



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ARTES

NYCHOLAS MAIA

**Um Modelo Computacional de Mixagem Automática para
Música Comercial**

***A Computational Model for Automatic Mixing for Pop
Music***

Campinas

2018

NYCHOLAS MAIA

**Um Modelo Computacional de Mixagem Automática para
Música Comercial**

***A Computational Model for Automatic Mixing for Pop
Music***

Dissertação apresentada ao Instituto de Artes da
Universidade Estadual de Campinas para obtenção
do título de Mestre em Música, na área de Música:
Teoria, Criação e Prática

*Dissertation presented to the Institute of Arts of the
State University of Campinas to obtain the degree of
Master of Music in the field of Music: Theory,
Creation, and Practice*

Orientador: Prof. Dr. José Eduardo Fornari Novo Junior

ESTE TRABALHO CORRESPONDE À
VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELO ALUNO NYCHOLAS
MAIA, E ORIENTADO PELO PROF. DR. JOSÉ
EDUARDO FORNARI NOVO JUNIOR.

Campinas

2018

Agência(s) de fomento e nº(s) de processo(s): Não se aplica.

ORCID: <https://orcid.org/0000-0002-8109-7304>

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Artes
Sílvia Regina Shiroma - CRB 8/8180

M28m Maia, Nycholas, 1989-
Um Modelo Computacional de Mixagem Automática para Música Comercial
/ Nycholas Maia. – Campinas, SP : [s.n.], 2018.

Orientador: José Eduardo Fornari Novo Junior.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de
Artes.

1. Gravação musical. 2. Percepção musical. 3. Música e tecnologia. I. Novo
Junior, José Eduardo Fornari, 1966-. II. Universidade Estadual de Campinas.
Instituto de Artes. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: A Computational Model for Automatic Mixing for Pop Music

Palavras-chave em inglês:

Music recorder

Musical perception

Music and Technology

Área de concentração: Música: Teoria, Criação e Prática

Titulação: Mestre em Música

Banca examinadora:

José Eduardo Fornari Novo Junior [Orientador]

Manuel Silveira Falleiros

Bruno Sanches Masiero

Data de defesa: 13-12-2018

Programa de Pós-Graduação: Música

BANCA EXAMINADORA DA DEFESA DE MESTRADO

NYCHOLAS MAIA

ORIENTADOR: Prof. Dr. José Eduardo Fornari Novo Junior

MEMBROS:

1. Prof. Dr. José Eduardo Fornari Novo Junior
2. Prof. Dr. Manuel Silveira Falleiros
3. Prof. Dr. Bruno Sanches Masiero

Programa de Pós-Graduação em Música do Instituto de Artes da
Universidade Estadual de Campinas.

A ata de defesa com as respectivas assinaturas dos membros da
Comissão Examinadora encontra-se no SIGA/Sistema de Fluxo de
Dissertação/Tese e na Secretaria do Programa da Unidade.

DATA DA DEFESA: 13/12/2018

*“...poems are made by fools like me
but only God can make a tree”
Joyce Kilmer (1913)*

AGRADECIMENTOS

A DEUS, pela inspiração em todos os dias.

Aos meus pais, Paulo e Sandra, a quem devo minha educação.

Ao meu orientador, Prof. José Eduardo Fornari Novo Junior (Tuti), pela paciência, confiança e orientação.

Ao Prof. Adolfo Maia Jr., meu tio, pela troca de ideias e sugestões na parte matemática desta dissertação.

RESUMO

Este trabalho propõe o desenvolvimento de um novo método de mixagem de áudio digital que visa automatizar dois principais processos comuns à tarefa de mistura de sons de instrumentos musicais em música comercial. O propósito prático desta automatização computacional é facilitar o trabalho do engenheiro de som no equilíbrio dos instrumentos musicais individuais, aproximando aos padrões de indústria fonográfica.

Primeiramente, discutimos os aspectos perceptuais do áudio, os equipamentos de manipulação sonora usualmente encontrados em estúdios de gravação, e também alguns aspectos estético-musicais. Como uma contribuição original a área, descrevemos este novo método de mixagem automática que visa à otimização musical ou sonora, sentido de dinâmica musical e distribuição espectral do áudio digital.

Nosso modelo é baseado na aproximação do espectro da mixagem a um espectro de referência via matrizes de volume e equalização para cada instrumento individualmente. Alguns tipos de espectros de referência propostos são baseados no princípio das curvas isofônicas de Fletcher & Munson.

O modelo, com todos os seus módulos, foi implementado computacionalmente em MATLAB e, potencialmente, pode futuramente vir a ser exportado como um programa *stand-alone* (independente) para ser utilizado por músicos, técnicos e engenheiros de som. O modelo computacional apresentado tem o propósito de validar nosso método de atingir o equilíbrio sonoro da mixagem pela automatização de três processos: a) otimização da dinâmica instrumental no domínio do tempo, b) correção tonal/espectral do áudio de cada instrumento gravado, c) melhoria de balanço tonal/espectral da soma final resultante da mixagem. Por fim, um exemplo completo da aplicação deste processo de mixagem automática é aqui apresentado e descrito.

Palavras-chave: Gravação Musical. Equilíbrio Perceptual. Áudio Digital. Mixagem de Áudio Automática.

ABSTRACT

In this work we propose the development of a new method of digital audio mixing that aims to automate two main usual processes for the task of mixing musical instrument sounds. The practical purpose of this computational automation is to facilitate the sound engineer's work for the balance of individual musical instruments, improving the sound of the recorded music approaching the phonographic industry standards.

Firstly, we discuss the perceptual aspects of audio, the sound manipulation equipment usually found in recording studios, as well as some aesthetic-musical aspects. As an original contribution to the field, we describe a new method of automatic mixing that aims the optimization of the musical dynamics and spectral distribution of digital audio.

Our model is based on the approximation of the mixing spectrum to a reference spectrum via sound volume matrices and individual equalization for each instrument. Some proposed types of reference spectra are based on Fletcher & Munson's equal-loudness contours.

The model, with all its modules was computationally implemented in MATLAB and eventually can be exported as a stand-alone program for use by musicians, technicians and sound engineers. This software achieves our propose of sound balance through the automation of three processes: a) optimization of the instrumental dynamics in the time domain, b) tonal / spectral correction of each recorded instrument, c) improvement of tonal / spectral balance of the final sum resulting from the mixing.

Keywords: Musical recording. Perceptual equilibrium. Digital Audio. Automatic Audio Mixing.

Lista de Ilustrações

Figura 1. A mixagem e as áreas do conhecimento	16
Figura 2. Comparação qualitativa das equações de Fechner e Stevens	21
Figura 3. Fonógrafo de Thomas Edison (1877).....	24
Figura 4. A orelha humana	35
Figura 5. O ouvido interno	37
Figura 6. Cápsula ótica	38
Figura 7. Labirinto ósseo da Cápsula ótica.....	38
Figura 8. Labirinto ósseo em rosa e no seu interior o labirinto membranoso em azul.....	38
Figura 9. Circulação das ondas sonoras na perilinfa através do labirinto da cóclea	39
Figura 10. O órgão de Corti.....	40
Figura 11. Localização do órgão de Corti do nervo coclear.....	41
Figura 12. A resposta por banda de frequência da cóclea humana.....	42
Figura 13. Curvas Isoaudíveis de acordo com a ISO 226:2003	44
Figura 14. Espectro do ruído branco e do ruído rosa.....	51
Figura 15. Cadeia padrão de Processamento de Áudio	54
Figura 16. Plugin de emulação do Channel Strip Neve 88RS.....	54
Figura 17. Função de Transferência de um Compressor	55
Figura 18. Plugin de compressão Avid Dynamics III.....	56
Figura 19. Compressor 1176 da Universal Audio	57
Figura 20. Gravador de Rolo Studer A800.....	57
Figura 21. Equalizador Gráfico Behringer FBQ1502HD.....	59
Figura 22. Equalizador Paramétrico SSL 611EQ	59
Figura 23. Equalizador Digital: Correção Espectral e Balanço Tonal	61
Figura 24. Métodos de Mixagem.....	62
Figura 25. Visão geral do nosso método de mixagem automatizada	64
Figura 26. Legenda dos Fluxogramas.....	65
Figura 27. Pré-processamento e Mixagem Inicial: Importa os arquivos de áudio	67
Figura 29. Mixagem: Soma amostra por amostra.....	68
Figura 30. Obtenção do Espectro Inicial	70
Figura 32. Obtenção do Espectro de cada Instrumento	73
Figura 33. Localização da Frequência de Corte	75
Figura 34. Controle da Dinâmica por Instrumento.....	82
Figura 35. Faders das mesas Yamaha CL Series.....	83
Figura 36. Mixagem Final: Soma dos instrumentos modificados	90
Figura 37. Análise e Controle da Dinâmica da Mixagem	91
Figura 38. Limite de Dinâmica da Mixagem.....	93
Figura 39. Mix Inicial da música “Rescue Me”	95
Figura 40. Mix Temporária da música “Rescue Me”	96
Figura 41. Mix Intermediária da música "Rescue Me"	97
Figura 42. Mix Final da música "Rescue Me"	98
Figura 43. Comparação entre todas as mixagens do método	99

Lista de Tabelas

Tabela 1 - Expoentes da função-potência para uma variedade de modalidades perceptivas baseados em julgamentos de razão	22
Tabela 2 - Proporções de Ouro de Sepmeyer (1965).....	45
Tabela 3 - Proporções de Ouro de Louden (1971)	46
Tabela 4 - Exemplo de armazenamento de dados em um vetor de dimensão 31×1	71
Tabela 5 - Exemplo de valor de correção comum a um instrumento	88
Tabela 6 - Valores de ganho atualizados	88

Lista de Abreviaturas e Siglas

SPL	Nível de Pressão Sonora
FFT	Transformada Rápida de Fourier
DTFT	Transformada de Fourier de Tempo Discreto
HI-FI	Alta definição
DB	Decibel
DBFS	Decibel de Escala Completa

SUMÁRIO

1 INTRODUÇÃO	14
1.1 BASES PARA AVALIAÇÃO SUBJETIVA DO SOM.....	16
1.2 GRAVAÇÕES MUSICAIS	23
1.3 O ESTADO DA ARTE	28
2 FUNDAMENTOS TEÓRICOS	29
2.1 ESCOPO ESTÉTICO-MUSICAL.....	29
2.2 A BUSCA PELO EQUILÍBRIO ESTÉTICO	30
2.3 A ESCRITA CORAL E ORQUESTRAL	30
2.4 O REGISTRO MUSICAL EM ÁUDIO	31
2.4.1 Música Erudita.....	32
2.4.2 Música Comercial.....	33
2.5 EQUILÍBRIO PSICOACÚSTICO	33
2.5.1 Fisiologia Auditiva	34
2.5.2 Curvas Isoaudíveis.....	43
2.5.3 Monitoramento e Sala de Mixagem.....	45
2.6 REFERÊNCIAS SONORAS.....	47
2.6.1 Gravações de alta fidelidade	49
2.6.2 Ruído Rosa (Pink Noise)	50
3 MIXAGEM.....	53
3.1 Principais Equipamentos	53
3.1.1 Compressor de Áudio	55
3.1.2 Equalizador de Áudio	58
3.2 FREQUÊNCIAS RESSONANTES	60
3.3 BALANÇO TONAL	60
3.4 PRINCIPAIS MÉTODOS DE MIXAGEM	61
3.4.1 Bottom Up Mixing.....	61
3.4.2 Top Down Mixing	62
4 MÉTODO	64
4.1 IMPORTAÇÃO DOS DADOS	65
4.1.1 Pré-Processamento e Mix Inicial	66
4.1.2 Obtenção do Espectro Inicial.....	69
4.1.3 Sinal de Referência	71
4.1.4 Comparação Espectral	72
4.1.5 Obtenção do Espectro de cada Instrumento	73

4.2 REMOÇÃO DE FREQUÊNCIAS INDESEJÁVEIS	73
4.2.1 Definição de Relevância por Banda	74
4.2.2 Frequências Graves: Abaixo de 100Hz	74
4.2.3 Frequências Ressonantes	75
4.2.4 Remoção das Frequências por Instrumento	78
4.3 CONTROLE DA DINÂMICA	78
4.4 EQUILÍBRIO ENTRE OS VOLUMES	82
4.5 CORREÇÃO ESPECTRAL	86
4.5.1 Contribuição Espectral de cada Instrumento	86
4.5.2 Comparação entre a Mix Intermediária e a Referência	87
4.5.3 Definição das Magnitudes de Correção	87
4.5.4 Remoção do ganho comum a todas as bandas	88
4.5.5 Limite de Distorção Espectral	89
4.5.6 Aplicação da Correção Espectral	89
4.6 MIXAGEM FINAL	90
4.7 MASTERIZAÇÃO	90
4.7.1 Limite de Dinâmica da Mixagem	91
5 RESULTADOS	94
6 CONCLUSÃO E TRABALHOS FUTUROS	100
REFERÊNCIAS	101

1 INTRODUÇÃO

O processo de gravação musical moderno consiste basicamente em três etapas: *captação* (*gravação/registo do fenómeno acústico*), *mixagem* (*soma de múltiplas gravações*) e *masterização* (*otimização final do áudio gravado*). Neste trabalho vamos focar no processo de mixagem com o qual se procura o equilíbrio sonoro entre os instrumentos musicais previamente gravados na etapa de captação. Os dois principais parâmetros da mixagem musical são o controle da dinâmica (amplitude de cada amostra sonora) e de resposta espectral do instrumento (amplitude de cada frequência do espectro) de cada instrumento. Mesmo sendo aparentemente apenas dois parâmetros técnicos, sua aplicação prática é complexa e demanda alto conhecimento técnico e equipamentos de manipulação de áudio de alta qualidade. De acordo com Nino et al. (2017) a mixagem é uma das principais etapas, como também uma das mais difíceis de ser realizada durante o processo de produção musical, trazendo implicações não apenas técnicas, mas também estéticas à música gravada.

Como os aspectos técnicos de áudio não são de conhecimento comum para grande parte dos músicos, o alto custo dos equipamentos de gravação e a pouca oferta de cursos formais na área de engenharia de áudio no Brasil dificultam a produção de material sonoro gravado com boa qualidade no país, bem como a sua investigação acadêmica e pesquisa científica. Por outro lado, o consenso e o equilíbrio entre a técnica de gravação e a arte musical marca o empenho de músicos e engenheiros de som, os quais buscam sempre novos recursos tecnológicos para atingir os seus objetivos artísticos e comerciais no meio fonográfico.

Iniciamos este trabalho discutindo os aspectos perceptuais do áudio, os equipamentos de manipulação sonora usualmente encontrados em estúdios de gravação, e também alguns aspectos estético-musicais. Em seguida, sendo esta uma contribuição original à área, a descrevemos como um novo método de mixagem automática que visa a otimização da dinâmica musical e distribuição espectral do áudio digital. O modelo é implementado computacionalmente em na linguagem e ambiente computacional MATLAB¹, o qual permite ao programador criar *softwares* de propósitos gerais através de scripts que são executados dentro do próprio ambiente MATLAB. Mesmo sendo uma linguagem interpretada e de baixa performance computacional, MATLAB é bem usando no meio acadêmico e industrial pela sua facilidade e rapidez para prototipar *softwares* de alta complexidade matemática.

¹ Disponível em: <<https://www.mathworks.com/products/matlab.html>>. Acesso em: out. 2018

Para que os softwares desenvolvidos em MATLAB possam ter uma melhor performance computacional há uma extensão do *software (Toolbox)* chamado “MATLAB CODER”, o qual permite converter o código MATLAB para C ou C++. Sendo C e C++ linguagens compiladas. De acordo com Andrews (2012), um software escrito em MATLAB pode ser executado 500 vezes mais rápido caso seja compilado para a linguagem C++. Há casos ainda em que parte do código MATLAB pode ser executado em paralelo. Nestes casos, a MathWorks diz em seu site que com a instalação física de uma GPU (*Graphics Processing Unit*) a velocidade de execução do *software* pode chegar a aproximadamente 760 vezes mais rápida² que a velocidade de execução original.

Para que esse *software* se torne uma ferramenta viável para a livre distribuição, é necessário que ele possa ser executado fora do ambiente de programação do MATLAB. Futuramente pretendemos criar uma interface gráfica para o usuário final e utilizar o MATLAB CODER para criar uma versão *stand-alone* (independente) deste *software*, versão a qual já estará compilada em C++ e que funcionará independente do ambiente de programação MATLAB e para múltiplos sistemas operacionais como Windows, Mac OSX e Linux. Neste momento apenas para mostrar a viabilidade do método o nosso software roda apenas por linha de comando.

Acreditamos que a linha de pesquisa em mixagem automatizada possa contribuir ao meio musical em três frentes:

1. Redução da complexidade técnica e subjetiva do processo de mixagem, fornecendo ao músico usuário uma sonoridade agradável de sua música, reduzindo assim o tempo dispendido em sua produção
2. Redução dos custos de uma mixagem adequada para a indústria fonográfica: a implementação destes conceitos técnico-científicos em um software irá propiciar a músicos leigos na prática de mixagem um bom resultado sonoro das suas próprias gravações, sem ter que recorrer à grandes estúdios ou à contratação de especialistas em mixagem. Essa redução de custos ajuda a viabilização de grupos com poucos recursos financeiros a terem o registro de sua prática musical com uma qualidade superior àquela mais comumente alcançada quando a mixagem não é realizada.

² Disponível em: <<https://www.mathworks.com/help/distcomp/examples/illustrating-three-approaches-to-gpu-computing-the-mandelbrot-set.html>>. Acesso em: out. 2018.

3. Otimização do trabalho do engenheiro de som: a partir do momento em que o software consegue fazer um equilíbrio inicial dos instrumentos gravados, o engenheiro de som pode se dedicar exclusivamente ao refinamento e ao seu gosto pessoal no momento da mixagem.

Esse método e ferramenta computacional não intenciona substituir o elemento humano no processo de mixagem, mas pretende tornar-se um mecanismo auxiliar ao engenheiro de som, contribuindo para melhorar a sua performance e a qualidade do seu trabalho. Por fim, é importante ressaltar que o foco principal deste projeto não é o de dar uma solução definitiva ao problema da mixagem, mas sim o de fornecer as bases científicas que estruturarão o futuro algoritmo de mixagem automatizada.

O estudo da mixagem automatizada neste trabalho transita entre 3 áreas do conhecimento: a) a música enquanto objeto estético, b) a psicoacústica como resultado da interação entre a fisiologia auditiva e interpretação cerebral do estímulo sonoro, c) e a engenharia de som e da computação como meio para descrever e manipular os fenômenos físicos estudados (amplitude e espectro sonoro) e como meio de articular novos métodos e soluções a este objeto de estudo.

Figura 1. A mixagem e as áreas do conhecimento



Fonte: Própria

1.1 BASES PARA AVALIAÇÃO SUBJETIVA DO SOM

Segundo Fastl; Zwicker (2007) e Stevens; Warshofsky (1968) a audição humana é um fenômeno psicoacústico, ciência na qual estuda-se a relação entre as sensações auditivas e os

fenômenos físicos do som. A audição se dá por um conjunto de fenômenos complexo envolvendo a física da geração e propagação do som no ar como ondas longitudinais de pressão, sua captação pelo ouvido humano, transformação do sinal acústico em impulsos elétricos e que, por sua vez, são interpretados no cérebro como som.

