK, 1963 apud VOGEL, 1993), permitindo ao ouvinte inferir sobre a profundidade espacial e também sobre a largura aparente das fontes sonoras. Com base no material sonoro nesta faixa de tempo somos capazes de perceber a dimensão acústica, e entender em que tipo de ambiente estamos.

A partir de aproximadamente 50ms múltiplas reflexões secundárias e difusões são percebidas. Alguns autores dividem as primeiras reflexões em “direcionais” e “difusas” (VÄÄNÄNEN, 2002). Reflexões chegando ao ouvinte a partir de 100ms contribuem mais para a percepção da reverberação do espaço, isto é, para a sensação de envolvimento. As múltiplas difusões sonoras que podem ocorrer numa sala acústica participam sobremaneira desta última faixa, informalmente também denominada a “cauda reverberante”. Alguns autores dividem esta faixa em reverberação inicial e posterior. Uma alta densidade de reflexões e também uma maior homogeneidade na distribuição energética entre elas é característica deste segmento.

⁵² principalmente em se tratando de ambientes com dimensões lineares superiores a 7m.

A resposta ao impulso pode ser medida de inúmeras formas em ambientes reais, dependendo da finalidade da medição. FARINA (2003), por exemplo, utiliza este procedimento para registrar o comportamento e a assinatura acústica de salas de concerto e casas de ópera, e armazenando-as para a posteridade. A resposta ao impulso é o principal recurso utilizado em máquinas simuladoras de reverberação e ecos para simular a acústica de ambientes e salas de concerto famosas, bem como de espaços cotidianos, como banheiros, espaços vazios, etc. através de uma simples convolução do sinal da fonte sonora com a resposta ao impulso.

A maior parte dos métodos de simulação acústica baseiam-se no conceito de produzir uma resposta ao impulso para um dado ambiente simulado e então convoluí-la com o sinal (anecóico) da fonte sonora que se deseja localizar e reverberar neste ambiente. A resposta ao impulso virtual de um ambiente artificial criado em computador pode ser calculada de diversas maneiras, utilizando-se abordagens via teoria das ondas, reflexões especulares, métodos numéricos e/ou estatísticos, ou abordagens híbridas.

Formalmente, em ambientes reais, a resposta impulsiva é dependente da frequência, isto é, a propagação em uma faixa de frequências produz uma resposta impulsiva distinta de outras faixas de frequências. Estes efeitos são devidos aos diferentes coeficientes de atenuação e absorção que os materiais apresentam em cada faixa de frequências. A construção de *IR's* cobrindo uma faixa larga de frequências envolve um custo computacional maior, e em muitos casos simplificações no modelo gerador ou relaxamento no atendimento a este requisito são perfeitamente admissíveis e produzem resultados satisfatórios. Uma maneira de se calcular *IR's* artificiais levando-se em consideração o comportamento previsto em várias faixas de frequência é abordado por EVEREST (2001).

Metodologias de modelamento acústico

A Figura 27 ilustra as principais metodologias utilizadas em modelamento acústico. Os principais métodos computacionais utilizados em acústica virtual e de qual escola ou abordagem de modelamento são derivados são mostrados a seguir.

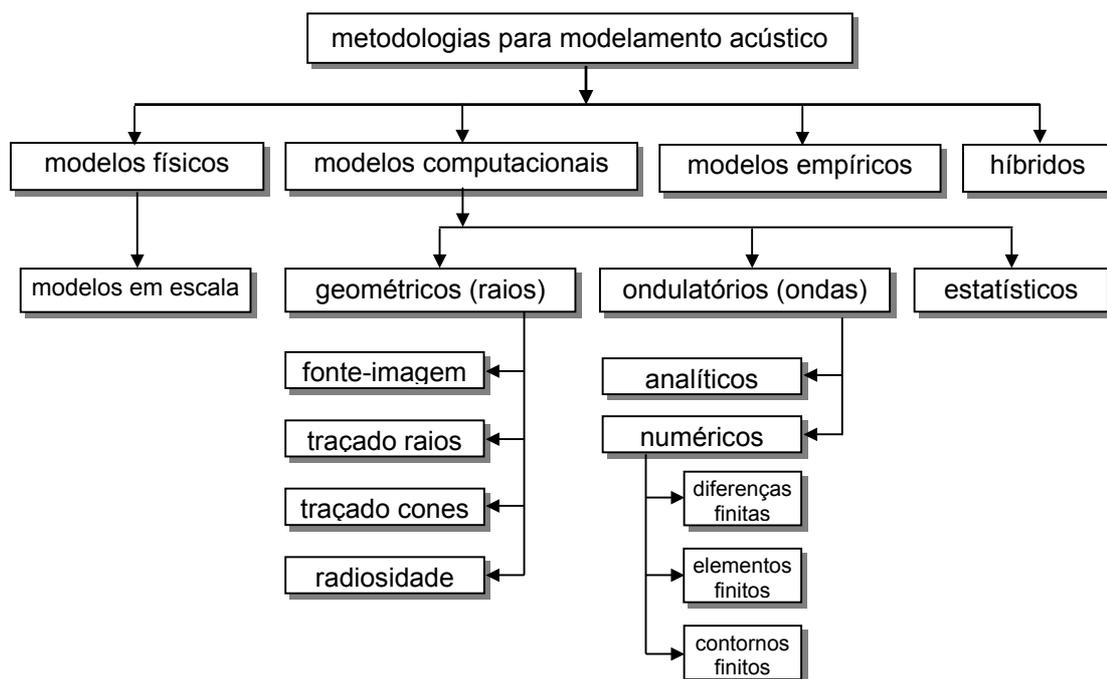


Figura 27 – Principais metodologias para modelamento acústico

Métodos analíticos e numéricos (baseado em ondas)

Os resultados mais precisos podem ser obtidos com técnicas baseadas na solução da propagação de ondas. Soluções analíticas, entretanto, podem ser achadas em casos raros, de geometrias simplificadas, como, por exemplo, no caso de salas retangulares com paredes rígidas (SAVIOJA, 1999).

Ainda no caso da abordagem via teoria das ondas, métodos de elementos de contorno, de elementos finitos e de diferenças finitas são utilizados. Citamos como exemplos as técnicas *FEM* (*finite-element method*), *BEM* (*boundary-element method*) e *FDTD* (*finite difference time domain*). Apesar de fornecerem resultados mais corretos matematicamente e fisicamente, por se basearem formalmente em princípios e leis básicas de propagação acústica (soluções da equação de onda), estes métodos possuem grande complexidade para implementação e implicam em custo computacional elevado, especialmente quando há requisitos de operação em tempo real envolvidos e o grau de precisão requerido (proporcional à resolução espacial

desejada) é alto, o que impraticava sua utilização na maioria das aplicações de tempo real⁵³.

Métodos geométricos (baseado em raios)

Os métodos de traçado de raios (*ray-tracing*) e de fonte-imagem (*image-source*) são os mais populares na abordagem via teoria de propagação especular. Eles modelam a propagação dos sons como se fossem raios, levando-se em consideração principalmente os fenômenos de oclusão e reflexão, e assim permitindo mapear os percursos dos “raios sonoros” desde a sua emissão até o ouvinte.

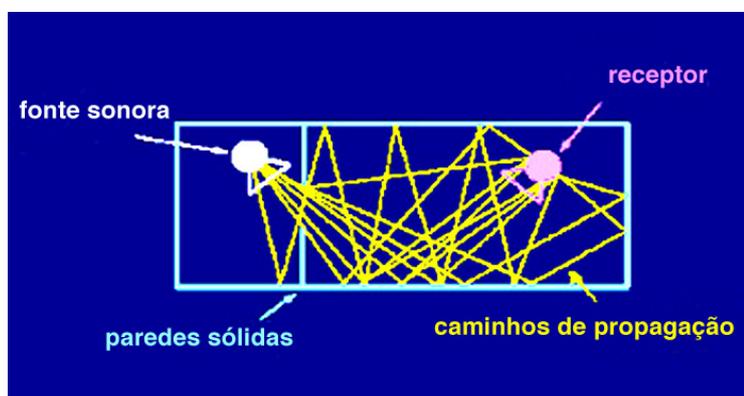


Figura 28 – Simulação de propagação das ondas sonoras por meio de raios de som (modelamento de reflexão especulares)

Os métodos de traçados de raios são bem conhecidos pela comunidade de computação gráfica, havendo diversas implementações para aplicações diversas nos anais do SIGGRAPH⁵⁴. Neste método, a energia total emitida pela fonte é distribuída de acordo com seu padrão de irradiação em certo número especificado de direções. Dependendo do tipo de superfície, o raio pode ser especular ou difusamente refletido. No primeiro caso o ângulo de incidência é igual ao ângulo de reflexão. No segundo, a direção do raio refletido é randômica. Estas características permitem o modelamento do padrão de irradiação da fonte sonora e da reflexão difusa. A energia

⁵³ o emprego de circuitaria dedicada como DSP's e aceleradores de hardware (como em placas de som para jogos) pode viabilizar o uso de métodos complexos em tempo real para muitas aplicações.

⁵⁴ SIGGRAPH – Special Interest Group on Graphics (<http://www.siggraph.org/>)

do raio é atenuada em cada reflexão devido à absorção e também pela propagação esférica. O ouvinte também deve ser modelado neste método, de forma a checar (por colisão) quais raios refletidos atingem uma célula (superfície ou volume) que o contém.

O grau de eficácia do método utilizado vai depender de quantos fenômenos acústicos ele é capaz de resolver, em particular de sua capacidade em achar todos os percursos possíveis, modelar absorção, transmissão, reflexão, difração e difusão, e também do modelo geométrico e tamanho do ouvinte considerado⁵⁵. Explicitamente não há garantias de que todos os raios sonoros que deveriam atingir o ouvinte sejam descobertos pela técnica. Usualmente sua eficiência é uma função do tempo de computação de que se dispõe e da resolução espacial (EVEREST, 2001). De uma maneira geral o método possui vantagens computacionais em comparação ao método de fonte-imagem, sendo preferido para modelar reflexões posteriores, enquanto o primeiro, embora tenha associada uma carga computacional maior, é mais adequado para se descobrir e calcular todas as primeiras reflexões, cuja importância na determinação da diretividade do som é fundamental.

No método de fonte-imagem, a fonte é refletida contra todas as superfícies para se achar todas as reflexões possíveis até o ouvinte (ALLEN, 1979); (BORISH, 1984). A Figura 29 ilustra o conceito do método, onde um reflexo especular é modelado como se proveniente de uma fonte (virtual) atrás da parede, que é uma imagem da fonte real. Reflexões de 2^a. ordem e de ordens superiores podem ser consideradas para enriquecer a resposta. A idéia é refletir todas as imagens reais e virtuais através de todos os contornos e superfícies e obter um conjunto de imagens especulares.

⁵⁵ a extensão da geometria modelada ao redor de um ouvinte pontual determinará quantos raios atingem o ouvinte e quantos, ainda que passem próximos, não o atingem. Uma margem de erro naturalmente associada a este modelo impõe limitações na capacidade do método em resolver todos os raios passíveis de audição.

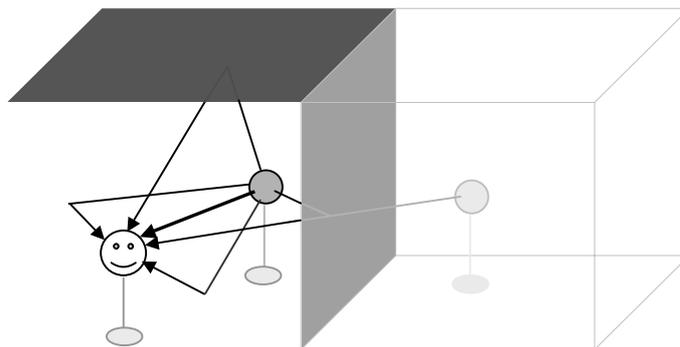


Figura 29 – Método da fonte-imagem para modelamento acústico discreto e especular

Após calcular todas as imagens (fontes virtuais, como indicado na Figura 29) uma *checagem de visibilidade* é realizada, para avaliar quais raios chegam ao ouvinte. O método é particularmente útil para se modelar as primeiras reflexões no ambiente. Para ambientes de paredes e geometria retangulares todas as fontes especulares obtidas são visíveis em cada posição do ambiente⁵⁶, e a computação pode ser rápida e eficazmente realizada (EVEREST, 2001), (ALLEN, 1979). Segundo SAVIOJA (1999) uma mudança de posição do ouvinte na cena neste método não muda o cenário de reflexões calculado: para uma nova posição do ouvinte somente uma nova checagem de visibilidade é necessária⁵⁷. BORISH (1984) estende o método de ALLEN (1979) a geometrias arbitrárias. Uma desvantagem do método estendido é o aumento de complexidade computacional para se calcular reflexões posteriores, uma vez que o número cresce exponencialmente (BORISH, 1984), o que o torna pouco eficaz para modelar a cauda reverberante do ecograma.

O método de traçado de raios tem grande semelhança. Cada fonte no algoritmo básico emite raios de som em todas as direções (padrão esférico de irradiação) que se refletem em todas as superfícies. Os raios que chegam ao ouvinte são contabilizados. Para tanto, o ouvinte (receptor) deve ser modelado como um objeto volumétrico interceptável pelos raios, usualmente uma esfera.

⁵⁶ exceção para os ambientes onde há objetos causando obstruções no caminho do som.

⁵⁷ esta situação, todavia, se aplica à sua abordagem/implementação.

Implementações mistas aproveitam vantagens de ambos os métodos: adequação do primeiro ao cálculo das reflexões primárias, e do segundo para as reflexões sucessivas, por apresentar melhor desempenho evitando-se explosões exponenciais de cálculos destas reflexões via método de imagens (SAVIOJA, 1999).

FUNKHOUSER (1999) explora o método de traçado de feixes em um sistema de auralização em quatro fases, sendo que as duas primeiras são processadas previamente (tempo diferido) e as duas últimas em tempo real. A mais importante contribuição é a idéia de pré-computar as estruturas de dados espaciais que mapeiam todos os possíveis caminhos de transmissões e reflexões especulares, e então utilizá-las para computar o caminho final da reverberação para aplicações interativas em tempo real, onde o ouvinte movimenta-se pelo ambiente.

O método de radiosidade, também utilizado em computação gráfica, é também conhecido por método de caminhos aleatórios, ou método de transição de energia. Particularmente útil na modelagem acústica em baixas frequências, este método baseia-se na troca de pacotes de energia entre a fonte, as superfícies no meio e o receptor que ocorre em intervalos de tempo iguais e regulares (tempo de transição), usualmente uma medida característica do ambiente. A fonte emissora (origem da energia acústica) é modelada segundo sua direcionalidade, e distribui sua energia pelas superfícies do ambiente, que são subdivididas em elementos triangulares. Cada elemento troca um número de pacotes de energia com os outros elementos, que vai depender da inclinação relativa e ângulo sólido entre os mesmos. O último passo consiste na modelagem da recepção dos pacotes (CAMILO, 2002).

EVEREST (2001) cita que na década de 1980 métodos híbridos utilizando mais de uma técnica foram introduzidos, por exemplo usando-se traçado de raios para se achar caminhos de reflexões potencialmente válidos, para se reduzir a quantidade de testes de validação no método de fonte-imagem. Alternativamente métodos utilizando-se feixes piramidais ou cones (traçado de cones) permitem verificar se o ouvinte está dentro da projeção do feixe cônico originado a partir de uma fonte (real ou virtual), dispensando então o teste de validação.

Além destas abordagens, que remetem a soluções precisas, há a possibilidade de se lançar mão de algoritmos alternativos para a computação de ecos, atrasos e reverberações, com a vantagem de uma complexidade algorítmica muito inferior e maior facilidade de programação. O modelamento do campo sonoro reverberante é frequentemente considerado difuso, e a resposta ao impulso correspondente neste trecho comporta-se como ruído aleatório exponencialmente decrescente (SAVIOJA, 1999).

Cadeias de atraso e circuitos de realimentação constituem os blocos construtores básicos de produtores artificiais destes efeitos, que para várias circunstâncias e ambientações podem produzir resultados muito convincentes. Filtros do tipo *comb*, *comb* passa-tudo, filtros passa-baixa e redes de atraso por retro-alimentação (combinações em série e em paralelo) são ferramentas comuns para implementações de reverberadores. Algumas características são desejadas neste tipo de circuitos:

- alta densidade de reflexões (como ocorre em ambientes reais), evitando-se flutuação no efeito reverberante,
- reprodução de uma alta densidade modal, para simular as características de resposta de frequência de ambientes reais, especialmente nas baixas frequências, mas sem valorizar um modo em detrimento de outros (para evitar coloração),
- tempo de reverberação dependente da frequência, simulando-se (com filtros passa-baixas) o efeito da absorção do ar e também por materiais absorventes,
- resposta ao impulso decrescente exponencialmente com o tempo
- produção de sinais incoerentes ou descorrelacionados, de forma a simular um campo difuso, isto é, uma situação aonde as reflexões chegam de diferentes direções com a mesma probabilidade.

Exemplos de implementação de um reverberador são mostrados em SAVIOJA (1999) e por SCHROEDER (1962).

Pela maior simplicidade da implementação frente aos métodos analíticos, estas técnicas alternativas são comumente empregadas em equipamentos processadores de efeitos, mesas de ecos, pedais, entre outros. Tais sistemas frequentemente possuem

interfaces de programação que permitem ao usuário intuitivamente selecionar a intensidade e características temporais do efeito em função de um mapeamento indireto dos tipos possíveis de ambiente a emular, como cânions, auditórios pequenos/médios/grandes, salas de aula, salas de concerto, salas de espetáculos de grande porte, estádios, etc.

No escopo do projeto AUDIENCE pretende-se investigar o uso de processadores deste tipo para produzir as reflexões posteriores e a reverberação do ambiente. Diversas mesas profissionais digitais possuem recursos para programar efeitos reverberantes em função do tamanho da sala, e para isso utilizam-se de padrões de retro-alimentação de sinal pré-definidos.

Modelamento do ouvinte ou receptor

A posição, o tamanho (dimensões) e a direção em relação ao ambiente são parâmetros fundamentais para o modelamento do receptor ou ouvinte. Para fins práticos o receptor é modelado como um objeto geométrico sobre o qual incidem raios sonoros (vide modelamento do ouvinte no método de traçado de raios) e a dimensão relativa desta geometria influencia na eficácia do método em resolver o que é ouvido naquela posição.

O modelamento do ouvinte considera principalmente o tipo de auralização que será efetuada. Para auralização bi-aural é importante considerar a recepção em dois pontos, produzindo-se saídas distintas para o ouvido esquerdo e para o direito. Posição e diretividade são determinantes para a seleção adequada dos parâmetros renderizadores.

Para propósitos de auralização multicanal, a região de audição produzida através de uma matriz de alto-falantes ampliada é envolvente, usualmente formando um volume dentro do qual situam-se um ou mais ouvintes fisicamente distribuídos. O modelamento do ouvinte no que diz respeito às características de diretividade de audição neste caso não se aplicam, pois os ouvintes devem poder dentro de uma

configuração multicanal mudar seu ponto de vista, a direção do plano mediano⁵⁸ e a elevação da cabeça, sem alterar as condições de auralização do volume.

A região de auralização ideal envolve o ouvinte, sendo que este define seu ponto de vista e plano de audição navegando dentro deste volume. Para o caso 3D todas as direções possíveis são permitidas ao ouvinte no volume de audição.

3.4.3 Síntese sonora e codificação espacial

As fontes sonoras podem ter seus sons acessíveis diretamente em arquivos de áudio ou sintetizados por algum sintetizador acoplado. O uso de sons armazenados em arquivos de áudio é mais adequado para experimentações simples. Os arquivos originais das fontes de som podem ser armazenados sem nenhuma informação direcional nem espacial (gravados em câmeras anecóicas, sem reflexões ou reverberações).

Após a aplicação dos métodos da camada 2 (simulação acústica) devemos proceder à codificação da informação espacial ao sinal sonoro que está associado à fonte simulada. Dois são os principais objetivos desta camada:

- a) adicionar informação espacial à informação temporal do som, e
- b) prover uma representação intermediária (um sinal codificado) adequado para transmissão à próxima camada.

Usualmente a convolução de um som anecóico com uma *IR* insere a informação espacial no som. Na maioria dos casos a codificação espacial está associada à aplicação de um filtro a um sinal anecóico, como por exemplo no caso da auralização bi-aural.

As respostas impulsivas bi-aurais (*HRTF's*) são exemplos de formatos de codificação sonora 3D intermediárias, requerendo somente uma convolução com os sinais originais anecóicos das fontes sonoras para produzir os sinais ouvidos na altura do ouvido direito ou esquerdo. Outro formato de representação intermediário é o *B-*

⁵⁸ Plano que atravessa o nariz dividindo a cabeça em lados esquerdo e direito

Format” do sistema Ambisonics, onde o campo sonoro 3D ouvido na posição do ouvinte num dado ambiente é codificado em 4 vetores ou canais: X, Y, Z, e W (para Ambisonics de 1^a. ordem) contendo simultaneamente os sons e sua informação espacial. A Tabela 4 mostra quantos canais são necessários para armazenar uma representação de *n-ésima* ordem. FARINA (1998) aborda o processo de construção de sinais *B-Format* a partir de cenas acústicas, bem como o processo de decodificação para N alto-falantes.

Esquemas de transmissão multicanal também consideram o envio dos sons combinados em canais (codificados conjuntamente) e as “receitas” para recuperação e recombinação para apresentação final⁵⁹.

Esta facilidade permite que possamos utilizar codecs diversos para representar espacialmente os sons em um dado ambiente virtual, possamos transmitir ou armazenar estas representações em forma codificada, utilizando-se de suas vantagens em ocupar um número padronizado de canais com características conhecidas, bem como decodificar e recuperar os sons com suas informações espaciais em decodificadores que sejam compatíveis e capazes de ler aqueles fluxos de dados multicanais, gerar os sinais que vão para cada caixa de som, segundo a necessidade ou demanda do usuário, ou segundo os requisitos e limitações do sistema de sonorização final.

A codificação espacial pode também ser efetuada através da combinação de canais por matriciação, como nos esquemas quadrifônicos e vários codificadores *surround*, como o *Dolby Surround*, ou pela produção de um número final de canais discretos num dos formatos multicanais da Tabela 2.

3.4.4 Decodificação espacial e sonorização da cena

A sonorização final da cena compreende a decodificação dos sons no formato de representação intermediária e a reprodução multicanal dos N sinais decodificados

⁵⁹ Exemplos de abordagens assim incluem a suite de ferramentas de composição e codificação *MPEG-4* e as tecnologias de codificação *surround* paramétrico para transmissão de rádio digital multicanal (*DAB*).

mapeados para N alto-falantes. O terminal responsável pela sonorização recebe ou recupera os sons codificados de algum meio de armazenamento ou transmissão (rede *IP*, *HD*, memória, *CD*, etc.). Dependendo do sistema/formato espacial em uso, pode ser possível e útil combinar previamente à decodificação os diversos sons (mixagem).

No caso de auralização com Ambisonics, o bloco sonorizador faz a decodificação dos 4 canais em *B-Format* para uma matriz de N falantes, dispostos preferencialmente numa configuração poligonal regular (FURSE, 1999).

Há a possibilidade de se utilizar um simples *mixer* digital para misturar sons de várias fontes codificadas em *B-Format* antes da decodificação. Os sons em *B-Format* também podem sofrer rotações e outras operações no espaço através de simples algorismos se isso for necessário (MALHAM, 1993).

O decodificador deve retornar sinal de áudio digital sem compressão, *PCM*-compatível ou em algum formato de alta resolução. A reprodução ou sonorização final é feita a partir de um *player* multicanal que recebe as amostras finais e envia cada canal separadamente para saídas independentes na placa de áudio multicanal. Em alguns casos, pode-se adicionar processamentos intermediários entre a decodificação e a reprodução multicanal, como por exemplo, a aplicação de filtros para uma finalidade específica, condicionando ou equalizando os canais antes de tocá-los.

