



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ARTES

JOÃO VICTOR FIGUEIREDO PINTO¹

**VERIFICAÇÃO DA ESTABILIDADE DA ESPACIALIDADE DE IMAGEM
SONORA MULTICANAL COM CAPTAÇÃO HEXAGONAL EM REDUÇÃO PARA
DOIS CANAIS E EMULAÇÃO EM AMBISONICS DE 3ª ORDEM**

Resumo de Iniciação Científica Voluntária - PICV
apresentado para o XXVIII Congresso de Iniciação
Científica da Universidade Estadual de Campinas.

Orientador(a): Prof. Dr. José Augusto Mannis²

Campinas
2020

1. Introdução

Com o desenvolvimento rápido de equipamentos e tecnologias digitais, a capacidade de sistemas de AR/VR (realidade aumentada e realidade virtual) também cresceu e passou a ser cada vez mais explorada. Para que a experiência imersiva seja reproduzida com fidelidade, é necessário que tais sistemas decodifiquem o sinal digital e reproduzam o espaço e o cenário desejado. Em relação ao áudio – tratamento do sinal sonoro – atualmente, ainda não há consenso para as técnicas de decodificação (CARPENTIER, 2017), que dependem diretamente de condições específicas de gravação e reprodução.

Dessa maneira, o presente projeto de iniciação científica buscou avaliar as condições específicas, do sistema de captação e reprodução hexagonal desenvolvido nas pesquisas de pós-doutorado do orientador (MANNIS, 2019). Todas as atividades foram realizadas no LASom

1.1 Objetivos

A pesquisa buscou avaliar as condições específicas e propor decodificações para novas disposições de reprodução. Foram duas frentes de trabalho:

- 1 Emulação em Ambisonics de 3ª ordem em duas dimensões, com oito alto-falantes, utilizando o software Max/MSP da *Cycling 74* e avaliação do desempenho da codificação Ambisonics para gravações não-coincidentes.
- 2 Redução a 2 canais com *head-related transfer functions* (HRTF) para audição em fones de ouvido, com o *plug in OpenAir* do *Studio One 4 Professional* e com os tipos de *panning* com a biblioteca SPAT no Max/MSP

¹ Email para contato: j170712@dac.unicamp.br

² Email para contato: jamannis@unicamp.br



2. Surround Sound e Áudio 3D

Hoje, o som surround é amplamente conhecido pelo público, sobretudo por estar nas salas de cinema e nos *home theaters*, no formato 5.1. Suas aplicações encontram-se frequentemente nas áreas de música e cinema, onde há maior liberdade criativa e interpretativa no domínio sonoro, além de ser um formato padronizado de broadcast para som multicanal em rádio digital³ e TV digital⁴. Por esse motivo, há grande diferença com relação ao áudio 3D, que está “relacionado a modelos físicos e um alto grau de realismo nos componentes de geração, transmissão e reprodução de som e vibrações” (VORLÄNDER, 2008, p. 281, tradução nossa).

A Recomendação ITU-R BS.2051-2: *Advanced sound system for programme production* (2018) explicita conceitos relevantes sobre sistemas de áudio 3D. Um sistema *channel-based* distribui cada canal em um alto-falante predefinido (como o sistema hexagonal de referência), enquanto um sistema *object-based* lida com os elementos sonoros de maneira individual e independente. Por fim, há os sistemas *scene-based*, que lidam com o campo sonoro como cena, como conjunto, sem limitar um elemento sonoro a determinado alto-falante ou distribuí-lo de maneira individual (HOA, ou *higher-order ambisonics*).

Com isso, observa-se que o sistema desenvolvido com redução a dois canais é *object-based*, devido ao *panning* com HRTFs. O sistema com emulação em ambisonics de 3ª ordem é, por sua vez, processado como *scene-based*. Ambisonics não foi aplicado para reprodução em fones de ouvidos devido à tendência da convolução para os dois canais torná-lo *object-based*, ao invés de *scene-based* (ZOTTER; FRANK, 2019).

3. Ambisonics

Ambisonics é a tecnologia resultante do trabalho de Gerzon (1973), em que gravava com as cápsulas dos microfones dispostos ortogonalmente, captando o campo sonoro em um ponto (gravação coincidente) e descrevendo-o através de harmônicos esféricos em uma teoria completa – da gravação até a reconstrução matemática e reprodução do campo sintetizado (ARTEAGA, 2015).

Como não há consenso na metodologia para processamento do sinal (CARPENTIER, 2017), optou-se por desenvolver programações que permitissem a comparação de todos os esquemas de normalização e as estratégias de decodificação.