A origem das pesquisas que relacionam os fenômenos físicos, como as ondas acústicas, com sua percepção subjetiva, como o som, advém da Psicofísica. Esta disciplina estuda a relação entre os estímulos físicos como a luminosidade, o peso ou excitação sonora e a experiência sensorial humana, levando uma grandeza física para uma dimensão psicológica. Perguntas como: “quão clara a sala está?”, “quão pesado é este objeto?” ou “quão alto é este som?” são exemplos simples de como as pessoas relacionam, intuitivamente, um fenômeno físico com à sua contraparte no âmbito psíquico. Neste sentido, Gescheider (1997) define a psicofísica como a ciência que estuda a relação entre o estímulo físico e a sensação ou percepção que eles produzem. De acordo com Boff et al. (1986) a psicofísica também se refere a uma classe geral de métodos que podem ser aplicados para estudar um sistema perceptivo. Para esta ciência o homem é um instrumento de mensuração que produz resultados (experiências, julgamentos, respostas, decisões) que podem ser analisados de uma forma sistemática.

Os primeiros experimentos psicofísicos foram realizados pelo médico alemão Ernst Heinrich Weber (1795-1878), cujos estudos abrangeram, principalmente, as sensações táteis. Já, Fechner (1860) em seu livro “Elementos de Psicofísica” mostrou os métodos e a teoria da mensuração da sensação, fornecendo as ferramentas básicas para o estudo da mente. Por sua vez, Stevens (1961) e Stevens; Volkman (1940) desenvolveram métodos indiretos para a mensuração da sensação de altura (*pitch*) e nível de audibilidade ou intensidade de nível sonoro percebida (*loudness*) como pode ser vista em Stevens (1959). Depois de Stevens, foram desenvolvidas outras formas mais precisas de se calcular *loudness*, como o Método de Zwicker, o qual foi responsável pela normatização de *loudness* em uma ISO em 1975. A versão revista e atualizada desta norma pode ser encontrada em International Organization for Standardization (2017). Como ainda não há um pleno consenso sobre como representar numericamente o nível sonoro percebido, Skovenborg; Nielsen (2004) comparam e discutem as variadas formas atuais de se calcular *loudness*.

Um dos temas mais atrativos da psicofísica é a relação entre a magnitude da resposta subjetiva e a magnitude física do estímulo. Os estudos relacionados a este tema visam obter equações que relacionem a intensidade dos estímulos com suas percepções para serem aplicadas em vários campos da ciência, como medicina, sistemas de comunicação direito e artes, especialmente, na música gravada e seu tratamento, a qual é a área deste estudo.

Cabral (1996) cita que Weber divulgou seus experimentos realizados entre 1829 e 1834 sobre as sensações cutâneas e musculares. No livro intitulado “*De Pulsu, Resorptione, Auditu et Tactu*”, Weber (1834) descreve a relação entre o sentido muscular sobre a estimação subjetiva de pesos, onde ele buscava entender se a discriminação de pesos era influenciada pelo sentido muscular. Em seus experimentos, os indivíduos comparavam os pesos com e sem o uso da musculatura, o que mostrou que a sensibilidade ao peso era muito mais acurada quando o sentido muscular participava ativamente, isto é, quando os próprios sujeitos levantavam os pesos. No livro “Dicionário Técnico de Psicologia”, Cabral resume a lei de Weber como:

A diferença apenas perceptível (DAP) da sensação ocorre quando o estímulo é aumentado ou diminuído por certa proporção de si mesmo, sendo tal proporção constante em qualquer sentido determinado.
Cabral (1996).

Em resumo a Diferença Apenas Perceptível (DAP) ou Limiar Diferencial é o aumento necessário na estimulação física para que uma pessoa perceba este aumento. Esta lei aplica-se aos 5 sentidos, mas as suas implicações são melhor entendidas quando se referem aos estímulos provocados pela luz e pelo som. Mais tarde, esta relação foi expressa em termos matemáticos por Gustav Fechner (1801-1887), médico e professor de Física da Universidade de Leipzig, a qual foi denominada como Lei de Weber (as vezes chamada de Lei de Weber-Fechner), conforme a equação abaixo.

$$k = \frac{\Delta e}{e} \quad (1)$$

Onde

k é a constante de Weber (adimensional)

Δe é a magnitude da diferença percebida entre o peso padrão e o peso de comparação

e é o peso tomado como padrão

Observe que k é adimensional, isto é, não tem uma unidade física (graus, cm, dB, Hz, etc.). No entanto, seu valor é diferente para cada tipo de estímulo e uma quantidade de referência usada como unidade de medida. Segundo Middlebrooks; Green (1991), no caso do

fenômeno sonoro, a diferença de intensidade mínima referente ao limiar diferencial (DAP) para humanos é de cerca de 1 dB.

Fernandes (2010) exemplifica a Lei de Weber em relação à frequência do estímulo sonoro dizendo que:

“[...] o intervalo entre as frequências não se mede pela diferença de frequências, mas pela relação entre elas. Desta maneira, se define uma oitava como sendo o intervalo entre frequências cuja relação seja igual a 2.” Fernandes (2010).

$$\frac{200}{100} = \frac{400}{200} = \frac{800}{400} = 2 \implies 1 \text{ oitava}$$

A partir da expressão acima, podemos deduzir que, na escala temperada temos:

$$1 \text{ semitom} = 2^{\frac{1}{12}} \text{ de oitava}$$

Segundo Benesty et al. (2007), no âmbito tonal, o limiar diferencial necessário para a comparação psicoacústica entre 2 tons abaixo de 500 Hz é de cerca de 3 Hz para ondas senoidais e 1 Hz para tons complexos. Acima de 1000 Hz (1 kHz), a DAP para ondas senoidais é de cerca de 0,6% da frequência. Neste caminho Olson (1967) diz que o número total de passos de tom perceptíveis na faixa de audição humana é de cerca de 1.400 e que o número total de meio tons na escala temperada, de 16 Hz a 16.000 Hz, é de 120 semitons.

Segundo o manuscrito de Silva; Rozestraten (2014) após as experiências de Weber abrirem o caminho para o nascimento da psicofísica, Fechner desenvolver seus fundamentos teóricos e experimentais. Experimentalmente, Fechner tinha como objetivo determinar a relação existente entre o estímulo, como entidade física mensurável em unidades físicas e a sensação causada por este estímulo. Como não é possível obter a magnitude exata desta sensação, procurou saber se, primeiramente, ela estava presente ou não, e então quantificá-la de alguma forma. A ideia foi entender quanto um estímulo tem que crescer para que uma mudança sensorial seja percebida, ou seja, determinar a diferença apenas perceptível. Foi esta diferença apenas perceptível que Fechner aceitou como unidade de medida da sensibilidade. Se quaisquer pares de estímulos são igualmente discrimináveis, então a distância psíquica entre estímulos em cada par será igual a zero. A lei baseia-se sobre duas hipóteses: a diferença em magnitude psíquica entre duas magnitudes físicas é proporcional à amplitude do estímulo

tomado como padrão no momento do julgamento. A segunda hipótese é a própria Lei de Weber, que a razão entre a DAP e a magnitude padrão é uma constante. Razões iguais entre os estímulos produziram intervalos de sensações iguais ou, ainda, quando a intensidade dos estímulos aumenta geometricamente a intensidade das sensações, por sua vez, aumenta aritmeticamente. Segundo Stevens (1975), a Lei de Fechner pode ser expressa como uma relação logarítmica em que:

$$R = k * \log(E) \quad (2)$$

Onde

R é a sensação/resposta subjetiva

k é a constante de Weber

E é o estímulo correspondente à grandeza física objetiva

Ou seja, a Lei de Fechner para a acústica nos diz que para sons simples de mesma frequência fundamental, a intensidade da sensação sonora cresce proporcionalmente ao logaritmo da intensidade física. Ou ainda: sons de frequência constante, cujas intensidades físicas variam em progressão geométrica, produzem sensações cujas intensidades subjetivas variam em progressão aritmética.

Embora a lei de Fechner não tenha sido contestada por aproximadamente 100 anos, Stevens (1961) publicou o artigo com o título “*To honor Fechner and repeal his law*” (em tradução livre: “Em honra a Fechner e revogação de sua lei”). Para Stevens ao invés de uma função logarítmica é uma função de potência que descreve a relação entre a magnitude da sensação e a intensidade do estímulo. Stevens chegou a esta conclusão após a realização de inúmeros experimentos e elaboração de vários métodos que podiam mensurar diretamente a sensação. A Lei de Potência de Stevens, conhecida como a terceira lei da psicofísica, é dada por:

$$R = k * E^n \quad (3)$$

Onde

R é a magnitude da resposta sensorial subjetiva

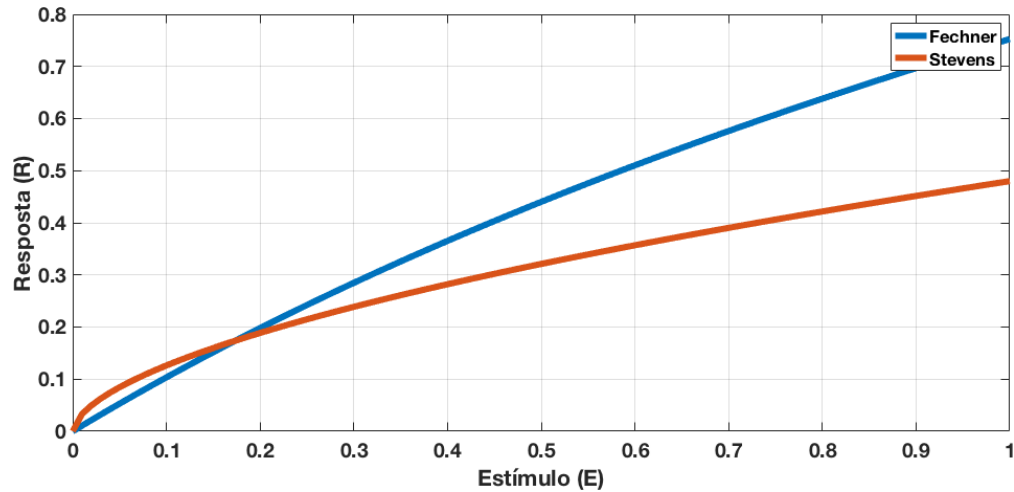
k é uma constante arbitrária que depende da unidade de medida utilizada

E é a magnitude física do estímulo objetiva

n é o expoente da função

Para o caso do som (vide Tabela 1), $n = 0,56$. Para $0 < n < 1$ as funções dadas pelas equações (2) e (3) têm valores muito próximos em um intervalo de pequenos estímulos. A Figura 2 mostra uma comparação qualitativa das equações de Fechner e Stevens.

Figura 2. Comparação qualitativa das equações de Fechner e Stevens



Na Tabela 1 vemos os dados da função-potência de Steven revisados por Silva; Macedo (1982). Nela encontramos os expoentes da função de potência para uma variedade de modalidades perceptivas.

Tabela 1 - Expoentes da função-potência para uma variedade de modalidades perceptivas baseados em julgamentos de razão

Modalidade ou Atributo	Expoente Médio (<i>n</i>)	Desvio-Padrão
Som	0,56*	0,13
Tonalidade	0,77	0,51
Sabor: sacarose	0,97	0,49
Sabor: sacarina	0,53	0,29
Sabor: salgado	1,07	0,41
Sabor: amargo	0,57	0,23
Sabor: azedo	0,94	0,28
Odor: amil-acetato	0,19	0,14
Odor: benzeno	0,56	0,01
Odor: heptano	0,35	0,26
Odor: octanol	0,14	0,11
Odor: geraniol	0,19	0,12
Tempo	0,91	0,18
Brilho	0,26	0,13
Temperatura: frio	1,04	0,05
Temperatura: quente	1,14	0,37
Numerosidade	0,84	0,22
Número	0,72	0,21
Peso	1,18	0,17
Aceleração angular	1,39	0,52
Choque elétrico	2,80	0,99
Aspereza de superfície	1,61	0,76
Comprimento visual	1,01	0,27
Área visual	0,77	0,16
Volume visual	0,72	0,10
Distância visual	0,97	0,22
*Valor correspondente a 0,28 em termos de intensidade sonora		

Fonte: Silva; Macedo (1982) - modificada

Em uma função de potência, sensações que exigem uma variação de energia muito grande têm baixos expoentes como ocorre com estímulos visuais (brilho, $n = 0,26$) e estímulos sonoros ($n = 0,56$) visando à proteção dos órgãos. É como se houvesse um transdutor sensorial

para evitar uma sobrecarga do sistema nervoso central, que atua como um compressor de áudio, o que faz com que o expoente seja menor que 1,0. Um aprofundamento desse assunto pode ser encontrado em Silva (1982), onde o autor faz uma descrição das funções de potência e das funções exponenciais usadas para estimar eventos psicofísicos.

Os estudos de Weber, Fechner e Stevens, além de enriquecerem campos de várias ciências, em especial da Psicofísica, abriram caminho para o conhecimento do fenômeno sonoro, sua qualificação e mensuração, desenvolvimento de equipamentos acústicos e métodos que permitem não só o aprimoramento das técnicas de gravação, mas fornecem as bases dos estudos técnicos de mixagem musical.

1.2 GRAVAÇÕES MUSICAIS

As gravações de áudio demoraram mais de 100 anos para atingir o nível de qualidade e precisão que apreciamos nos dias de hoje, na música digital e nos serviços de *streaming* online. No entanto, os anos entre 1877 e 1925 foram marcados pelo aparecimento de pesquisas e desenvolvimento de mecanismos puramente mecânicos de gravação e reprodução sonora. O fonógrafo de Thomas Edison, criado em 1877, foi o primeiro invento relevante desse período. O protótipo consistia em um recipiente cônico ou corneta (Figura 3) que recebia o som em uma extremidade, fazia vibrar uma membrana fixada na outra, que estava articulada com uma pequena agulha que riscava papéis, cera ou metais macios, dispostos em forma de cilindro. Estes riscos, ou ranhuras, podiam ser novamente convertidos em som através de processo inverso, ou seja, enquanto o cilindro gravado girava, a agulha passava por entre as ranhuras fazendo o cone vibrar e transformar a energia mecânica em energia acústica (som). A eficiência energética desse processo era baixa e por isso o som reproduzido tinha baixa intensidade e pouca fidelidade de gravação. Este período culminou com o aparecimento do disco, que se tornou o meio padrão para gravação e principalmente de reprodução do som gravado, dominando o mercado de áudio até o final do século XX.

Figura 3. Fonógrafo de Thomas Edison (1877)



Fonte: Breda (2018)³

O segundo período de desenvolvimento (1925 a 1945) foi marcado pelo uso da eletrônica no processo de gravação. Esta tecnologia permitiu que os processos de gravação se expandissem não apenas na área do áudio, mas também para filmes. De acordo com Adams; Butler (1999), surgiram os microfones elétricos, amplificadores de sinal elétrico e os gravadores eletromecânicos da Western Electric, empresa norte americana fundada por Elisha Gray em 1869. O som agora podia ser capturado, amplificado, filtrado e balanceado eletronicamente, e o dispositivo de corte (ranhuras) do disco era eletricamente acionada, porém ainda era fisicamente gravado em um disco. Discos de áudio para a população em geral foram produzidos mecanicamente em massa por estampagem de um composto à base de goma-laca e, posteriormente, plástico polivinílico, comumente chamado de vinil.

Os microfones elétricos melhoraram dramaticamente a captação da performance dos cantores, enquanto que os amplificadores eletrônicos permitiam que instrumentos musicais de menor projeção sonora, como violão e os instrumentos orquestrais da família das cordas, fossem tocados em equilíbrio de volume com os instrumentos de sopro, os quais possuem maior projeção sonora. Ao mesmo tempo músicos e compositores também começaram a compor e gravar músicas eletrônicas utilizando instrumentos eletrônicos/elétricos tais como o Teremim (descrito a seguir), órgão eletrônico e o Hammond Novachord, o qual segundo Morris (2014) foi o primeiro sintetizador analógico polifônico comercial do mundo.

³ Disponível em: <<https://www.portalsaofrancisco.com.br/biografias/thomas-edison>>. Acesso em mai. 2018.

De acordo com Glinsky (2000), o Teremim (ou *Theremin*) foi um dos primeiros instrumentos musicais eletrônicos do mundo. Ele foi inventado por Léon Theremin em 1928 e seus controles são basicamente duas antenas perpendiculares nas quais o músico controla a frequência e a amplitude de tons puros através da posição e movimentação de suas mãos no sistema.

Neste período, mais notadamente na década de 1920, surge o transdutor eletroacústico ou alto-falante, o qual possui o mesmo princípio de funcionamento de um microfone dinâmico, porém no sentido reverso. Quando uma corrente elétrica, representando o sinal de áudio amplificado passa pela bobina do alto-falante, esta cria um campo eletromagnético variável, o qual faz com que o cone anexado à um ímã se mova para trás e para frente, de modo diretamente proporcional à variação desta corrente. Este movimento gera ondas de pressão acústica de várias frequências que viajam pelo ar até nossos ouvidos, que as "ouvem" (interpretam) como som. Embora tenha havido inúmeros aperfeiçoamentos, o alto-falante dinâmico não mudou substancialmente em quase 100 anos, e continua sendo o meio mais comum, preciso e confiável para converter sinais de áudio eletrônicos em som audível.

O terceiro período relevante da história da gravação se dá entre 1945 a 1975 com o surgimento da fita magnética, em 1930 na Alemanha. A descoberta proporcionou uma qualidade de áudio gravados praticamente indistinguível das transmissões ao vivo. Seu princípio de funcionamento se dá pela reorientação do material ferromagnético na película. Através de um processo eletromagnético similar ao descrito acima, porém o leitor da fita gera uma variação de tensão de acordo com a orientação do material ferromagnético da película, gerando o sinal de áudio.

A facilidade em gravar e regravar áudio na fita magnética fez com que, a partir de 1950, ela se fosse tida como o meio padrão de gravação de áudio nas indústrias de rádio e música, levando ao desenvolvimento das primeiras gravações de alta definição (*hi-fi*)⁴ estéreo para o mercado doméstico. Logo após temos o desenvolvimento da técnica de gravação de fita *multi-track*⁵ para música, superando disco de vinil como o principal meio de gravação musical.

A fita magnética também trouxe uma reformulação radical do processo de gravação. Possibilitou gravações de longa duração e fidelidade muito maior do que as tecnologias anteriores e ofereceu aos engenheiros de gravação a mesma plasticidade excepcional que o filme deu aos editores de cinema. A partir desse momento sons capturados em fita poderiam

⁴ *High Fidelity*

⁵ Multipista

ser mais facilmente manipulados, editados e combinados, o que era impossível em gravações em disco. Segundo Paludo (2010), os experimentos baseados em fitas magnéticas atingiram um pico na década de 1950 com as gravações do casal Les Paul (guitarra) e Mary Ford (vocal), pioneiros no uso de edição de fita e gravação multipista (*multi-tracking*) para criar grandes conjuntos "virtuais" de vozes e instrumentos, construídos inteiramente a partir de gravações de suas próprias vozes e instrumentos.