No caso de redes de distribuição de áudio digital utilizando-se protocolos como *ADAT*, *AES/EBU* (AES, 1997), *S/PDIF*, *MADI*, etc. (POHLMANN, 1995) os sinais de áudio seguem digitais até os alto-falantes, onde então sofrem a conversão final para o domínio analógico.

4 UMA SOLUÇÃO PARA AURALIZAÇÃO EM RVI COM AMBISONICS

4.1 Introdução

O foco desta pesquisa está na investigação do uso de técnicas para simulação e geração de campos sonoros 2D/3D em ambientes audiovisuais imersivos, particularmente em *CAVE's* como ambiente mais avançado nesta categoria, partindo-se da hipótese de que estas técnicas são as mais adequadas para se criar ou recriar ambientações sonoras com níveis de realismo e graus de liberdade e imersão sonora compatíveis com o necessário para aplicações de realidade virtual imersiva completa.

Os sistemas baseados em estéreo, panoramas (SVPA) e os paradigmas *n.m* (5.1, 6.1, 7.1, 10.2) não fornecem o nível de imersão defendido para este tipo de aplicação (nível 5). Objetivamos a produção de um ambiente acústico virtual coerente e realista, com a correta percepção da localização dos objetos sonoros e do ambiente acústico ao redor, independente da configuração de alto-falantes em uso (número e posições).

É desejável que os usuários percebam os sons formando-se em posições que não correspondam à posição física dos alto-falantes, impedindo a associação destes como origem primária dos sons, e criando a ilusão de que objetos sonoros estão realmente distribuídos em pontos diversos da cena audiovisual. Os alto-falantes serão assim meras fontes secundárias produzindo a frente de onda acústica gerada por fontes (virtuais) primárias.

Um objetivo claro neste estudo de caso é investigar se há aumento do realismo percebido e quebra da relação direta alto-falante/origem sonora com o uso de uma técnica de auralização reconhecida no meio científico: o sistema Ambisonics.

Experimentos prévios com campos sonoros em *CAVE's* não utilizaram a técnica Ambisonics (ECKEL, 1998), (OGI, 2003). Queremos investigar se é possível alcançar um nível de imersão 5 (veja **Tabela 6**) na CAVERNA Digital com esta técnica de auralização.

Apresentamos a seguir um experimento piloto realizado para implementação de um auralizador Ambisonics na CAVERNA Digital. Inicialmente apresentamos conceitualmente o projeto do módulo auralizador integrado aos demais subsistemas de RVI, para situá-lo no contexto global. Para tanto, o diagrama de blocos geral de um sistema de produção sonora desejado é apresentado, identificando-se os blocos principais que serão alvos deste experimento, quais sejam os módulos orientados à camada de simulação acústica e codificação espacial, pela geração de saídas impulsivas (*IR*) codificadas em *B-Format*.

A forma como cada camada da arquitetura de produção sonora espacial foi projetada e realizada é apresentada nos itens seguintes, destacando-se contribuições, detalhes de implementação, e limitações. Uma descrição ampla da infra-estrutura geral projetada e implantada para sonorização multicanal é apresentada. As operações de codificação final do sinal sonoro em *B-Format* e a decodificação Ambisonics para 8 caixas acústicas são realizadas utilizando-se *softwares* suplementares.

4.2 Um Auralizador Ambisonics em camadas

Nosso estudo de caso tem como metas principais: a) investigar a implementação de um auralizador espacial, utilizando os conceitos da arquitetura modular proposta no item 3.4, b) prototipar um simulador acústico e codificador Ambisonics, para auralizar cenas audiovisuais virtuais descritas em *X3D*, c) e finalmente investigar a eficiência e validade da auralização na localização e percepção espacial sonora em um ambiente audiovisual imersivo como a CAVERNA Digital.

O experimento foi executado em quatro fases:

- montagem da infra-estrutura de áudio e sistema de sonorização multicanal
- projeto da integração do auralizador à camada de descrição de cena e de decodificação e sonorização multicanal
- prototipação do auralizador (simulador acústico e codificador Ambisonics)
- testes de auralização multicanal com Ambisonics.

4.2.1 Diagrama de Blocos Geral do Sistema: Projeto

A execução e controle da aplicação de auralização é projetada para ser realizada em três fases, a saber, uma de *inicialização* do sistema, outra, em *regime permanente* onde a auralização está em curso e admite-se o controle da aplicação (navegação pelo mundo virtual e comandos do usuário) e dos programas associados, e outra de *finalização*, quando encerra-se a auralização. O controle e a gerência da navegação audiovisual é realizado pelo navegador de mundos virtuais, no caso o navegador *X3D Jinx* (SOARES, 2004a). A auralização em si é um processo concorrente ao de navegação e projeção visual, que pode ser validado inclusive de maneira independente. A demonstração da funcionalidade de gerência da auralização e sua integração com o mecanismo de gerência visual e demais subsistemas da CAVERNA Digital não é, entretanto, objeto do nosso primeiro experimento.

Conjuntos de tarefas ou processos contidos num mesmo subsistema podem ser agrupados como sugere a hierarquia de tarefas para a produção e apresentação da cena acústica na Figura 22. A Figura 30 mostra um resumo das tecnologias aplicáveis a cada camada da arquitetura introduzida no capítulo 3.

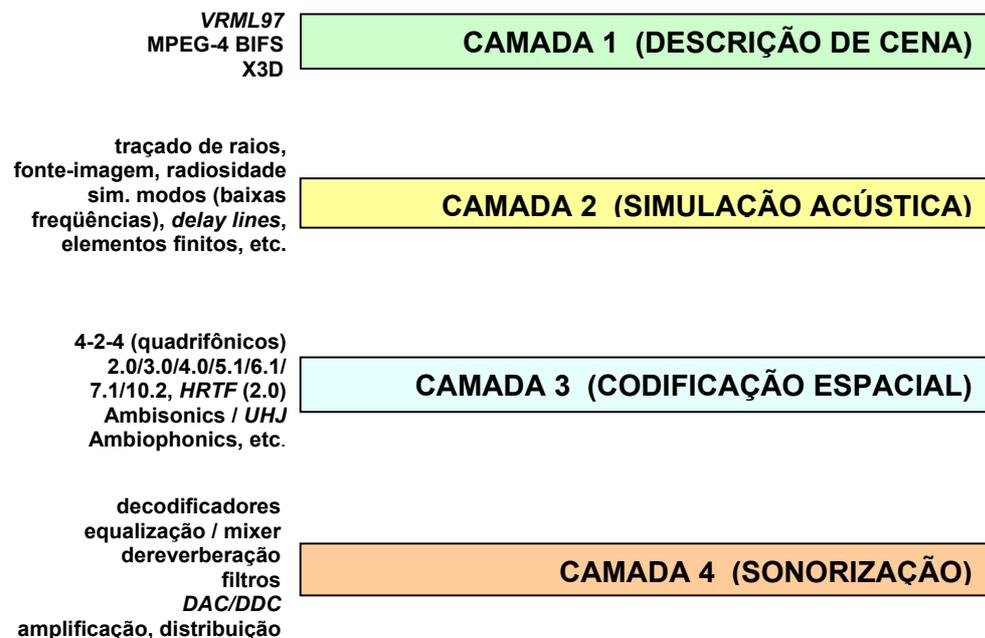


Figura 30 – Resumo de tecnologias disponíveis por camada da arquitetura

O diagrama de blocos geral do sistema de produção sonora espacial agrupando as tarefas do aplicativo de RVI com o auralizador é apresentado na Figura 31.

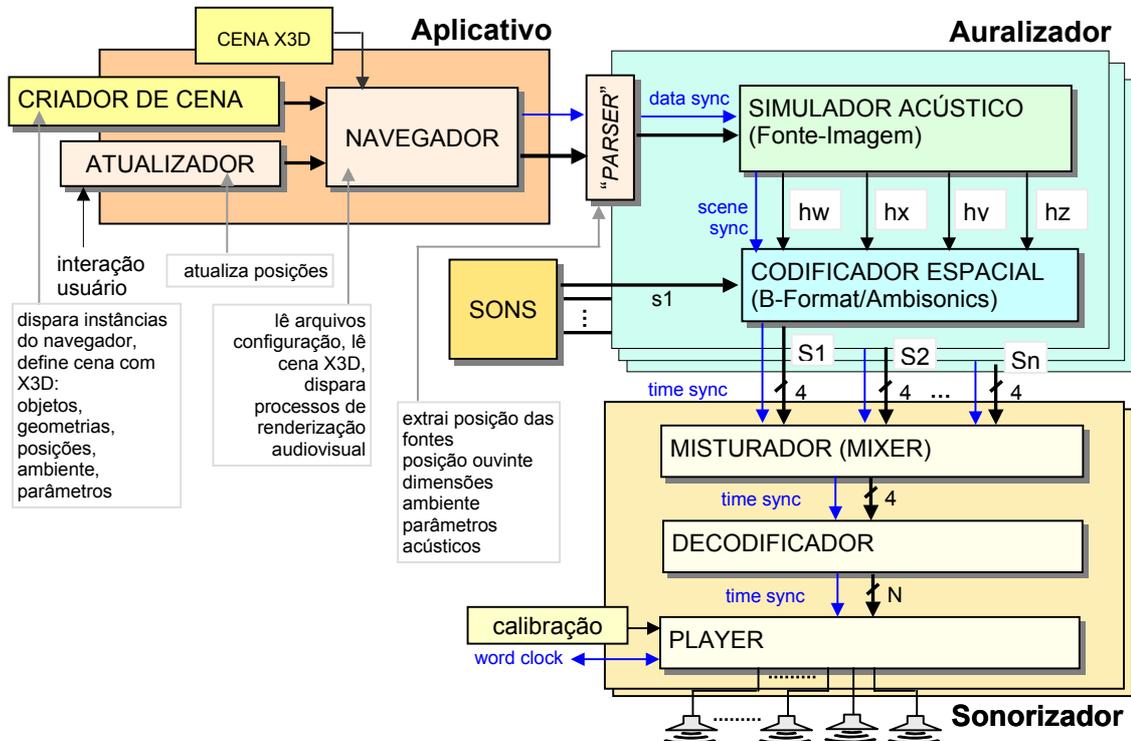


Figura 31 – Diagrama de blocos geral de sistema para produção sonora espacial

O aplicativo de RV mostrado na Figura 31 é uma instância do navegador Jinx que invoca um arquivo de configuração inicial, lê a cena (visual e acústica) descrita em *X3D* e inicia a distribuição de tarefas através dos nós do aglomerado computacional para produzir o ambiente virtual. Em síntese, uma primitiva do MPI inicia uma instância do aplicativo em cada nó. Os arquivos de configuração determinam quais funções serão alocadas para este nó e como será o cenário de comunicação deste com os demais nós/processos distribuídos. Alguns nós podem então ser alocados para tarefas de processamento gráfico, outros, para tarefas de processamento sonoro.

Projetamos a integração do navegador Jinx ao sistema de auralização e sonorização, concebendo uma seqüência de comandos que se inicia da forma acima, dispara o módulo auralizador e sonorizador, carrega os dados da cena acústica, e estabelece uma comunicação de sinais de sincronismo. A extração de parâmetros acústicos da

cena pode ser realizada automaticamente pelo Jinx e passado ao aplicativo auralizador.

O navegador é controlado pelo usuário. Atualizações de posição do usuário no mundo virtual são repassadas na cena armazenada de trabalho, que também atualiza os dados da cena acústica para o auralizador. No nosso primeiro experimento, entretanto, os objetos da cena não têm movimento. A investigação da qualidade e precisão da projeção do campo sonoro é o principal interesse nesta fase, deixando a integração completa do mundo visual e interação do usuário para fases posteriores.

Na Figura 31 o módulo auralizador (em verde) realiza duas tarefas principais:

- a) a simulação da propagação acústica (empregando o modelo acústico proposto) calculando o \mathbf{IR}_j (resposta ao impulso artificial calculada para o objeto j) e \mathbf{B}_{ji} (matriz de direção de cada raio sonoro i gerado pela propagação do objeto j em *B-Format*), e
- b) a codificação do som do objeto j em *B-Format*, (módulo de codificação 3D para o objeto j) gerando-se (para Ambisonics de 1ª ordem) 4 canais de som por objeto.

Uma proposta para distribuição de carga computacional é invocar uma instância de auralizador para cada objeto O_j (fonte sonora j) sendo calculada na cena, e alocar um nó do aglomerado responsável pela cadeia de auralização de uma fonte sonora presente na cena. A Figura 31 ilustra esta situação, considerando um nó com várias instâncias de auralização (vide blocos auralizadores em paralelo). Um ou mais objetos (fontes sonoras) seriam auralizados no nó.

No nosso experimento concentramos a auralização em um nó. Entretanto, concebemos em projeto a possibilidade de explorar as melhores condições de distribuição de carga, inclusive realizando a auralização (simulação mais codificação espacial) em um nó, e enviando o som codificado para outro nó – o nó sonorizador – onde sons provenientes de outros nós seriam misturados, decodificados e sonorizados. Também há a possibilidade de se utilizar dois nós sonorizadores,

endereçando conjuntos de fontes diferentes para cada um, mas mantendo-se o sincronismo entre as placas de som (com *word-clock* central).

O codificador Ambisonics virtual é a implementação central deste trabalho, e consta da integração de um simulador acústico baseado em fonte-imagem (ALLEN, 1979) e implementação inédita de um codificador de *B-Format* para fontes virtuais.

4.2.2 Ambiente acústico virtual de testes

Construímos uma cena audiovisual simples descrita em *X3D*, contendo uma sala de geometria regular e simples (retangular) com uma fonte sonora (objeto sonoro), e um ouvinte localizado no centro de um cubo, representando a CAVERNA Digital e sua respectiva área de audição, como indicado na Figura 32.

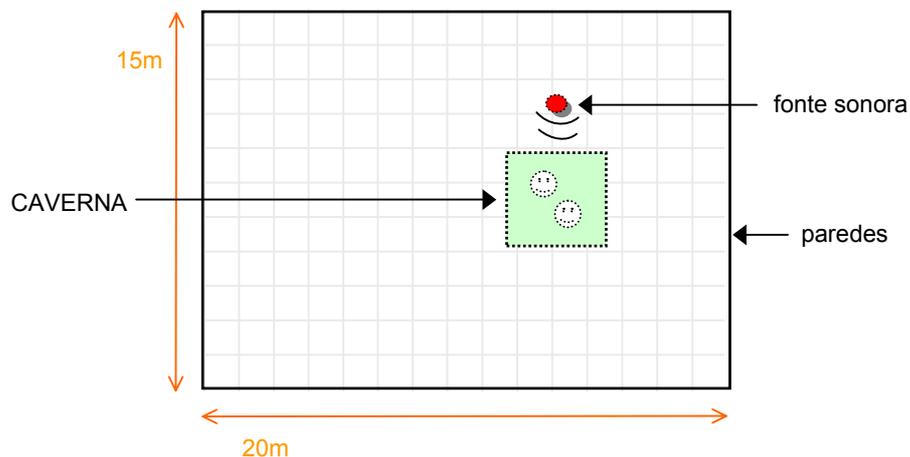


Figura 32 – Sala acústica virtual para testes de auralização (planta baixa)

O protótipo do auralizador Ambisonics consiste no principal algoritmo implementado neste experimento prático. A camada de simulação acústica e de codificação espacial estão incorporadas nesta implementação piloto. Em sua entrada o módulo auralizador admite posições das fontes sonoras e ouvinte e os parâmetros acústicos da cena de testes ilustrada na Figura 32.

4.2.3 Mapeamento dos objetos na cena acústica

Todos os objetos na cena audiovisual são mapeados utilizando-se *X3D*. Atualmente o padrão *X3D* cobre uma descrição dos atributos visuais da cena e possui alguns recursos sintáticos para descrição de objetos e processos de sonorização simples. Os

objetos sonoros têm descritos além de seus atributos geométricos e de localização os seus atributos acústicos.

Atributos acústicos da cena

A descrição da cena deve levar em consideração a individualidade acústica dos objetos pertencentes à cena que sejam atuantes no processamento acústico. Do ponto de vista de construção sintática da cena, as geometrias principais formadoras do ambiente, como paredes, teto, piso, detalhes arquitetônicos (como balcões, colunas, anteparos, fosso de orquestra, etc.) devem, pelo menos do ponto de vista acústico, ter suas características físicas e acústicas individualizadas, para uma extração de atributos hierarquizada e completa.

Os objetos passivos, como anteparos, obstáculos, paredes, teto e chão possuem os seguintes atributos acústicos minimamente definidos:

- coeficiente de transmissão
- coeficiente de reflexão β
- coeficiente de absorção α
- padrão de difusão (caso um objeto difusor)

A geometria dos objetos também deve ser informada – como dimensões e posição espacial na cena (frequentemente já embutidos na descrição visual, mas eventualmente sem a individualidade necessária para fins acústicos). Os objetos produtores de som – as fontes sonoras – devem ter os seguintes parâmetros definidos:

- padrão de irradiação (diretividade)
- fonte de som: arquivo de som, endereço do *stream* (*URL*), entrada da placa de som (canal de entrada)
- coeficientes de transmissão, reflexão e absorção

Formalmente, visando à completude da descrição, a cena ou aplicação deveria também admitir atributos gerais do ambiente, para efeito de posterior processamento por parte do sistema de auralização que vai produzir a cena (principalmente os módulos simuladores acústicos e sonorizadores) tais como:

- nível de imersão desejado/ideal e nível de imersão mínimo requerido
- sugestão de módulos algorítmicos ou técnicas preferenciais (se disponíveis) para realizar a simulação acústica, a codificação sonora, a decodificação e a reprodução sonora final (e.g. *players*).

Dadas as ausências de campos no padrão *X3D* para descrição de parâmetros acústicos, propomos extensões sintáticas para os nós descritores de som do *X3D*, para encampar alguns parâmetros fundamentais mínimos para a descrição de uma cena acústica, para descrever atributos acústicos dos objetos, e eventualmente endereçar informações pertinentes aos módulos de simulação acústica, codificação, decodificação e renderização final.

Para os propósitos de nosso experimento, adicionamos os seguintes nós e parâmetros à sintaxe tradicional do *X3D*:

- `<AcousticScene>` : nó para acomodar descritores de cena (como um todo)
- `<AcousticMaterial>`: nó para acomodar descritores de parâmetros acústicos de material
- `coefref`: coeficiente de reflexão de material
- `freqref`: faixa de frequência que se aplica.

Estamos realizando testes para saber em quais lugares estes dados poderiam melhor ser colocados, e que estruturas seriam mais adequadas como proposta de extensão sintática ao *X3D*. Isso é relevante em função da forma como objetos complexos da cena podem ser descritos pelo programador, de forma que os dados acústicos sejam atribuídos a uma superfície ou objeto na concepção humana de objeto, e não na concepção programática de uma associação de polígonos e formas. Quando ocorrer assim, acreditamos que uma forma adequada de posicionar os parâmetros acústicos seja sob um nó pai do tipo `<AcousticScene>` que é montado pelo programador especificamente pensando no extrator de cena acústica que vai ler o código. Quando a descrição do objeto aproxima-se de uma primitiva simples, como um cubo, acreditamos que a melhor forma de posicionar os dados acústicos é direto sob um nó

do tipo `<AcousticMaterial>` filho desta primitiva. A Tabela 8 mostra um exemplo dos dois casos, em ordem (o primeiro em cima, o segundo, embaixo).

Tabela 8 – Exemplos de sintaxe descritora de propriedades acústicas para objetos no X3D

```

<X3D>
  <Scene>

    <AcousticScene DEF="auditorio" center="0 0 0" size="30 15 8"
      coefref=".8 .6 .6 .6 .6 .7"
      freqref="1000 1000 1000 1000 1000 1000"/>

    <Transform DEF="pis0" center="7.5 0.05 10">
      <Shape>
        <Box size="15 0.1 20"/>
        <Appearance>
          <Material diffuseColor="1 1 1"/>
        </Appearance>
      </Shape>
      <Sound>
        <AcousticMaterial coefref=".8" freqref="1000"/>
      </Sound>
    </Transform>
  
```

O Jinx fará o *parsing* da cena acústica. Um componente de *software* objeto do *PD* faz a interface com o Jinx, recebendo os dados acústicos e geométricos da cena e seus objetos. Este módulo passará os dados necessários para a simulação acústica da camada seguinte.

O módulo descritor de cena (em *X3D*) está embutido no navegador Jinx. Sua tarefa é descrever a cada momento a geometria, posição e propriedades acústicas dos objetos na cena. Sua meta é manter em memória cada objeto O_j , a posição de O_j (x,y,z), seu *status* (se fonte sonora ou objeto passivo), ponteiro para o arquivo de som (se fonte sonora), e dados acústicos básicos (coeficiente reflexão, etc.).

A figura mostra uma renderização 3D da cena de testes introduzida em 4.2.2 (cuja planta baixa é mostrada na Figura 32).

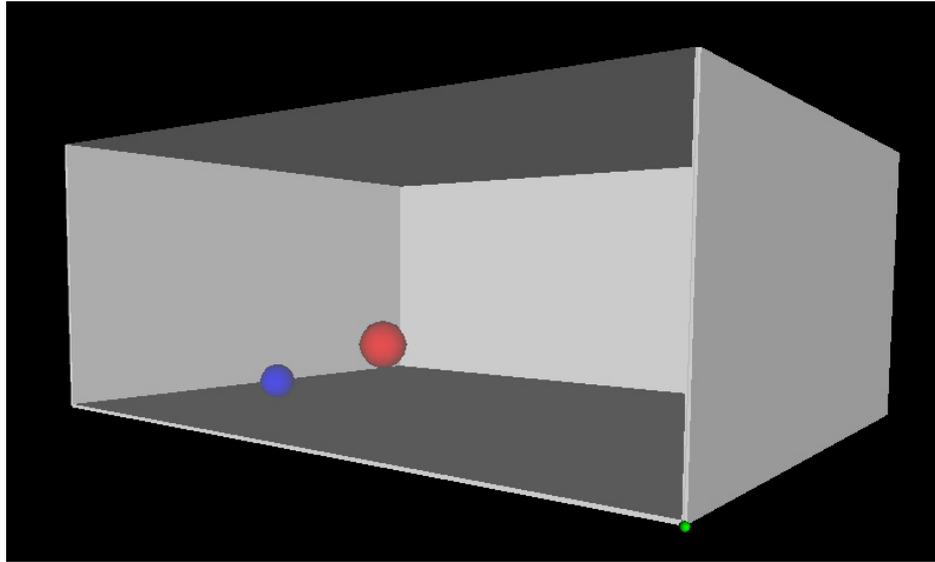


Figura 33 – Cena testes em 3D

Um pequeno auditório com estas proporções (15x20x8m) está dentro da “área de Bolt”, um critério para salas retangulares exibirem uma distribuição modal favorável (EVEREST, 2001, p.276). A origem das coordenadas está indicada por um ponto verde no canto mais próximo do leitor à frente. O eixo que corre da origem horizontalmente para a esquerda da figura é o eixo “y”, o “x”, correndo pela direita em direção ao fundo. O ouvinte está localizado no centro ao fundo da sala. A fonte sonora está à esquerda do ouvinte. Ela modela um instrumentista sentado.

O módulo responsável pela descrição da cena acústica executa o seguinte algoritmo básico:

- para cada alteração na posição relativa entre fontes sonoras e ouvinte:
 - atualizar a descrição da cena acústica para cada nó rodando um auralizador

4.2.4 Simulação acústica

O som direto vindo do objeto é o principal indicativo da direção e posição do objeto na cena, sendo o componente direcional mais importante usado pelo sistema perceptual humano para inferir a direção do objeto, cabendo às reflexões e à cauda reverberante complementar a descrição acústica da cena, perceptualmente informando ao ouvinte os atributos do ambiente onde se encontra.