Com relação às normalizações, por mais que as programações permitam a escolha, o foco ficou limitado à normalização N2D, visto que o sistema é de duas dimensões e foi o indicado por Carpentier (2017), que faz parte do IRCAM, laboratório de referência para o LASom e cooperou em uma oportunidade com a pesquisa.

3.1 Estratégias de decodificação

Para completar a programação, é necessário ainda definir os métodos matemáticos de reconstrução da matriz de decodificação HOA (de acordo com o posicionamento dos alto-falantes do sistema de reprodução) e os tipos de otimização. A combinação método+tipo ideal para a captação hexagonal será possível com a continuação da pesquisa e desenvolvimento dos experimentos.

³ Padrão Digital Radio Mondiale (DRM).

⁴ O padrão de televisão digital adotado no Brasil é o ISDB-TB uma adaptação do ISDB-T (Integrated Services Digital Broadcasting Terrestrial) – padrão japonês acrescido de tecnologias desenvolvidas no Brasil.



Abaixo consta uma tabela resumindo os métodos e tipos disponíveis nas programações e suas principais características.

| | | |
|---------|--|---|
| MÉTODOS | AllRAD | publicado por (ZOTTER; FRANK, 2012) e descrito como o “mais robusto e flexível método de decodificação para altas ordens conhecido” (p. 54, tradução nossa). |
| | EPAD (energy-preserving) | Recomendado por Carpentier (2018), tem volume independente de panning quando a ortonormalidade da decodificação é garantida com a disposição ótima de alto-falantes. |
| | Direct-sampling e mode-matching | garantem ortonormalidade e preservação da energia quando reproduzidos em t-designs ótimos (ZOTTER; FRANK, 2019). Sendo esse o caso, tornam-se EPAD. |
| TIPOS | BASIC | reconstrução física do campo sonoro. Otimização mais simples e limitada, com bons resultados para frequências abaixo de 500Hz (ARTEAGA, 2015; MURILLO; FAZI; SHIN, 2014). |
| | MAX-Re | reconstrução do campo sonoro por propriedades psicoacústicas. Visa preservar a energia geral no <i>sweet spot</i> (BERTET et al., 2009). Bons resultados para frequências médias e altas. |
| | INPHASE | também através de psicoacústica. Além das características de MAX-Re, exige sinais em fase. Bons resultados fora do <i>sweet spot</i> . |
| | BASIC+MAX-Re | BASIC para frequências abaixo de 500Hz e MAX-Re acima. É o principal candidato para o sistema hexagonal. |
| | MAX-Re +INPHASE | combinação mista com inphase para baixas frequências e max-Re para médias e altas. |

3.2 Resultados: programação em Max/MSP

A programação em Max/MSP está disponível online para visualização⁵. Observa-se o *slider* para controle da reprodução, junto com os 6 canais originais, na região alta. Ao centro e à esquerda estão o *encoder* e o *decoder* para HOA do SPAT (biblioteca da equipe *Espaces Acoustiques et Cognitifs* (CARPENTIER, 2018) do *Institute of research and Coordination in Acoustics/Music – IRCAM*). Ao lado há as opções de normalização, métodos e tipos, possibilitando a alteração imediata durante a escuta, juntamente com as coordenadas das fontes e dos alto-falantes. Finalmente abaixo, alguns elementos visuais.

4. Reprodução binaural

Foi abordada também a reprodução em fones de ouvido, devido a praticidade incomparável. No entanto, vários outros problemas aparecem, prejudicando principalmente a ambiência (difusão, espacialidade e profundidade sonora) ao alterar os parâmetros psicoacústicos ILD (*interaural level difference*) e ITD (*interaural time delay*) (VORLÄNDER, 2008).

As *head-related transfer functions* (HRTFs) descrevem as diferenças de escuta entre os dois ouvidos devido às dimensões da cabeça e tronco, a depender da direção do estímulo sonoro. Nas programações em Max/MSP, as HRTFs são representadas pelos *SOFA Files*, um “formato de representação de dados orientados espacialmente [...] para promover intercambialidade e extensibilidade” (BOREN et al., 2014, p. 4, tradução nossa).

4.1 Pannings Binaurais

Há disponível na biblioteca SPAT vários modelos de HRTF, que podem ser gravados através de medições em pessoas, manequins ou por modelos matemáticos aproximados. São eles:

⁵ Disponível em: <https://drive.google.com/file/d/1tXvrXU67G-C6MwofOiS0necpFH2VwiNY/view?usp=sharing>



- **Binaural:** tipo de panning que permite a inserção de qualquer medição de HRTF.
- **Spherical head model:** assume o formato da cabeça como uma esfera, os resultados pioram a medida que a elevação da fonte sonora aumenta (BOGELEIN et al., 2018).
- **Snowman:** este panning inclui à cabeça esférica um tronco também esférico. Algazi, Duda e Thompson (2002) garantem sua consistência para frequências baixas e contribuição para a melhora na localização de fontes virtuais elevadas.
- **Near field binaural:** este panning procura compensar os efeitos de localização que surgem quando a fonte está próxima da cabeça (menos de 1 metro) (BRUNGART; RABINOWITZ, 1999).