Outros avanços simultâneos na tecnologia de áudio levaram à introdução de uma gama de novos formatos e dispositivos de áudio, tanto em disco quanto em fita, incluindo a invenção dos discos de 33 RPM⁶ de 12 polegadas, também conhecidos como *Long Play* (LP) e os *Single* ou Compacto Simples, de 7 polegadas tocados em 45 RPM. Por fim, a portabilidade dos gravadores de fita cassete (ou K7) viabilizaram gravação de alta fidelidade de performances ao vivo, tornando-as uma boa alternativa aos enormes discos de vinil.

A gravação de áudio passa ainda hoje pelo seu quarto período de desenvolvimento iniciado em 1975 e denominado como o Período Digital. Esta fase, praticamente substituiu as tecnologias de gravação anteriores em apenas 40 anos. Segundo Immink (2010), a principal diferença entre áudio analógico e áudio digital está no método de captura e armazenagem do sinal. Enquanto no áudio analógico, a variação de tensão vinda do microfone é registrada continuamente no disco ou na película, no áudio digital é feito um processo intermediário entre o sinal analógico e o gravador, esse processo é chamado de amostragem digital onde o conversor analógico-digital (ADC⁷) captura amostras do sinal (*samples*) em um período de tempo regular. No caso do CD são capturadas 44.100 amostras por segundo, ou seja, uma amostra a cada 22.67 microssegundos. No caso do Blu-ray podemos chegar a capturar 192.000 amostras por segundo, ou seja, uma amostra a cada 5.20 microssegundos.

Depois de ter segmentado o sinal analógico contínuo em um conjunto de amostras discretas, o ADC envia esse conjunto de amostras (sinal digitalizado) para um gravador digital, sendo o principal deles o computador. Quando é solicitado a reprodução desse sinal digital, o computador envia esse conjunto de amostras para o conversor digital-analógico (DAC⁸) e lá elas são interpoladas no tempo para formar um sinal elétrico contínuo e variável no tempo. Quando esse sinal elétrico passa pelo alto-falante, o cone se movimenta formando variações de pressão atmosféricas correlacionadas à variação do sinal elétrico. Temos assim a energia

⁶ Rotações Por Minuto

⁷ *Analog to Digital Converter*

⁸ *Digital to Analog Converter*

elétrica se transformando em energia mecânica, ou seja, energia acústica, a qual, ao escutarmos, interpretamos como som.

Nos anos 80 as empresas Sony e Philips introduziram o seu novo sistema de gravação digital: o formato de áudio para disco compacto digital (CD⁹). O disco compacto rapidamente substituiu os outros sistemas porque eram pequenos, portáteis e duráveis, e podiam reproduzir o áudio em todo espectro sonoro, com uma larga faixa dinâmica sem distorção. Como os CDs são codificados e lidos óticamente usando um feixe de laser, onde (ao contrário dos LPs e dos *tapes*) não há contato físico direto entre o disco e o mecanismo de reprodução, de modo que um CD bem cuidado poderia ser reproduzido repetidamente, sem degradação ou perda de fidelidade. Os CDs também representaram um avanço considerável, tanto no tamanho físico do meio quanto em sua capacidade de armazenamento. Segundo Immink (1999), outra vantagem interessante é que os CDs têm menos da metade do tamanho geral do antigo formato LP de 12", mas ofereciam cerca do dobro da duração do LP médio, com até 80 minutos de áudio. Segundo Immink (2007), a convenção da duração de um CD de áudio foi influenciada pela necessidade de se ter a Nona Sinfonia de Beethoven gravada em CD, o que atrairia a atenção e eventual aprovação do audiófilos alemães.

Na década de 90 o surgimento do arquivo digital (mp3 e outros formatos) e os algoritmos de compressão de sinal digital revolucionam o modo de distribuição musical no mercado. Novos reprodutores musicais, como o Apple iPod (2001), permitiram que os a população em geral pudesse adquirir e armazenar grandes quantidades de mídia digital de alta qualidade (áudio, vídeo, jogos e outras aplicações) e acumular bibliotecas de mídia de dezenas ou mesmo centenas de milhares de músicas.

Assim, o arquivo de áudio digital marcou o fim de uma era na gravação e o início de outra, eliminando efetivamente a necessidade de criar ou usar uma mídia de gravação física específica e se tornando o principal meio de capturar, processar, gerar e distribuir gravações comerciais de som. Por fim, surge neste último período os serviços de transmissão (*streaming*) de música digital, ganhando popularidade a partir do ano de 2010. De acordo com Harold Richardson (2014), esses serviços não têm interesse que o ouvinte baixe ou possua os arquivos de áudio, mas, em vez disso, que eles ouçam música pela Internet. Para conseguir este objetivo, os serviços de *streaming* normalmente oferecem um consumo limitado de música gratuitamente para, em seguida, tentarem vender serviços *premium* pagos ilimitados Swanson (2013).

⁹ Compact Disc

A história da evolução da gravação e reprodução dos sons e, particularmente, da música gravada, é repleta de casos muito interessantes de ciência, uso de tecnologias tais como a invenção do disco de vinil, fita K7, laser, de inventores visionários como Thomas Edison e grandes indústrias fonográficas que fomentaram o comércio musical e consequentemente o avanço na qualidade de áudio gravado. Por esse motivo o avanço das pesquisas acadêmicas nesta área pode contribuir para a descoberta e criação de ferramentas que tragam melhorias para o processo de gravação e mixagem na música comercial.

1.3 O ESTADO DA ARTE

Em linhas gerais, podemos dividir o estado da arte do processo de mixagem em duas grandes áreas: a da mixagem convencional e da mixagem automatizada por computador. Como descrito anteriormente, nos últimos 50 anos os estudos eletrônicos e acústicos elevaram tanto a qualidade dos equipamentos de áudio quanto o tratamento acústico das salas de gravação e mixagem atingindo hoje gravações e mixagens de altíssima qualidade. O estudo formal e acadêmico de mixagem em instituições de ensino estrangeiras (como *Queen Mary University* e *Berklee College*) possibilita a formação de profissionais altamente capacitados e preparados para trabalhar com uma grande gama de equipamentos e técnicas de gravação, o que ajuda a consolidar a área de engenharia de áudio e mixagem como uma área do conhecimento acadêmico.

A área que trata da mixagem automatizada por computador é mais recente devido à complexidade de simular a audição e decisões humanas no processo de mixagem, os quais serão detalhados nos próximos capítulos. No próximo capítulo citamos alguns trabalhos acadêmicos que visam simular computacionalmente partes da audição humana e sua interação com a mixagem automatizada.

2 FUNDAMENTOS TEÓRICOS

Neste capítulo iremos apresentar as bases teóricas deste trabalho, as quais serviram de estudo e entendimento da fisiologia auditiva humana e também contribuíram para a criação do nosso novo método de mixagem automatizada. Como o foco deste trabalho é exclusivamente a mixagem musical descrevemos abaixo a evolução do pensamento sobre o equilíbrio musical na música ocidental.

2.1 ESCOPO ESTÉTICO-MUSICAL

Visto que este trabalho, como outros anteriores, não visa dar uma solução definitiva ao problema da mixagem automatizada, definimos um escopo musical reduzido de maneira que os conceitos e métodos nele apresentados possam ser aplicados e validados dentro deste limitado domínio musical. Assim, os limites desta pesquisa restringem-se apenas ao âmbito da chamada música comercial, podendo este método ser aplicado em vários estilos musicais correlatos como: *pop*, *rock*, *country*, *jazz* e outros. Escolhemos este domínio musical devido ao fato que o processo de mixagem dos seus variados estilos tem muitos fatores, ou semelhanças, técnicas como, por exemplo, a alta taxa de compressão dos sinais musicais e preenchimento sonoro de todo espectro auditivo, o que facilita o nosso teste de validação do método proposto. O detalhamento dessas características e processos de manipulação sonora serão apresentados nos próximos capítulos.

Complementando nosso argumento da restrição de domínio de aplicação de nosso método, fora do escopo deste trabalho está a mixagem de músicas nas quais estes fatores/semelhanças técnicas não estão sempre presentes. Um exemplo claro disso é o âmbito da música erudita de concerto, onde quase não há compressão de dinâmica do áudio gravado, pois o ouvinte espera sentir a real variação de dinâmica da orquestra ou do solista, como se ele estivesse numa sala de concerto. Sobre o preenchimento espectral, não há necessidade de manipulação espectral pois o compositor erudito, em princípio, já atingiu o balanço espectral desejado no processo de orquestração da obra, onde ele anotou na partitura quais instrumentos musicais iriam tocar cada trecho musical, obtendo uma gama espectral rica desde o grave (Contrabaixos, *Gran Cassa*, Tubas, etc...) até o agudo (*Piccolo*, Flautas, Violinos, etc...).

É importante ressaltar que este trabalho não está excluindo a possibilidade deste método ser aplicado em sonoridades orquestrais, como por exemplo a música cinematográfica orquestral, mas sim a obra musical criada com a concepção de não ser manipulada

artificialmente de nenhuma forma ou de não ser puramente acústica. Mais detalhes sobre esses fatores técnicos serão abordados nas próximas sessões.

2.2 A BUSCA PELO EQUILÍBRIO MUSICAL

O equilíbrio musical é alvo de investigação desde a antiguidade grega. O regente Tilson Thomas (2012)¹⁰ em sua palestra no TED intitulada “*Music and emotion through time*” descreve essa busca grega e os conceitos de equilíbrio musical inspirado no ideal da chamada “Música das Esferas”, em analogia ao equilíbrio dos céus com seus planetas e estrelas. Ainda no caminho do controle e equilíbrio musical Bhatara et al. (2011) demonstra experimentalmente a importância do controle da dinâmica e tempo para a expressividade musical no seu trabalho intitulado “*Perception of Emotional Expression in Musical Performance*”. Por fim, em um sentido mais amplo e cognitivo, Peretz (2009) em sua palestra no “*Center for Academic Research and Training in Anthropogeny*” descreve como músicas com frases e estruturas bem balanceadas (“*Mozart Effect*”) podem melhorar levemente o desempenho cerebral do ouvinte para algumas atividades dentro de um pequeno período de tempo como pode ser visto em Jenkins (2001).

2.3 A ESCRITA CORAL E ORQUESTRAL

Na história da música ocidental podemos observar uma evolução cronológica da escrita coral e orquestral desde o cantochão do século III até as grandes peças orquestrais modernas. Segundo Grout; Palisca (2014), o desenvolvimento da escrita coral, principalmente durante os séculos XIV e XV, é um exemplo de como os compositores eruditos daquela época buscavam o equilíbrio sonoro combinando naipes de vozes diferentes, escrevendo linhas homofônicas ou contrapontísticas, padronizando a proporção de integrantes por naipe e criando linhas vocais que enfatizassem o timbre de cada naipe dentro da sua tessitura vocal. Dessa forma, a notação e o equilíbrio sonoro coral constituiu-se como a base do desenvolvimento da música ocidental instrumental. Um exemplo desse equilíbrio sonoro pode ser visto na coletânea de partituras de corais de J. S. Bach (1685 - 1750) publicada por Richter (1966). Por mais que já houvesse produção de música instrumental antes do século XVI, é a partir da obra de Bach que notamos uma quantidade maior de obras escritas para diversas formações instrumentais e também quando a orquestra toma uma forma mais definida em relação à sua formação instrumental. Na

¹⁰ Tilson Thomas, “Music and Emotion through Time”.

orquestra, com a expansão dos registros possíveis, os compositores do século XVI em diante tiveram que buscar novas formas de equilíbrio sonoro (tessituras, densidades, volumes, etc), desde características simples como a proporção de integrantes por naipe de cada instrumento da orquestra como apresentado, por exemplo, em Rimsky-Korsakov (1873), até processos mais complexos como a combinação harmônica e timbrística de dois ou mais instrumentos orquestrais tocando simultaneamente, como mostrada em Piston (1955).

De qualquer forma, tanto na escrita coral quanto na escrita orquestral o compositor é quem majoritariamente determina a sonoridade da peça e consequentemente seu “equilíbrio musical/sonoro”, pois ele conhece o som dos instrumentos escolhidos (qualidade sonora do instrumento) e possui uma referência estética de como esses instrumentos deverão soar em um determinado espaço de audição (sala de concerto, igreja, sala de teatro, etc.). A partir do momento em que se perde a noção da qualidade sonora do instrumento musical ou quando não há conhecimento prévio das características do ambiente em que a obra será executada, perde-se consequentemente a materialidade sonora da correspondente notação musical escrita na partitura, ou seja, não há mais como saber como a peça descrita apenas em notação musical irá soar ao ser executada.

É a partir deste ponto, ou seja, a insuficiência da notação musical para garantir o “equilíbrio sonoro”, que atualmente nos deparamos com um dos problemas enfrentados pelos compositores de música comercial, a saber, onde o equilíbrio sonoro desejado da obra não basta ser notado no papel, mas sim gravado em forma de áudio para que assim ela possa ser reproduzida corretamente de acordo com a idealização do produtor musical.

2.4 O REGISTRO MUSICAL EM ÁUDIO

Como já descrito no Capítulo 1, a partir da primeira metade do século XX, o avanço tecnológico possibilitou diversos meios de registro musical iniciando também a busca pela maximização da qualidade sonora gravada. Com o tempo, novas técnicas de gravação, mixagem e masterização foram incorporadas a essa atividade resultando não apenas em um registro sonoro mais fidedigno da execução musical, mas também a criação de novas sonoridades que não existiam na música acústica tradicional, tais como o chamado “efeito de proximidade do microfone à fonte sonora”, *delays* (atrasos), distorções (saturações dos sinais) intencionais, modificações espectrais e outras manipulações sonoras.

Novos equipamentos de áudio abriram possibilidades de manipulação do mesmo no domínio do tempo, no domínio da frequência e em relação a magnitude da envoltória do sinal,

ressaltando ou encobrendo detalhes da execução musical, de modo a gerar novas sensações, emoções e expectativas nos ouvintes. Neste momento é notável a influência da tecnologia no processo de idealização e concepção de uma obra musical, visto que a notação tradicional em partitura já não representa na totalidade a sonoridade da obra gravada, mas se torna um dos elementos que compõe o resultado sonoro total, em outras palavras, a notação musical é o mapa e não o caminho.

Abaixo discutimos o modo com que os recursos tecnológicos são utilizados de diferentes formas no momento de se gravar um repertório tipicamente erudito e outro tipicamente comercial.

2.4.1 Música Erudita

Ao fazer o planejamento e o preparo de uma sessão de gravação de música erudita, o profissional de áudio precisa decidir sobre diversos detalhes que influenciarão no resultado sonoro final como: qual seria a sala ideal para gravar esse repertório, qual a melhor posição dos músicos nesta sala, quantos e quais microfones utilizar, qual o melhor posicionamento de cada microfone, etc. A escolha da distância de cada microfone à fonte sonora pode implicar na percepção do tamanho da sala na qual a música está sendo executada. Caso o engenheiro de som queira obter um resultado sonoro compatível com a prática de um período musical ou estética pré-estabelecida, seu planejamento de gravação precisará ser adequado a esse quesito musical.

Por um lado, ao gravar um repertório sinfônico romântico como “Nona Sinfonia de Beethoven” ou “Oitava Sinfonia de Mahler”, a estética já pré-estabelecida de outras gravações e a percepção sonora dessas obras executadas em grandes salas de concerto direciona a escolha de grandes salas de gravação e o uso de microfones distantes das fontes sonoras para que o fonograma reflita, ao invés de uma percepção mais individual e destacada, uma sonoridade de “massa orquestral”. Por outro lado, ao gravar um repertório clássico para piano solo, como a “Sonata nº 1 em Dó Maior de Mozart”, espera-se uma sala de gravação menor e com microfones mais próximos à fonte sonora, implicando numa sonoridade mais detalhada e intimista, a qual representa o espaço acústico típico de uma sala para música de câmara.

Feito um registro adequado da performance musical coral/orquestral, ou seja, um registro que respeite estas idiossincrasias, não são aplicados processos de mudança drástica no áudio, como “corretores de afinação” ou altas taxas de compressão, já que para a música erudita o foco é ter um registro de áudio que se aproxime da percepção que o ouvinte tem ao vivo na

sala de concerto. Neste caso, a maior parte do que podemos incluir como “mixagem” já foi realizada, inicialmente pelo compositor na sua notação orquestral da partitura e, em segundo, pelo engenheiro de som com as suas escolhas técnicas com implicações estético-musicais. Resta um trabalho mais detalhista de mixagem na qual o engenheiro de som irá apenas ressaltar efeitos particulares e detalhes determinados pelo compositor/intérprete e pelas técnicas de gravação empregadas.

2.4.2 Música Comercial

No âmbito da música comercial temos um cenário particularmente diferente do anteriormente descrito. Aqui não temos uma determinação estética clara como na música de concerto. Em muitos casos não há uma notação clara de quais instrumentos serão utilizados e nem quais os modelos, tipos ou naipes desses instrumentos são mais adequados a cada tipo de música. Também não há uma referência de qual seria a sala ideal para se ouvir determinado tipo de música e ainda, não há um equilíbrio natural entre as fontes sonoras majoritariamente utilizadas (por exemplo: bateria e vocal). Enfim, o número de sonoridades dentro da música comercial se torna imensa e o uso de equipamentos de manipulação de áudio abre ainda mais possibilidades musicais.

Holmes (2012) cita o uso do *vocoder* “*Siemens Synthesizer*” (equipamento originalmente voltado ao ramo de telecomunicações) como elemento de criação musical no “*Siemens Studio for Eletronic Music*”. Artistas como Stevie Wonder (“*Send one your love*”, “*A Seed’s a Start*”), Michael Jackson (“*Pretty Young Thing*”) e Coldplay (“*Major Minus*”) são exemplos de músicos que utilizaram esse tipo de tecnologia a favor da sua criação musical.

Com essa enorme gama de possibilidades sonoras, o trabalho de combiná-las de modo a agradar a maior parte dos ouvintes se torna uma tarefa não trivial, a qual demandará não apenas conhecimento teórico-musical, mas grande experiência com os equipamentos de áudio, com o tratamento da sala de mixagem, e também um conhecimento intuitivo/empírico da fisiologia do sistema auditivo humano e da percepção sonora. Nos próximos capítulos discutiremos os detalhes de cada uma dessas áreas do conhecimento necessárias para uma boa mixagem musical no âmbito da música comercial.

2.5 EQUILÍBRIO PSICOACÚSTICO

Ouvir é perceber a informação acústica (normalmente, as ondas de variação de pressão atmosférica) na forma de audição (potenciais de ação que caminham pelo nervo auditivo, da cóclea ao cérebro). Escutar é interpretar, através da atenção, aquilo que ouvimos. Escutar é um fenômeno perceptual, estudado pela psicoacústica. Escutar é um fenômeno mental, psicológico (portanto subjetivo) estudado pelas ciências cognitivas.