Pretendemos mostrar a validade de se abordar a produção sonora através de implementações de funcionalidades por camadas, segundo os conceitos da proposta de arquitetura modular. Nossa meta é implementar um simulador acústico simples, mas eficiente na camada de simulação acústica, que receba os atributos da cena e de seus objetos, calcule a propagação acústica devida a fonte irradiando neste ambiente para um ouvinte, e retorne os parâmetros que contenham informação espacial desta propagação (as *IR's*) para serem usados na codificação espacial do som.

Vamos usar um modelo simples nesta implementação, para fins de validação somente, abrindo então espaço para que futuramente modelos mais sofisticados sejam implementados, integrados ou incorporados nesta camada.

Para o presente experimento implementamos o método de fonte-imagem (*image-source*) proposto em (ALLEN, 1979). A sala proposta acima (cena de testes) está de acordo com a geometria coberta por esta técnica. A simulação acústica embutida no auralizador fornece a localização dos sons⁶⁰ para o codificador de *B-Format* de 1ª ordem, em forma de 4 vetores resposta ao impulso (h_w , h_x , h_y e h_z), um para cada canal *B-Format*.

Os métodos mais completos de simulação acústicas disponíveis⁶¹ não são adequados para uma integração com o auralizador proposto em um primeiro experimento, devido à complexidade de suas implementações, resultante dos elevados requisitos de qualidade de simulação exigidos pela aplicação principal a que se destinam, a saber, a simulação acústica para fins de avaliação de propriedades acústicas e projeto de salas de audição.

O programa original em FORTRAN de (ALLEN, 1979) foi adaptado com as premissas acima.

⁶⁰ a diretividade de cada reflexão simulada será fornecida juntamente com a respectiva intensidade (amplitude) ponderada pela distância percorrida e atenuada pelas reflexões ocorridas nas superfícies absorvas (no caso do experimento em questão, somente as paredes).

⁶¹ como o RAIOS, desenvolvido na UFRJ, o Aurora, de Angelo Farina, entre outros.

A simulação acústica da cena audiovisual será promovida através do cálculo de sua resposta artificial ao impulso (*IR*) levando-se em consideração o som direto e um número finito de reflexões no percurso entre cada fonte sonora e o ouvinte, mais as respectivas características de diretividade associadas. Isto permitirá obter uma boa aproximação para os primeiros 100ms de resposta do espaço acústico, tempo onde as primeiras reflexões ocorrem.

A partir de 80-100ms verifica-se um adensamento de reflexões distribuídas, que produzem um campo difuso e reverberante. A partir desta faixa as reflexões não encerram informação direcional importante nem apresentam correlação relevante entre si. Assim, não precisam ser calculadas individualmente, podendo ser obtidas, por exemplo, através de algoritmos utilizando estruturas de filtros recursivas. Diversos algoritmos para reverberação existem propostos na literatura, não sendo nosso foco abordar esse tópico. Reverberadores comerciais podem ser empregados em conjunto com o auralizador Ambisonics proposto, especialmente para complementar o efeito de envolvimento a partir de 100ms, isto é, produzindo-se a reverberação (cauda reverberante).

Segundo experimentos perceptuais, uma excelente representação do campo reflexivo e reverberante pode ser alcançado considerando-se as 10 primeiras reflexões de uma simulação acústica (DEVRIES, 1999). Em nosso experimento, todavia, buscamos investigar a qualidade e o grau de percepção espacial alcançado considerando-se uma resposta ao impulso que englobasse o raio direto, as primeiras reflexões e o início da parte difusa e reverberante da mesma. Trabalhamos portanto com respostas impulsivas entre 80 e 200ms de comprimento efetivo.

Por este método identifica-se um número finito de raios de som que partindo da fonte sonora O_j percorreriam caminhos permitidos na sala acústica e chegariam ao ouvinte, ao mesmo tempo em que se calcula a atenuação e o tempo de atraso para cada raio de som entre a fonte e o ouvinte. A Figura 34 mostra as fontes-imagens geradas pelas reflexões especulares no espaço acústico.

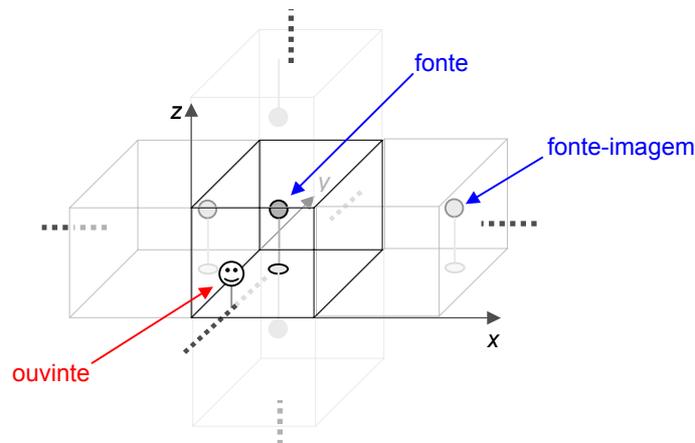


Figura 34 – Disposição das fontes imagens no espaço acústico

O modelo acústico do método de fonte-imagem identifica além do som direto todas as reflexões que chegam até o ouvinte num certo espaço de tempo. A resposta ao impulso onidirecional teórica prevista para a sala estimulada a partir da posição de O_j e ouvida na posição do ouvinte é obtida calculando a amplitude (intensidade) e tempo de propagação (atraso) para o som direto e para todas as reflexões.

A resposta impulsiva pode ser armazenada num vetor \mathbf{IR}_j contendo a descrição de I raios associados ao objeto O_j : o raio direto v_0 , as reflexões v_i e respectivos parâmetros de atraso (t_i) e amplitude (A_i):

$$\mathbf{IR}_j = (v_0(A_0, t_0) \quad v_1(A_1, t_1) \quad \dots \quad v_I(A_I, t_I)), \text{ ou simplesmente } \mathbf{IR}_j = (v_0 \quad v_1 \quad \dots \quad v_I) \quad \text{eq. 1}$$

Podemos escrever a equação 1 acima de uma outra forma mais conveniente para representar cada par *amplitude-atraso* por uma linha numa matriz como abaixo:

$$\mathbf{IR}_j = \begin{pmatrix} v_0 & 0 & \dots & 0 \\ 0 & v_1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & v_I \end{pmatrix} \quad \text{- eq. 2}$$

que é a **matriz de intensidade e atraso dos raios de O_j** .

Até este ponto, estaria assegurada para aquele ambiente a descrição formal do comportamento temporal da propagação sonora entre a fonte O_j e o ouvinte, isto é, a descrição de sua **resposta impulsiva onidirecional**. Entretanto, para incorporar

também as características espaciais da propagação, é necessário considerar a posição e direção virtual de origem do raio direto e de cada reflexão (raio i).

Para codificar a diretividade para cada raio i a posição $r(x,y,z)$ da fonte-imagem associada a ele será usada para calcular a contribuição deste raio em cada direção do sistema de coordenadas. A posição da fonte-imagem do raio i é dada pelo vetor $r_i = (r_{ix}, r_{iy}, r_{iz})$. Podemos expressar este vetor mais convenientemente em um formato *B-Format* vetorial, assim:

$$\mathbf{B}_{ji} = \begin{pmatrix} r_{iX} \\ r_{iY} \\ r_{iZ} \\ r_{iW} \end{pmatrix}, \text{ onde } r_{iW} = 1 \quad - \text{eq. 3}$$

Estendendo este raciocínio para todos os raios da fonte O_j obtemos a matriz Ambisonics que codifica a direção de todos os raios de O_j , que é dada por:

$$\mathbf{B}_{ji} = \begin{pmatrix} r_{0X} & r_{1X} & \dots & r_{iX} \\ r_{0Y} & r_{1Y} & \dots & r_{iY} \\ r_{0Z} & r_{1Z} & \dots & r_{iZ} \\ r_{0W} & r_{1W} & \dots & r_{iW} \end{pmatrix} = \begin{pmatrix} B_{jX} \\ B_{jY} \\ B_{jZ} \\ B_{jW} \end{pmatrix} \quad - \text{eq. 4}$$

que é a **matriz de direção dos raios de O_j** , em *B-Format*.

O módulo simulador acústico tem como tarefa calcular todos os caminhos entre o objeto j e o ouvinte, utilizando para isso um modelo acústico para tal (método da imagem ou traçado de raios). Sua meta é calcular as matrizes \mathbf{IR}_j e \mathbf{B}_{ji} acima, sendo que $v_i = (A_i, t_i)$ em \mathbf{IR}_j contém a amplitude do raio i e o tempo de chegada (atraso), e $r_i = (r_{ix}, r_{iy}, r_{iz})$ em \mathbf{B}_{ji} contém a direção do raio i em coordenadas cartesianas.

A simulação da propagação acústica entre a fonte sonora O_j e o ouvinte fica então completamente descrita pela resposta impulsiva da sala adequadamente codificada em 3D. Para Ambisonics, isto é feito gerando-se o vetor \mathbf{IR}_j^B utilizando a resposta impulsiva onidirecional (\mathbf{IR}_j) juntamente com os dados da direção $r_i(x,y,z)$ de cada raio sonoro (previamente codificado em \mathbf{B}_{ji}).

A operação pode ser representada por uma multiplicação matricial, como abaixo:

$$\mathbf{IR}_j^B = \mathbf{B}_{ji} \cdot \mathbf{IR}_j = \begin{pmatrix} r_{0X} & r_{1X} & \dots & r_{IX} \\ r_{0Y} & r_{1Y} & \dots & r_{IY} \\ r_{0Z} & r_{1Z} & \dots & r_{IZ} \\ r_{0W} & r_{1W} & \dots & r_{IW} \end{pmatrix} \begin{pmatrix} v_0 & 0 & \dots & 0 \\ 0 & v_1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & v_I \end{pmatrix} = \begin{pmatrix} v_0 r_{0X} & v_1 r_{1X} & \dots & v_I r_{IX} \\ v_0 r_{0Y} & v_1 r_{1Y} & \dots & v_I r_{IY} \\ v_0 r_{0Z} & v_1 r_{1Z} & \dots & v_I r_{IZ} \\ v_0 r_{0W} & v_1 r_{1W} & \dots & v_I r_{IW} \end{pmatrix} \text{ - eq. 5}$$

sendo que $\mathbf{IR}_j^B(4xI) = \mathbf{B}_{ji}(4xI) \cdot \mathbf{IR}_j(IxI)$ é o tamanho das matrizes, onde I é o número máximo de reflexões adotado no modelo acústico⁶².

\mathbf{IR}_j^B é a matriz em *B-Format* que codifica completamente no tempo e no espaço a resposta impulsiva do ambiente obtida na posição do ouvinte devido à fonte O_j .

O algoritmo em alto nível para o módulo responsável pela simulação de propagação acústica do objeto j é portanto:

para cada fonte sonora j no ambiente A

- pega-se dimensões {X,Y,Z} do ambiente A
- pega-se posição do ouvinte e fonte j {x,y,z}
- determina-se duração T desejada de cálculo (tamanho do IR) em número de amostras NPTS
- inicializa-se 4 vetores de IR (IRx,IRy,Irz,IRw) um para cada sinal de uma IR composta em *B-Format*
- dispara-se o algoritmo para achar todas as fonte-imagem envolvidas no tempo T que disparam raios até o ouvinte (inclusive a própria fonte)
- para cada tempo t (para as 4 IR's) obtém-se todas as reflexões (raios) que chegam ao ouvinte neste tempo
- para cada reflexão válida:
 - i. obtém-se sua direção 3D,
 - ii. calcula-se sua atenuação pela distância percorrida e reflexões absorptivas ocorridas no ambiente,
 - iii. calcula-se sua contribuição em cada IR (isto é, codifica-se a amplitude em *B-Format*)
- repete-se para todos os tempos t até T (extensão max)
- acham-se todos os M raios (reflexões) no tempo T
- ao final, cada IR calculada corresponde a uma linha da matriz B_{ji} de localização (em *B-Format*), contendo a amplitude e o tempo de cada reflexão (raio i) (i=1 até M)

⁶² I pode por exemplo indexar o próprio tempo, considerando uma situação em que uma reflexão/raio está associada a cada unidade de tempo. Na equação 5 o termo de v_i que é multiplicado na operação matricial é somente a amplitude A_i . O tempo t_i de cada raio não é multiplicado.

O módulo de propagação acústica foi implementado em FORTRAN a partir de modificações no programa original de fonte-imagem para gerar além da resposta ao impulso onidirecional mais 3 respostas ao impulso, correspondendo ao registro do deslocamento (direção e sentido) produzido por cada raio na posição do ouvinte nas três coordenadas espaciais (x , y , z). O programa modificado está no Anexo I.

A Figura 35 mostra as 4 *IR*'s (h_w , h_x , h_y e h_z) resultantes de 200ms de simulação para a cena de testes proposta (vide Figura 33).

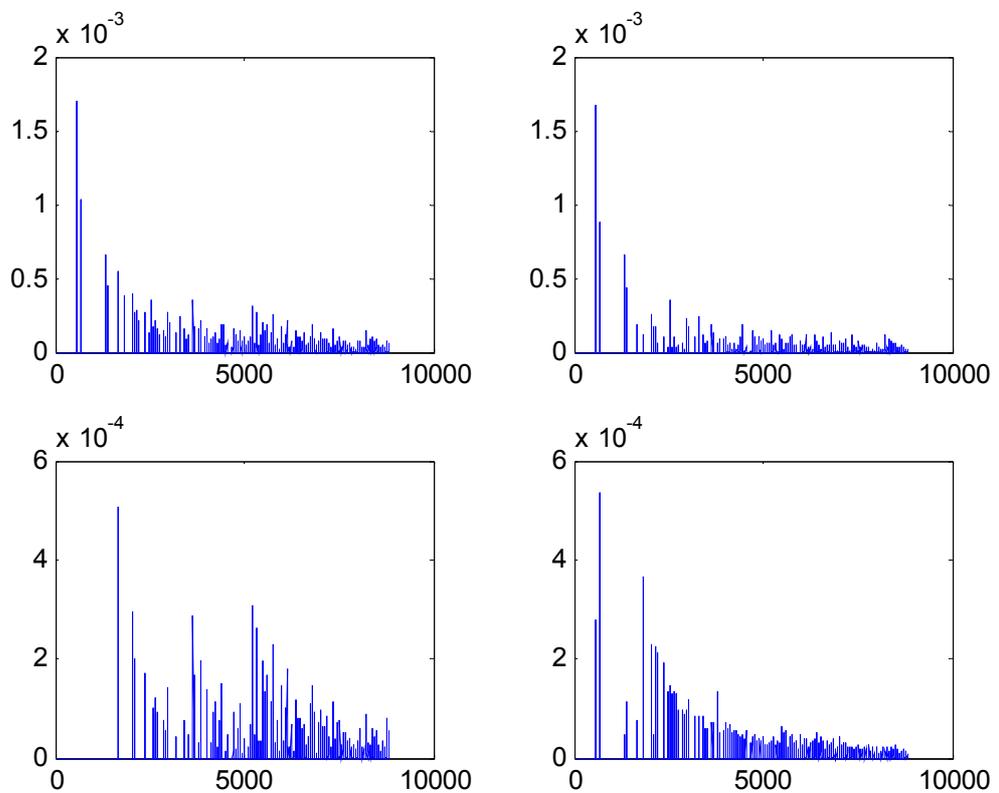


Figura 35 – Respostas impulsivas h_w , h_x , h_y , e h_z da cena de testes para 200ms

A resposta ao impulso obtida pelo simulador acústico baseado no método de imagens não considera um comportamento acústico dependente da frequência. A equação principal de modelagem em (ALLEN, 1979) não incorpora os coeficientes acústicos dos materiais em função da frequência, nem os parâmetros de filtragem distintos em cada faixa.

4.2.5 Codificação sonora espacial

Utilizamos a técnica de perifonia Ambisonics para codificar os sons das fontes sonoras distribuídas num campo sonoro tridimensional. A localização dos objetos no espaço acústico virtual será feita com respeito à sua posição em relação ao ouvinte⁶³ através do mapeamento tradicional da técnica Ambisonics, onde se parametriza a distância ao centro (raio), o ângulo de azimute (θ) e o ângulo de elevação (φ), ou, alternativamente, os vetores cartesianos respectivos de sua posição em respeito aos eixos x , y , e z .

Esta técnica foi a escolhida por vários motivos, a saber: a) permite a reprodução de um campo sonoro 3D por meio de alto-falantes (sem fones de ouvido) diretamente projetado no cubo da CAVERNA, b) admite escalabilidade no número de alto-falantes que se deseja utilizar, c) admite flexibilidade no posicionamento dos alto-falantes ao redor da área de audição, i.e., o cubo da CAVERNA, d) é mais bem documentada, e) e possui diversos componentes de *software* disponíveis no mercado, (como por exemplo *plugins VST* e binários livres) implementando codificadores e decodificadores para *B-Format*, que serão usados como referência para implementações e testes de um auralizador Ambisonics para a CAVERNA.

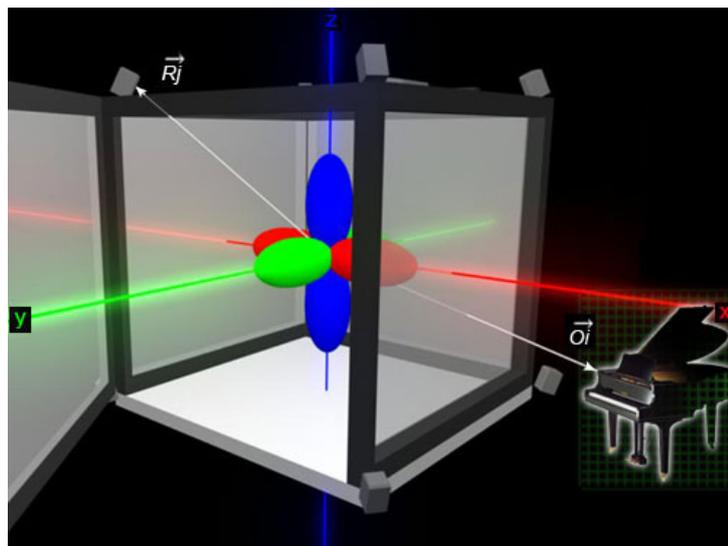


Figura 36 – Vetor de posição de um objeto sonoro O_j e do falante R_j

⁶³ origem do sistema de coordenadas também coincidente com o centro da CAVERNA

A Figura 36 ilustra uma configuração Ambisonics tradicional com 8 alto-falantes (geometria cúbica) para sonorizar uma cena audiovisual na CAVERNA contendo um objeto sonoro (um piano virtual). Ao centro mostra-se uma representação do padrão de cobertura das esféricas harmônicas em cada eixo (x, y, z). A referência da área de audição (origem dos eixos) é o centro da CAVERNA. Tantos os objetos sonoros virtuais (vetor O_j na figura) constantes da cena que se está projetando quanto os alto-falantes fisicamente instalados no sistema (vetor R_j na figura) têm suas posições corretamente parametrizadas em relação à origem. O primeiro vetor permite codificar em *B-Format* a correta localização dos objetos na cena. O segundo vetor permite corretamente decodificar o som em *B-Format* para cada alto-falante, conhecendo-se sua posição em relação à origem.

O módulo auralizador tem como tarefa gerar ou sintetizar o som do objeto j como seria ouvido na posição do ouvinte central. Sua meta é gerar o som do O_j em *B-Format* de 1ª ordem, isto é, em 4 canais, a saber: O_{jX}, O_{jY}, O_{jZ} e O_{jW} .

A \mathbf{IR}_j^B da equação 5 traz os 4 *IR*'s que precisam ser convoluídos com O_j :

$$\mathbf{IR}_j^B = \begin{pmatrix} v_0 r_{0X} & v_1 r_{1X} & \dots & v_I r_{IX} \\ v_0 r_{0Y} & v_1 r_{1Y} & \dots & v_I r_{IY} \\ v_0 r_{0Z} & v_1 r_{1Z} & \dots & v_I r_{IZ} \\ v_0 r_{0W} & v_1 r_{1W} & \dots & v_I r_{IW} \end{pmatrix} = \begin{pmatrix} h_X \\ h_Y \\ h_Z \\ h_W \end{pmatrix} \quad \text{- eq. 6}$$

A Figura 37 ilustra a convolução final para codificar o som O_j e produzir sua representação intermediária O_j^B . Isso é feito convoluindo-se o som de O_j por cada vetor resposta ao impulso h_X, h_Y, h_Z e h_W , (equivalente a convoluir O_j por cada linha da matriz \mathbf{IR}_j^B).

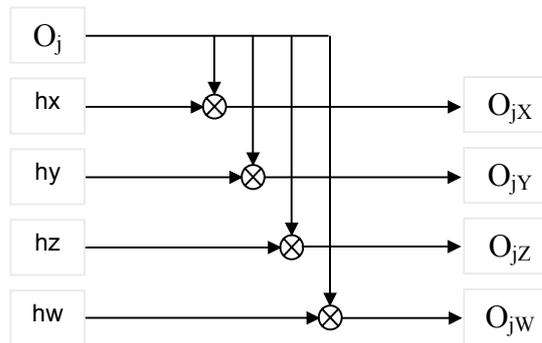


Figura 37 – Esquema de convolução do som O_j pelas IR 's em B -Format

O módulo responsável pela codificação espacial Ambisonics do som do objeto j realiza o seguinte algoritmo:

para objeto j :

pega-se IR_j e B_{ji} :

- gera uma IR para cada sinal do B -Format: hx , hy , hz e hw (modelo da propagação 3D)
- pega-se o som do objeto O_j em arquivo (wave)
- gera o som 3D de O_j em B -Format (convolui o som do objeto j com hx , hy , hz e hw , gerando 4 canais O_{jX} , O_{jY} , O_{jZ} , O_{jW})
- condiciona sinal de saída em B -Format para transmissão ou armazenamento ou consumo imediato (pelo nó/módulo sonorizador)

No *pipeline* da produção de áudio espacial, a convolução é umas das operações de maior custo computacional, cuja implementação tem impacto direto no desempenho do sistema e sua capacidade de auralização em tempo real. Técnicas de convolução rápidas são necessárias para atender a esta finalidade.

A maneira mais rápida de se realizar convolução é através de multiplicações no domínio da frequência. Para tanto, requer-se uso intensivo de FFT . Usualmente os filtros utilizados em auralização são longos, assim como os sinais sonoros (usualmente em fluxo contínuo) implicando FFT 's dispendiosas e baixo desempenho. Duas técnicas para convolução para dados longos tradicionalmente são descritas na literatura: a *overlap-and-save* e *overlap-and-add* (PRESS, 1994).

Uma abordagem bastante atraente é a convolução particionada para acelerar este processo. A máquina *BruteFIR*⁶⁴ é uma biblioteca para realizar convolução particionada e filtragem *FIR* concebida para produzir o maior *throughput* possível para filtros longos (maiores que 10.000 *taps*) e convoluir muitos canais em tempo real. Para isso utiliza características especiais de algumas instruções dos processadores Intel e *AMD* (instruções do tipo *SIMD*) para acelerar os cálculos multiplicando-se o desempenho através da aplicação de uma instrução *SIMD* para vários operandos ao mesmo tempo, e principalmente tratando a convolução de forma particionada (particionando-se longos filtros) o que permite o uso de *FFTs* menores, onde portanto a maior parte de processamento seja utilizada na multiplicação dos fatores no domínio da frequência (multiplicação dos coeficientes do filtro com os do sinal de entrada) do que na própria *FFT*. *FFT*'s menores requerem menos tempo de processamento.