4.2 Resultados: programação em Max/MSP e *plug in OpenAir*

A programação em Max/MSP para a reprodução binaural está disponível online⁶. Acima há o *slider* para controle da reprodução, junto com os 6 canais originais. Na região à direita, observa-se as coordenadas das fontes sonoras e, mais abaixo, os tipos de panning, que também podem ser alterados durante a escuta. Por último, há a caixa de escolha do arquivo SOFA (o HRTF que será utilizado) e alguns elementos visuais.

É importante ressaltar que os modelos *spherical head model* e *snowman* já têm suas medições prédefinidas, ou seja, não há a possibilidade de adicionar um SOFA para esses pannings.

Com relação ao *OpenAir*⁷ da *DAW Studio One Professional*, foram inserido HRTFs nas direções específicas do sistema hexagonal de referência. Dessa maneira, é possível comparar o comportamento *object-based* dos pannings no Max/MSP com o *channel-based* no *Studio One*.

5. Considerações finais

Com as programações será possível prosseguir com a pesquisa com o foco na avaliação subjetiva e comparativa de cada decodificação, verificando a qualidade e estabilidade do resultado com as gravações do sistema de captação e reprodução hexagonal. Para a realização dos experimentos, além da leitura da bibliografia, também já foram desenvolvidas planilhas para a coleta e manipulação de dados. Tais materiais estão disponíveis online na descrição do vídeo-poster dessa pesquisa⁸. Infelizmente, essa etapa foi adiada devido à pandemia.

⁶ Disponível em:

<https://drive.google.com/file/d/1KsobrM0rQBKLRceelcvUdhHGGWQPYWGA/view?usp=sharing>

⁷ Disponível em: https://drive.google.com/file/d/1a4_wDDsfx5eGf3a9r_rO6PLF22PVM0cb/view?usp=sharing

⁸ Disponível em: <https://youtu.be/-LVQOkvAcoA>



6. Referências Bibliográficas

ALGAZI, V.; DUDA, R.; THOMPSON, D. The use of head-and-torso models for improved spatial sound synthesis. **113th Convention of the Audio Engineering Society**, 2002.

ARTEAGA, D. **Introduction to Ambisonics**. Audio 3D – Escola Superior Politècnica Universitat Pompeu Fabra, 2015.

BERTET, S. et al. Influence of microphone and loudspeaker setup on perceived higher order ambisonics reproduced sound field. **Ambisonics Symposium**, 2009.

BÖGELEIN, S. et al. Localization cues of a spherical head model. **Proceedings of the 44th DAGA**, 2018.

BOREN, B. et al. PHOnA: a public dataset of measured headphone transfer functions. **137th Audio Engineering Society Convention**. Audio Engineering Society, 2014.

BRUNGART, D.; RABINOWITZ, W. Auditory localization of nearby sources. Head-related transfer functions. **The Journal of the Acoustical Society of America**, v. 106, n. 3, p. 1465–1479, 1999.

CARPENTIER, T. Normalization schemes in ambisonic: Does it matter? **142nd Audio Engineering Society International Convention**. Audio Engineering Society, 2017.

CARPENTIER, T. A new implementation of Spat in Max. **15th Sound and Music Computing Conference (SMC)**, p. 184–191, 2018.

GERZON, M. Periphohy: With-Height Sound Reproduction. **Journal of the Audio Engineering Society**, v. 21, n. 1, p. 2–10, 1973.

MANNIS, J. Development of a Multi-Channel Sound Recording and Reproduction System. **IRCAM Forum**, Paris. 2019.

MURILLO, D.; FAZI, F.; SHIN, M. Evaluation of ambisonics decoding methods with experimental measurements. **EAA Joint Symposium on Auralization and Ambisonics**, 2014.

Recommendation ITU-R BS.2051-2: Advanced sound system for programme production, 2018.

VORLÄNDER, M. **Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality**. Springer Science & Business Media, 2007.

ZOTTER, F.; FRANK, M. All-round ambisonic panning and decoding. **Journal of the Audio Engineering Society**, v. 60, n. 10, p. 807–820, 2012.

ZOTTER, F.; FRANK, M. **Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality**. Springer Open, 2019.