De uma forma geral, uma boa mixagem musical é aquela em que o ouvinte tem uma percepção de “equilíbrio” entre as fontes sonoras. Este equilíbrio perceptual tem, a princípio, diversas variáveis envolvidas. As principais delas são o chamado “volume percebido” (*loudness*), altura (*pitch*) e timbre. Além dessas, temos outras variáveis secundárias como o conteúdo espectral de cada fonte, localização ou “imagem” espacial dos instrumentos e outros parâmetros psicoacústicos sinestésicos como “brilhante”, “opaco”, “macio”, “metálico” e outros. Sabendo disso, o trabalho do técnico ou engenheiro de som no momento da mixagem é manipular o áudio gravado de forma a aproximar o fonograma do equilíbrio psicoacústico e estético desejado pelo artista.

Uma das dificuldades encontradas no processo de mixagem é que a escuta humana não responde de forma linear, e a ainda não se tem um completo entendimento do seu funcionamento devido a alta complexidade do sistema auditivo. No entanto, mesmo a compreensão básica que temos do funcionamento do sistema auditivo nos permite não apenas obter melhores mixagens musicais, mas também desenvolver métodos matemáticos e computacionais que possam emular, mesmo que parcialmente, esse sistema sensório-perceptual, como é o caso deste trabalho.

Nas subseções abaixo descrevemos os ingredientes fundamentais sobre os quais estão baseadas as técnicas de mixagem e que, portanto, são importantes também para a construção do nosso modelo.

2.5.1 Fisiologia Auditiva

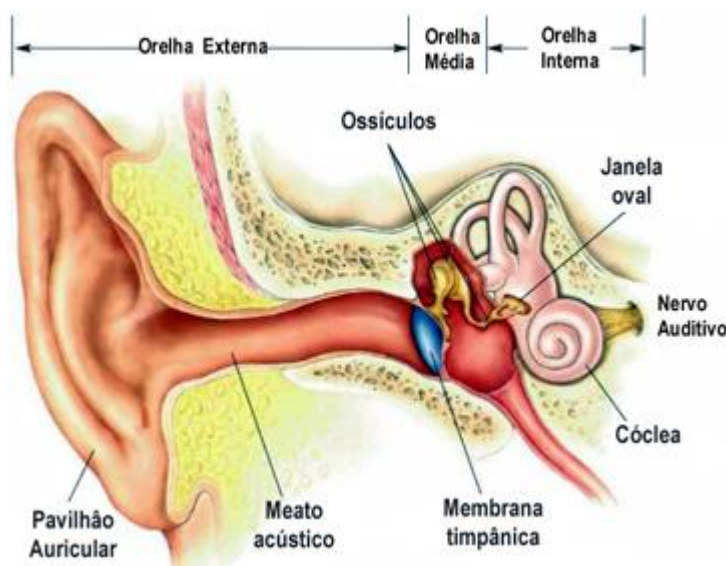
Como dito anteriormente, a mixagem é um processo que depende de ambiente adequado, equipamentos sofisticados e profissional competente e que possua boa e bem treinada audição. No entanto, até que os processos computacionais se desenvolvam ao ponto de dar à mixagem automatizada a qualidade exigida pelo mercado, o uso do sentido de audição neste processo torna-se indispensável para dar à música o equilíbrio psicoacústico necessário.

O papel do ouvido é captar o sinal acústico, o qual carrega energia acústica das moléculas do ar, em ondas longitudinais de pressão, amplificá-lo, transformá-lo em sinais

elétricos e encaminhá-los para o cérebro, o qual os interpreta como sons. Este processo é extremamente complexo e preciso. Dele participam as membranas timpânicas, oval e redonda, os ossículos, vestíbulo, líquidos (perilíngua e endolíngua), membrana basilar e células ciliadas (neurônios mecanoreceptores) que transformam e transferem a energia acústica em sinais elétricos. Referente à pressão sonora, a sensibilidade auditiva se estende de 20 μPa até 20 Pa, o que corresponde ao intervalo de 0 a 120 dB. Referente ao espectro de frequências, o ouvido humano é capaz de captar frequências no intervalo de 20 Hz a 20.000 Hz, o que é equivalente a cerca de 10 oitavas. Segundo Maia (2002), nenhum equipamento inventado pelo homem é tão abrangente e preciso ao mesmo tempo.

A orelha humana (ou ouvido humano), conforme mostrada na Figura 4, é constituída de três partes: orelha externa, média e interna.

Figura 4. A orelha humana



Fonte: Nishida et al. (2011)¹¹

Antigamente o conjunto das estruturas anatômicas relacionadas com a audição e o equilíbrio era conhecida como ouvido. Esta nomenclatura foi substituído pelo termo “orelha”, a qual é subdividida em orelha externa, média e interna. A orelha interna pode ser denominada de órgãos vestibulo coclear já que ela pode ser dividida em uma parte vestibular, responsável pelo equilíbrio e uma parte coclear relacionada com a audição. Na Figura 4 temos uma visão esquemática de toda estrutura que forma a orelha.

¹¹ Disponível em: <

http://www2.ibb.unesp.br/Museu_Escola/2_qualidade_vida_humana/Museu2_qualidade_corpo_sensorial_audica_o1.htm>. Acesso em: out. 2018

A orelha externa é formada pelo pavilhão aditivo, ou pino, e pelo meato acústico. Conforme Zorzetto (2006) a pino é formada por uma lâmina de cartilagem elástica de formato irregular, recoberta por uma fina camada de pele. Possui várias depressões e elevações, sendo a concha a maior depressão. Bento et al. (2013, p. 37) descreve a responsabilidade da Pina na captura das ondas sonoras do exterior e direcioná-las para dentro do meato acústico. Seu formato de concha permite uma amplificação do som de cerca de 14 dB em 4000 Hz. O meato acústico é uma estrutura circular de, aproximadamente, 25 mm de comprimento de extensão. É recoberto por um epitélio que possui pelos e glândulas de cerume que auxilia na captação de partículas estranhas e nocivas. Se consideramos o meato acústico como uma coluna de ar aberta em uma extremidade e fechada na outra, de comprimento 2,4 cm as frequências de ressonância na faixa audível são três: $f_1 = 3400$ Hz, $f_3 = 10.1$ KHz e $f_5 = 16.9$ KHz. De acordo com Bistafa (2011) a primeira ressonância do canal auditivo explica a maior sensibilidade do ouvido entre 3 e 5 KHz.

A orelha externa termina na membrana timpânica, que limita uma cavidade dentro do osso temporal que denominamos orelha média. A membrana timpânica é uma membrana fibrosa recoberta de pele, a qual pode ser vista pelo médico através do aparelho *otoscópio*. Ela é importante para transmissão da onda sonora externa, ou seja, a membrana timpânica vibra sob as pressões da onda acústica externa e transmite suas vibrações para a orelha média. Ela situa-se entre a orelha externa e interna e é o local onde estão alojados os três menores ossículos do corpo humano conhecidos como martelo, bigorna e estribo, os quais recebem a vibração da membrana timpânica e transmitem esse sinal, amplificado cerca de 20 vezes, para os líquidos da cóclea alojada na orelha interna.

A orelha média possui dois músculos: o tensor do tímpano e o músculo estapédio que tem como função controlar a intensidade da energia mecânica transmitida para essa sessão da orelha. Quando uma alta pressão acústica chega ao tímpano ou vai aumentando gradativamente estes músculos se contraem limitando a vibração da membrana timpânica e dos ossículos, evitando assim a ruptura desta membrana e também a desarticulação dos ossículos. A diminuição da energia transmitida à cóclea também protege as células ciliares localizadas na membrana basilar responsáveis pela emissão de sinais elétricos para o cérebro.

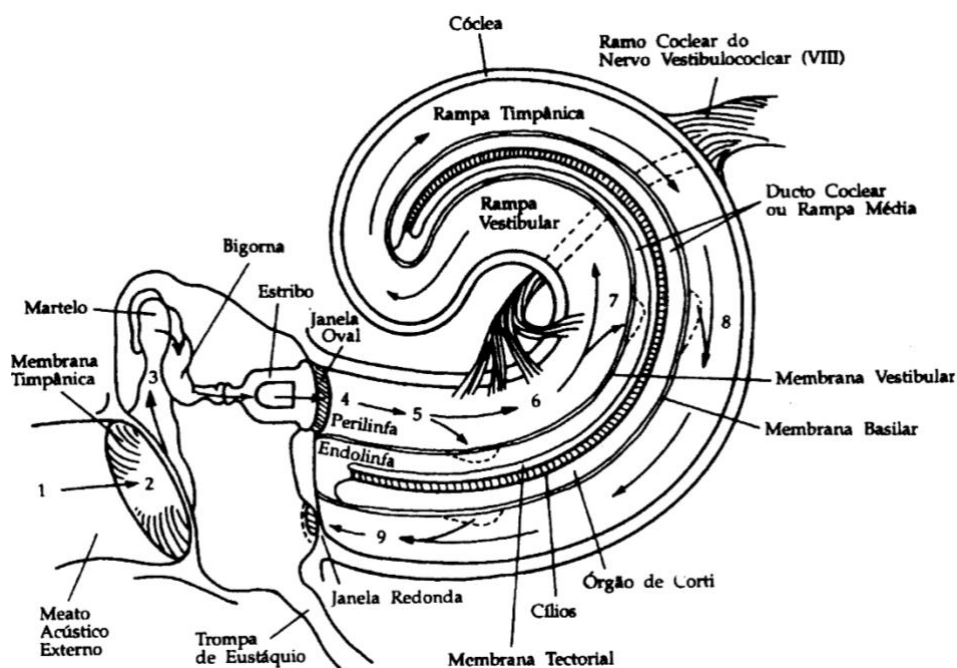
Como não existe comunicação aérea entre a orelha externa e média, e como é necessário que o tímpano vibre livremente para uma transmissão sonora adequada, é necessário que a pressão dos dois lados da membrana timpânica seja igual. Este equilíbrio de pressão é feito pela tuba auditiva ou Trompa de Eustáquio que é um canal ósseo-cartilaginoso que liga a orelha média à faringe, já que esta tem comunicação com o meio externo. Normalmente a tuba auditiva

se encontra fechada no repouso, mas se abre pela ação do músculo tensor do véu palatino e do músculo levantador do véu palatino os quais são acionados na deglutição e/ou no bocejo, permitindo passagem de ar da parte nasal da faringe para a orelha média. Isso permite a equalização da pressão do ar externo com a pressão da cavidade timpânica e, ainda, o arejamento da orelha média. De acordo com Makibara et al. (2010) este mecanismo protege a orelha de mudanças rápidas de pressão, mantém a mucosa conservada e permite que a unidade tímpano-ossicular possa vibrar sem intercorrências. A tuba auditiva também possui a função de drenagem, protegendo a orelha média do acúmulo de secreções.

Observa-se a pressurização do ouvido médio quando em voo, ou descida ou subidas de serras, quando surgem dores na orelha e/ou perda temporária de audição. Como é bem conhecido, a abertura da tuba auditiva e consequentemente a descompressão pode ser facilitada pela abertura da boca (bocejo) ou mesmo mascando chicletes.

A orelha interna mostrada na Figura 5 é a parte mais complexa do sistema auditivo e comunica-se com o ouvido médio através de dois orifícios fechados por membranas: A janela oval localizada junto à base da rampa ou escala vestibular e sobre a qual se articula o estribo e a janela coclear ou redonda que limita a base da rampa timpânica e que por ser flexível permite a variação da pressão, e por tanto das onda sonoras que produzem impulsos elétricos para cérebro.

Figura 5. O ouvido interno



Fonte: Russo (1999, p. 162)

A orelha interna é constituída de uma estrutura oca e rígida denominada capsula ótica (localizada dentro do osso temporal a qual possui várias escavações formando o labirinto coclear, o labirinto vestibular e três canais semicirculares, como pode ser visto na Figura 6 e Figura 7.

Figura 6. Cápsula ótica

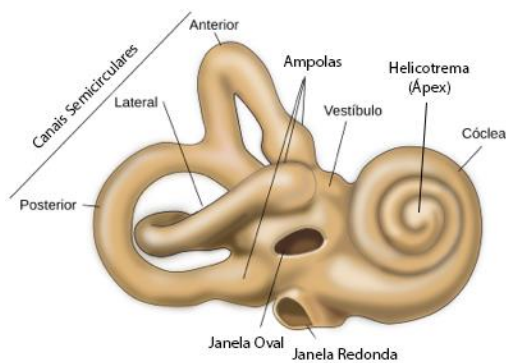
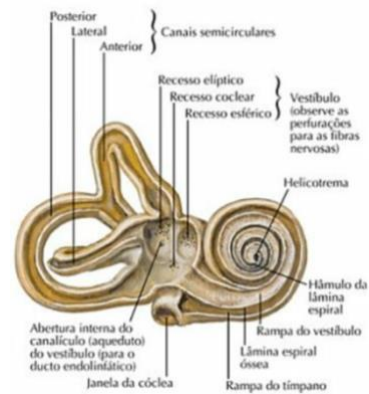


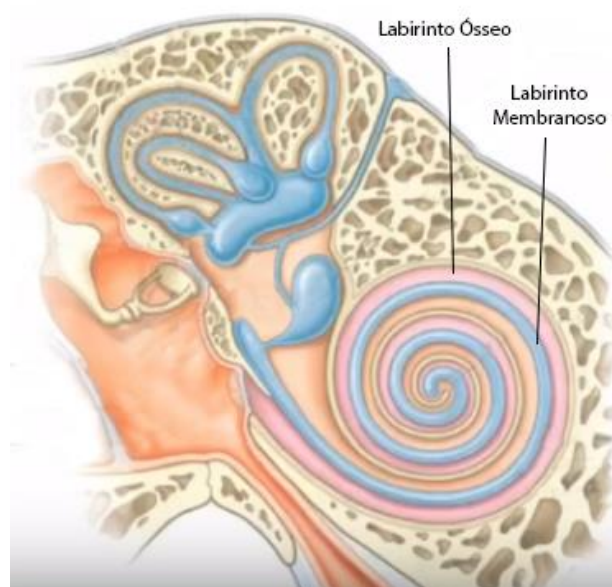
Figura 7. Labirinto ósseo da Cápsula ótica



Fonte: Netter (2015)

Dentro destas escavações se encontra o labirinto membranoso. O volume vazio do labirinto ósseo (cor rosa) é preenchido por um líquido denominado perilinfa. Da mesma forma, ocupando o espaço vazio do labirinto membranoso se encontra outro líquido denominado endolinfa, como ilustrado na Figura 8 na cor azul.

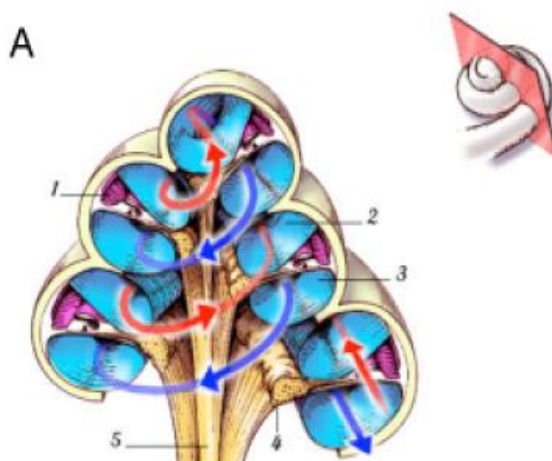
Figura 8. Labirinto ósseo em rosa e no seu interior o labirinto membranoso em azul



Fonte: Mauler (2015) - Modificada

No labirinto ósseo há duas aberturas: a janela oval, a qual é articulada com o estribo para transmissão da vibração para a perilínfa fazendo que ondas hidráulicas percorra o espaço entre o labirinto ósseo e membranoso (Figura 9). O labirinto membranoso recebe as ondas movimentando o seu líquido interno (endolinfa) que provoca e aciona células especializadas na captação dessas vibrações criando polarização e despolarização elétricas nas estruturas próximas. Essas células possuem um tipo de conjunto de cílios chamado de estereocílios, cuja movimentação gera potenciais de ação nas células ciliadas os quais são responsáveis pela transmissão e ampliação dos sinais elétricos que seguem para pelo nervo coclear até o sistema nervoso central levando as informações que serão interpretadas como som.

Figura 9. Circulação das ondas sonoras na perilínfa através do labirinto da cóclea



Fonte: Miranda-Vilela (2018)¹²

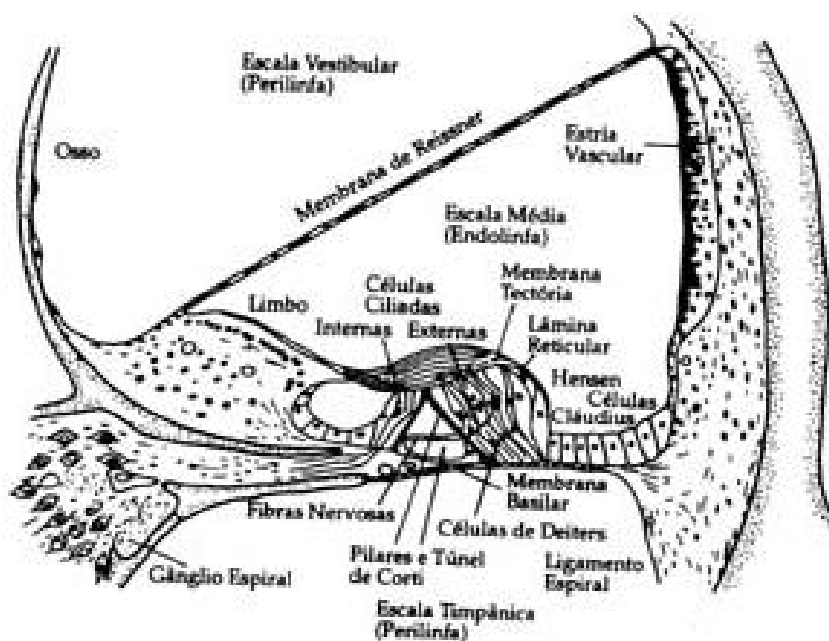
- 1- escala ou rampa média ou coclear
- 2- escala ou rampa vestibular
- 3- escala ou rampa timpânica
- 4- gânglio espiral
- 5- nervo coclear (partindo da membrana basilar)

Como pode ser visto na Figura 10 a escala vestibular separa-se do canal coclear pela membrana de Reissner e o da escala timpânica pela membrana basilar, onde está situado o órgão de Corti, o qual é formado por células ciliares e/ou nervosas responsáveis pelo envio dos impulsos eletroquímicos ao cérebro. De acordo com Oliveira (1997), a membrana basilar

¹² Disponível em: <<https://afh.bio.br/sistemas/sensorial/3.php>>. Acesso em ago. 2018.

alarga-se da sua base (0,04 mm) até o ápice (0,4 mm), aumentando, portanto, a sua massa. Sobre o órgão de Corti está a membrana tectorial com seu bordo interno fixo no modíolo e seu bordo externo livre. A cóclea no seu estado natural mede 35 mm quando desenrolada e possui apenas 5 mm do ápice à base. No entanto, é a parte vital do sistema auditivo porque é nela que se realiza a conversão de variações de pressão em sinais elétricos, que enviados ao cérebro são interpretados como sensação auditiva.