Convolução particionada também é citada como adequada para reduzir erros de quantização em longas extensões de *FFT* (que exigem elevado número de multiplicações e adições). Um problema tradicional em *hardware* não-*DSP* são os atrasos de entrada/saída (I/O) devido a arquitetura de acesso à memória nos computadores tradicionais, e podem, por exemplo, causar atrasos superiores aos esperados na análise de complexidade algorítmica, que muitas vezes não levam em consideração os atrasos relativos a acesso de memória na arquitetura empregada.

4.2.6 Decodificação do formato espacial e sonorização multicanal

O módulo de sonorização tem a função de receber os sons em *B-Format* de todos os nós do aglomerado que estiverem sonorizando alguma fonte sonora. Antes de enviar os canais para a placa multicanal porém, o programa misturador (mixer em *B-Format*) combinará linearmente todos os sons em *B-Format* gerando uma saída *B-Format* única, com 4 canais (para 1ª ordem). O decodificador Ambisonics será invocado então para decodificar 1 vetor de 4 canais para N saídas (alto-falantes). As

⁶⁴ Disponível em http://www.ludd.luth.se/~torger/brutefir.html#bruteconv_1 (acesso em: 30 maio 2005).

equações de decodificação Ambisonics levam em conta o número (N) e a configuração ou disposição geométrica dos alto-falantes ao redor da área de audição.

No presente experimento optamos por adotar aplicativos abertos de terceiros para esta camada na hierarquia, utilizando componentes já disponíveis na arquitetura proposta. Utilizamos o aplicativo *Ambisonic Player* v.1.12 de Richard Furse (FURSE, 1999).

Configurações de alto-falantes

O posicionamento dos alto-falantes na CAVERNA apresenta diversas limitações, devido às áreas proibidas pela projeção visual. Isso restringe, por exemplo, uma configuração com 4 caixas apontando para o centro das telas à altura do ouvido do usuário (algo entre 1,70 a 1,80m). É possível com mais facilidade explorar configurações cúbicas e configurações mistas contando com alto-falantes de reforço para incrementar a percepção da elevação, o que resulta em geometrias não regulares.

Para decodificação Ambisonics, testamos 3 configurações regulares:

- anel octogonal planar (sem elevação) fora da CAVERNA (no *hall* de entrada da mesma) com 8 alto-falantes distribuídos em anel ao redor da área de audição para se avaliar a qualidade da auralização sem a interferência das telas
- anel octogonal montado nos bastidores por trás das telas da CAVERNA, para avaliar a qualidade da auralização numa situação real de uso dentro da CAVERNA,
- uma configuração cúbica tradicional (posicionando os alto-falantes levemente transladados no plano horizontal para evitar que ficassem obstruídos pela estrutura da CAVERNA.

Na primeira configuração (anel octogonal fora da CAVERNA) um arranjo de 8 caixas sobre pedestais em octógono foi disposto no hall de entrada da CAVERNA. A Figura 38 mostra uma foto deste experimento.

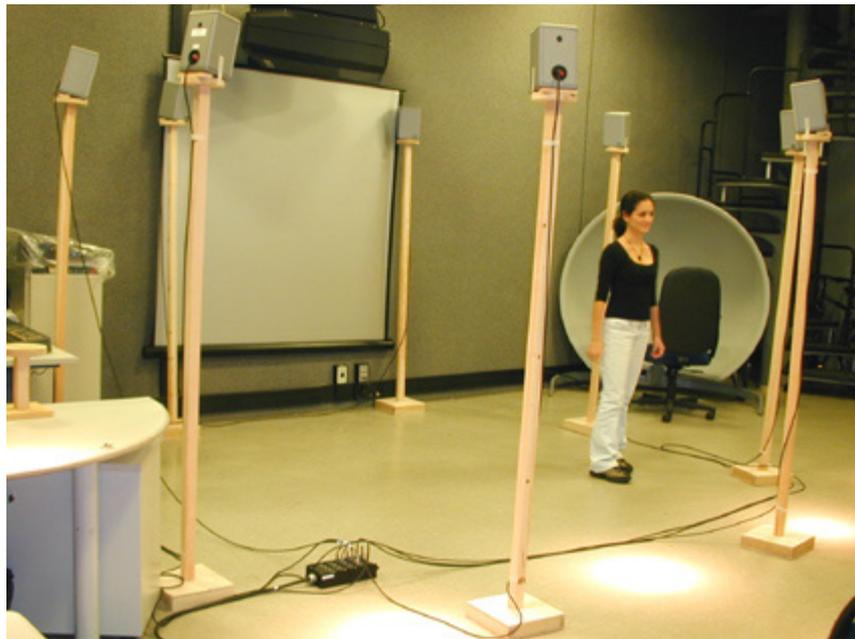


Figura 38 – Foto de uma configuração octogonal planar externa à CAVERNA

Na segunda configuração, o mesmo anel octogonal foi montado atrás das telas da CAVERNA. A Figura 39 mostra este arranjo dos 8 alto-falantes instalados em anel na CAVERNA.

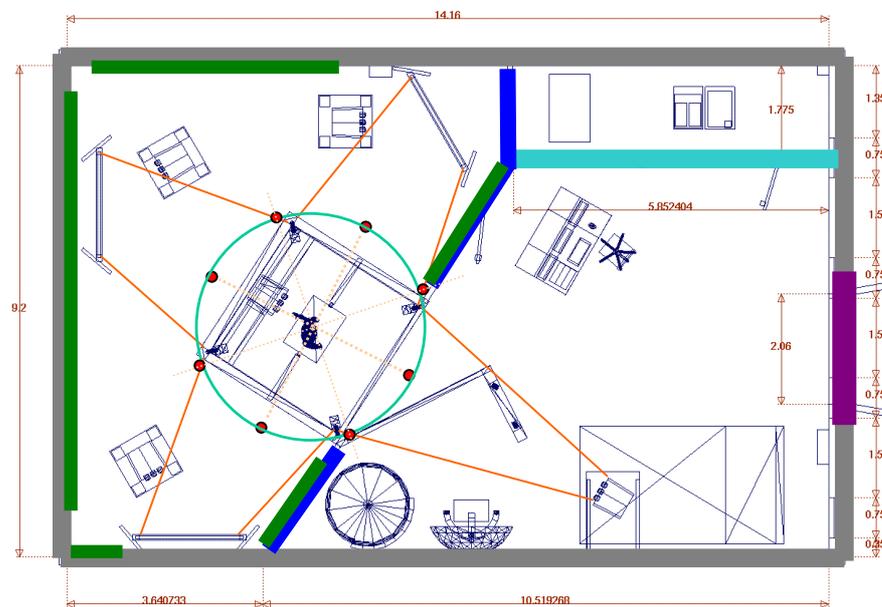


Figura 39 –Montagem de alto-falantes em anel octogonal (interno)

A terceira foi a configuração cúbica típica, com a matriz de alto-falantes instalada nos vértices do cubo de multiprojeção da CAVERNA, atrás das telas. Este arranjo foi

o único testado com capacidade 3D (elevação). A Figura 40 indica o posicionamento escolhido para os alto-falantes, priorizando a simetria típica deste arranjo.

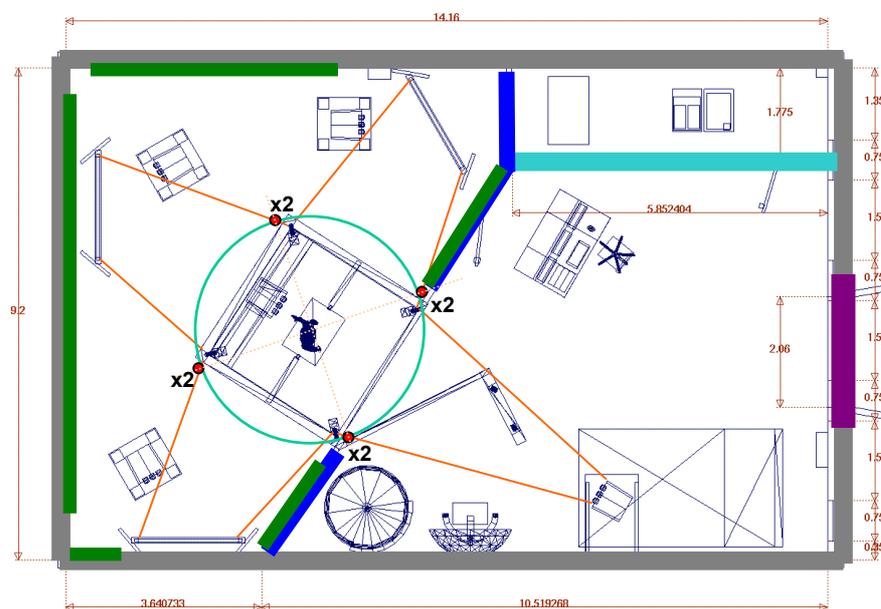


Figura 40 – Matriz de alto-falantes em configuração cúbica típica (3D)

A Figura 41 mostra uma foto da montagem dos alto-falantes nas proximidades do canto inferior esquerdo da CAVERNA Digital.



Figura 41 – Foto ilustrativa da montagem e cabeamento de alto-falante nos cantos inferiores da CAVERNA.

Decodificação para N canais

A etapa de sonorização consiste primeiro em decodificar os sons das fontes sonoras armazenados em *B-Format* para N canais independentes, e então preparar (condicionar) e enviar os sinais para reprodução a cada uma das N caixas de som da matriz multicanal.

O algoritmo em alto nível desta camada está detalhado abaixo:

o módulo responsável pela reprodução multicanal dos objetos sonoros:

- recebe ou acessa os sons 3D codificados em *B-Format* de cada objeto O_j
- mixa os sons de cada O_j em *B-Format*
- decodifica o som mixado (de *B-Format* para N canais de saída)
- sonoriza (toca) N canais em sincronismo (N saídas para N alto-falantes)
- (amplificação externa se caixas acústicas passivas)

O diagrama de blocos do módulo de sonorização é mostrado na Figura 31.

Decodificação Ambisonics

As equações para decodificação Ambisonics para as configurações adotadas são apresentadas por (GERZON, 1976). FURSE (1999) documenta os coeficientes para decodificação de muitas configurações regulares pré-calculados.

O diagrama de blocos na Figura 42 mostra as operações de mixagem, decodificação e sonorização. O exemplo se refere à decodificação de *B-Format* para Ambisonics de 1ª ordem (4 canais: X, Y, W, Z). Objetos sonoros codificados em *B-Format* (O_j^B) são misturados previamente antes da decodificação global. Ganhos adequados dependente da frequência (G_n) são utilizados como filtros psicoacústicos (do tipo *shelf*) aplicados aos canais X, Y, Z e W para tratar diferenças na percepção humana quanto aos mecanismos de localização psicoacústicos distintos operando nas altas frequências dos sons (>700Hz) e nas baixas frequências (<300-400Hz). No arranjo cúbico, para calcular a saída de uma caixa pondera-se cada sinal *B-Format* pelo cosseno do ângulo de posicionamento daquela caixa em relação ao eixo em questão (veja detalhe na Figura 42).

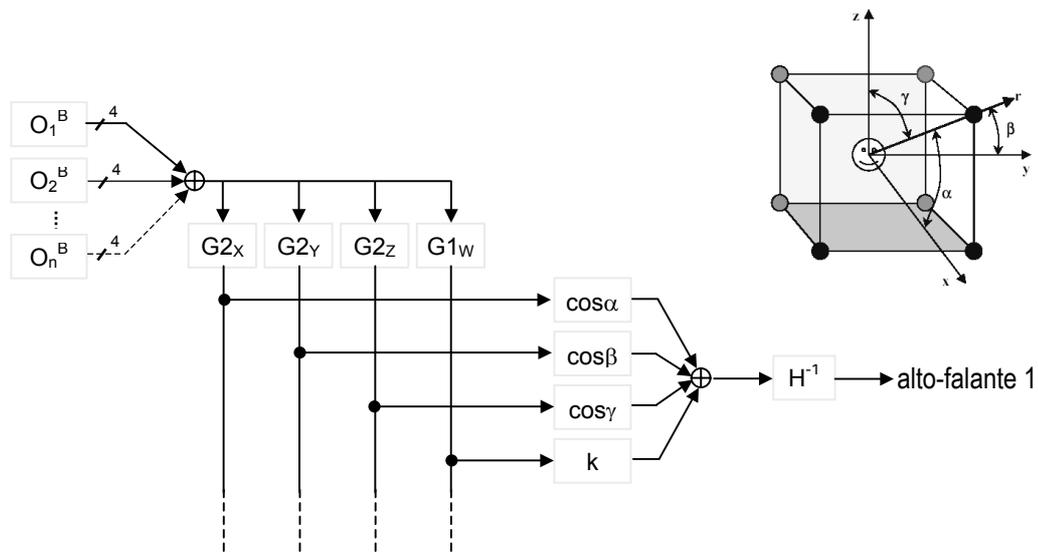


Figura 42 – Decodificação de som 3D (em *B-Format*) para um canal de saída

Finalmente, eventuais filtragens necessárias para pós-condicionar ou equalizar o sinal para o alto-falante podem ser aplicadas (estrutura H^{-1} na figura).

4.2.7 Implantação da infra-estrutura de áudio multicanal

O projeto e construção de uma infra-estrutura para sonorização multicanal na CAVERNA Digital consiste numa das principais contribuições deste trabalho. O sistema implementado é equivalente ao planejado para as duas primeiras fases do projeto AUDIENCE, para experimentos baseados em Ambisonics.

Para a realização do experimento piloto e validação do auralizador Ambisonics procedemos à montagem, aquisição, construção, instalação e integração de diversos equipamentos, estruturas e subsistemas, descritos na Tabela 9 a seguir.

Tabela 9 – Infra-estrutura geral do sistema de produção sonora espacial

Nó dedicado a processamento de áudio e sonorização	
Duplo processador Intel® Xeon 3.06GHz, 2GB DDR-RAM, HD SCSI 36GB.	Um nó do aglomerado de computadores (<i>cluster</i>) foi configurado em “dual-boot” Linux (Fedora) e Windows (XP) para hospedar uma placa multicanal, para executar programas auralizadores diversos, realizar testes, programação, simulação, sonorização
Placa de áudio multicanal	
Delta 1010 (M-Audio) 24bits/96kHz, c/ MIDI e word-clock	Módulo externo de 8 saídas balanceadas/não-balanceadas analógicas, 2 saídas S/PDIF, conectado ao nó de áudio via placa hospedeira em barramento PCI

Software para medidas acústicas, processamento de áudio e auralização	
Aurora, <i>plugins</i> para Audition, de Angelo Farina	Sistema completo para medidas de resposta ao impulso e auralização de sinais acústicos anecóicos, contendo diversos módulos processadores implementados na forma de <i>plugins</i> para o <i>software</i> CoolEdit (atualmente Adobe Audition)
Ambisonics players	
Ambisonic Player (Furse) Ambisonics VST Plugins	<i>Player</i> para Windows. <i>VST's</i> são módulos de <i>software</i> de função específica que funcionam conectáveis a programas hospedeiros de áudio/música (como Cubase e outros) compatíveis com esta tecnologia, para processamento e geração de áudio (como decodificadores Ambisonics da Voxengo)
Software para desenvolvimento em Fortran	
Force 2.0 (Ambiente Integrado de Desenvolvimento FORTRAN)	IDE utilizado para desenvolvimento do algoritmo do módulo simulador acústico e geração de <i>IR's</i> em <i>B-Format</i> (1ª ordem)
Ambiente gráfico para aplicações em música e processamento de áudio	
<i>PD – Pure Data</i> , de Miller Puckette	Ambiente de programação gráfico para montagem e configuração de cadeias de processamento de áudio, conexões, desenvolvimento e prototipagem de aplicações em áudio/música
Software para simulação de sistemas e processamento de sinais	
MATLAB versão 6.0.0.88	Utilizado para análise de sinais
Software para áudio multicanal	
CoolEdit Pro – Audition, da Adobe	Utilizados para gravação, edição, processamento, aplicação de efeitos nativos e importáveis (através de <i>plugins</i>), para reprodução e testes de audição
Software de navegação em RVI	
Jinx, da CAVERNA Digital	Utilização prevista para projeção do espaço virtual (cena audiovisual de testes consistindo de um auditório retangular de 15x30x8m) Utilização futura prevista na fase de validação da navegação no mundo virtual
Alto-falantes (caixas acústicas)	
Caixa acústica LANDO HiFi, modelo Merlin 100	8 caixas acústicas analógicas de “alta-fidelidade”, de duas vias (<i>bass reflex</i>), potência nominal 75W rms, potência musical 100W, resposta de frequência 60Hz-20kHz, impedância 4 a 8 Ohms), utilizadas em testes pantofônicos (octógono no plano horizontal) e perifônicos (cubóide) (Fase II do projeto AUDIENCE prevê até 16 caixas)
Pedestais	
Pedestais montáveis de altura variável, em madeira, fabricados pelos Serviços Gerais da POLI-USP	Pedestais montáveis, de altura e posição reconfiguráveis, com enchimento de concreto na base para maior estabilidade

Estruturas de sustentação	
Estruturas de suporte superior, em madeira, fabricadas pelos Serviços Gerais da POLI-USP	Estruturas de suporte montáveis para sustentação e posicionamento de caixas acústicas na parte superior da CAVERNA, com amplos graus de liberdade na cobertura horizontal e angulação da direção de irradiação das caixas, por meio de braços articulados
Cabearamento e conectorização de áudio	
Multicabos blindados e cabos individuais conectorizados em P10, fabricação da CABOS GOLDEN	2 multicabos blindados de 12 vias balanceadas para mandada longa de sinal para 2 pontos nos bastidores da CAVERNA; 32 cabos individuais, padrão microfone, conectorizados P10 (TRS) para distribuição de sinal para as caixas acústicas
Amplificadores de áudio para caixas analógicas	
Marca SANKYA, modelo ST 4.0, unidades experimentais da série	2 amplificadores de 4 canais (total 8 canais) de 30W rms por canal, resposta de 10Hz-30kHz de frequência, DHT ⁶⁵ 0,01%, impedância nominal de 4-6 Ω , entrada em nível de linha não-balanceada, e saídas amplificadas não-balanceadas

Posicionamento dos alto-falantes

As caixas poderiam ser fixadas na estrutura, entretanto cremos que esta opção só aumenta as restrições nos experimentos iniciais, dificultando o reposicionamento dos alto-falantes para testar arranjos e técnicas diferentes. Dentro dos objetivos de se permitir escalabilidade e flexibilidade na escolha dos métodos de sonificação bem como nas configurações de posição para os alto-falantes, optamos por uma solução baseada no uso de pedestais móveis e reconfiguráveis (em altura). A figura mostra os pedestais que foram desenvolvidos para a CAVERNA Digital, em madeira.

⁶⁵ DHT – Distorção Harmônica Total



Figura 43 – Pedestais de madeira reconfiguráveis para posicionamento dos alto-falantes

Os pedestais possuem três partes montáveis e encaixáveis por meio de pinos: uma base (com estrutura de concreto para uma maior estabilidade), uma coluna de madeira, e um topo também de madeira encaixável, em formato circular (para se evitar acidentes com pontas em caso de eventual queda sobre as telas da CAVERNA). A altura nominal do pedestal é 1,95m, e eles foram projetados para serem posicionados na parte de trás da CAVERNA, por trás das telas, instalados nos cantos da estrutura, fora do alcance dos projetores e apontando para dentro da CAVERNA. Devido à elevação de 20cm do piso interno, a altura nominal dos pedestais para quem está dentro da CAVERNA é de 1,75m, o que alcança a altura dos ouvidos para uma grande parte da população.

Esta solução é flexível o suficiente para permitir posicionar os alto-falantes para configurações *n.m* (5.1/6.1/7.1), Ambisonics, SVPA e Ambiophonics.

Esquema de conexões físicas de áudio

A Figura 44 apresenta um diagrama geral de conexões e distribuição de áudio do sistema construído, mostrando os equipamentos envolvidos (nó processador de áudio, placa multicanal, amplificadores de saída, e multicabo para transmissão dos canais a cada alto-falante montado ao redor da CAVERNA).

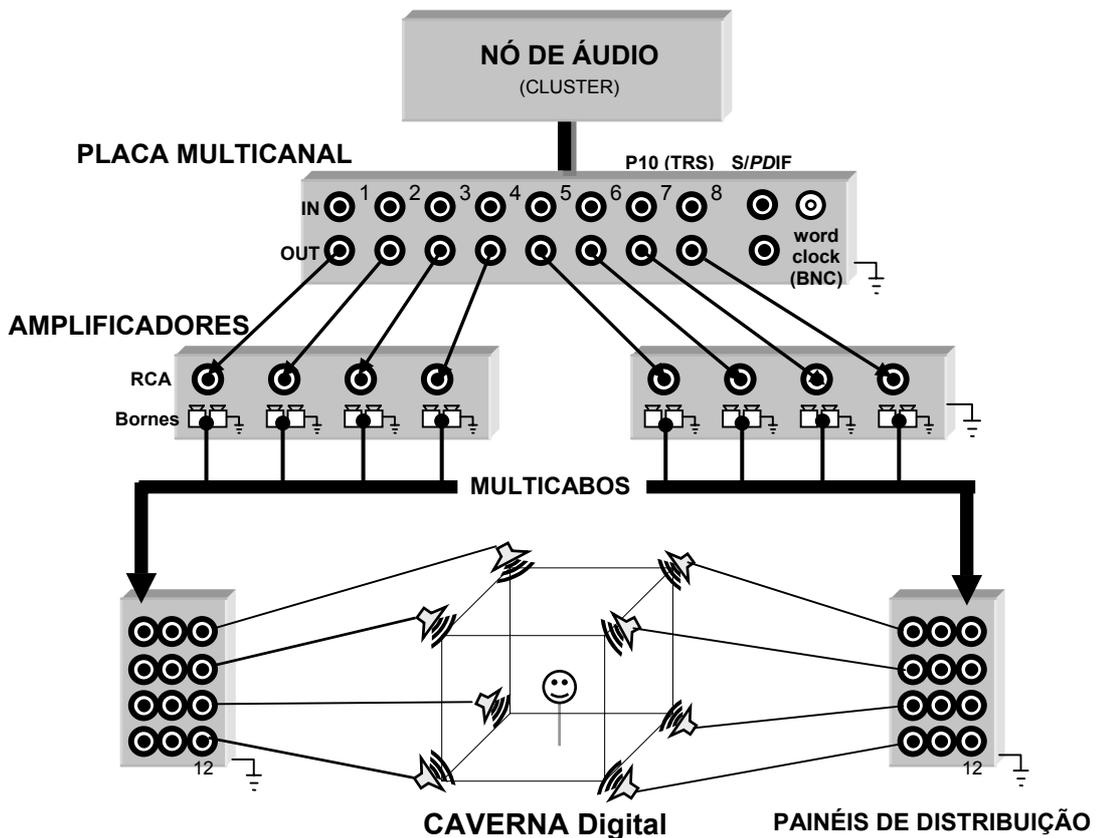


Figura 44 – Diagrama geral de conexões e distribuição do sistema de áudio

4.3 Simulações e testes (casos de uso)

Utilizamos a metodologia ilustrada na Figura 45 para implementar os testes de auralização com o módulo auralizador e codificador Ambisonics desenvolvido.