Figura 10. O órgão de Corti



Fonte: Russo (1999, p. 163)

Segundo Oliveira (1997), o potencial elétrico das células ciliares externas pode chegar a 150 mV e é possível graças à diferença da composição iônica dos líquidos que circulam na cóclea adicionado ao efeito da composição do meio intracelular. Esse potencial elétrico aumenta bastante a sensibilidade da célula, que pode ser representada pela capacidade de rápida despolarização de sua membrana com pequenos deslocamentos ciliares. O processo de polarização ou despolarização, cujos efeitos elétricos são registrados por eletrodos implantados nas proximidades da cóclea, é mantido por meio dos movimentos dos cílios que acionam os canais de entrada e de saída de íons que alteram a concentração de potássio, sódio e cálcio nas células.

Para Bento et al. (2013), as células ciliadas externas realizam dois tipos de contração: rápida e lenta. Essas contrações provocam amplificação e acurada análise frequencial dos sons permitindo o aumento de até 50 dB o que implica numa grande amplificação da vibração da

membrana basilar, resultando numa maior sensibilidade das células ciliares internas que assim podem ter seus estereocílios estimulados, alcançando nas suas extremidades apicais a membrana tectórica.

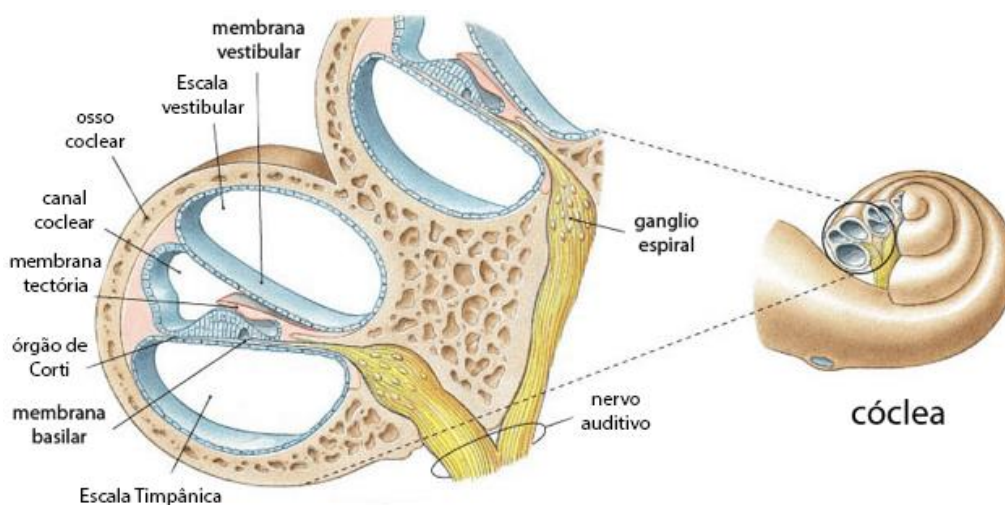
Por este mecanismo, um pequeno número de células ciliadas externas seria estimulado pela frequência específica de um som, permitindo uma acurada discriminação de frequências.

Em suas palavras, o autor conclui:

As células ciliadas externas constituiriam o amplificador coclear, sendo importantes no mecanismo de amplificação do estímulo para determinar o funcionamento das células ciliares internas, que seriam as unidades receptoras e codificadoras cocleares e teriam papel importante na seletividade frequencial da cóclea. Oliveira (1994)

Estas células ocupam toda a cóclea formando o órgão de Corti como pode ser visto na Figura 11.

Figura 11. Localização do órgão de Corti do nervo coclear



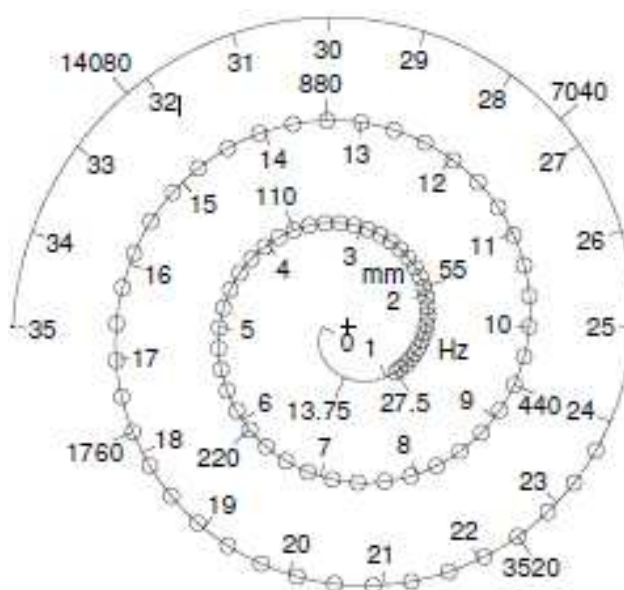
Fonte: Netter (2015) - Modificada

Conforme Gazzaniga et al. (2013), a localização ou seção da célula ciliada na cóclea determina a frequência de som a que ela responde, pois a diferente espessura do canal restringe como a membrana basilar se moverá em resposta as diferentes frequências das vibrações. Assim, diferentes frequências (e consequentemente comprimentos de onda) são captados por diferentes regiões da cóclea. Segundo Carter (2014), os sons de baixa frequência (os sons graves) são captados na região mais próxima ao centro da espiral, enquanto a alta frequência

(os sons agudos) é captada na base da cóclea, próximos à janela vestibular. Essa distribuição espacial dos receptores de som é chamada tonotopia.

Lyon (2017) apresenta o mapa de localização de frequência de Greenwood ilustrado em uma espiral que se aproxima da forma da cóclea humana. Distâncias do ápice em milímetros são rotuladas dentro da espiral e as frequências de oitavas rotuladas do lado de fora. As frequências fundamentais, ou notas, das notas das 88 teclas de um piano são marcadas por círculos. Observa-se neste mapa que as frequências geometricamente espaçadas - oitavas e notas - estão aproximadamente espaçadas, quase 5 mm por oitava, nas regiões basal e média, mas estão agrupadas perto do ápice, com apenas cerca de 1 mm para a oitava mais baixa do piano. A cóclea humana tem cerca de dois e três quartos de voltas; o último quarto de giro mostrado no centro (o último 1 mm), que mapeia frequências abaixo de zero, deve ser interpretado como o helicotrema.

Figura 12. A resposta por banda de frequência da cóclea humana



Fonte: Lyon (2017, p. 127)

Conforme o número de seções acionadas, sinais complexos são emitidos através do nervo coclear para o cérebro nos dando a sensação da fala, música, ruído e todos os sons da natureza.

Como vimos no capítulo 2 a audição é um processo não linear, isto é, a sensação não é proporcional ao estímulo, pelo contrário ela privilegia os sons relacionados às frequências relacionadas à comunicação humana. Este realce se dá desde que as ondas sonoras são captadas

na pina, na passagem pelo conduto auditivo e na cóclea, ampliando e filtrando de forma a ser entendida pelo cérebro.

Vê-se do exposto acima que o sistema auditivo humano capta, conduz, transforma, amplia e transmite a energia de origem acústica em sinais elétricos para o cérebro, num processo complexo, ainda não completamente entendido, mas muito eficiente. Estudos de mixagem musical automática, como os de Scott et al. (2011) e Mansbridge et al. (2012) utilizam técnicas estatísticas tradicionais e aprendizado de máquina (*machine learning*) para análise e controle dos sinais musicais, de modo com que a mixagem automatizada possa simular, pelo menos em parte, o ouvido humano e assim obter uma sonoridade agradável em suas mixagens para a maioria dos ouvintes dentro do âmbito da música comercial.

Ainda sobre a relação da fisiologia da orelha e as dificuldades encontradas para simulá-la destacamos os estudos do Prof. Joshua D. Reiss do *Centre for Digital Music - Queen Mary - University of London*, com uma excelente produção científica, tanto na mixagem totalmente automatizada (sem interferência humana) quanto em ferramentas de auxílio mixagem tradicional como mencionado em Reiss; Perez-Gonzalez (2009). Os trabalhos de mixagem automática de Reiss seguem métodos diferentes e matematicamente mais complexos do que este que desenvolvemos neste trabalho, porém não encontramos resultados ou ferramentas que pudessem dar um exemplo audível da sua pesquisa.

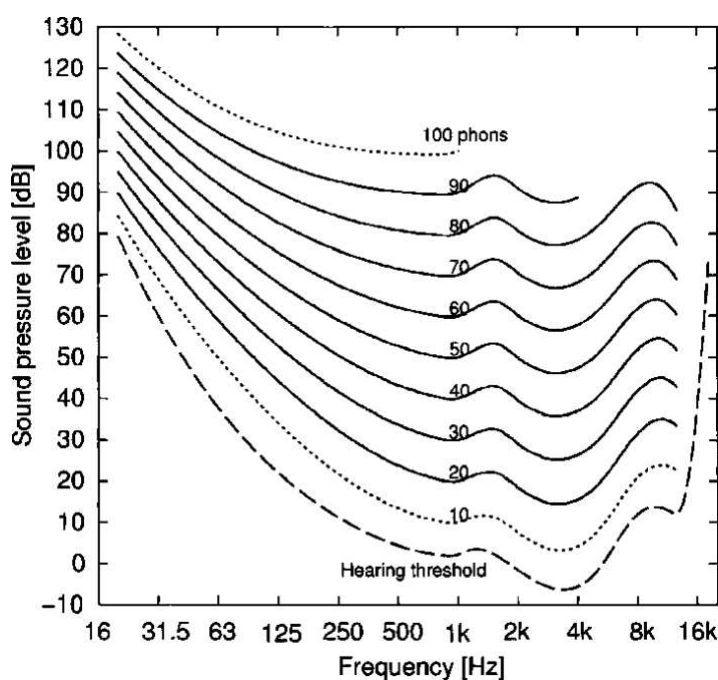
2.5.2 Curvas Isoaudíveis

As chamadas curvas isoaudíveis ou isofônicas (*Equal-Loudness Curves*) descrevem os níveis de pressão sonora (em dB) por frequência (em Hz) que são percebidos com o mesmo volume (*loudness*) para um conjunto de ouvintes. Essa curva pode variar com o tipo de população, o método realizado de medição, e outras condições externas. Cada curva mostra o nível de pressão sonora necessário em frequências diferentes para se produzir uma sensação de volume igual em todas elas. Estas curvas, por convenção, elas são rotuladas a partir de seu valor de intensidade na frequência de 1000 Hz.

Os primeiros estudos de curvas isoaudíveis foram feitas por Fletcher; Munson (1933) porém de acordo com Suzuki; Takeshima (2004) a primeira curva isoaudível aceita internacionalmente foi baseada apenas em dados da população britânica em meados de 1950. Essa curva gerou a primeira versão da norma ISO 226. Com o decorrer dos anos, mais dados e novas populações foram estudadas, o que favoreceu a atualização da norma internacional promovendo dados mais acurados da percepção auditiva humana. No momento da escrita deste

trabalho a versão da norma é ISO 226:2003¹³, porém a própria organização informa que em breve divulgará uma nova versão da norma. A Figura 13 mostra as curvas isoaudíveis de acordo com a ISO 226:2003. Estas curvas têm grande importância conceitual e computacional na construção do nosso método de mixagem automática. Os detalhes serão expostos no Capítulo 4.

Figura 13. Curvas Isoaudíveis de acordo com a ISO 226:2003



Fonte: Suzuki; Takeshima (2004)

Observa-se da Figura 13 que elas também expressam a não linearidade da audição em relação a frequência, ou seja, os tons ou sons muito graves, abaixo de 250 Hz, necessitam de muito mais energia para serem ouvidos em relação aos sons de frequência de 250 a 4000 Hz. Neste sentido a norma ISO acima citada define o nível de *loudness* (volume percebido) como o valor do nível de intensidade sonora em *phons*¹⁴ que possui o mesmo valor numérico que o nível de pressão sonora em decibéis de um som de referência, consistindo de uma onda sinusoidal incidente frontal de frequência de 1000 Hz, que tem o mesmo volume ou intensidade que som fornecido.

¹³ International Organization for Standardization, “ISO 226:2003 - Normal equal-loudness-level contours”.

¹⁴ Unidade de medida de *loudness* para tons puros. 0 *phons* corresponde ao limiar de audibilidade em 1kHz

2.5.3 Monitoramento e Sala de Mixagem

Hoje, na mixagem convencional, faz-se necessário o uso de salas preparadas para este fim, seja quanto dimensionamento adequado, seja no isolamento e absorção em paredes e teto; porém grande parte dos profissionais da área de mixagem da atualidade dificilmente possuem salas que cumpram os requisitos mínimos para uma boa escuta do material gravado, o que pode dificultar a percepção musical. Frequentemente nesses ambientes inadequados ocorre um fenômeno acústico chamado “modo(s) de ressonância” ou “modo(s) de sala” (principalmente nas frequências graves) os quais são causados por ondas estacionárias presentes na sala.

Ondas estacionárias são formadas pelo resultado da superposição de duas ondas de mesma frequência, mesma direção e sentidos opostos. Em uma sala inadequada com geometria de um paralelepípedo podem ocorrer três tipos de ondas estacionárias: axiais, tangenciais e oblíquas. As axiais, que acontecem entre duas paredes paralelas, as tangenciais que se formam entre quatro superfícies, e, as oblíquas, entre todas as seis superfícies de uma sala. As ondas estacionárias acentuam certas frequências e atenuam outras, distorcendo o espectro sonoro do ouvinte.

Um dos principais erros de construção de salas de mixagem está no fato de não se atentar as recomendações referentes às dimensões da sala, como por exemplo as chamadas “Proporções de Ouro” (Golden Ratios). Everest (2001) apresenta as 3 principais proporções¹⁵ de salas estudadas por Sepmeyer (1965) e Louden (1971), as quais podem ser vistas na Tabela 2 e na

Tabela 3 respectivamente.

Tabela 2 - Proporções de Ouro de Sepmeyer (1965)

	Altura	Largura	Comprimento
A	1.00	1.14	1.39
B	1.00	1.28	1.54
C	1.00	1.60	2.33

¹⁵ Valores adimensionais

Tabela 3 - Proporções de Ouro de Louden (1971)

	Altura	Largura	Comprimento
A	1.00	1.40	1.90
B	1.00	1.30	1.90
C	1.00	1.50	2.50

É importante ressaltar que o tamanho da sala é de grande importância para o aumento/diminuição dos modos de ressonância. Em salas pequenas, mesmo as que respeitam os “*golden ratios*”, há uma maior chance de que os modos fiquem audíveis em comparação com salas grandes.

Há também problemas do condicionamento acústico das salas, os quais não permitem obter uma curva de resposta plana (ou seja, aquela que possui a mesma intensidade para todas as frequências) no ponto de escuta do *mixer*, localização da sala comumente chamada de “*sweet spot*”. Soma-se a isso o tempo de reverberação fora da especificação normativa, tornando o ambiente muito seco ou muito reverberante, o que diminui a clareza e definição do som. Osman (2010) apresenta o tempo ideal de reverberação de acordo com as variadas normas internacionais, nas quais encontramos que esse valor de RT60¹⁶ varia entre 0.3 e 0.7 segundos.

Em locais inapropriados para a tarefa de mixagem, o ruído e vibração externa (ruído de trânsito, bares, vibrações do solo) podem causar mascaramento suficiente para influenciar o trabalho do técnico ou engenheiro de áudio. O mascaramento por sons externos à gravação não permite ao operador a audição de detalhes sonoros que podem influenciar fortemente na qualidade final do trabalho.

Atualmente, com o avanço do conhecimento científico e da tecnologia, é possível construir salas adequadas com fundamento nas melhores proporções dimensionais, estabelecer o isolamento e tratamento acústico para obtenção de uma determinada resposta espectral e tempo de reverberação, fabricar monitores de referência que respondem de forma plana em todo o espectro de frequência e desenvolver microfones e *softwares* que possam facilitar o trabalho de mixagem.

¹⁶ Indica o tempo que leva para um pulso sonoro decair 60 dB em sua intensidade

Embora o avanço tenha sido grande, o custo de implementação de toda esta estrutura ainda é muito alto para a grande maioria de técnicos e engenheiros de áudio, ou até mesmo para micro e pequenas empresas. O foco deste trabalho é contribuir para o desenvolvimento de softwares que possam auxiliar o processo de mixagem abstraindo os problemas físicos e acústicos de grande parte das salas de pequenos e médios estúdios atuais.

2.6 REFERÊNCIAS SONORAS

Nesta seção apresentamos diversas visões de autores mostrando o caráter interdisciplinar da mixagem bem como a complexidade dos procedimentos altamente dependentes da habilidade e experiência do técnico de áudio. A seguir nos concentramos na definição característica de uma boa mixagem segundo Bazil (2008) e alguns outros autores.

Primeiramente, para Rumsey (2013) a mixagem musical não se configura como uma ciência. Ela é um conjunto de boas práticas, algumas delas verdadeiros segredos para técnicos e até mesmo para engenheiros de som. Embora muitos profissionais tenham desenvolvido a habilidade para esse tipo de trabalho e utilizem uma referência de mixagem personalizada e efetiva, a sua experiência não foi traduzida e confirmada por métodos científicos, isto é, não obtiveram reprodutibilidade comprovada em laboratório. Portanto, o aspecto mais relevante da mixagem é a capacidade e habilidade técnica do profissional de obter intencionalmente certos resultados através do manuseio de seu ambiente acústico e de seus equipamentos de áudio.

Essa tarefa é difícil pois, para Bazil (2008), a mixagem é um trabalho associado à subjetividade e ao conhecimento técnico específico e intuitivo. O trabalho de mixagem exige conhecimento técnico e controle de vários equipamentos, como: mesa de som, monitores de referência, placas de áudio, endereçamento dos cabos, processadores de efeitos externos, *softwares*, etc. Izhaki (2011) reforça a necessidade do treinamento auditivo deste profissional de modo que ele possa perceber a relação de causa e efeito dos ajustes a serem realizados, o que demanda sensibilidade estética, artística e conhecimento musical.

Uma observação: o profissional de mixagem deve resguardar a sua saúde auditiva já que o ouvido é principal ferramenta utilizada na sua avaliação sonora. De acordo com Francelin et al. (2010) enquanto perdas temporárias de audição podem prejudicar a mixagem, perdas permanentes, dependendo do grau, podem inviabilizar o exercício da profissão.

Moylan (2006) completa dizendo que a mixagem deve ser acompanhada por produtores ou/e intérpretes para as tomadas de decisões referentes ao projeto musical, de modo a exprimir nele a verdadeira concepção sonora e artística desejada. Neste sentido, espera-se que os

engenheiros de mixagem resolvam questões técnicas, como garantir a audibilidade das fontes, bem como fazer escolhas criativas para implementar a visão musical do artista, produtor ou de si mesmos.