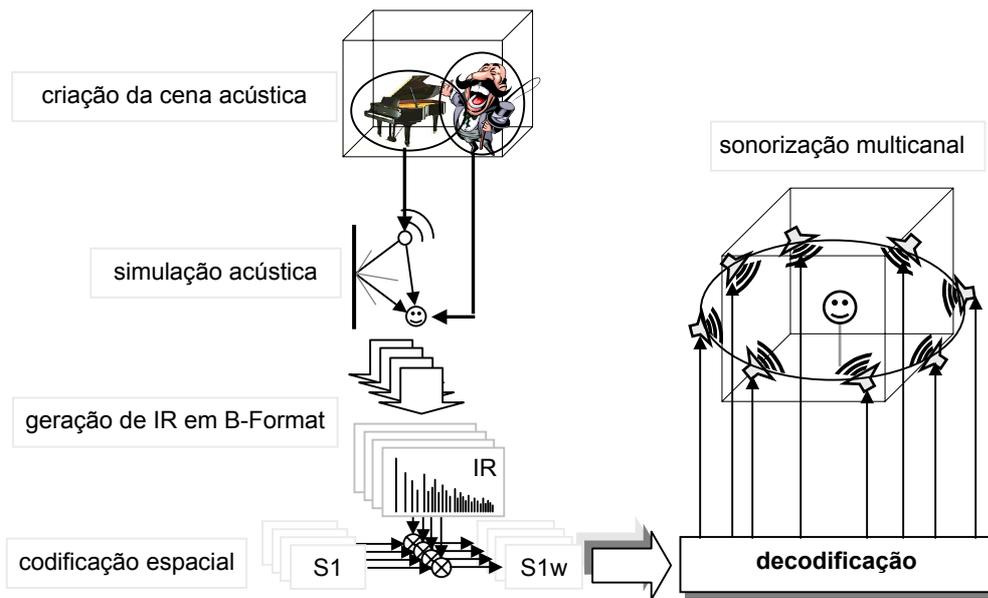


Figura 45 – Metodologia de testes de auralização

Criamos uma cena de testes codificada em *X3D* contendo tanto a descrição visual quanto acústica do meio, os objetos e seus atributos e parâmetros acústicos de interesse para a auralização.

A cena de testes é descrita nos itens 4.2.2 e 4.2.3, mostrada nas figuras 32 e 33, e consiste em:

- uma sala retangular de 15m (largura) x 20m (profundidade) x 8m (altura), descrita no item e mostrada na figura
- um objeto sonoro (fonte) posicionado à esquerda do ouvinte (a 4,5m deste) nas coordenadas (3;14;1), simulando a posição de um instrumentista sentado (a 1m do solo)
- ouvinte, posicionado nas coordenadas (7,5; 14;1,75)

Esta cena encontra-se no Anexo II (em código *X3D*).

Alimentamos o simulador acústico com os dados da cena sobre (i) dimensões do ambiente, (ii) posição da fonte sonora (objeto O_j), (iii) posição do ouvinte (receptor, referência centro da CAVERNA), e (iv) parâmetros acústicos das paredes (coeficientes de reflexão β), gerando 4 *IR's* em *B-Format* em sua saída.

Aplicamos a acústica pretendida ao sinal da fonte sonora, convoluindo-o com as *IR's* em *B-Format*, gerando assim o sinal da fonte codificado (espacialmente) em Ambisonics.

O último passo consistiu na decodificação do sinal espacial para 8 sinais, endereçados aos alto-falantes das configurações testadas, e respectiva sonorização final.

Utilizamos *IR's* de 200ms de comprimento (8820 pontos para taxa de amostragem de 44.1 kHz). Duas fontes sonoras de teste foram auralizadas com estes filtros, na mesma posição e para a mesma sala:

- uma flauta, gravada anecóica,
- uma bateria sintética, produzida sem efeitos ambientais (anecóica).

Tomamos o cuidado no trecho de testes para a flauta de considerar alguns gestos musicais cujos conteúdos espectrais variantes e reconhecida manifestação em faixas de frequências específicas são importantes para avaliar o desempenho de codificadores em geral (FARIA, 1997), como por exemplo:

- frase ligada, veloz, decrescente
- modulações do tipo vibrato
- notas mal entonadas (conteúdo harmônico denso)
- notas secas, ataques acentuados e articulados
- sequência de notas oitavadas
- notas suavemente entonadas

Doze pessoas de uma amostragem aleatória ouviram as cenas acústicas auralizadas, e avaliaram dois quesitos principais (direção do som e dimensão da sala), bem como deram depoimentos livres sobre a qualidade espacial e sônica percebida. Os resultados são apresentados no próximo item.

Adicionalmente aos itens de teste preparados, outros itens⁶⁶ foram avaliados na montagem de testes em anel fora da CAVERNA (Ambisonics livre). Esta ficou funcional e disponível por 2 semanas, em que foi ouvida por várias pessoas.

Testes preliminares com o simulador acústico implementado mostraram bons resultados perceptuais para modelar as reflexões primárias e difusas.

Nosso objetivo com a proposição de cenas audiovisuais simples nestes primeiros experimentos com auralização Ambisonics na CAVERNA é avaliar diversas questões em aberto, a saber:

- qualidade sonora na área de audição pretendida,
- artefatos sonoros ouvidos e em que circunstâncias,
- avaliação perceptual do grau de imersão obtido,
- desempenho do simulador acústico simples proposto,
- desempenho da auralização com fontes (timbres) diferentes,
- precisão na localização percebida dos objetos sonoros,
- distância percebida da fonte,
- comportamento/estabilidade das imagens em baixas frequências,
- desempenho e avaliação de sucesso em diferentes casos de uso, cenas audiovisuais ou circunstâncias específicas, entre outros.

Para os propósitos iniciais de avaliação da qualidade espacial, nível de imersão sonora e estabilidade do campo auralizado na CAVERNA não foi necessário o uso de navegação interativa pela cena audiovisual virtual. Embora esta possa por um lado ampliar o alcance do julgamento global da qualidade da auralização, por outro lado poderia também prejudicar esta tarefa, retirando a atenção na avaliação subjetiva do plano sonoro.

Por essa razão não utilizamos para o experimento teste de auralização o sistema de navegação audiovisual interativa, nem integramos o sistema gerenciamento da

⁶⁶ como uma gravação Ambisonics de um órgão tocando em uma catedral de grande porte.

CAVERNA. O Jinx foi usado para projetar a experiência visual do auditório virtual criado para os testes, renderizando a parte gráfica da cena audiovisual.

4.4 Resultados

No momento em que esta tese foi concluída a infra-estrutura de áudio do sistema de auralização estava plenamente instalada, e foram realizados testes quanto à qualidade espacial percebida, tanto numa montagem fora da CAVERNA (no *hall* de entrada) quanto dentro da mesma, conforme descrito em 4.3.

Infra-estrutura

A infra-estrutura de áudio implantada para a CAVERNA Digital mostrou-se adequada e satisfatória para atender as tarefas de sonorização e auralização, e fornece a flexibilidade e escalabilidade necessárias para diferentes montagens, configurações e conexões, garantindo ao mesmo tempo uma qualidade e resolução de áudio excelente. Mostrou-se acertada a decisão por equipamentos de alta resolução, baixo nível de ruído e excelente desempenho acústico e elétrico.

O sistema de suportes reconfiguráveis também mostrou-se eficaz e flexível o suficiente para os requisitos de pesquisa e montagens temporárias de *patches* de áudio, que são uma demanda tradicional na área.

Nível de imersão

Verificamos que a qualidade espacial percebida varia entre nível 3 e 4, dependendo do programa auralizado (timbre) e posição relativa na área de audição. Nas audições externas à CAVERNA, longe da influência das telas, algumas pessoas perceberam indicadores de elevação⁶⁷, e o nível de imersão instantâneo atingia 5. A escala de imersão proposta em 3.1 mostrou-se adequada para mensurar julgamentos de qualidade, mas algo difícil para julgar a falta de qualidade. É necessário o desenvolvimento de ferramentas para utilização da mesma, como por exemplo para o

⁶⁷ isso pode ser explicado em função da inclinação dos alto-falantes em direção a um centro virtual aproximadamente a 1,75m de altura, mas irradiando-se de 1,95m (altura dos pedestais), o que causava a percepção dos sons vindo de cima.

equacionamento de métricas objetivas e subjetivas de qualidade e parâmetros acústicos.

Codificação espacial Ambisonics

A codificação Ambisonics proposta e incorporada no algoritmo modificado de (ALLEN, 1979) foi bem sucedida. O algoritmo codifica corretamente a posição e diretividade dos reflexos do reflectograma gerado, e produz como saída as respostas impulsivas do ambiente acústico para os quatro vetores *B-Format*.

As fontes sonoras foram corretamente localizadas, isto é, o mecanismo de codificação da direção forneceu o resultado esperado. Em muitos casos a indicação correta da direção se deu melhor com a combinação de notas mais graves (dó3 e dó4 na flauta) tocadas com algum vibrato e com duração maior, o suficiente para poder movimentar a cabeça enquanto a nota perdurava e resolver a confusão na localização. Percebemos portanto que a indicação da direção correta, nas condições deste experimento, mostrou-se dependente do programa (notas, timbres e duração).

Simulação acústica e auralização

A simulação acústica utilizando o método de fonte-imagem de ALLEN (1979) mostrou plenamente satisfatória, superando as expectativas em alguns casos. Ela incorpora mecanismo de atenuação com a distância de forma correta, e permitiu obter com precisão todos os reflexos na faixa de tempo desejada, e num tempo computacional curto o suficiente para permitir seu uso em aplicações de tempo real, mesmo considerando-se faixas superiores a 100ms, quando vários autores sugerem o uso de reverberadores para produzir o efeito difuso.

Doze pessoas avaliaram dois quesitos de qualidade espacial nos testes de auralização: (i) a direção de onde o som vinha no espaço, (ii) a percepção da sala virtual (qual dimensão da sala virtual era mais comprida). Com relação ao experimento em anel octogonal do lado de fora da CAVERNA, que não mapeava elevação, os resultados estão detalhados na Tabela 10.

Tabela 10 – Resultados do enquete subjetivo de qualidade espacial percebida

Direção do som	
Acertaram a posição correta codificada	4, sem sombra de dúvida
Acertaram, mas indicaram elevação	6 apontaram a direção horizontal correta, mas indicaram alguma elevação
Acertaram parcialmente	1 pessoa registrou a direção correta no início do trecho ouvido, mas acusou um movimento da fonte enquanto o som evoluía para uma posição posterior da sala virtual 1 pessoa registrou duas posições para o som, uma correta (“som agudo”) e outra incorreta (“som grave”), esta última tendendo à parte posterior da sala virtual.
Erraram	0
Percepção da sala virtual (dimensão maior)	
Acertaram que a maior dimensão é a frente-trás, de 20m	3
Erraram ao indicar a dimensão esquerda-direita (15m) como maior	3
A sala lhes pareceu pequena ou uma “cozinha”	1
Erraram a dimensão correta e comentaram sobre o tamanho da sala	1 pessoa, que afirmou que a sala era pequena
Não souberam avaliar ou tiveram dúvidas para afirmar	4

Nos testes realizados dentro da CAVERNA, a configuração cúbica forneceu melhores resultados, indicando em momentos instantâneos um nível 5 de imersão. O quadrante correto onde situava-se a fonte virtual do som no espaço foi corretamente avaliado por todos que ouviram os trechos de testes dentro da CAVERNA, embora a taxa de acerto quanto à direção correta tenha sido menor que no experimento feito fora da CAVERNA, e também foi maior a percepção de uma fonte “circulante” no ambiente durante trechos da execução dos dois timbres. Com relação à espacialidade ou dimensões do espaço virtual, os avaliadores não souberam dizer como era o espaço.

De uma maneira geral, a qualidade da percepção espacial com Ambisonics se deu melhor fora da CAVERNA do que dentro. Dentro da CAVERNA, os resultados foram melhores na configuração cúbica.

Os resultados mostraram-se dependentes do programa musical, em especial sensíveis aos atributos timbre, altura (*pitch*) e duração. Para os trechos de flauta, que englobaram um universo mais diversificado de gestos, a taxa de acerto da localização foi superior à verificada com a bateria. O tempo que os avaliadores demoraram em tomar uma decisão quanto à localização da bateria foi maior do que para a flauta.

Alguma distorção na formação do plano em passagens de flauta, principalmente na frase decrescente que atravessa uma oitava de cima a baixo, foi percebida circulante no espaço 3D. Este trecho foi preparado com notas dó3 (~262Hz), dó4 (~523Hz) e dó5(~1046Hz), iniciando com uma frase decrescente ligada de dó5 a dó3, trafegando das faixas de frequência alta (>700Hz) e baixa (<300Hz) de referência no formalismo de faixas de frequência do Ambisonics.

Para as baixas freqüências o grau de realismo é menor que para as freqüências médias, em função da técnica de fontes-imagens utilizada, sabidamente boa para modelagem de reflexões especulares (geométrica). A ausência de outros métodos complementares atuando concomitantemente com o de fonte imagem é fonte de degradação no nível de realismo e grau de imersão percebido. A percepção correta da localização dos sons de dentro da CAVERNA em algumas posições é falseada, devido à ausência de tratamento anecóico adequado da mesma.

Em certas posições, “vazamentos” de som através da estrutura da CAVERNA produziam interferência com o campo sonoro esperado. Para altas freqüências, encontramos o ambiente bem difusor, efeito possivelmente criado pela transmissão sonora através das telas. As caixas não respondem a baixas freqüências, inferior a 60Hz. O modelamento correto da propagação sonora, fenômenos modais e difração nas baixas freqüências acrescido de um *sub-woofer* na CAVERNA pode incrementar o grau de realismo para estas freqüências. Esta adição é esperada no futuro.

Um resultado unânime no julgamento de todos foi a capacidade do sistema em produzir um campo sonoro que não parece vir das caixas, isto é, as fontes secundárias permaneceram ocultas, os ouvintes acessaram somente as fontes primárias (virtuais).

5 DISCUSSÃO DOS RESULTADOS

5.1 Introdução

A seguir procedemos a algumas discussões relevantes sobre o sistema desenvolvido, sua implementação, sua aplicabilidade nas diversas finalidades onde a sonificação tridimensional é importante, e sobre os resultados obtidos.

É importante reconhecer que a simulação e reprodução realista de um campo sonoro em ambientes de realidade virtual imersiva é uma meta somente alcançável através de refinamentos e melhoramentos sucessivos, tanto nos modelos utilizados quanto nas técnicas de implementação. Um resultado de alta qualidade é alcançável após iterados ciclos de implementações, testes, e melhoramentos.

5.2 Qualidade espacial percebida

Embora preterida em implementações anteriores de auralização em *CAVE's*, a técnica Ambisonics mostrou-se promissora para aplicações em ambientes audiovisuais do tipo imersivos como a CAVERNA Digital, e, para situações simples de teste, mostrou-se robusta e reproduzível, ainda que resultados no nível de imersão e estabilidade das imagens tenham apresentado variações e artefatos.

Este resultado, de certa forma, era esperado, em função das condições acústicas da CAVERNA, que estão longe de idealmente anecóicas para hospedarem aplicações de auralização 3D (interferência da acústica local e efeitos devido às telas). Dentro da CAVERNA os sons mostram-se com uma qualidade quase metálica, mostrando-se um ambiente reflexivo e difusivo, com atenuação progressiva com o aumento da frequência. Medidas acústicas realizadas na CAVERNA mostram que o tempo de reverberação da sala tem aproximadamente 1s, o que é longo demais para hospedar atividades de auralização sem que haja algum tratamento. A princípio, o caminho mais acertado parece ser uma combinação de tratamento acústico com equalização ativa. As atividades do projeto AUDIENCE prevêm abordar esta questão.

Com relação à percepção da localização das fontes virtuais, é interessante notar que, para o material sonoro utilizado nos testes, a direção dos objetos sonoros virtuais era

melhor identificada com a ajuda de componentes de baixa frequência, enquanto para alguns componentes de alta frequência o som parecia às vezes circular pelo ambiente. Isso encontra uma provável explicação devido à decodificação da direção (e.g. o azimute do objeto, no plano horizontal) utilizando-se uma aproximação para o cálculo do chamado azimute de localização do vetor velocidade, ou azimute de Makita (GERZON, 1992).

Este procedimento de decodificação da direção utiliza a somatória dos ganhos individuais de cada alto-falante para obter a pressão acústica no ponto de audição e os componentes do vetor velocidade, obtendo-se a partir daí a direção do vetor ganho de velocidade que aponta para o (aparente) azimute do objeto sonoro. Estas quantidades são indicativas da direção de um objeto segundo as teorias de localização pela diferença de fase, que opera melhor nas baixas frequências. Possivelmente uma calibração dos parâmetros de filtros do tipo *shelf* comumente utilizados em codecs Ambisonics possam melhorar estes efeitos.

Com relação às diferenças de qualidade espacial percebida com timbres diferentes, particularmente com relação a taxa de acerto na localização da flauta ser maior do que da bateria, isso se dá por dois motivos: um, como um instrumento composto de vários itens estes irradiam em alturas (tonalidades) diferentes e os artefatos verificados provocavam a separação dos itens, alguns mais dispersos no ambiente ou percebidos vindos de muitos lugares (pratos) e outros mais localizados (bumbo), e também devido a difusibilidade natural da sala virtual, que, vazia e retangular, com paredes com coeficiente de reflexão de 0,9, apresenta mesmo um padrão de reflexos denso, como pode-se ver na Figura 35 que mostra o reflectograma ou as *IR's* em *B-Format* utilizadas no experimento.

5.3 Arquitetura de auralização

Verificamos que a estratégia “dividir para conquistar” embutida na arquitetura em camadas para produção sonora espacial mostrou-se um caminho correto para abordar problemas e propor soluções numa área tão extensa, complexa e multidisciplinar quanto áudio e acústica. Em especial, verificamos que uma especificação definida e conhecida dos sinais na interface entre uma camada que se dedica a resolver um

problema e outra, destinada a outro problema, é extremamente útil para ampliar as alternativas em termos de ferramentas passíveis de uso. Isto torna o trabalho de desenvolvimento mais robusto e rápido.

Por exemplo, tendo-se um conjunto de respostas impulsivas bem conhecidas e modeladas referindo-se a atributos bem definidos da camada 2 (simulação acústica) como saídas desta camada, podemos utilizar qualquer ferramenta compatível com a camada 3 para ler e processar estes dados. Assim, pudemos utilizar várias ferramentas de análise e processamento de *IR's* parametrizadas para montar nossa representação espacial (codificação espacial), e manter a compatibilidade e validade do sinal por toda a cadeia. A arquitetura ainda poderá se mostrar muito mais útil no caso de desejarmos a integração de ferramentas diversas disponíveis para a camada 2, e mesmo utilizar *API's* e bibliotecas proprietárias como ferramenta empregada por exemplo na camada 4 (sonorização) como reproduutor multicanal.

Com relação ao estudo do tema auralização em RVI verificamos que existem muitos desenvolvimentos, eles não estão conectados, a maior parte das implementações utilizam técnicas bi-aurais ou de panoramas por amplitude (SVPA), e desenvolveram solução de áudio próprias, implementando algumas técnicas, bibliotecas, e integrando os módulos de processamento de áudio com os outros subsistemas de RVI, como o *DIVA* (SAVIOJA, 1999) e o *blue-c* (NAEF, 2002) e o *Wave Engine* de (OGI, 2003). O CSS (ECKEL, 1998) por exemplo utiliza o Avango, que usa MAX/FTS do IRCAM. Parcialmente o CSS também usa as bibliotecas do SPAT/IRCAM que são implementada em MAX/FTS, rodando sobre SGI. Os demais utilizam soluções de áudio não otimizadas para reprodução 2D ou 3D.

Não identificamos *a priori* uma tendência ou preferência clara por um pacote, uma biblioteca, uma arquitetura de *hardware* ou de uma técnica específica. Isso vem a confirmar uma grande diversidade da área e a falta de uma arquitetura integradora. Entretanto, a julgar pela disponibilização progressiva de maior capacidade computacional nos sistemas e uma maior exigência quanto à qualidade de áudio espacial, as indicações mais promissoras em termos de técnicas para áudio 3D em ambientes audiovisuais apontam para o uso de *WFS* e novos ambientes imersivos priorizando a auralização através de áudio multicanal e níveis elevados de imersão.

5.4 Limites técnicos

O processamento do áudio num sistema de sonorização é em essência um processo serial. Os limites técnicos existentes em uma etapa deixam sua assinatura na mídia que entra no próximo bloco do sistema. A propagação de erros, artefatos indesejados, ou alterações nos atributos de qualidade do áudio, como distorção harmônica, nível de ruído de fundo, presença de componentes de frequências espúrias, alterações de fase ou atenuações de amplitude dos sinais devido a parâmetros exógenos ao processo de auralização podem ser críticos e desviar o resultado final do esperado.

Por exemplo, os parâmetros de relaxamento e curvas de resposta de frequência e fase de alto-falantes novos se modificam com o tempo de uso, quando se diz que os mesmos “se amaciam”. Isto influencia na resposta de frequência do conjunto falante-sala acústica, e portanto na equalização do resultado final. Alterações na fase dos sinais devido ao amaciamento da resposta acústica dos transdutores podem ser críticos quando o alvo do processo principal é o cancelamento de componentes, causando artefatos críticos e inaceitáveis.

5.4.1 Problemas desafiadores

Acústica local

Semelhante a (ECKEL, 1998) verificamos problemas críticos com relação à montagem da infra-estrutura física de áudio (em especial o posicionamento dos alto-falantes num ambiente restritivo) e a necessária compensação dos efeitos produzidos pelas telas e pelo ambiente reflexivo da CAVERNA. A filtragem dependente da frequência das telas e a superposição da acústica local são dois dos maiores problemas a serem solucionados para CAVERNAs em geral, independente do tipo de auralização, para melhores resultados.

Outro trabalho que cita o desenvolvimento de filtros de compensação para reduzir a influência da atenuação da tela é (OGI, 2003). No nosso caso, utilizando-se uma técnica de simulação de campo sonoro como Ambisonics, estes efeitos são mais pronunciados, no sentido em que impedem em várias circunstâncias (dependendo do programa sonoro) a correta síntese do campo sonoro e a formação estável dos

indicadores auditivos. A técnica também mostra-se mais propensa a produzir artefatos espectrais nos casos de má-síntese.

Modelamento acústico

Os resultados obtidos superaram a expectativa em certos aspectos, devido à precisão do método de fonte-imagem para a sala de testes trabalhada. O simulador acústico requer melhorias para modelar efetivamente a dependência da frequência, como por exemplo permitir a produção de *IR's* com cobertura de larga faixa de frequências, com custo computacional otimizado.

Entretanto, para geometrias mais complexas, será inevitável integrar outros métodos de simulação acústica, que modelem de baixas a altas frequências, e que exibam ótimo desempenho computacional, para permitir sempre que possível auralização em tempo real.

Usualmente as técnicas mais precisas de modelamento acústico baseadas nos métodos de ondas e geométricos são utilizadas nos programas para projeto acústico, mas não têm muito apelo para sistemas interativos, devido ao custo computacional elevado e dificuldades de atendimento a requisitos estritos de tempo real.

Uma infra-estrutura de computação baseada em aglomerado computacionais pode viabilizar o uso de várias técnicas de modelamento baseada na solução de ondas e a integração de técnicas antes impraticáveis em aplicações de tempo real em sistemas monoprocessados. As implementações podem ser integralmente baseadas numa arquitetura via *software*, o que permite e estimula uma maior integração dos subsistemas de áudio com os subsistemas gráficos e de gerenciamento.

O custo computacional de modelamento de ambientes com precisão e realismo ainda é muito alto, mas as técnicas vão se aprimorando, técnicas diversas para modelamento acústico, inclusive as estatísticas contribuem para a simulação arquitetural acústica, técnicas de radiosidade ou luminosidade, modelamento para geometrias especiais, técnicas de traçado de cones/feixes, e outras.

Eficácia da auralização

A localização espacial auditiva ainda não é completamente compreendida. Há pelo menos meia dúzia de indicações auditivas de localização (atraso interaural, sombra da cabeça, ecos dos ombros, efeitos da orelha, etc.) sem ter sua importância relativa conhecida, e ainda há os fatores psicoacústicos ou (melhor dizendo) neuroacústicos (uma vez que a audição e sua cognição é mais neural que mecânica/física). Novas possibilidades com relação a parametrização da síntese e da produção de campos sonoros poderão ainda mostrar-se úteis em particular para a simplificação de sistemas complexos que envolvam centenas de canais, como é o caso de *WFS*.