Conforme Macedo (2007) explica, na mixagem, é importante que todos os elementos técnicos, tecnológicos e criativos estejam a serviço de uma linguagem expressiva. De acordo com Araújo (2015), o objetivo principal da mixagem é garantir que toda a informação sonora essencial (desejada) esteja inserida na extensão do espectro de frequência e dinâmica da mídia de reprodução para a qual a mixagem se destina.

Nesta forma de encarar o processo de mixagem, Izhaki (2011) apresenta três princípios para obter de uma mixagem criativa: visão, ação e avaliação. Pela visão se dá a “visualização” da sonoridade, enquanto profissionais iniciantes obtêm resultados por tentativas de erros e acertos, os experientes conseguem imaginar (“visualizar”) os resultados e posteriormente obtê-los. A ação se refere à decisão sobre que equipamento utilizar e como utilizá-lo. Iniciantes testam muitas ferramentas e equipamentos o que pode alterar a ideia original. A avaliação consiste na habilidade do profissional de responder acertadamente as questões como “a mixagem está soando como desejado?”; “Está soando corretamente?” e “Tem algo de errado e o quê?”, “Este áudio irá soar de forma similar em diferentes tipos de reprodução? Como em sistemas automotivos, ou tocadores de MP3, etc?” conduzindo cada fase do trabalho ao resultado esperado.

De acordo com Bazil (2008), para se obter uma boa mixagem deve-se observar sete elementos que ressaltam a natureza subjetiva de mixar:

- a) *Clareza*: O áudio deve estar livre da presença de qualquer frequência sonora (parcial harmônica¹⁷ ou inarmônica¹⁸) exagerada ou que não esteja contribuindo para a definição do resultado final sonoro.
- b) *Limpeza*: As faixas de áudio devem estar sem ruídos ou sons indesejados
- c) *Separação*: Deve-se ouvir cada pista de áudio sem vazamentos, ou seja, não deve haver captação indesejada do som de outros instrumentos entre faixas.
- d) *Nível*: As faixas de áudio não devem variar desnecessariamente o seu “volume percebido” (*loudness*) durante a música, para evitar dificuldade no tratamento da variação de amplitude do sinal, ou seja, na dinâmica musical.

¹⁷ Componente espectral que pertence a série harmônica da frequência fundamental do som

¹⁸ Componente espectral que não pertence a série harmônica da frequência fundamental do som

- e) *Balanço*: Deve existir um bom balanceamento espectral e espacial do material mixado, não havendo direcionamento desnecessário para um dado lado do campo estereofônico.
- f) *Gênero*: Para aumentar a chance de sucesso no mercado consumidor, a mixagem final deve soar de acordo com a sonoridade já pré-estabelecida do gênero musical específico.
- g) *Teste em sistemas*: Deve possuir uma sonoridade aceitável em qualquer sistema de reprodução, desde os mais baratos até sistemas de alta definição, incluindo a escuta em várias circunstâncias como na direção de veículos, durante exercício físico, no almoço ou mesmo descansando.

Embora muitos aspectos da subjetividade do artista e do profissional da área de áudio estejam envolvidos na mixagem, existe um consenso comum no apontamento de bons trabalhos de mixagem, ou seja, existem referências sonoras as quais podem ser tidas como “alvos” ou “padrões” de bons trabalhos de mixagem. Atualmente existem dois tipos de referências sonoras para esse tipo de trabalho, são elas as “*gravações de alta fidelidade*” e o *Ruído Rosa*. Abaixo detalhamos essas duas referências.

2.6.1 Gravações de alta fidelidade

Os engenheiros de som utilizam gravações de referência de alta qualidade e definição nas quais se apoiam tanto para o aprendizado quanto para a aplicação. Esses arquivos de áudio têm características sonoras que apresentam fidelidade espectral intacta e sem perdas. Mesmo não havendo um referencial teórico que defina as características de uma chamada “gravação de referência”, cada profissional da área de mixagem seleciona, dentro do estilo musical que vai mixar, gravações de artistas famosos do mesmo estilo musical. Para garantir a qualidade do material gravado, esses artistas realizam seus trabalhos musicais em grandes gravadoras, as quais possuem os melhores engenheiros de som, salas de mixagem adequadas e equipamentos de alta precisão. Exemplos atuais dessas empresas são a Universal Music, Sony Music, Warner Music e a EMI Group.

De acordo com Izhaki (2011), faz-se uso dessas gravações ao compará-las à música que vêm sendo mixada pelos estúdios famosos através de análises espectrais e auditivas. Nessa atividade, dentro do possível, avaliam-se os elementos conforme descritos acima por Bazil

(2008) e manipula-se o áudio a ser mixado no intuito de atingir um resultado sonoro próximo ao padrão utilizado como referência.

O engenheiro de som Moss (2018a) reforça a importância da escolha das gravações de referência de modo com que elas estejam de acordo com o trabalho que se propõe mixar, levando-se em consideração aspectos como gênero, dinâmica e época. Na mixagem de um naipe de cordas, por exemplo, deve-se usar como referencial uma faixa que possua uma boa mixagem de um conjunto similar de cordas, e não uma que apresente uma instrumentação diferente.

2.6.2 Ruído Rosa (Pink Noise)

Conforme citado no capítulo 1, criamos um método de mixagem automática o qual abstrai as necessidades e preocupações sobre as dimensões físicas da sala e seu tratamento e monitores de áudio, pois o nosso modelo parte de princípios fisiológicos e psicoacústicos da audição humana. O cerne do método está na utilização do sinal chamado ruído rosa (*pink noise*) como nossa referência psicoacústica, pois o seu espectro se aproxima da resposta de frequência do ouvido (ou orelha) humano.

Essa referência já vem sendo utilizada pela maioria dos softwares de calibração de salas (ex. Smaart¹⁹, REW²⁰, etc) e também se tornou uma ferramenta técnica de mixagem amplamente utilizada nos últimos anos, conforme Silverstein (2018) e Bazil (2014). A diferença do seu uso neste trabalho é que, a partir da referência do ruído rosa, nós automatizamos o processo todo de mixagem, imitando a prática atual de estúdio, incluindo os processos de compressão, equalização e remoção de frequências ressonantes.

Alguém poderia sugerir o uso do ruído branco (*white noise*) como sinal de referência, já que apresenta uma resposta plana, isto é, apresenta a mesma energia em cada frequência do espectro, mas dada as características do ouvido já apresentadas acima, este tipo de ruído não adequado para o nosso estudo.

Da maneira que escutamos julgamos aumentos de altura (*pitch*) conforme a escala logarítmica, ou seja, a razão entre a variação da frequência e a frequência inicial, e não por incrementos iguais de frequência. Por exemplo:

$$\frac{(40 - 20)}{20} = \frac{(2000 - 1000)}{1000} = 1 \text{ oitava}$$

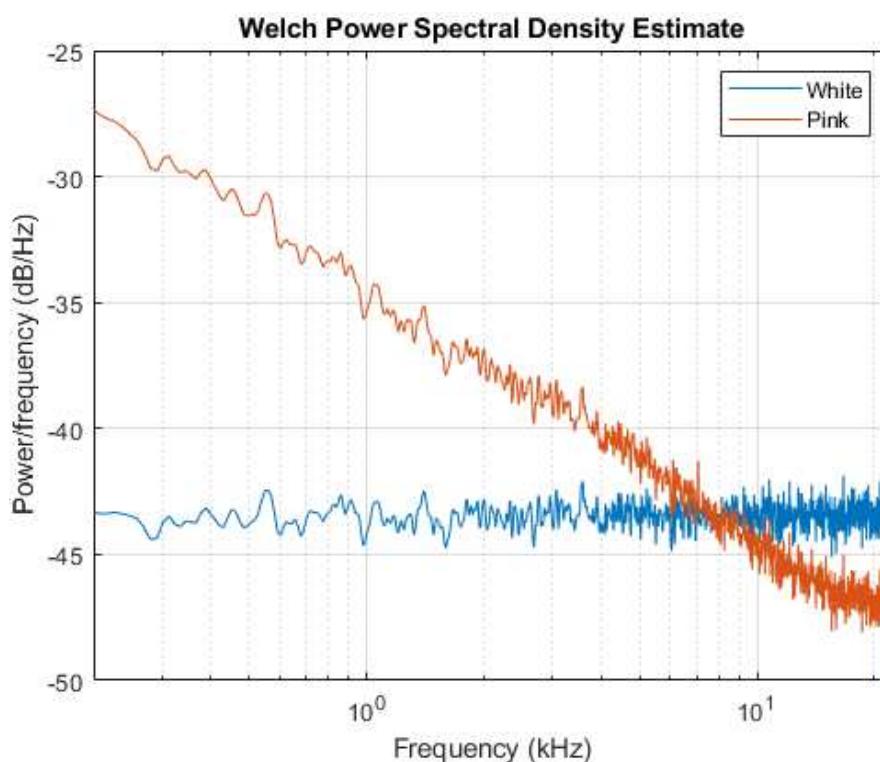
¹⁹ Disponível em: <<http://www.rationalacoustics.com/smaart/about-smaart/>>. Acesso em set. 2018.

²⁰ Disponível em: <<https://www.roomeqwizard.com/>>. Acesso em set. 2018.

Isso significa que quando ouvimos o um sinal que contém todas as componentes em frequência com a mesma amplitude, como ruído branco, não temos a percepção de um sinal “equilibrado espectralmente”, mas sim de um sinal que aumenta de 3 dB por oitava, ou seja, dobrando de amplitude a cada oitava.

Uma distribuição de energia mais semelhante à da audição humana resulta quando filtramos o ruído branco, de modo que este perca 3 dB por oitava para frequências crescentes. Desta forma, cada banda de oitava passa a ter a ser percebida como tendo o mesmo *loudness*. Quando escutamos esse som filtrado, temos a percepção de um som “equilibrado espectralmente”. O resultado desta filtragem é o que chamamos de o ruído rosa (*pink noise*). A Figura 14 mostra a relação dos espectros do ruído branco e ruído rosa.

Figura 14. Espectro do ruído branco e do ruído rosa



Fonte: MathWorks (2018)²¹

Neste trabalho escolhemos trabalhar com a resolução espectral de 1/3 de oitava tanto para o sinal de referência (ruído rosa) como para todos os instrumentos musicais a serem mixados pois, de acordo com Rossing et al. (2001), esta seria uma resolução próxima à resolução das bandas críticas da cóclea humana. Deste modo, modificações espectrais que

²¹ Disponível em: <<https://www.mathworks.com/help/signal/ref/poctave.html>>. Acesso em set. De 2018.

sejam menores que $1/3$ de oitava estariam dentro de uma mesma banda crítica, não sendo percebidas devido ao mascaramento inerente a fisiologia da cóclea humana. No próximo capítulo iremos apresentar os métodos e ferramentas atuais do processo de mixagem tradicional, seus equipamentos, sequência de trabalho e dificuldades comuns à essa atividade. Através da observação e tentativa de imitação deste processo tradicional, desenvolvemos o nosso método de mixagem automática visando automatizar alguns processos repetitivos e encontrar soluções para as dificuldades encontradas atualmente neste processo.

3 MIXAGEM

De acordo com Gibson; Petersen (1997), o processo de mixagem de áudio pode ser definido como a mistura adequada (soma) de pistas²² de áudio, sejam elas previamente gravadas (estúdio) ou ao vivo (sonorização de P.A.²³).

O desejo de alcançar melhores resultados musicais, tanto sonoros quanto estéticos, motivaram o desenvolvimento de equipamentos, técnicas e métodos de manipulação de áudio, desde o início do século XX até os tempos atuais, como indicado no capítulo 1. A seguir descrevemos os principais equipamentos e técnicas usadas atualmente no processo de mixagem (subseção 3.1), o problema das frequências ressonantes (3.2), e do balanço tonal (3.3).

3.1 Principais Equipamentos

Atualmente há uma grande variedade de tipos de equipamentos para manipulação de áudio, bem como diversas marcas e modelos dos mesmos. Empresas como SSL²⁴, API²⁵, Neve²⁶, Avid²⁷, Neumann²⁸ conseguiram desenvolver *hardwares* e *softwares* que se tornaram indispensáveis para um estúdio de gravação e mixagem moderna, como a mesa de som SSL 4000, o pré-amplificador Neve 1081, o software de gravação e mixagem Avid Pro Tools e outros.

Como herança do período analógico, permanece em boa parte dos equipamentos e softwares atuais a cadeia de processamento padrão provinda do modo com que as mesas de som analógicas tratavam o áudio chamada “*Channel Strip*”. Essa sequência de processamento era iniciada por um filtro passa-alta para remoção de baixas frequências indesejadas seguido por um equalizador paramétrico para remoção de frequências ressonantes e para melhorar o balanço tonal, um compressor para diminuir a variação da amplitude do sinal, *gate* para silenciar sinais de baixa intensidade, e por fim um limitador para limitar a amplitude máxima do sinal, como mostra a Figura 15.

²² No âmbito da mixagem digital, as “pistas de áudio” são os arquivos de áudio previamente gravados no estúdio

²³ *Public Audition*

²⁴ <https://www.solidstatellogic.com/>

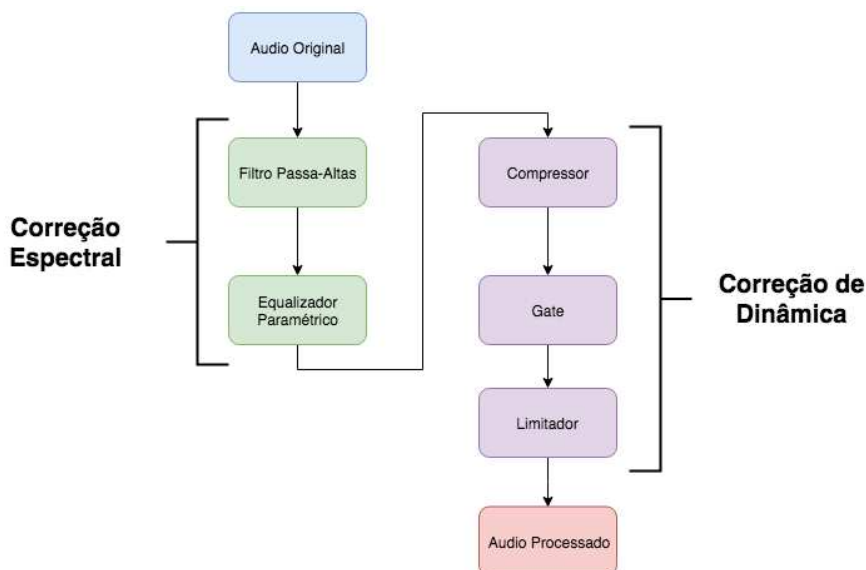
²⁵ <https://www.apiaudio.com/>

²⁶ <https://ams-neve.com/>

²⁷ <https://www.avid.com/>

²⁸ <https://en-de.neumann.com/>

Figura 15. Cadeia padrão de Processamento de Áudio



Fonte: Própria

Como já apresentado no capítulo 1, com o advento da era digital muitos *hardwares* reconhecidos mundialmente por sua sonoridade agradável foram emulados em *software*. Na Figura 16 temos um exemplo de um *plugin*²⁹ de “Channel Strip”.

Figura 16. Plugin de emulação do Channel Strip Neve 88RS



Fonte: Universal Audio³⁰

²⁹ Um pequeno software que estende as capacidades de outro software maior

³⁰ Disponível em: <<https://www.uaudio.com/uad-plugins.html>>. Acesso em ago. 2018.

Além destes equipamentos básicos oferecidos pelo “*Channel Strip*” para a manipulação, há também uma segunda classe de *hardwares* e *softwares* de áudio destinada a criar efeitos de áudio. Os principais efeitos são *reverb*, *delay*, *flanger*, *chorus* e saturação. Diferentemente da cadeia do “*Channel Strip*” não há uma sequência padrão ou recomendada para o uso dos efeitos, deixando aberto à criatividade e experiência do operador.

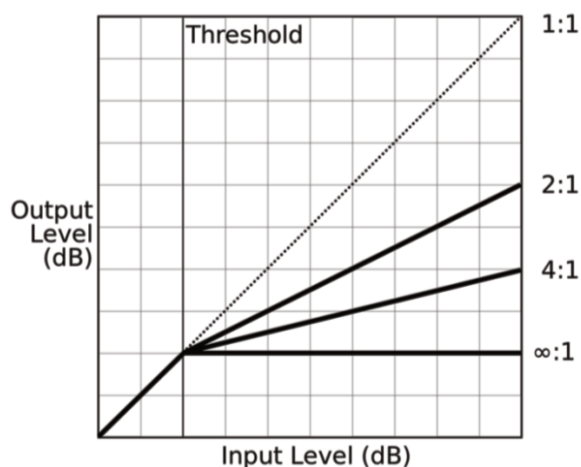
Nas próximas sessões vamos detalhar o funcionamento dos dois principais equipamentos de manipulação de áudio importantes em relação ao nosso modelo computacional: o compressor e o equalizador.

3.1.1 Compressor de Áudio

Dado um sinal de áudio gravado de um instrumento musical, fica evidente as variações de amplitude do sinal no tempo, sendo algumas delas propositalis ao discurso musical e outras não intencionais e inerentes à tecnologia e/ou deficiência da técnica de captação. O excesso de variação de amplitude do sinal musical dificulta o processo de mixagem, diminuindo a inteligibilidade do instrumento gravado.

De acordo com Gibson; Petersen (1997), um compressor de áudio tem duas funções principais: obter uma melhor relação de sinal-ruído da fonte sonora e também estabilizar a “imagem virtual” da fonte sonora ao ser reproduzida por um par de monitores de áudio. Podemos entender os dois parâmetros básicos de um compressor através de uma curva característica, onde o sinal de entrada é dado no eixo horizontal e o sinal de saída (sinal comprimido) é dado no eixo vertical, como mostra a Figura 17.

Figura 17. Função de Transferência de um Compressor



Fonte: Iain Fergusson (2007)³¹

³¹ Disponível em: <https://commons.wikimedia.org/wiki/File:Compression_ratio.svg>. Acesso em ago. 2018.

Quando a amplitude instantânea do sinal é maior do que o valor do parâmetro *threshold*, o compressor começa a atuar diminuindo a amplitude do sinal, de acordo com o parâmetro *ratio*. Caso o *ratio* seja 1:1, não há compressão e o sinal de saída é idêntico ao de entrada. Caso o *ratio* seja muito alto (próximo de infinito), a nova amplitude máxima do sinal será igual ao valor do *threshold*.

A escolha da combinação dos valores de *threshold* e *ratio* vão depender de outros fatores, como a amplitude média do sinal de entrada, tipo de instrumento musical analisado, quantidade de variação desejada e até estética musical a ser alcançada. Na Figura 18 podemos ver a versão de *plugin* de um compressor com os seus parâmetros de controle.

Figura 18. Plugin de compressão Avid Dynamics III



Fonte: Avid (2018)³² - modificado

É importante ressaltar que alguns compressores não possuem um controle ajustável de *threshold*, pois nestes aparelhos esse valor já está definido internamente, seja por circuito eletrônico, seja por software. Podemos ver um exemplo disso na Figura 19 abaixo.