Engenharia de *software*

Um cenário relativamente pouco integrado existente nos sistemas de áudio espacial foi abordado neste trabalho. Verificamos outros pesquisadores tecendo comentários sobre as barreiras dos sistemas vigentes. BURGESS (1993) diz que um sistema de controle para áudio espacial deve ter pelo menos as seguintes características para ser útil em servidores de áudio de propósito múltiplo: interfaces de controle de rede, suporte para clientes múltiplos, suporte para reprodução de múltiplos canais simultâneos, priorização de requisições para o caso de recursos limitados, mecanismos para sincronização, mecanismos para mobilidade de fontes sonoras interativas e em tempo real, e idealmente um renderizador acústico para modelar a audição de ambientes automaticamente.

A esta lista, adiciono o uso de técnicas diferentes, de diferentes complexidades, integradas para superposição de efeitos e soma de esforços, como por exemplo um *reverb* comercial para sintetizar a cauda reverberante com menor custo computacional. O projeto AUDIENCE prevê estudos e implementações em comunicação entre módulos (em nós diferentes), suporte a ouvintes múltiplos e interatividade (pelo paradigma de auralização multicanal), e arquitetura flexível e escalável, para permitir a integração de técnicas e métodos diversos, em qualquer camada do processo.

6 CONCLUSÕES

Aparentemente numa análise inicial ou simplista o número de trabalhos, contribuições e tecnologias em áudio espacial pode parecer pequeno, mas ledão engano. O tema floresce, na literatura há inúmeras provas de contribuições espalhadas no tempo, muitas vezes obscurecidas dentro de anais de eventos cuja temática principal não era áudio ou música ou tecnologia de áudio, mas tratavam de ocupar um espaço secundário oferecido ao áudio.

Os trabalhos teóricos, ferramentas e implementações práticas existentes são muitos, e eventualmente tão diversos, que não conseguem se tocar num escopo mais amplo, porque grande parte abordam problemas pontuais ou específicos quanto ao processamento de áudio em sistemas preponderantemente orientados à temática visual ou de sistemas. O que identificamos é que o problema preponderante não é a falta de trabalhos, mas como eles se conectam ou não se conectam, pela falta de uma estrutura sistêmica mais abstrata ou uma estrutura de referência global, através da qual se conectar.

Esta falta de coesão pode ser atacada de maneira sistemática através da proposição de padrões. Essa estratégia mostrou-se extremamente vitoriosa no caso do sucesso do MPEG e suas tecnologias codificadoras de mídias. Padrões aceleram e facilitam não somente a inserção de produtos no mercado, por parte da indústria, mas também estimulam a colaboração entre pesquisadores individuais, instituições, colaboradores pontuais, programadores, cientistas e profissionais de áreas multidisciplinares afetadas, etc.

Um modelo de referência para produção de áudio espacial seria útil não somente para montar uma árvore de conectividade de tecnologias para áudio 2D e 3D, mas também para encampar a produção de áudio como um todo, e conectar sua história e ferramentas. Creemos contribuir com um passo avançado neste sentido ao defender um modelo de referência para produção de áudio espacial hierárquico, orientado a 4 camadas, aberto, onde principalmente se tornem abertos e acessíveis os dados que trafegam entre camadas, em outras palavras, identificando-se e fornecendo acesso aos sinais da interface entre camadas.

6.1 Conclusões gerais

Estudou-se uma seleção de técnicas e métodos para auralização aplicáveis a ambientes do tipo CAVE, e realizou-se um estudo de caso utilizando-se uma delas como técnica de codificação e reprodução de áudio 2D/3D acoplada a uma aplicação visual de realidade virtual na CAVERNA Digital da Escola Politécnica da USP.

O trabalho envolveu a montagem de toda uma infra-estrutura de *hardware*, sistema de cabeamento e matrizes de alto-falantes, instalação de *softwares*, desenvolvimento de programas para simulação acústica, para codificação espacial sonora, e reprodução multicanal final.

No desenrolar deste trabalho identificamos os processos de maior dificuldade de implementação, os pontos frágeis e robustos associados ao modelo empregado. Identificamos limitações e restrições pertinentes, como os problemas e contingências associados a requisitos de tempo real.

Os resultados obtidos são um primeiro passo para se avaliar a qualidade e o grau de realismo que se pode obter com a arquitetura de produção sonora espacial proposta. Partindo-se de um modelo acústico simplificado, avaliamos qual grau mínimo de complexidade e completude o modelo acústico deveria realmente atender para permitir uma simulação satisfatória de um ambiente físico, e quais as dificuldades para escalonar e distribuir a carga computacional dos diversos módulos operando em tempo real.

Conjuntamente com a visualização virtual, a auralização forma a base para uma nova tecnologia para simulação de salas. KLEINER (1993) entretanto ressalta que a habilidade da tecnologia em reproduzir precisamente a impressão acústica das características audíveis de uma sala ainda permanece por ser verificada, limitando a credibilidade da auralização como ferramenta de projeto (*design*). Isso se deve claramente à grande quantidade de fatores e variáveis a se considerar e controlar durante o processo, uma vez que, da fonte ao ouvinte, o som pode passar por uma multiplicidade de reflexões, transmissões e caminhos de difração.

Uma simulação precisa é sempre computacionalmente intensa. KLEINER (1993) sugere que a combinação de auralização com reprodução trans-aural, equalização da sala e controle ativo de ruído poderiam expandir as aplicações desta tecnologia bem além dos laboratórios. Neste sentido, a equalização da sala, de fato, também é prevista aqui, como forma de se compensar a interferência da tela da CAVERNA bem como minimizar os efeitos e interferência da acústica local sobre o campo sonoro produzido. Se bem sucedida, a dereverberação da área interna da CAVERNA poderá tornar a tela acusticamente “invisível” aos ouvintes.

A montagem da infra-estrutura necessária (aquisição de equipamentos, montagem, instalação e configuração) foi iniciada de maneira limitada em 2004, mas só concluída em meados da 1ª fase do AUDIENCE. Uma configuração para testes só foi possível construir aliando-se aquisições pontuais (placa multicanal) e a cessão de equipamentos por parte de colaboradores empresariais, interessados em participar do projeto de áudio imersivo da CAVERNA Digital.⁶⁸

A auralização multicanal mostra-se uma ferramenta poderosa e adequada para ampliar significativamente a experiência de imersão nos ambientes e sistemas audiovisuais. Em particular, verificamos que em ambientes de realidade virtual baseados em multiprojeção estereoscópica, como a CAVERNA Digital, o emprego da auralização para projetar campos sonoros em uma região de audição ampliada pode proporcionar o equivalente auditivo à estereoscopia a um pequeno número de ouvintes, sem a necessidade do uso de fones de ouvido.

A proposição de uma técnica multicanal, que faça uso de alto-falantes ao invés de fones de ouvido, impõe de fato alguns desafios importantes no caso da auralização de *CAVE's*, como por exemplo as restrições de posicionamento dos falantes como também a maior dificuldade de se produzir um *sweet-spot* satisfatório, sem a introdução de muitos artefatos de áudio. Entretanto, para os propósitos de realidade imersiva, o paradigma de alto-falantes atende plenamente à primeira necessidade de

⁶⁸ Projeto AUDIENCE (*AUDio Immersion Experience by Computer Emulation*) em <http://www.lsi.usp.br/interativos/nem/audience/>

liberdade e independência de dispositivos adicionais junto ao corpo, evitando-se o uso de fones de ouvido, além de que é adequado à premissa de multiusuários dentro do mesmo ambiente imersivo.

As técnicas de auralização são as mais avançadas ferramentas que poderiam propiciar uma simulação correta de ambientes acústicos, com o grau de sofisticação e precisão compatíveis com os critérios utilizados no mundo visual. Elas encontram aplicações na engenharia, ciências biológicas, entretenimento e serviços multimídia, embora não tenham sido ainda exploradas em profundidade, não só porque muitas estão em estágios iniciais de desenvolvimento (como *WFS*) mas também devido a limitações na capacidade de processamento das plataformas reprodutoras e limitações no número de canais de saída das mesmas, além, é claro, da viabilidade econômica em se utilizar múltiplos alto-falantes no caso de auralização multicanal.

A técnica de Ambisonics mostrou-se a mais adequada no momento para um experimento piloto de auralização integrado a um aplicativo de RVI na CAVERNA Digital. Ao mesmo tempo em que permite o mapeamento e produção de um campo sonoro 2D/3D ela apresenta uma formulação matemática elegante e custo computacional relativamente baixo. Os pontos críticos no seu uso, entretanto, estão na avaliação da melhor configuração possível para os alto-falantes que garantam uma reconstrução 3D satisfatória enquanto minimizando o alcance de artefatos e distorções, principalmente devido aos problemas na acústica local e na precisão do cancelamento (ou compensação) dos efeitos (distorções e difusão) provocado pela transmissão através das telas da CAVERNA.

Verificamos que os obstáculos que impedem hoje a aceleração da massificação e disseminação do uso da espacialização sonora em larga escala são transitórios e naturais, senão recorrentes, no amadurecimento tecnológico, industrial e cultural de qualquer nova tecnologia de áudio que já tenha sido introduzida no passado, não constituindo-se de limites intangíveis, especiais ou de natureza desconhecida, mas limitando-se a tratáveis questões a longo prazo relacionadas ao amadurecimento do modelo de negócios, da padronização na manufatura de produtos, e na afirmação de uma cultura consumidora dos inúmeros produtos e aplicações que estas tecnologias poderão incubar.

Sistemas de realidade virtual baseados em multiprojção estereoscópica e áudio imersivo utilizando o paradigma de CAVERNA Digital são ferramentas importantes utilizáveis em várias aplicações de engenharia, em investigações científicas avançadas, em entretenimento, e na prototipação de serviços inovativos e sistemas multimídia, como por exemplo a investigação de modelos para a TV do futuro.

Este trabalho apresentou diversas inovações e propostas para a sonificação avançada em *CAVE's*, e demonstrou a viabilidade e validade da síntese de campos sonoros para fomentar uma geração mais completa de aplicações de realidade virtual imersiva, abrindo novas possibilidades de exploração do som em sistemas multimídia.

A utilização de aglomerados de computadores convencionais para a construção de simuladores acústicos e geradores de campos sonoros tridimensionais de alta qualidade mostra-se uma alternativa viável para realizar a investigação científica do potencial de aplicação futuro destes sistemas, de uma forma como nenhuma outra solução poderia atender no mercado.

Finalmente, este trabalho apresentou várias técnicas possíveis de sonificação multicanal e espacial, e permitiu delinear algumas tendências tecnológicas e de aplicações neste segmento e nos segmentos onde estas tecnologias demonstram impacto.

6.2 Inovações

Citamos as contribuições mais importantes resultantes deste trabalho:

- Primeira CAVERNA (*CAVE*) orientada para auralização com campos sonoros 2D/3D a implementar uma infra-estrutura para áudio multicanal e desenvolver uma solução geral para Ambisonics integrável a um navegador de realidade virtual no mundo
- Arquitetura hierárquica modular e aberta (orientada a camadas funcionais) para produção de áudio espacial
- Simulador acústico e codificador espacial, com entrada de dados descritos em *X3D* e saída em *B-Format*

- Uso de *cluster computing* para simulação acústica e auralização 3D
- Uso de sintaxe de descrição de cena acústica em *X3D*
- Proposição de uma escala para medir o grau de imersão sonora.

6.3 Trabalhos Futuros

Em verdade o universo de trabalhos futuros agigantou-se diante das inúmeras possibilidades de evolução, melhorias e aperfeiçoamentos abertos por este trabalho.

6.3.1 Configurações avançadas para alto-falantes

Anel octogonal virtual

Na configuração octogonal testada na CAVERNA os alto-falantes foram posicionados na frente dos cones de luz, o que é uma situação proibitiva no uso normal.

Estamos implementando uma configuração octogonal virtual (horizontal, planar) para contornar esta limitação. Nesta, quatro alto-falantes são posicionados mais atrás, fora do cone de projeção, como sugere a Figura 46. Para estes temos de compensar o atraso da chegada do som e a atenuação devido a um percurso maior que os outros quatro até o centro da CAVERNA. Neste teste avaliaremos as condições de auralização considerando que o padrão de irradiação e dispersão das fontes secundárias (alto-falantes) não mais é uniforme, e que uma compensação possa ser necessária para corrigir eventual degradação do campo.

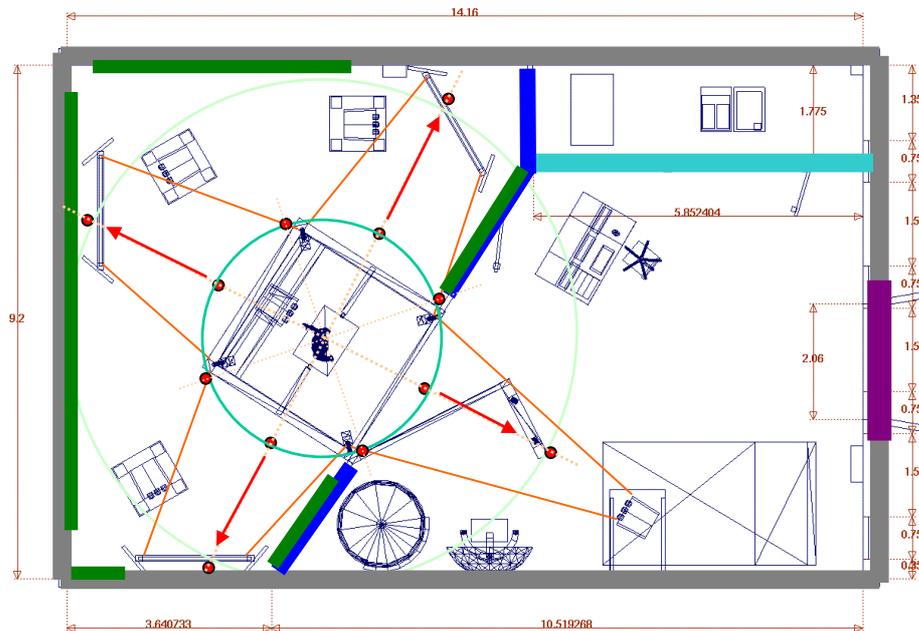


Figura 46 – Matriz de alto-falantes em configuração anel octogonal virtual (2D)

Para a segunda fase do projeto AUDIENCE pretende-se prover uma instalação flexível com 16 alto-falantes na CAVERNA Digital (FARIA, 2004), para experimentos com Ambisonics de até 3ª ordem.

Pretende-se avaliar diversas configurações possíveis para os alto-falantes ao redor da CAVERNA. Sabemos antemão que as configurações simétricas e geometricamente regulares são preferíveis pela maior facilidade na seleção das equações de decodificação. Por outro lado, sabemos também de algumas configurações regulares reconhecidas por alguns autores como ruins do ponto de vista de resultado acústico, como por exemplo a cúbica (VAN ASSELT, 2002).

Um exemplo de posicionamento previsto para investigação é mostrado na Figura 47.

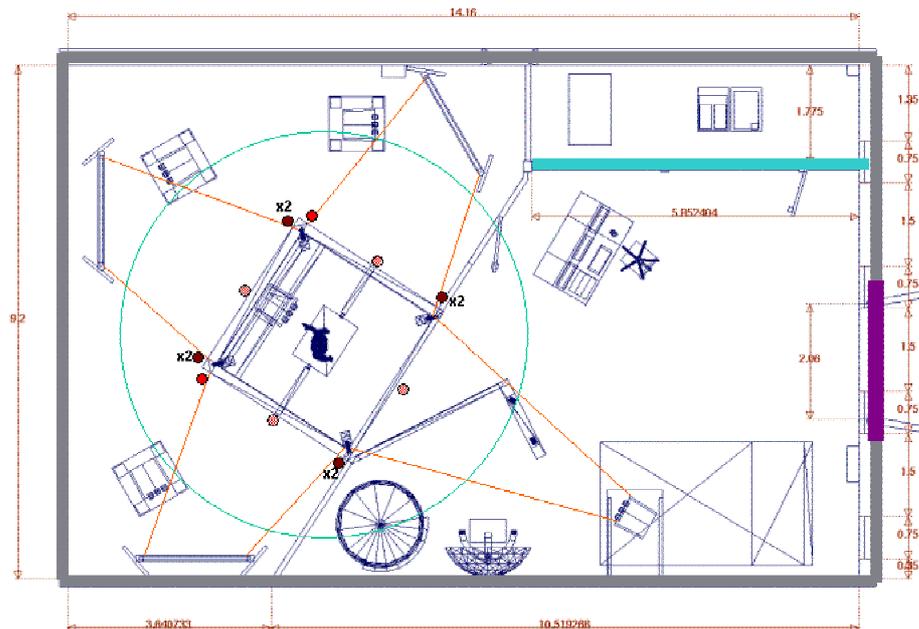


Figura 47 – Configuração não-regular prevista para avaliação futura

Para se evitar os indesejáveis efeitos das telas, uma das primeiras propostas é investigar configurações posicionando os alto-falantes nas partes superiores, com visada direta ao centro da CAVERNA.

Equalização e calibração

A última operação de processamento importante antes de gerar a saída de cada canal final seria a aplicação de um filtro inverso para cancelar o efeito do sistema alto-falante-ambiente (tela) até o centro da CAVERNA (H^{-1} na Figura 42). A rigor, este filtro inverso corrige a distorção e filtragem introduzida pelo transdutor e pela tela no percurso acústico até o centro do eixo de coordenadas Ambisonics. Este filtro pode ser *FIR* ou *IIR*, e é obtido através do levantamento da resposta impulsiva da função de transferência alto-falante/tela, medida com um microfone adequado. Para implementar a filtragem será utilizado um *software* de convolução que permita aplicar um filtro de tamanho adequado e executado em paralelo em todos os canais, de forma a garantir o sincronismo de um mesmo atraso igual por todos. Várias implementações de máquinas e algoritmos de convolução poderão ser utilizadas. Em especial nos referenciamos às técnicas citadas em (FARINA, 1998), e às ferramentas

computacionais *BruteFIR* e *ALMUS*, otimizadas para convoluções em tempo real (TORGER, 2004).

KIRKEBY (1999) e FARINA (1998) abordam a teoria de equalização e a aplicação de técnicas de filtragem inversa para cancelamento de não-linearidades de transdutores e dereverberação da acústica local.

Finalmente, para um correto acoplamento das impressões perceptuais sonora e visual, uma etapa adicional de calibração é necessária, para principalmente adequar a intensidade sonora com respeito ao esperado para uma fonte virtual O_j situada a uma distância L do centro da área de audição. Adicionalmente, uma sintonia fina no posicionamento e na direção dos alto-falantes e uma equalização global no caminho de cada canal completam a calibração. Estas tarefas são esperadas para um momento posterior, a partir do momento em que algumas configurações fixas de alto-falantes sejam definidas.

Tratamento da acústica local

Em tese, o cubo da CAVERNA deve idealmente estar contido num espaço o mais anecóico possível, de forma a reduzir as possíveis interferências de ruídos e sons externos nos campos sonoros simulados.

Além do tratamento ativo de equalização e calibração, um tratamento acústico do ambiente é absolutamente necessário. Dois problemas maiores devem ser abordados: (i) um alto nível de ruído contínuo de máquinas e do ar condicionado, e (ii) a acústica “viva” por trás das telas.

Uma solução inicial paliativa seria aumentar a absorção atrás da CAVERNA utilizando-se placas, espumas e diversos materiais absorventes de baixo custo, aplicando-os nas superfícies da estruturas, e também nas paredes, reforçando a absorção. Uma proposta mais consistente a ser investigada é a aplicação de cortinas acústicas dependuradas na estrutura do galpão, isolando o máximo possível a acústica dos bastidores da acústica interna ao cubo.

Um problema adicional a ser tratado é a falta de isolamento no teto. A CAVERNA foi construída junto ao atual prédio da Administração da EPUSP, tomando-se uma parte de um grande galpão em forma quadrilátera ali existente. O galpão é um espaço volumétrico vazio, de paredes e teto reflexivos e portanto altamente reverberante. Atividades que gerem pressão sonora apreciável ali dentro ressoam dentro da CAVERNA, que ocupa um volume na lateral deste galpão. Um tratamento acústico poderá melhorar o nível de isolamento do galpão adjacente principalmente tratando-se o teto da CAVERNA e a parede lateral comum a ambos os espaços.

A porta de entrada da CAVERNA deverá ser substituída por uma porta acústica. Ajustes menores utilizando-se materiais acústicos simples podem ser aplicados à janela e à porta que dá acesso à sala de máquinas, onde ficam os aglomerados de computadores, sendo uma fonte de ruído contínuo.

6.3.2 Auralização integrada com o PD

A nossa proposta para integração do bloco auralizador global é utilizar o programa *Pure Data (PD)*, de Miller Puckette (PUCKETTE, 2005) como suporte para estabelecer as conexões entre módulos de *software*, e ordenar seqüências de operação. O *PD* eventualmente não representa o melhor compromisso de desempenho para aplicações de auralização finais e críticas, mas se apresenta como uma boa solução integradora de módulos diferentes e excelente plataforma para experimentos, dada a flexibilidade de se alterar e intervir na cadeia de processamento introduzindo-se ou retirando-se blocos (objetos de processamento) e alterando-se a rota dos sinais de áudio. Em outras palavras, como plataforma experimental o *PD* funciona como um painel através do qual inserimos e conectamos os módulos para se processar a auralização.

A sessão de auralização no *PD* começa disparada no nó sonorizador pelo aplicativo Jinx servindo a uma aplicação qualquer de navegação audiovisual na CAVERNA. O Jinx realiza a extração dos parâmetros acústicos da cena, e repassa-os para o bloco *acoustsim*, um objeto do tipo tilde (~) do *PD*, responsável pela simulação acústica utilizando o método de fonte-imagem (primeira versão). O *acoustsim* poderá receber os dados necessários para a simulação de um processo de interface com o

Jinx, ou de um processo de leitura dos dados da cena extraídos e armazenados em arquivo. Atualizações na cena deverão invocar uma respectiva atualização neste processo de leitura de dados, para uma imediata atualização da cena acústica. O `acoustsim` recebe em sua entrada as dimensões do ambiente $[r_l(X,Y,Z)]$, as posições do ouvinte $[r_0(X,Y,Z)]$ e da fonte sonora $[r(X,Y,Z)]$, e os coeficientes de reflexão de cada parede do ambiente $[\beta_1, \dots, \beta_6]$, e produz em sua saída 4 *IR's* em *B-Format* $[h_w, h_x, h_y, h_z]$ relativas à simulação da cena acústica vigente. Uma máquina ou bloco de convolução perfaz a codificação espacial Ambisonics das fontes sonoras, convoluindo-as com as *IR's*. Um mixer atuará quando mais de uma fonte sonora codificada em *B-Format* estiver presente, misturando-as linearmente e cuidando para evitar *clipping*. Um vetor “mixado” final em *B-Format* é enviado ao módulo decodificador (`spatialdec`) para ser decodificado para N canais de saída, referentes à configuração de alto-falantes que se estiver usando.

A máquina de estados básica do sistema de produção sonora espacial inclui um inicia, um estado auralizando (incorporando eventos de atualizações) e finaliza. A geração de sincronismo entre os blocos e sinalização de estados será implementada gradualmente, segundo a arquitetura proposta na Figura 22 e Figura 31.

6.3.3 Descrição avançada de cena

O AudioBIFS do MPEG-4 não é normativo no sentido de especificar as técnicas de espacialização a serem utilizadas pelo terminal ou sistema de auralização, nem tampouco quanto ao grau de imersão desejado. Cremos ser de importância a adição de alguns parâmetros extras para que se possa completar e estender ainda mais o alcance das ferramentas composicionais do MPEG-4 para produzir efeitos seletivos importantes na ativação de certas funcionalidades, ferramentas e/ou circuitos de sonorização das camadas subjacentes do processo de produção sonora no terminal ou plataforma de auralização. Assim, adicionaríamos 1) a passagem do parâmetro que especifica o nível de imersão mínimo e o ideal para auralização da cena, e 2) a indicação de técnicas ou métodos de síntese e simulação acústica, formatos de codificação, ferramentas ou bibliotecas de decodificação e suporte a sonorização final. Desta forma, pode-se especificar elementos de controle atuantes em qualquer camada do sistema de auralização através dos meta-dados da camada composicional.

6.3.4 Modelamento acústico

O procedimento proposto para captura das propriedades direcionais dos sons e modelagem do ambiente acústico é baseado na inferência da propagação dos sons no ambiente através de algoritmos como o de traçado de raios (*ray-tracing*) ou fonte-imagem (*image source*). Os métodos mais tradicionais utilizam esta abordagem, seguida pela codificação da informação temporal e espacial através de respostas impulsivas.

Tais algoritmos entretanto não levam em consideração as características de onda do som propagado, não cobrindo os efeitos de borda, difrações e também os estados estacionários e modais associados à propagação de ondas sonoras no ambiente. Métodos mais sofisticados para o modelamento acústico de espaços fechados e para a simulação de respostas impulsivas foram propostos que permitem levar em consideração estas características complexas das frentes de onda, como por exemplo as técnicas baseadas na análise de propagação do campo de onda (BERKHOUT, 1999). Na segunda fase do projeto AUDIENCE são previstas investigações mais profundas destes métodos, que estão associados à técnica de *Wave Field Synthesis (WFS)* (FARIA, 2004).

6.3.5 Casos de uso progressivos

Após as avaliações iniciais da auralização Ambisonics com cenas audiovisuais simples nos primeiros experimentos, é previsto uma série de casos de uso em complexidade crescente para a montagem de cenas, ampliação de funcionalidades e requisitos de qualidade e imersão. Diversos itens são abordados, a saber:

- sincronismo entre vídeo e áudio,
- continuidade nas transições entre uma situação acústica e outra no tempo seguinte,
- performance em tempo real dos módulos de *software*,
- ampliação da área de audição estável
- redução de artefatos sonoros,
- técnicas para aumentar o grau de imersão percebido,
- melhoria na modelagem da distância dos objetos,

- correção e desempenho da filtragem inversa em cancelar o efeito do sistema caixa/tela, inclusive no que se refere à correção no posicionamento real das caixas em relação à origem de acordo com o que foi parametrizado no sistema,
- calibração correta de diversos parâmetros,
- adição de mais fontes sonoras no ambiente,
- incorporação de efeitos, como Doppler,
- aumento na resolução da grade espacial que mapeia o ambiente (discretização do espaço) e na suavidade de transição de filtros em tempo real,
- uso de auralização distribuída (nós diferentes auralizando em paralelo)
- cobertura de fenômenos acústicos adicionais, como modos de baixa frequência
- uso de configurações não regulares de alto-falantes para auralização,
- integração de técnicas de síntese sonora,
- suporte a diferentes padrões de irradiação de instrumentos, etc.

6.3.6 Tendências gerais

O mercado profissional e de consumo tem introduzido novidades nesta área tradicionalmente em função das demandas da indústria do cinema, e mais recentemente das áreas de multimídia, informática e tecnologia da informação. Como em toda área tecnológica, a oferta de recursos gráficos, computacionais e de áudio constitui grande estímulo para a pesquisa e desenvolvimento de sistemas antes impossíveis. A pré-existência de tecnologia naturalmente impulsiona o desenvolvimento de outras.

Verifica-se uma tendência na utilização cada vez maior de múltiplos canais e programas de áudio nos sistemas de armazenamento mais modernos (e.g., *DVD-Audio*, *SACD – Super Audio CD*). A oferta de maior poder de processamento também nas plataformas reprodutoras finais, como *set-top boxes* e outras (FARIA, 2002a) estimula o emprego de técnicas decodificadoras multimídia computacionalmente mais pesadas, como codecs *MPEG-4*.

As atividades de padronização são fundamentais para a concepção integrada e “universal” de novos produtos, permitem a portabilidade, interoperabilidade, e alavancam a indústria de eletrônica de consumo, sendo ao mesmo tempo estimulantes e necessários ao desenvolvimento tecnológico numa escala de economia global (FARIA, 2002b).

Há uma tendência à utilização de métodos computacionais mais complexos para a produção ou reprodução sonora, como auralizadores multicanais, cuja viabilidade de sucesso na indústria de consumo está associada à maior oferta de poder de processamento digital nas plataformas.

Não há uma arquitetura de referência ou padrão preponderante de áudio em realidade virtual imersiva. OGI (2003) também comenta que embora sistemas para produção acústica como *Avango* e *blue-c* tenham sido desenvolvidos para uso conjunto com projeção imersiva, não há uma tecnologia padronizada ainda estabelecida na área.

O que existe é uma diversidade de propostas e soluções baseadas em *software* ou em arquiteturas mistas de *hardware/software* proprietárias e de alto-desempenho. Mas o assunto é relativamente novo, e há diversas técnicas disponíveis para áudio 3D na sua maioria ainda não integradas a sistemas de realidade virtual, embora algumas tendências possam estar se delineando, as melhores propostas tendendo a apostar nos formatos capazes de simular campos sonoros 3D.

Entre a comunidade científica que se dedica ao tema da auralização e pesquisas aplicadas em acústica há intenso debate sobre a qualidade das diversas técnicas, e também sobre suas aplicações em realidade virtual. Neste contexto, alguns grupos informalmente têm declarado sua intenção em construir *CAVE's* audiovisuais e aplicações baseadas em *WFS*, mas ainda não há resultados publicados (comunicação pessoal, obtida na conferência internacional da *Audio Engineering Society* sobre áudio multicanal em 2003) (AES, 2003).

Para os sistemas de auralização são necessários subsistemas para realizar o mapeamento dos objetos, o modelamento acústico do ambiente virtual, a codificação 3D (formato de representação de áudio 3D intermediário), a transmissão, a

decodificação ou renderização para um conjunto de alto-falantes ou fones de ouvido (*player* ou reproduutor que perfaz a sonorização).

6.3.7 Ampliação da infra-estrutura e Auralização interativa em tempo real

A aquisição de novos equipamentos e a integração de vários subsistemas são previstos paulatinamente, de forma a proporcionar melhorias sucessivas. Os seguintes subsistemas serão ampliados e/ou integrados:

Tabela 11 – Ampliação prevista na infra-estrutura para auralização

Aglomerado de Computadores
Cluster novo de áudio. Auralização distribuída Uso de biblioteca para envio de fluxo de áudio codificado em tempo real pela rede ao nó auralizador. Sistema com 6 nós. Dois nós a princípio rodarão instâncias dos auralizadores para diversos objetos. Um nó será utilizado para hospedar o sonorizador (mixer, decodificador e filtragem inversa para cada caixa).
Interfaces de áudio digital
Novas interfaces de áudio digital (Delta 1010 ou novas placas de 16 canais) Nó sonorizador terá até 2 placas/módulos multicanais de 8 ou 16 saídas.
Hierarquia de sincronismo
Implementação completa da hierarquia de sincronismo em áudio para cobrir o caso de auralização distribuída e sincronismo com subsistemas de gerenciamento e vídeo/gráfico.
Dispositivo para captura de gestos/posicionamento
<i>Flock of Birds</i> (integração de sistema existente) Sistema de rastreamento e captura de posição do usuário na CAVERNA, não utilizado nos testes de audição e validação de auralização multicanal.
Dispositivo para rastreamento e controle de navegação em RVI
“joystick”, “mouse” e <i>Flock of Birds</i> Uso previsto para fase II do AUDIENCE, para interação do usuário com mundo virtual, para possibilitar navegação da “área de audição” aumentada pelo cenário (cena audiovisual).
Sistema de gerenciamento integrado
Integração do modelo de auralizador construído sobre o <i>PD</i> com subsistema de gerenciamento da CAVERNA Digital. Integração do Jinx e outras ferramentas (como o <i>Glass</i>) ao subsistema de auralização.

6.3.8 Sistema flexível para auralização multiformato

Um trabalho futuro desejado é o desenvolvimento de um sistema de gerenciamento exclusivamente voltado para sonificação espacial em *CAVE's* com objetivos semelhantes ao sistema de gerenciamento SIRIUS da CAVERNA. O SIRIUS hoje é

uma aplicação Java, portátil para sistemas portáteis de mão (*palm-top*) capaz de realizar o gerenciamento e monitoramento dos subsistemas da CAVERNA.

Um sistema de gerenciamento para sonificação espacial poderia estar sempre funcionando (*stand-by*) aguardando por alguma aplicação solicitando acesso aos sistemas de áudio (simulação, processamento, codificação/decodificação multiformato, sonorização multicanal), determinar qual o formato desejado de áudio, alocar os recursos do sistema de forma transparente à aplicação, e repassar o comando de uso dos recursos para a aplicação, enquanto monitora a execução e estabilidade do sistema, para agir em casos de erro ou mau funcionamento (tratamento de erros).

Neste caso, uma dada configuração de alto-falantes “mais universal” e capaz de ser utilizada por diversos formatos de áudio 3D deveria estar sempre disponível, para que o gerenciador simplesmente acione ou não os alto-falantes desejados na configuração de irradiação desejada. Na impossibilidade de atender a uma reconfiguração via *software*, o sistema poderá emitir a sinalização apropriada por intervenção do usuário. Adicionalmente, tarefas de autocalibração, localização e medida de distância/posicionamento correto dos alto-falantes e cálculo de filtros para *setups* específicos deveriam ser automatizados no sistema, para agilizar o preparo e entrega do sistema pronto para uso por uma dada técnica.

6.3.9 Sistemas para síntese acústica dinâmica

Alguns autores sugerem que a superimposição de efeitos espaciais nas gravações estéreo, como reverberações e reflexões sugerindo um ambiente acústico, são indesejáveis, podendo mesmo chegar a incômodos (FARINA, 2001). De fato, isto é um procedimento virtualmente presente em qualquer gravação fonográfica disponível no mercado, bem como faz parte do tradicional trabalho do produtor e do “engenheiro de som” em masterizar o trabalho final, registrando uma assinatura acústica. Entretanto há vantagens interessantes em oferecer um material sonoro livre de uma assinatura acústica específica, deixando essa possibilidade para o usuário final.

Do ponto de vista comercial, quebrar com esta tradição pode mostrar-se bastante impopular num primeiro instante, e mesmo inviável, haja visto que as plataformas de reprodução finais não estão ainda preparadas para tal capacidade de edição acústica. A superimposição de um ambiente acústico qualquer à gravação hoje é uma realidade, porém ainda requer pesquisa e desenvolvimento adicionais para se propor um sistema genérico, simples e de baixo custo tanto computacional quanto financeiro para que se popularize. É nosso intuito ao conduzir investigações dessa natureza em ambiente de pesquisa acadêmico buscar formas de viabilizar esta passagem entre o paradigma de consumo do som estático (i.e., registrado e transmitido com uma assinatura acústica imutável) e o paradigma de consumo do som dinâmico, i.e., contendo os componentes instrumentais e fontes sonoras livres numa cena neutra, passíveis de serem combinados no palco sonoro que se deseja e no ambiente acústico que se deseja.

6.3.10 Protocolos e *hardware* para distribuição de áudio multicanal

Para a terceira fase do projeto AUDIENCE, prevê-se a necessidade de propor e desenvolver protótipos de matrizes de transdutores de som de alta densidade e resposta de frequência plana numa larga faixa cobrindo das baixas frequências até aproximadamente 10kHz para emprego da técnica *WFS* na CAVERNA Digital. Não há monitores profissionais ou caixas acústicas multicanais disponíveis no mercado que possam ser montadas lado a lado formando, em diversas geometrias desejadas, uma matriz densa de cones de alto-falantes. Ainda, não existe nenhum sistema de caixas de som padronizado para instalação em *CAVE's* para uso em *WFS*.

Adicionalmente, matrizes densas de alto-falantes além de necessitarem de um projeto acústico que permita o menor tamanho possível do invólucro associado a um excelente desempenho em baixas frequências, também necessitam de um sistema de distribuição de áudio multicanal que possa transmitir desde os módulos de som até as caixas todos os canais dentro de um conceito de cabeamento integrado.

Uma boa maneira de fazer isso seria utilizar um protocolo de distribuição de áudio multicanal digital de alta resolução (192kHz/24 ou 32-bit), como por exemplo o

AES10 (*MADI – Multichannel Audio Digital Interface*) que permite até 56 canais PCM (*Pulse Coded Modulation*) transmitidos serialmente em cabeamento elétrico terminado por conector tipo *BNC* até 50m de distância (POHLMANN, 1995), ou o protocolo *Cobranet* de distribuição de áudio não comprimido por redes Ethernet convencionais (<http://www.cobranet.com>), ou ainda alguma nova proposta melhor adaptada às demandas e restrições de aplicações avançadas como *WFS*, e que combinem diversas vantagens presentes em diversos sistemas de codificação e distribuição de áudio multicanal, como:

- a) empacotamento digital com alta compressão sem perdas,
- b) distribuição em meio elétrico balanceado ou par trançado simples, utilizando tecnologias como *LVDS (low voltage differential signaling)* capazes de alta velocidade, *jitter* reduzido e baixa latência,
- c) conectorização final padronizada e simples, como por exemplo utilizando RJ45 como o *Cobranet*,
- d) um receptor/distribuidor dos canais digitais para cada conversor D/A embutido na matriz de alto-falantes.

A matriz de alto-falantes neste caso deveria ser dotada de conversores D/A e amplificada.

7 ANEXOS

Anexo I

Código do simulador acústico para fonte-imagem gerando *B-Format* (FORTRAN)

```

program sroom2bformat
c PGM sroom-to-B-Format
c Regis Faria version 1.3 (2005)
c Adapted from Allen (1979) e hb's version of the program
c subroutine to calculate four room impulse responses (hw,hx,hy,hz)
c r=vector radius to receiver in sample periods
c r0=vector radius to source in sample periods
c rl=vector of box dimensions in sample periods
c beta= vector of six wall reflection coefs (0<beta<=1)
c ht= impulse resp array
c npts= # of points of ht to be computed
c zero delay is in ht(1)
c
integer npts
dimension r(3),r0(3),rl(3),beta(2,3),ht(8820,4)
a=44100/340
r(1)=7.5*a
r(2)=14*a
r(3)=1.75*a
r0(1)=3*a
r0(2)=14*a
r0(3)=a
rl(1)=15*a
rl(2)=20*a
rl(3)=8*a
beta(1,1)=.9
beta(1,2)=.9
beta(1,3)=.7
beta(2,1)=.9
beta(2,2)=.9
beta(2,3)=.7
npts=8820
call sroom(r,r0,rl,beta,ht,npts)
open (1,file='htwp2.raw')
open (2,file='htxp2.raw')
open (3,file='htyp2.raw')
open (4,file='htzp2.raw')
write(1,2)(ht(l,1), l=1,npts)
write(2,2)(ht(j,2), j=1,npts)
write(3,2)(ht(k,3), k=1,npts)
write(4,2)(ht(l,4), l=1,npts)
20 format(F15.14,/)
close(1)
close(2)
close(3)
close(4)
stop
end

subroutine sroom(r,r0,rl,beta,ht,npts)
implicit integer (i-n)

```

Anexo II

Cena de Testes em X3D

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE X3D PUBLIC "http://www.web3d.org/specifications/x3d-3.0.dtd"
    "file:///www.web3d.org/TaskGroups/x3d/translation/x3d-3.0.dtd">
<X3D>
  <Scene>
    <AcousticScene DEF="auditorio" center="0 0 0"
      coefref=".7 .9 .9 .9 .7"
      freqref="1000 1000 1000 1000 1000 1000" size="20 15 8"/>
    <Transform DEF="piso" center="7.5 0.05 -10">
      <Shape>
        <Box size="15 0.1 20"/>
        <Appearance>
          <Material diffuseColor="1 1 1"/>
        </Appearance>
      </Shape>
      <Sound>
        <AcousticMaterial coefref=".7" freqref="1000"/>
      </Sound>
    </Transform>
    <Transform DEF="paredeX_fundo" translation="0 4 10">
      <Shape>
        <Box size="15 8 0.1"/>
        <Appearance>
          <Material diffuseColor="1 1 1"/>
        </Appearance>
      </Shape>
    </Transform>
    <Transform DEF="paredeOX_palco" translation="0 4 -10">
      <Shape>
        <Box size="15 8 0.1"/>
        <Appearance>
          <Material diffuseColor="1 1 1"/>
        </Appearance>
      </Shape>
    </Transform>
    <Transform DEF="parede_Y" translation="-7.5 4 0">
      <Shape>
        <Box size="0.1 8 20"/>
        <Appearance>
          <Material diffuseColor="1 1 1" transparency="0.6"/>
        </Appearance>
      </Shape>
    </Transform>
    <Transform DEF="parede_OY" translation="7.5 4 0">
      <Shape>
        <Box size="0.1 8 20"/>
        <Appearance>
          <Material diffuseColor="1 1 1"/>
        </Appearance>
      </Shape>
    </Transform>
    <Transform DEF="Teto" center="7.5 0.05 10">
      <Shape>
        <Box size="15 0.1 20"/>

```