³² Disponível em: <<https://www.avid.com/plugins/dynamics-iii>>. Acesso em ago. 2018.

Figura 19. Compressor 1176 da Universal Audio



Fonte: Universal Audio (2018)³³

Um terceiro tipo de compressão, menos usual, seria passar o áudio por um gravador de rolo (*tape recorder*) com alto valor de ganho de entrada, gerando um tipo específico de compressão e saturação no áudio de saída, já que esse tipo de equipamento possui características mecânicas e eletrônicas que podem causar esses tipos de resultados em determinadas configurações.

Alguns produtores musicais têm preferência por esse tipo de compressão quando também é necessário acrescentar harmônicos ao áudio original. Esses harmônicos se tornam mais audíveis no momento em que a variação da dinâmica é muito restringida pelo compressor e a diferença de nível de volume entre as componentes espectrais predominantes e os harmônicos mais fracos se torna pequena.

Na Figura 20 podemos ver um exemplo de um gravador de rolo analógico usado em estúdios de gravação profissionais.

Figura 20. Gravador de Rolo Studer A800



Fonte: Tile House Studios³⁴

³³ Disponível em: <<https://www.uaudio.com/hardware/1176ln.html>>. Acesso em ago. 2018.

³⁴ Disponível em: <<https://www.tilehousestudios.com/photos?lightbox=image7pp>>. Acesso em ago. 2018.

Apenas uma nota para melhor esclarecimento, conforme Reiss (2016) o julgamento da qualidade do áudio analógico e digital através da percepção humana tem sido amplamente debatido por muitos anos, mas ainda se trata de um assunto de pesquisa pois não há uma conclusão clara do assunto. No trabalho intitulado “*Which of the Two Digital Audio Systems Matches the Quality of the Analog System?*” Woszczyk et al. (2009) detalha os resultados experimentais da comparação perceptual entre o áudio analógico e áudio digital os quais tendem a conclusão de que os ouvintes não percebem a diferença entre o áudio analógico e o áudio digital amostrado em altas taxas de *sample rate*.

A nossa proposta se mostra inovadora na forma de uso do conceito de compressão de áudio, na qual propomos analisar a dinâmica do áudio previamente gravado (*a priori*) com a finalidade de alcançar melhores taxas de compressão *a posteriori*. No capítulo 4 iremos descrever como podemos obter uma dinâmica do sinal mais controlada e otimizada à nossa percepção musical.

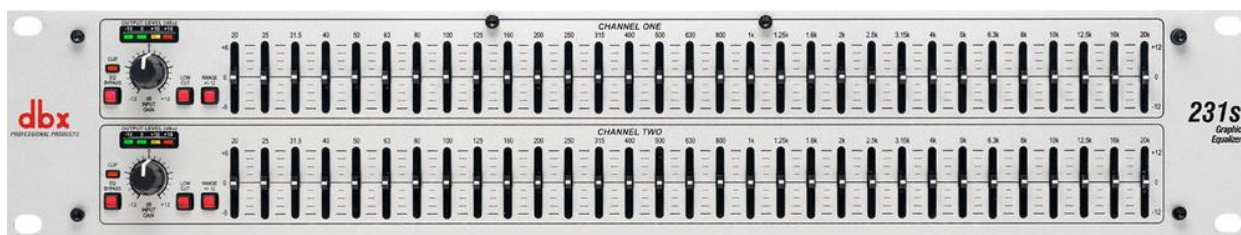
3.1.2 Equalizador de Áudio

O conteúdo espectral do áudio gravado de um instrumento musical pode sofrer grandes variações dependendo de alguns parâmetros, tais como: tipo de microfone usado na captação, posicionamento do microfone em relação ao instrumento, tamanho da sala de gravação e outros.

Para alcançar a sonoridade desejada de cada instrumento musical, os profissionais da área de mixagem utilizam um equalizador de áudio, o qual manipula o espectro do áudio de acordo com alguns parâmetros de controle. Um equalizador comercial pode ser de dois tipos básicos: gráfico ou paramétrico.

O equalizador gráfico é aquele em que seus parâmetros fixos são: número de bandas, largura de banda e frequência central. Assim o único controle ajustável do equalizador gráfico é o ganho a ser aplicado em cada banda. Na Figura 21 podemos ver um exemplo de equalizador gráfico com dois canais (esquerdo e direito) e 15 bandas de equalização por canal.

Figura 21. Equalizador Gráfico DBX 231s



Fonte: DBX³⁵

Os equalizadores gráficos são menos utilizados em estúdio pela falta de flexibilidade dos parâmetros descritos acima.

Os equalizadores paramétricos possuem os 3 controles ajustáveis, ou seja, o operador pode controlar em cada banda qual será a frequência central, largura de banda³⁶ e ganho da banda específica.

Alguns equalizadores são também chamados de semi-paramétricos quando não permitem o ajuste da frequência central ou da largura de banda. Na Figura 22 podemos ver um equalizador que possui duas bandas totalmente paramétricas nas regiões de médio-grave (*Low Medium Frequency - LMF*) e médio agudo (*High Medium Frequency - HMF*) destacadas respectivamente em azul e verde e duas bandas semi-paramétricas para as regiões de graves (*Low Frequency - LF*) e agudos (*High Frequency - HF*) destacadas respectivamente em marrom e vermelho, as quais não permitem o ajuste da largura de banda.

Figura 22. Equalizador Paramétrico SSL 611EQ



Fonte: Solid State Logic - SSL³⁷

Como dito anteriormente, equalizadores são usados para manipular o espectro do áudio e os dois principais processos de manipulação espectral do material gravado são: remoção das frequências ressonantes e ajuste do balanço tonal geral através do equalizador de áudio.

³⁵ Disponível em: <<https://dbxpro.com/en/products/231s>>. Acesso em jan. 2019.

³⁶ A largura de banda é também descrita como “Fator Q”. Mais detalhes no capítulo 4.

³⁷ Disponível em: <<https://www.solidstatellogic.com/studio/500format-modules>>. Acesso em ago. 2018.

3.2 FREQUÊNCIAS RESSONANTES

Devido a problemas de captação da fonte sonora, como o mal posicionamento dos microfones, a construção defeituosa do instrumento musical, ou mesmo problemas com a ressonâncias/modos de sala, ocorre reforço indesejado de algumas frequências pontuais. Isto gera ressonâncias que podem ser ouvidas no sinal gravado, as vezes em pequenos trechos ou durante todo o áudio, diminuindo a clareza e a definição timbrística do instrumento musical gravado.

Essa pequena distorção espectral ocorre em bandas de frequências muito estreitas, e na maioria das vezes passa despercebida pelo ouvinte. Porém, sua correção é indispensável para obter um bom resultado sonoro.

Uma técnica de estúdio conhecida como “*Sweep and Destroy*” utiliza um equalizador paramétrico para a remoção dessa distorção espectral. Moss (2018b) descreve e demonstra essa técnica em sua vídeo-aula na internet. A Figura 23 apresenta um equalizador paramétrico no qual a primeira banda está sendo utilizada para remover uma frequência ressonante específica.

O detalhamento da técnica para a remoção das frequências ressonantes será apresentado no capítulo 4, onde mostramos um novo algoritmo criado para simular a técnica “*Sweep and Destroy*”.

3.3 BALANÇO TONAL

Enquanto a remoção de frequências ressonantes atua em bandas muito estreitas de frequência para corrigir um problema gerados na etapa de captação, o processo de correção tonal atua na melhora da qualidade timbrística global do instrumento musical, ressaltando ou atenuando bandas largas de frequência.

A palavra “tonal” aqui não deve ser confundida com “tonalidade musical” ou “tom puro”, mas deve ser entendida como uma qualidade global da percepção do espectro. Essa qualidade global refere-se a distribuição adequada energia por frequência que representa o instrumento tocado. Palavras usuais em estúdio como “som agudo/grave”, “claro/escuro”, “nítido/opaco”, tentam qualificar e dimensionar o balanço tonal do áudio escutado.

Ao final desse processo de correção é comum dizermos que alcançamos um bom balanço tonal, ou seja, através do ajuste do equalizador foi atingido uma sonoridade psicoacusticamente equilibrada no domínio da frequência. A Figura 23 mostra a tela de um

*plugin*³⁸ de equalização de 6 bandas de frequência. Neste exemplo, na banda 1 está sendo corrigida uma frequência de ressonância (banda estreita) e a banda 6 está sendo utilizada para compor parte do balanço tonal do instrumento musical (banda larga).

Figura 23. Equalizador Digital: Correção Espectral e Balanço Tonal



Fonte: Waves Audio (2017)³⁹

3.4 PRINCIPAIS MÉTODOS DE MIXAGEM

Não há uma forma canônica ou padrão de mixagem de áudio. Cada profissional varia o seu método para atingir o equilíbrio sonoro dependendo da sua experiência profissional, estilo musical, preferência do artista e equipamento disponível. Mesmo assim, Moss (2018a) apresenta em seu curso online de mixagem “*Behind the Speakers*” dois modelos genéricos que englobam a maioria desses métodos. São eles o “*Bottom Up Mixing*” e o “*Top Down Mixing*”. Cada um desses métodos possui três etapas fundamentais distintas, as quais são detalhadas abaixo.

3.4.1 Bottom Up Mixing

O título “*Bottom Up*” vem do conceito piramidal de ouvir e trabalhar primeiro nos instrumentos de modo individual (base da pirâmide) e gradativamente ir trabalhando no resultado da soma desses instrumentos (topo da pirâmide). Sendo assim, as etapas do método “*Bottom Up*” são:

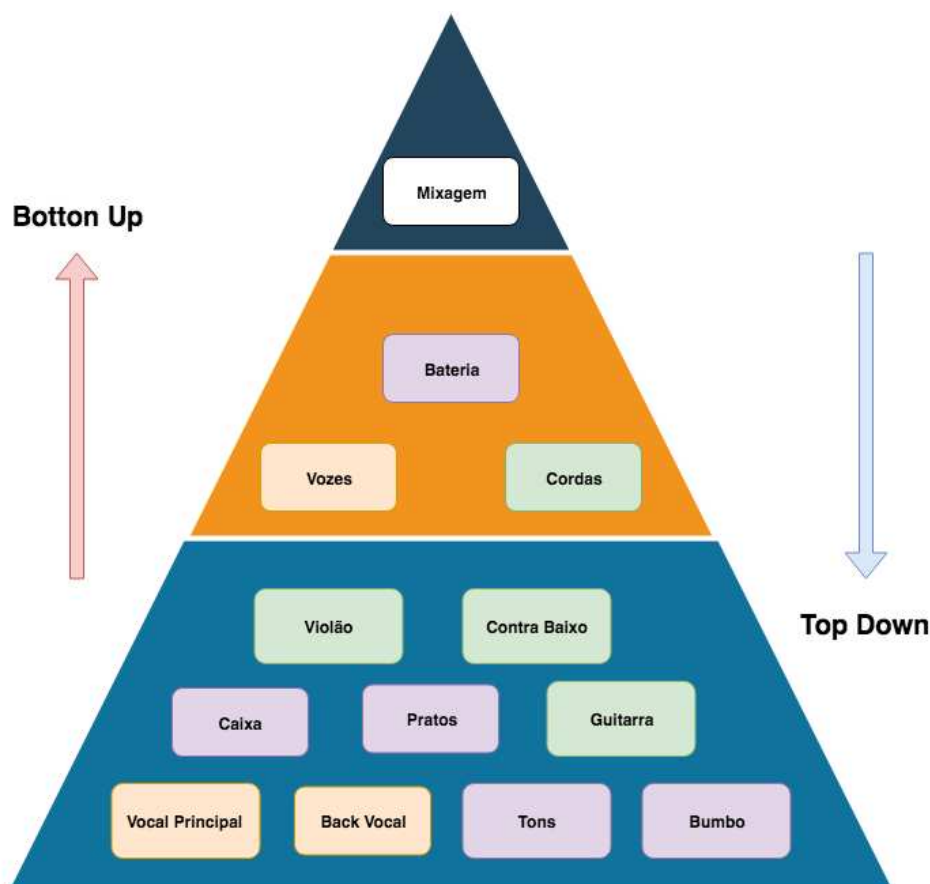
³⁸ Componente de *software* que adiciona uma funcionalidade específica para um outro *software* existente

³⁹ Disponível em: <<https://www.prosoundweb.com/channels/live-sound/waves-audio-now-shipping-f6-floating-band-dynamic-eq-plugin/>>

- Ajustar um volume arbitrário para cada instrumento em “*solo mode*⁴⁰” de modo que consiga ouvir bem a sonoridade do instrumento, porém ignorando a relação dele com os demais instrumentos da mixagem.
- Realizar o tratamento da sonoridade de cada instrumento musical individualmente, equalizando, comprimindo, adicionando efeitos a cada um deles de modo a aprimorar o som individual de cada fonte sonora.
- Realizar o processo de equilíbrio de volume de cada instrumento tratado em relação aos outros instrumentos, de modo a obter uma sonoridade equilibrada da soma (*mix*).

A Figura 24 nos exemplifica o conceito de mixagem “*Bottom Up*” e “*Top Down*”.

Figura 24. Métodos de Mixagem



Fonte: Própria

3.4.2 Top Down Mixing

⁴⁰ Quando ativado o sistema de som reproduz apenas aquele canal desejado

Este método realiza o procedimento inverso, não se atentando à sonoridade individual de cada instrumento, mas sim à sonoridade da soma resultante, ou seja, a sonoridade da mixagem atual naquele momento. De acordo com esta sonoridade, tenta-se atingir os objetivos técnicos e estéticos atuando diretamente no áudio já mixado. O que não for possível corrigir atuando no áudio mixado, é corrigido pontualmente no instrumento individual.

Sendo assim, as etapas do método “*Top Down*” são:

- a) Realizar um equilíbrio mediano de volumes iniciais entre os instrumentos de modo que a soma resultante (*mix*) seja considerada medianamente equilibrada, mesmo havendo necessidade de compressão e equalização dos canais individuais.
- b) Aplica-se os equipamentos de processamento de sinais (equalizadores, compressores, etc.) no sinal mixado, visando-se aproximar o máximo possível do ideal estético-musical desejado
- c) São realizados pequenos ajustes nos instrumentos individuais de modo a aperfeiçoar a sonoridade final da mixagem

Em seu curso de mixagem, Moss (2018a) recomenda que seja usada a técnica “*Top Down*”, pois ela atinge a sonoridade desejada com mais rapidez e requer menos equipamentos ou *plugins* de áudio, já que a maior parte do processamento está sendo realizada diretamente no áudio já mixado, restando pouco para ser reprocessado nos canais/instrumentos individuais. A Figura 24 exemplifica também o método “*Top Down*”.

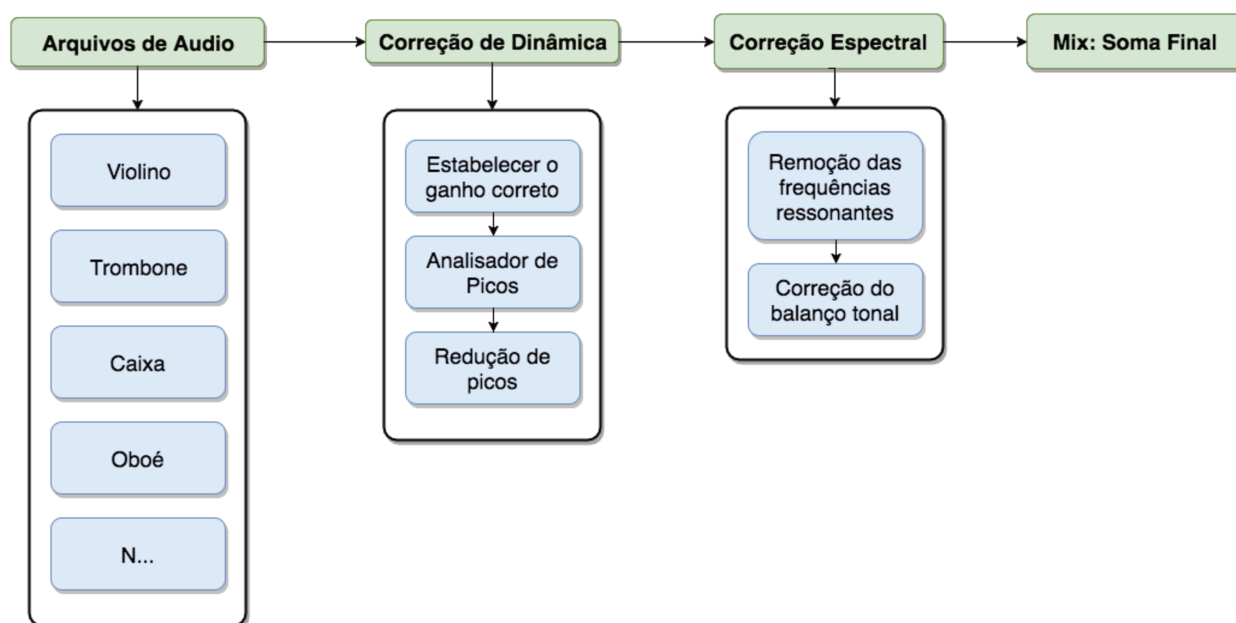
Até aqui apresentamos os conceitos básicos dos processos de gravação e mixagem, a complexidade do sistema auditivo, os principais equipamentos de mixagem e como seus parâmetros podem alterar a nossa percepção de aspectos psicoacústicos do som gravado. No próximo capítulo utilizaremos os conceitos desta revisão na criação e implementação de um novo método de mixagem automática, o qual tem como base a curva de resposta de frequência da cóclea humana.

4 MÉTODO

Como mencionado anteriormente, os principais parâmetros de uma mixagem são referentes à análise e ao controle da variação de dinâmica (amplitude) e a análise e controle do conteúdo espectral (frequência) do áudio. Sendo assim, um engenheiro de som consegue fazer as correções de dinâmica usando métodos de compressão e as correções espectrais usando métodos de equalização para cada instrumento da mixagem.

Na Figura 25 temos uma visão geral da proposta conceitual do nosso modelo de mixagem automatizada tendo como entrada um conjunto de arquivos de áudio (instrumentos musicais) e, como saída, um arquivo de áudio resultante da mixagem automatizada de todas os instrumentos musicais.

Figura 25. Visão geral do nosso método de mixagem automatizada



Como a escuta é um fenômeno psicoacústico, propomos neste trabalho usar o conceito da não linearidade auditiva como um meio de encontrar um sinal genérico que represente o equilíbrio psicoacústico. As curvas isofônicas nos informam como o ouvido humano médio escuta frequências em um determinado nível de pressão sonora (SPL^{41}). A ideia então é aproximar o espectro da mixagem ao espectro de um sinal genérico que tenha em si as

⁴¹ Sound Pressure Level

características não lineares do ouvido, otimizando assim o equilíbrio espectral (distribuição de energia) do material gravado e, conseqüentemente, melhorando a qualidade sonora.

O método proposto neste trabalho é dividido em 7 partes sequenciais, as quais contém vários processos imitativos da práxis atual de estúdio. Essas partes foram implementadas computacionalmente usando a linguagem e ambiente de programação MATLAB. Na Figura 26 mostramos a legenda dos fluxogramas e nas seções abaixo descrevemos cada uma das partes do processo computacional.