```
<Appearance>
  <Material diffuseColor="1 1 1" transparency="0.6"/>
</Appearance>
</Shape>
</Transform>
<Transform DEF="Ouvinte" translation="0 1.8 -4">
  <Shape>
    <Sphere radius=".8"/>
    <Appearance>
      <Material diffuseColor="1 0 0"/>
    </Appearance>
  </Shape>
</Transform>
<Transform DEF="ZeroZeroZero" translation="-7.5 -0.05 10">
  <Shape>
    <Sphere radius=".1"/>
    <Appearance>
      <Material diffuseColor="0 1 0"/>
    </Appearance>
  </Shape>
</Transform>
<Transform DEF="fonte_esq" translation="-4.5 1 -4">
  <Shape>
    <Sphere radius=".5"/>
    <Appearance>
      <Material diffuseColor="0 0 1"/>
    </Appearance>
  </Shape>
  <Sound DEF="instrumento">
    <AudioClip loop="true" url="instrumento.wav"/>
  </Sound>
</Transform>
</Scene>
</X3D>
```

Anexo III

Equações de Codificação e Decodificação Ambisonics (para 1ª ordem e 2D)

Ambisonics é uma técnica para reprodução de frentes de ondas planas em ambiente livre utilizando-se conjuntos de alto-falantes para simular eventos ou fontes sonoras. Uma abordagem básica à teoria ilustra a reprodução de uma frente de onda em um ponto de audição através da interferência ou combinação dos sinais dos alto-falantes. Formalmente, assume-se que a fonte sonora esteja à distância suficiente do ouvinte para produzir ondas planas, assim como os alto-falantes, que numa situação artificial simulam a mesma frente de onda produzida pela fonte virtual.

Uma comparação da situação real (figura superior) e da situação artificial (figura inferior) é utilizada para derivar os sinais que cada alto-falante deverá tocar.

Uma onda plana da fonte sonora até o ponto de audição pode ser descrita por $S_\psi = P_\psi e^{j\vec{k}\vec{r}}$, onde P_ψ é a pressão sonora, $\vec{k} = 2\pi / \lambda$ é o vetor número de onda com a direção de propagação, $\vec{r} = re^{j\phi}$ é o vetor posição do ponto de audição, ψ é o ângulo da frente e ϕ é o ângulo do ouvinte em relação às coordenadas.

Esta onda plana pode ser expressa por meio de uma série de cossenos usando-se as funções cilíndricas de Bessel:

$$S_\psi = P_\psi (J_0(kr) + 2 \sum_{m=1}^{\infty} j^m J_m(kr) \cos m(\phi - \psi)) \quad \text{-- (eq. 1), onde } m \text{ é a ordem do sistema.}$$

Na situação artificial, o sinal resultante produzido no ponto de audição por uma matriz regular de alto-falantes pode ser escrito como uma soma de ondas planas:

$$S = \sum_{n=1}^N P_n J_0(kr) + \sum_{m=1}^{\infty} j^m J_m(kr) (2 \sum_{n=1}^N P_n \cos m(\phi_n - \phi)) \quad \text{-- (eq. 2), onde } \phi_n \text{ é o ângulo do alto-falante } n \text{ no sistema de coordenadas.}$$

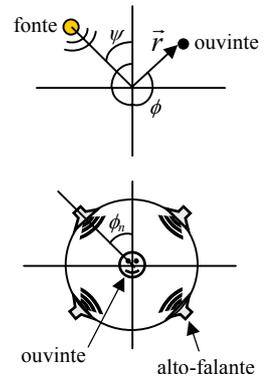
Igualando-se as equações (1) e (2) cria-se a condição para que a matriz de alto-falantes produza o mesmo efeito no ponto de audição que a irradiação real da fonte sonora. Examinando-se os termos nas equações, verificam-se então as seguintes igualdades:

$$X = P_\psi \cos m\psi = \sum_{n=1}^N P_n \cos m\phi_n, \quad Y = P_\psi \sin m\psi = \sum_{n=1}^N P_n \sin m\phi_n, \quad e \quad W = P_\psi = \sum_{n=1}^N P_n.$$

Para um arranjo bidimensional (2D) são necessários $N=2m+1$ sinais (canais) para codificar o campo sonoro. Para um sistema de 1ª ordem ($m=1$) $N=3$ canais são necessários, e obtêm-se as equações codificadoras à pagina 37 (item 2.2.3) sem a elevação (canal Z).

Para a decodificação utiliza-se o número de alto-falantes desejado (mínimo igual a N). A equação de decodificação de 1ª ordem para cada alto-falante n leva em consideração sua

posição em relação às coordenadas, e é dada por $P_n = \frac{1}{N} (W + 2X \cos \phi_n + 2Y \sin \phi_n)$.



8 REFERÊNCIAS

AES INTERNATIONAL CONFERENCE, 24; MULTICHANNEL AUDIO: THE NEW REALITY, Banff, 2003. **Proceedings**. New York: AES, 2003.

AHNERT, W.; TENNHARDT, H-P. Acoustics for auditoriums and concert halls. In: BALLOU, G.M. (Ed.) **Handbook for sound engineers**. 3ed. Woburn, Focal Press, 2002. p.111-155.

ALESIS. **QuadraVerb2: Reference Manual**.1995. Disponível em: <<http://www.alesis.com/index.php?id=56,30,0,0,1,0>> Acesso em: 22 mar. 2005.

ALLEN, J. B.; BERKLEY, D. A. Image method for efficiently simulating small-room acoustics. **Journal of the Acoustical Society of America**, v.65, n.4, Abr. 1979, p.943-950.

AUDIO ENGINEERING SOCIETY. **AES Recommended practice for digital audio engineering – serial transmission format for two-channel linearly represented digital audio data- AES3**. AES, 1992 (rev. 1997).

AUDIO ENGINEERING SOCIETY. **Multichannel surround sound systems and operations – Technical Document AESTD1001.1.01-10**. AES, 2001. 23p. Disponível no site da AES em: <<http://www.aes.org/>> Acesso em: 30 maio 2005.

BAMFORD, J. S. **An analysis of Ambisonics sound systems of first and second order**. 1995. 270p. Dissertação (Mestrado). University of Waterloo, Dept. Physics. Waterloo, Canada, 1995. Disponível em: <<http://audiolab.uwaterloo.ca/~jeffb/thesis/thesis.html>>. Acesso em: 21 nov. 2004.

BARGAR, R.; DAS, S. Virtual Sound Composition for the CAVE. In: 1993 International Computer Music Conference, Tokyo, 1993. **Proceedings** of the addendum. Tokyo: International Computer Music Association, 1993.

BEGAULT, D.R. – **3-D Sound for Virtual Reality and Multimedia**, Academic Press, Cambridge, MA, 1994.

BEGAULT, D.R. Virtual acoustics, aeronautics, and communications. **Journal of the Audio Engineering Society**, v.46, n.6, June 1998. p.520-530.

- BERG, J.; RUMSEY, F. Systematic evaluation of perceived spatial quality. In: AUDIO ENGINEERING SOCIETY CONFERENCE ON MULTICHANNEL AUDIO, 24. Banff, 2003. **Proceedings**. AES, Nova York, 2003. p.184-198.
- BERKHOUT, A. J. et al. A wave field extrapolation approach to acoustical modeling in enclosed spaces. **Journal of the Acoustical Society of America**, v.93, n.5, p.2764-2778, May 1993.
- BERKHOUT, A. J.; DE VRIES, D.; SONKE, J. J. Array technology for acoustic wave field analysis in enclosures. **Journal of the Acoustical Society of America**, v.102, n.5, pt.1, p.2757-2770, Nov. 1997.
- BERKHOUT, A. J.; DE VRIES, D.; VOGEL, P. Acoustic control by wave field synthesis. **Journal of the Acoustical Society of America**, v.93, n.5, p.2764-2778, May 1993.
- BORISH, J. Extension of the image model to arbitrary polyhedra. **Journal of the Acoustical Society of America**, v.75, n.6, June 1984, p.1872-1836.
- BRIX, S. et al. CARROUSO – An European approach to 3D audio. In: AES INTERNATIONAL CONVENTION, 110., Amsterdam, 2001. **Proceedings**. New York: AES, 2001.
- BUCHNER, H.; KELLERMANN, W. An acoustic human-machine interface with multi-channel sound reproduction. In: MMSP'01, WORKSHOP ON MULTIMEDIA SIGNAL PROCESSING, IV. Cannes, 3-5 Out, 2001. **Proceedings**. IEEE, 2001.
- BURGESS, D. A.; VERLINDEN, J. C. **A first experience with spatial audio in a virtual environment**. 1993. Technical Report. Disponível em: <<http://www.cc.gatech.edu/gvu/reports/1993/>>. Acesso em: 06 dez.2004.
- BURGESS, D. A.; VERLINDEN, J. C. An architecture for spatial audio servers. In: VR SYSTEMS FALL'93 CONFERENCE, 1993. **Proceedings**. New York: 1993. Disponível em: <<http://www.cc.gatech.edu/gvu/reports/1993/>>. Acesso em: 06 dez. 2004.
- CAMILO, T.S.; TENENBAUM, R.A.; BENTO COELHO, J.L. Engenharia acústica auxiliada por computador: um método híbrido para simulação de acústica de salas. In: SemEA 2002 – Primeiro Seminário Brasileiro de Engenharia de Áudio, Belo Horizonte, 2002. **Anais**. Belo Horizonte, 2002. Disponível em: <<http://www.cpdee.ufmg.br/~semea/anais/>>. Acesso em: 28 maio 2005.

CHÉENNE, D.J. Modeling and auralization. In: BALLOU, G.M. (Ed.) **Handbook for sound engineers**. 3ed. Woburn, Focal Press, 2002. p.179-202.

CHEUNG, N. et al. Head-related transfer function modeling in 3-D sound systems with genetic algorithms. **Journal of the Audio Engineering Society**, v.46, n.6, June 1998. p.531-539.

COOK, P. et al. N>>2: Multi-speaker display systems for virtual reality and spatial audio projection. In: ICAD'98 – INTERNATIONAL CONFERENCE ON AUDITORY DISPLAY, 5. Glasgow, 1998. **Proceedings**. Glasgow: ICAD, 1998.

COULTER, D. **Digital audio processing**. Lawrence, Kansas (EUA): R&D Books, 2000. 413p.

CRUZ-NEIRA, C.; D.J. SANDIN; T.A. DEFANTI. Surround-screen projection-based virtual reality: The design and Implementation of the CAVE. In: SIGGRAPH 1993. **Proceedings**. Anaheim: ACM SIGGRAPH, July 1993.

CYCLING'74. **Max/MSP for Mac and Windows**. Disponível em: <<http://www.cycling74.com/products/maxmsp.html>>. Acesso em: 30 maio 2005.

DANIEL, J. et al. Further investigations of high order Ambisonics and Wavefield Synthesis for holophonic sound imaging. In: Audio Engineering Society Convention, 114., Amsterdam, 2003. **Proceedings**. Nova York: AES, 2003.

DE VRIES, D.; BOONE, M.M. Wave field synthesis and analysis using array technology. In: 1999 IEEE WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS. New Paltz, 1999. **Proceedings**. New Paltz: IEEE, 1999. p.15-18.

ECKEL, G. A spatial auditory display for the CyberStage. In: ICAD'98 – INTERNATIONAL CONFERENCE ON AUDITORY DISPLAY, 5., Glasgow, 1998. **Proceedings**. Glasgow: ICAD, 1998.

ELEN, R. **Ambisonic.Net**. Disponível em: <<http://www.ambisonic.net>>. Acesso em: 22 nov. 2004.

ELEN, R. **Ambisonics: the surround alternative**. 4p. Disponível em <http://www.ambisonics.net>. 2001. Acesso em: 09 maio 2003.

EVEREST, F.A. **Master handbook of acoustics**. 4a. ed. Nova York, McGraw-Hill, 2001. 615p.

FARIA, R. R. A. AUDIENCE – *Audio Immersion Experience by Computer Emulation*. Disponível em: <<http://www.lsi.usp.br/interativos/nem/audience/>>. Acesso em: 22 nov. 2004.

FARIA, R. R. A. **Aplicação de wavelets na análise de gestos musicais em timbres de instrumentos acústicos tradicionais**. 1997. 182p. Dissertação (Mestrado). Escola Politécnica, Universidade de São Paulo. São Paulo, 1997.

FARIA, R. R. A. et al. A Novel Approach for Multimedia Rendering Platforms. In: FSKD'02 – INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS AND KNOWLEDGE DISCOVERY, 1., Singapura, 2002. **Proceedings**. Singapura: Nanyang Technological University, 2002, v.2, p.763-767.

FARIA, R. R. A. Evolução tecnológica e os paradigmas vigentes: um ensaio existencialista. Ensaio disciplina “Filosofia da Ciência”, USP, 2002. São Paulo, 2002. 12p. /não publicado/

FARIA, R. R. A.; et al. A multichannel spatial sound modular architecture for auralization in immersive virtual reality. Submetido ao WASPAA 05 – 2005 IEEE WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, New Paltz, 16-19 Out. 2005.

FARIA, R. R. A.; et al. Improving spatial perception through sound field simulation in VR. In: VECIMS 2005 –IEEE CONFERENCE ON VIRTUAL ENVIRONMENTS, HUMAN-COMPUTER INTERFACES, AND MEASUREMENT SYSTEMS, Giardini Naxos, 2005. **Proceedings**. Giardini Naxos, 2005, p.103-108.

FARINA, A. et al. Ambiophonic principles for the recording and reproduction of surround sound for music. In: AES INTERNATIONAL CONFERENCE, 19., Schloss Elmau, 2001. **Proceedings**. New York: AES, 2001.

FARINA, A.; AYALON, R. Recording concert hall acoustics for posterity. In: AUDIO ENGINEERING SOCIETY CONFERENCE ON MULTICHANNEL AUDIO, 24., Banff, 2003. **Proceedings**. Nova York: AES, 2003. p.223-236.

FARINA, A.; UGOLOTTI, E. Software implementation of B-Format encoding and decoding. In: AES INTERNATIONAL CONFERENCE, 109. **Proceedings**. Nova York: AES, 1998. Disponível em <<http://pcfarina.eng.uni-pr.it/>>

FIELDER, L. D. Analysis of traditional and reverberation-reducing methods of room equalization. **Journal of the Audio Engineering Society**, Nova York, v.51, n.1/2, p. 3-26, Jan./Fev. 2003.

FIGUEIREDO, F. L. et al. Análise de parâmetros acústicos subjetivos: critérios para avaliação da qualidade acústica de salas de música. In: REUNION ANUAL DE LA SOCIEDAD ARGENTINA PARA LAS CIENCIAS COGNITIVAS DE LA MUSICA, 4. Tucumán, 2004. **Proceedings**. Tucumán: SACCOM, 2004. 10p.

FIRELIGHT TECHNOLOGIES. **FMOD: music & sound effects system**. Release 3.74, 2004. Disponível em: <<http://www.fmod.org/>> . Acesso em: 15 nov. 2004.

FOUAD, H. et al. An extensible toolkit for creating virtual sonic environments. In: ICAD 2000 – INTERNATIONAL CONFERENCE ON AUDITORY DISPLAY, Atlanta, 2000. **Proceedings**. Atlanta: ICAD, 2000.

FUNKHOUSER, T. et al. Real-time acoustic modeling for distributed virtual environments. In: SIGGRAPH 99, Los Angeles, 1999. **Proceedings**: ACM Siggraph, 1999. p. 365-374.

FURSE, R. **First and second order Ambisonic decoding equations**. Disponível em: <<http://www.muse.demon.co.uk/ref/speakers.html>>. Acesso em: 30 maio 2005.

GERZON, M. Ambisonics decoders for HDTV. In: CONVENTION OF THE AUDIO ENGINEERING SOCIETY, 92., Viena, 1992. **Proceedings**. Preprint 3345 (Audio Engineering Society). 42p. Disponível em: <<http://www.aes.org/>>

GERZON, M. Ambisonics in multichannel broadcasting and video. **Journal of the Audio Engineering Society**, v.33, n.11, nov.1985. p.859-871. /Apresentado an 74th. CONVENTION OF THE AUDIO ENGINEERING SOCIETY, Nova York, out.1983/.

GERZON, M. Periphery: with-height sound reproduction. **Journal of the Audio Engineering Society**, v.21, n.1, 1973. p.2-10. /Apresentado na CONVENTION OF THE CENTRAL EUROPE SECTION OF THE AUDIO ENGINEERING SOCIETY,2., Munique, mar. 1972/.

GERZON, M. Surround-sound psychoacoustics: criteria for the design of matrix and discrete surround-sound systems. **Wireless World**, Dez. 1974, p.483-486.

GLASGAL, R. Ambiophonics. Disponível em: <<http://www.ambiophonics.org/>>. Acesso em: 8 ago. 2004.

GLASGAL, R. Surround ambiophonic recording and reproduction. In: AES INTERNATIONAL CONFERENCE ON MULTICHANNEL AUDIO, 24., Banff, 2003. **Proceedings**. New York: AES, 2003, p.11-20.

GOBBETTI, E.; BALAGUER, J. VB2 – An architecture for interaction in synthetic worlds. In: UIST'93, Atlanta, 1993. **Proceedings**: ACM, 1993. p. 167-178.

GOEBBELS, G. LALIOTI, V. Co-presence and co-working in distributed collaborative virtual environments. In: INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS, VIRTUAL REALITY AND VISUALIZATION, 1., Cidade do Cabo, 2001. **Proceedings**. ACM Siggraph, 2001. p.109-114.

GOUDESEUNE, C. **Virtual Sound Server Reference Manual**. 2003. Disponível em: <<http://www.isl.uiuc.edu/Software/software.htm>> Acesso em: 30 maio 2005.

HÂRMÂ, A. et al. Drawing quality maps of the sweet spot and its surroundings in multichannel reproduction and coding. In: AES INTERNATIONAL CONFERENCE, 21., São Petesburgo, 2002. **Proceedings**. Nova York: AES, 2002, 9p.

HORBACH, U. et al. Real-time rendering of dynamic scenes using wave field synthesis. In: ICME'02 – 2002 IEEE International Conference on Multimedia and Expo, Lausanne, 2002. **Proceedings**. vol. 1. p.517-520.

INFORMATION SOCIETY TECHNOLOGIES (IST) PROGRAMME. **Project CARROUSO**: Creating, assessing and rendering in real time of high quality audio-visual environments in MPEG-4 context: system specification and functional architecture (public version). 12 Jun. 2001. Disponível em: <<http://www.idmt.de/projects/carrouso/index.html>>. Acesso em: 30 maio 2005.

INTERNATIONAL TELECOMMUNICATIONS UNION – RADIOCOMMUNICATION. **Multichannel stereophonic sound system with and without accompanying picture – Rec. ITU-R BS.775-1**. ITU, 1994. (rev. 2002)

KAUFF, P.; SCHREER, O. An immersive 3D video-conferencing system using shared virtual team user environments. In: CVE'02, Bonn, 2002. **Proceedings**. ACM, p. 105-112.

- KENDAL, G. S. A 3-D sound primer: directional hearing and stereo reproduction. **Computer Music Journal**, v.19, n.4, p.23-46, Winter 1995.
- KIRKEBY, O. Digital filter design for inversion problems in sound reproduction. **Journal of the Audio Engineering Society**, v.47, n.7/8, July/Ago. 1999. p.583-595.
- KLEINER, M. et al. Auralization – an overview. **Journal of the Audio Engineering Society**, v.41, n.11, p.861-875, Nov. 1993.
- KURASHIMA, C. S. **Proposta de Um Sistema de Teleconferência Imersiva Panorâmica**. 2005. Edição Revisada. Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos. São Paulo, 2005.
- KUTTRUFF, K.H. Auralization of impulse responses modeled on the basis of ray-tracing results. **Journal of the Audio Engineering Society**, v.41, n.11, nov. 1993. p.876-880. /Apresentado na 91ª Conferência da AES, Nova York, out. 1991/.
- LOKKI, T; JÄRVELÄINEN, H. Subjective evaluation of auralization of physics-based room acoustics modeling. In: ICAD01 – 2001 INTERNATIONAL CONFERENCE ON AUDITORY DISPLAY, Espoo, 2001. **Proceedings**. Espoo, 2001. p.26-31.
- MALHAM, D. **Music Technology Group: Sound in Space**. Disponível em: <http://www.york.ac.uk/inst/mustech/3d_audio/>. Acesso em: 30 maio 2005.
- MALHAM, D.G. **3-D sound for virtual reality systems using Ambisonic techniques**. Disponível em: <http://www.york.ac.uk/inst/mustech/3d_audio/artindex.htm> Acesso em: 15 jun.2004. /Apresentando na VR93 Conference, 1993./.
- MALHAM, D.G. Computer control of Ambisonic soundfields. In: 82nd. Convention of the Audio Engineering Society, Londres, 1987. **Preprint 2463**. 12p. Disponível em: <<http://www.aes.org/>>
- MALHAM, D.G.; MYATT, A. 3-D sound spatialization using Ambisonic techniques. **Computer Music Journal** , v.19, n.4, p.58-70, Winter 1995.
- MARTIN, G. **Introduction to Sound Recording**. Livro Eletrônico (em desenvolvimento). Disponível em: <<http://www.tonmeister.ca/main/textbook/>>. Acesso em: 14 mar. 2004.

MCGRATH, D.; REILLY, A. **Creation manipulation and playback of soundfields.**

Technical Paper. Disponível em: <http://www.headwize.com/tech/lake2_tech.htm>. Acesso em: 30 maio 2005.

MILLER, R. E. Transforming ambiophonic + ambisonic 3D surround sound to & from ITU 5.1/6.1. In: AES INTERNATIONAL CONFERENCE, 114. Amsterdam, 22-25 Mar. 2003. **Proceedings.** New York: AES, 2003, 16p.

MOUCHTARIS, A. et al. Head-related transfer synthesis for immersive audio. In: IEEE WORKSHOP ON MULTIMEDIA SIGNAL PROCESSING, 2., Redondo Beach, 1998. **Proceedings.** Redondo Beach: IEEE, 1998, p.155-160.

NAEF, M. et al. Spatialized audio rendering for immersive virtual environments. In: VRST'02, 2002, Hong Kong. **Proceedings.** ACM, 2002.

NATIONAL RESEARCH DEVELOPMENT CORPORATION (London, EN). Michael Anthony Gerzon. **Multidirectional sound reproduction systems.** US 3,997,725. December 14, 1976.

NEHER, T. et al. Unidimensional simulation of the spatial attribute 'ensemble depth' for training purposes, part 1: pilot study into early reflection pattern characteristics". In: AUDIO ENGINEERING SOCIETY CONFERENCE ON MULTICHANNEL AUDIO, 24., Banff, 2003. **Proceedings.** Nova York: AES, 2003. p.123-137.

OGI, T.; KAYAHARA, T. et al. Immersive sound field simulation in multi-screen projection displays. In: INTERNATIONAL IMMERSIVE PROJECTION TECHNOLOGIES WORKSHOP, 7., EUROGRAPHICS WORKSHOP ON VIRTUAL ENVIRONMENTS, 9., Zurich, 2003. **Proceedings.** Zurich: Eurographics, 2003. p.135-142.

PAPE, D. **Bergen Sound Server & Library, Version 0.4.1.** Disponível em: <<http://www.evl.uic.edu/pape/sw/bergen/>> Acesso: 15 nov. 2004.

PATRONIS, E. Stadiums and outdoor venues. In: BALLOU, G.M. (Ed.) **Handbook for sound engineers.** 3ed. Woburn, Focal Press, 2002. p.159-165.

POHLMANN, K. C. **Principles of digital audio.** 3.ed. Nova York: McGraw-Hill, 1995. 622p.

PRESS, W.H. et al. **Numerical recipes in C: the art of scientific computing**. 2^a. edição (corrigida) Cambridge University Press, New York, 1994. 994 p.

PUCKETTE, M. **Software by Miller Puckette – Pd: real-time music and multimedia environment**. Disponível em: <<http://www-crca.ucsd.edu/~msp/software.html>>. Acesso em: 30 maio 2005.

PULKKI, V. et al. Analyzing virtual sound source attributes using a bi-aural auditory model. **Journal of the Audio Engineering Society**, New York, v.47, n.4, p. 203-217, Apr. 1999.

PULKKI, V. Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning. **Journal of the Audio Engineering Society**, v.49, n.9, p.753-767, Sept. 2001.

PULKKI, V. **Spatial sound generation and perception by amplitude panning techniques**. 2001. 42p. Dissertação (Doutorado). Helsinki University of Technology, Department of Electrical and Communications Engineering. Helsinki, 2001.

PULKKI, V. Virtual sound source positioning using vector base amplitude panning. **Journal of the Audio Engineering Society**, Nova York, v.45, n.6, p. 456-466, June 1997.

PULKKI, V.; KARJALAINEN, M. Localization of amplitude-panned virtual sources I: Stereophonic panning. **Journal of the Audio Engineering Society**, v.49, n.9, p.739-752, Sept. 2001.

ROYAL PHILIPS ELECTRONICS. **Super Audio CD Format. Catálogo Informativo**, versão 2002-30. 2002.

RUMSEY, F. Controlled subjective assessments of two-to-five channel surround sound processing algorithms. **Journal of the Audio Engineering Society**, v.47, n.7/8, July/Ago. 1999. p.563-582.

RUMSEY, F. **Spatial audio**. Oxford: Focal Press (Elsevier), 2003. 240p.

RUSCHIONI, R. A.; WEY, J. D. R.; SANTOS, E. T.; ZUFFO, M. K. Some Experiences Implementing Virtual Worlds: The Aria Project. In: IFIP WG9.4/WG9.5 Conference, Florianopolis, 1997. **Proceedings**. 1997 (CD-ROM). Disponível em HTML em: <<http://docentes.pcc.usp.br/toledo/pdf/ifip97/aria.html>>. Acesso em: 15 mar. 2005.

SAVIOJA, L. et al. Creating interactive virtual acoustic environments. **Journal of the Audio Engineering Society**, New York, v.47, n.9, p.675-705, Sept. 1999.

SAWHENEY, N.; MURPHY, A. ESPACE 2: An experimental hyperaudio environment. In: CHI 96 (Interactive Posters), 1996. **Proceedings**. ACM, 1996. p.105-106.

SCHROEDER, M.R. Natural-sounding artificial reverberation. **Journal of the Audio Engineering Society**, New York, v.10, n.3, p.219-223, July 1962.

SOARES, L. P. et al.. Virtual Hang-gliding over Rio de Janeiro. In: IEEE VR WORKSHOP VR FOR PUBLIC CONSUMPTION. **Proceedings**. Chicago: 2004.

SOARES, L. P., CABRAL, M. C., BRESSAN, P. A., et al. Powering Multiprojection Immersive Environments. In: SIACG 2002, Guimarães, 2002. **Proceedings**. Guimarães: 2002

SOARES, L. P.; ZUFFO, M. K. Jinx: an *X3D* browser for VR immersive simulation based on clusters of commodity computers. In: INTERNATIONAL CONFERENCE ON 3D WEB TECHNOLOGY, 9., Monterey, 2004. **Proceedings**. Monterey: 2004. p.79-86.

STROBEL, N.; SPORS, S.; RABENSTEIN, R. Joint audio-video object localization and tracking: A presentation of general methodology. **IEEE Signal Processing Magazine**, Jan. 2001, p.22-31.

STUART, J. R. et al. MLP lossless compression. In: AES REGIONAL CONVENTION, 9., Tokyo, 1999. **Proceedings**. Tokyo: AES, 1999. 11p.

SUSNIK, R. et al. Spatial sound generation using *HRTF* created by the use of recursive filters. In: EUROCON 2003: COMPUTER AS A TOOL. Ljubljana, Eslovênia, 2003. **Proceedings**. v.1. Ljubljana: IEEE (região 8), 2003, p.449-453.

TORGER, A. **AlmusVCU user manual**, version 0.85. 18 Jul. 2004. Disponível em: <<http://www.ludd.luth.se/~torger/almusvcu.html>> Acesso em: 23 abr. 2005.

VÄÄNÄNEN, R. **Parametrization, auralization, and authoring of room acoustics for virtual reality applications**. 2003. 167p. Dissertação (Doutorado). Helsinki University of Technology. 10 mai. 2003. ISBN 951-22-6544-3. Disponível em: <<http://lib.hut.fi/Diss/>>. Acesso em 10 maio 2005.

- VÄÄNÄNEN, R.;HUOPANIEMI, J. SNHC audio and audio composition. In: PEREIRA, F.C.N.; EBRAHIMI, T. **The MPEG-4 book**. New Jersey, IMSC Press/Prentice Hall, 2002. p.545-581.
- VAN ASSELT, A. **Recording sound for 3/2: introduction & critical review**. Jun. 2002. 39p. CITAÇÃO INCOMPLETA
- VOGEL, P. **Application of wave field synthesis in room acoustics**. 1993. 302p. Tese (Doutorado) – Delft University of Technology. Delft, 1993.
- VORLÄNDER, M. International round robin on room acoustical computer simulations. In: INTERNATIONAL CONGRESS ON ACOUSTICS, 15., Trondheim (Noruega), 1995. **Proceedings**. p.689-692. Disponível em: <<http://www.at.oersted.dtu.dk/~odeon/pdf/vorlander.pdf>>. Acesso em 28 maio 2005.
- WEB3D CONSORTIUM. **Open Standards for Real-Time 3D Communication: X3D**. 2005. Disponível em: <<http://www.web3d.org/>>. Acesso em: 30 maio 2005.
- WHITE, J.V. Synthesis of 4-2-4 matrix recording systems. **Journal of the Audio Engineering Society**, v.24, n.4, p.250-257. May 1976.
- ZÖLZER, U. **Digital audio signal processing**. Chichester, (Inglaterra): John Wiley & Sons, 1997. 279p.
- ZUFFO, J. et al. CAVERNA Digital – Sistema de Multiprojeção Estereoscópico Baseado em Aglomerados de PCs para Aplicações Imersivas em Realidade Virtual. In: SYMPOSIUM OF VIRTUAL REALITY, 4., Florianópolis, 2001. **Proceedings**.