Figura 26. Legenda dos Fluxogramas



4.1 IMPORTAÇÃO DOS DADOS

O método aqui apresentado importa um conjunto de arquivos de áudio de uma música previamente gravada para dentro do *software* desenvolvido neste trabalho. Cada um destes arquivos de áudio corresponde a um único instrumento musical. Exemplo: guitarra, vocal, contrabaixo, etc.

Neste processo de importação, cada arquivo de áudio é representado por um vetor inicial, ou seja, uma sequência numérica que, em cada entrada numérica, representa o valor de amplitude do sinal numa determinada fração de tempo. É importante lembrar que o tamanho desse vetor varia de acordo com a duração da música gravada. A Eq. (4) nos mostra como o tamanho do vetor pode ser alterado de acordo com a duração da música e a taxa de amostragem escolhida.

$$S = D * \tau \quad (4)$$

Onde:

S é a quantidade total de amostras dentro do vetor

D é a duração da música, em segundos (s)

τ é a taxa de amostragem escolhida no momento da gravação, em Hertz (Hz)

Por exemplo, numa música com duração de 3 minutos e que foi gravada numa taxa de amostragem padrão de CD (44.100 Hz) temos: $(3 * 60) * 44100 = 7.938.000$ valores de amplitude por instrumento.

Além da taxa de amostragem, durante a gravação é também escolhida a resolução (*bit depth*) do áudio a ser gravado. A resolução diz respeito à precisão do ADC durante a amostragem. No caso do CD, a resolução padrão é de 16 bits por amostra, ou seja, $2^{16} = 65.536$ níveis possíveis em que a amplitude do sinal analógico pode ser representada. No caso de gravação em estúdio profissional, DVD ou Blu-ray a resolução padrão é de 24 bits por amostra, ou seja, $2^{24} = 16.777.216$ níveis em que a amplitude do sinal analógico pode ser representada. Atualmente, 24 bits é a resolução máxima para áudio. Por convenção, neste trabalho iremos usar os valores de amplitude normalizados entre -1 e +1.

Apenas para exemplificar a relação do vetor inicial de áudio com a taxa de amostragem e a resolução, mostramos na Eq. (5) como calcular o tamanho de um arquivo de áudio digital em um formato não comprimido:

$$T = \frac{S * R}{8 * 10^6} \quad (5)$$

Onde:

T é o tamanho do arquivo de áudio, em megabytes (Mb)⁴²

S é a quantidade total de amostras dentro do vetor

R é a resolução por amostra, em bits

A partir do momento que importamos os arquivos dentro do nosso *software* e temos a representação de cada um deles como vetores de amplitudes podemos dar início à etapa de pré-processamento.

4.1.1 Pré-Processamento e Mix Inicial

O pré-processamento consiste na normalização e redução de ganho de cada arquivo de áudio (vetor inicial). A normalização é necessária para que não haja grande predominância de um instrumento sobre o outro devido a diferenças de intensidade sonora no momento da gravação/captação do som gerado pelo instrumento musical. Depois da normalização, cada

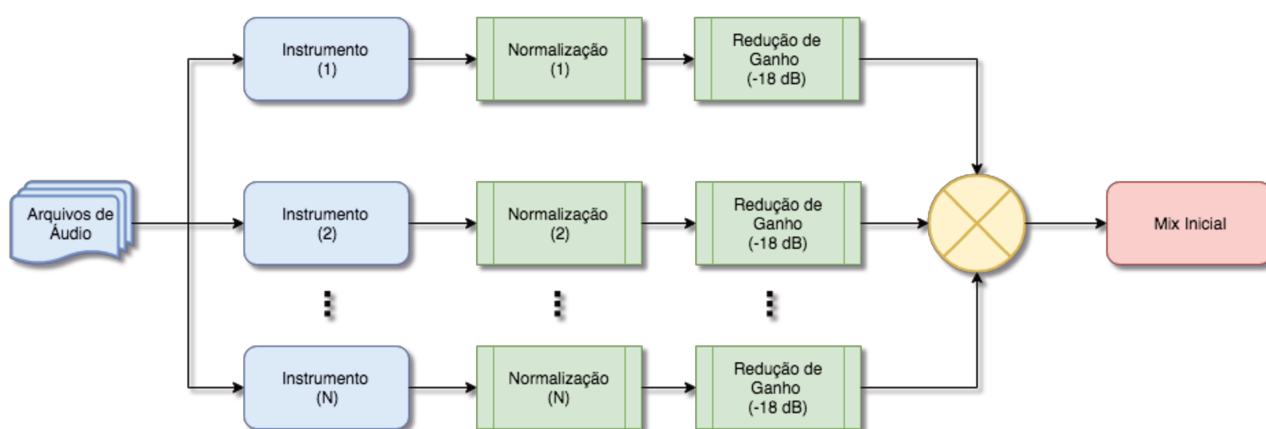
⁴² 1 byte = 1 bits

signal é representado digitalmente em sua máxima intensidade possível, atingindo nos picos de intensidade amostras com valores de +1 ou -1, os quais são os valores máximos da amostragem digital equivalente a 0 dBFS⁴³.

Para que possamos trabalhar com esses vetores de forma a não distorcer o áudio digital original e otimizar o sinal para futuros processamentos, cada vetor passa por uma redução de ganho padrão (-18 dB⁴⁴ ou -20 dB⁴⁵) conforme a norma da *European Broadcasting Union* (2000) mantendo a relação de volume obtido na normalização, porém possibilitando a mixagem desses vetores sem que o vetor resultante da soma tenha distorções (tais como o *digital clipping*⁴⁶).

Uma vez que os vetores de áudio foram normalizados e otimizados para posterior processamento, nós somamos todos eles e obtemos um novo vetor de amostras aqui intitulado como “Mix Inicial”, o qual será uma das referências sonoras utilizadas pelo nosso método para decidir a relação ideal de volume, equalização e compressão entre os instrumentos. Podemos observar isto no fluxograma apresentado na Figura 27:

Figura 27. Pré-processamento e Mixagem Inicial: Importa os arquivos de áudio



O termo “mixagem” neste pré-processamento é usado no sentido literal de soma dos arquivos individuais que gerar um novo arquivo de áudio contendo todos os instrumentos musicais previamente importados. Podemos ver um exemplo de soma amostra a amostra na

⁴³ dBFS = Decibel Full Scale

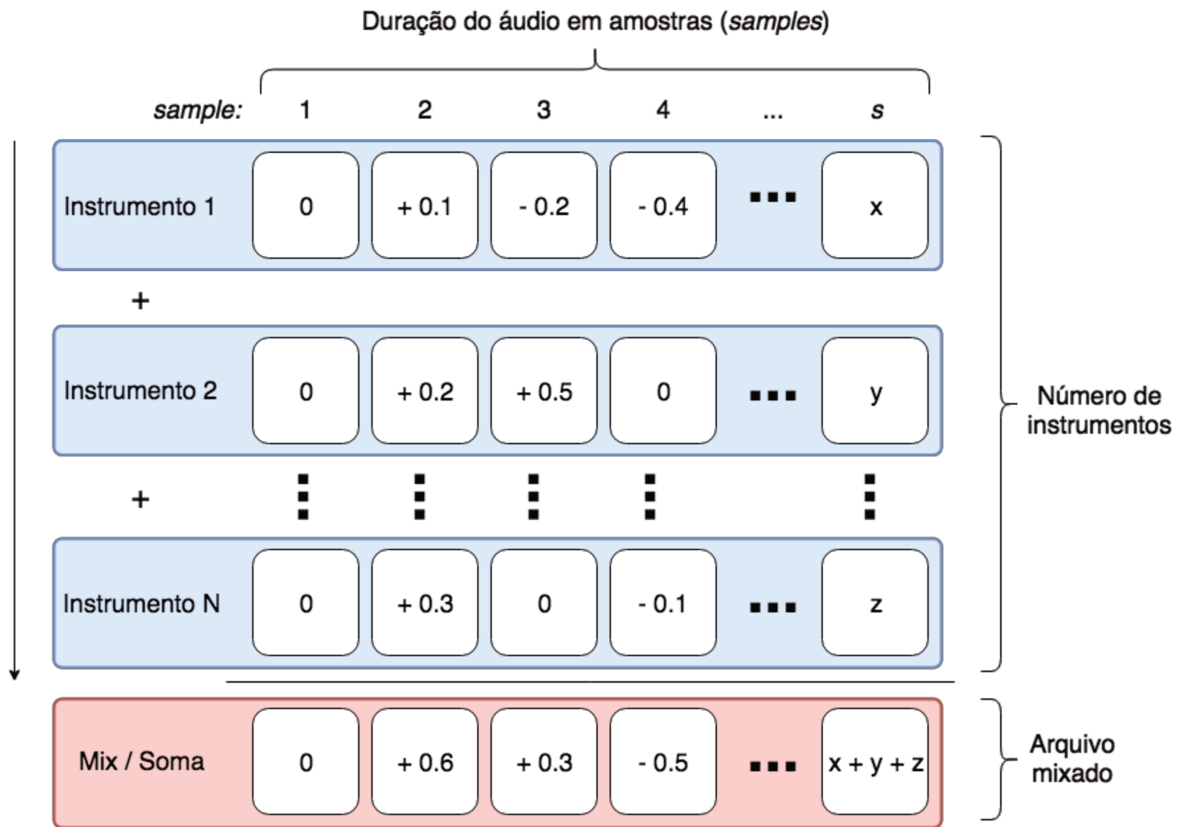
⁴⁴ Padrão europeu de calibração. 0dB VU = -18 dBFS

⁴⁵ Padrão americano de calibração. 0dB VU = -20 dBFS

⁴⁶ Refere-se a distorção do sinal de áudio quando a amplitude do sinal tende a ultrapassar o limite de resolução de intensidade do conversor analógico-digital, fazendo com que o sinal digital seja ceifado nas intensidades extremas (1 e -1) o que corresponde a gerar parciais no domínio da frequência que deturpam a qualidade sonora deste arquivo.

Figura 28, onde cada arquivo de áudio é representado por um vetor horizontal de amostras e cada amostra é representada por um quadrado arredondado contendo um valor de amplitude normalizada entre -1 e +1.

Figura 28. Mixagem: Soma amostra por amostra



Fonte: Própria

Em termos matemáticos, podemos expressar esta soma usando a fórmula abaixo:

$$M(s) = \sum_{n=1}^N A(n, s) \quad \text{para } s = 1, 2, \dots, N \quad (6)$$

Onde:

M é o vetor temporal que armazena a soma das amostras (*samples*)

s é o índice da amostra (*sample*)

n é o índice do instrumento

N é o número total de instrumentos

A é a matriz que armazena os valores de amplitude de cada instrumento

Observe que, devido à redução de ganho prevista pela norma European Broadcasting Union (2000), os valores do vetor M estão no intervalo $[-1 \ +1]$.

Como estamos trabalhando com músicas em estéreo, o conteúdo musical é registrado sempre em dois canais independentes (esquerdo e direito), assim o processo de mixagem acima deve ser feito tanto para o canal direito quanto para o canal esquerdo, o que resulta em dois vetores distintos. Para simplificar a nomenclatura e incluir os arquivos mono (apenas 1 canal), neste trabalho vamos nos referenciar ao nome dos vetores sempre no singular, mesmo que eles contenham informações para o caso estéreo.

4.1.2 Obtenção do Espectro Inicial

Dentro do universo da música comercial, podemos facilmente observar que, devido ao tipo da forma musical, muitas músicas iniciam com poucos instrumentos tocando efetivamente, o que resulta em um conteúdo espectral mais limitado nesse trecho. Também é fácil de observar que partes como “refrão” ou “ponte” normalmente possuem uma densidade maior de instrumentação, o que resulta em um conteúdo espectral mais rico e extenso, em toda a extensão do espectro de frequências (do grave ao agudo).

Assim, na prática de estúdio, o engenheiro de som foca a sua atenção no trecho musical em que maioria dos instrumentos estão tocando simultaneamente, dos mais graves aos mais agudos, para que, ao ouvir todo o intervalo de frequências dessa região, possa manipular digitalmente todo o arquivo, a fim de otimizar o equilíbrio psicoacústico da mixagem, não somente deste intervalo temporal, mas de toda a música.

Seguindo esse fundamento psicoacústico e a prática do engenheiro de som, obtemos a intensidade de nível percebida (*loudness*) através do cálculo do *Short Term Loudness Descriptor*. Resumidamente, o cálculo do *Short Term Loudness Descriptor* segmenta o sinal de áudio em janelas de curta duração (padrão de 3 segundos) e calcula o *loudness* para cada janela usando a escala LU⁴⁷ ou LUFS⁴⁸. O resultado desse processo é um conjunto de valores escalares armazenados em um vetor temporal. Mais detalhes sobre o cálculo do *Short Term Loudness* pode ser encontrado em International Organization for Standardization (2017).

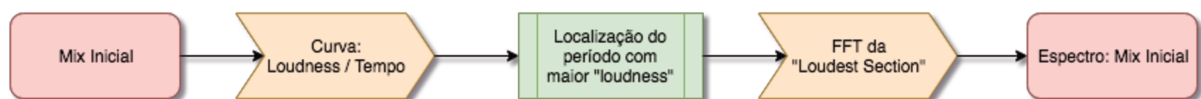
⁴⁷ *Loudness Unit*: 1 LU = 1 dB

⁴⁸ *Loudness Unit Full Scale*: 0 LU = -23 LUFS, conforme a norma European Broadcasting Union (2016)

A partir do vetor temporal “Mix Inicial”, calculado acima, obtemos o seu espectro no domínio da frequência através da Transformada de Fourier de Tempo Discreto ($DTFT$ ⁴⁹), porém executamos a $DTFT$ apenas no intervalo temporal com maior “volume percebido” (*loudness*). Para o propósito deste trabalho vamos apenas usar o conceito de que a $DTFT$ nos dá o espectro de um sinal temporal e o *loudness* nos dá a intensidade de nível sonoro percebida pelo ouvinte. Mais detalhes sobre a $DTFT$ podem ser encontrados em Oppenheim et al. (1999).

O vetor de *loudness* obtido nos dá uma série temporal na qual podemos localizar trechos onde há maior número de instrumentos tocando com dinâmica forte, sendo compatíveis com um “refrão” ou uma “ponte”. A este trecho musical (intervalo temporal) denominamos “*loudest section*”. Por fim, aplicamos a $DTFT$ na “*loudest section*” usando o algoritmo FFT (*Fast Fourier Transform*) e obtemos o espectro da Mix Inicial como pode-se ver na figura abaixo.

Figura 29. Obtenção do Espectro Inicial



Normalmente a resolução do resultado da $DTFT$ depende da duração do sinal analisado, ou seja, para sinais temporais curtos, temos baixa resolução espectral, e para sinais temporais longos, temos alta resolução espectral. De acordo com Rossing et al. (2001) a resolução biológica da cóclea humana é aproximadamente equivalente à segmentação do espectro auditivo na escala de 1/3 de oitava. Além disso, o padrão de análise espectral em terço de oitava é amplamente utilizado nas normas internacionais, como por exemplo na International Organization for Standardization (2003). Por esses motivos, neste trabalho iremos analisar espectralmente os sinais usando a resolução de 1/3 de oitava.

O processo de $DTFT$ (1/3 de oitava) nos dá um valor amplitude normalizada entre 0 e 1 para cada uma das 31 bandas de frequência analisadas e seu resultado é armazenado em um vetor de tamanho 31×1 . A tabela abaixo exemplifica o armazenamento deste vetor, sendo que computacionalmente gravamos apenas a última coluna da tabela (Valor Armazenado).

⁴⁹ Discrete-Time Fourier Transform

Tabela 4 - Exemplo de armazenamento de dados em um vetor de dimensão 31×1

Índice do Vetor	Banda de Frequência Correspondente	Valor Armazenado
1	20 Hz	0.8
2	25 Hz	0.2
3	31.5 Hz	1.0
...	...	...
31	20 kHz	0.5

É importante ressaltar que mesmo usando o espectro do trecho musical com maior intensidade instrumental (refrão, ponte, etc.) após a conclusão total do processo de mixagem os trechos com menor intensidade instrumental também estarão dentro de um equilíbrio sonoro desejável. Isso acontece, pois, ao equilibrar as seções mais densas, é possível obter uma relação de volume e equilíbrio espectral adequada entre cada instrumento nas seções menos densas da música. A relação de volume e equilíbrio espectral será melhor detalhado nos próximos capítulos.

4.1.3 Sinal de Referência

A qualidade de uma mixagem é um conceito, seja do ponto de vista físico ou psicoacústico, complexo e multivariado. Por isso, quando falamos da qualidade de áudio dentro do domínio musical, não usamos parâmetros absolutos, mas sim relativos. A maneira mais comum de se fazer essa comparação é utilizando uma gravação musical “de referência” visando a mesma qualidade e estética para o processo de mixagem de uma determinada música.

Essa gravação de referência deve ser musicalmente compatível com a música que se pretende mixar, ou seja, deve ser do mesmo gênero musical, ter instrumentação próxima, andamento aproximado e outras características musicais e estéticas que aproximem a música que se deseja mixar da referência escolhida. Quando o engenheiro encontra uma gravação de referência válida para o seu caso específico, ela o ajuda a direcionar os processos da mixagem de modo a conseguir uma sonoridade parecida com a da gravação de referência, pelo uso da comparação auditiva dos dois arquivos de áudio.

O problema desse método é a subjetividade do ouvinte, a qual dificulta a decisão de como e quais parâmetros da mixagem devem ser alterados para atingir a sonoridade da referência. Por mais que os parâmetros musicais citados acima sejam importantes para a validação de uma gravação de referência, neste trabalho acreditamos que se possa ter uma outra

fonte de referência sonora que possibilite o controle dos parâmetros mais significantes do processo de mixagem, independente do gênero musical, instrumentação, andamento ou outros aspectos sonoros ou características musicais.

Partindo das curvas isofônicas detalhadas no capítulo 2, pensamos inicialmente que a nossa referência psicoacústica poderia ser um som que tivesse o seu espectro moldado de modo que todas as frequências pudessem ser ouvidas igualmente. Seguindo este raciocínio fizemos alguns experimentos gerando o ruído cinza (*grey noise*, detalhado no capítulo 2) o qual possui características espectrais aqui desejadas. Logo no início dos experimentos constatamos experimentalmente que o sistema auditivo humano trabalha de forma logarítmica, tanto na percepção da quantidade de energia quanto na percepção de altura (*pitch*). Por esse motivo notamos que o ruído cinza não seria uma referência psicoacústica adequada, pois nele ouvimos cada frequência do espectro com a mesma sensação de intensidade. Como há uma maior quantidade de frequências agudas no espectro auditivo (exemplo: só a banda 10kHz a 20kHz ocupa metade do nosso espectro auditivo), a quantidade de informação aguda se torna exagerada sendo impossível ter um parâmetro de equilíbrio espectral.

De acordo com Bazil (2014) e Silverstein (2018), o ruído rosa (*pink noise*, conforme definido no capítulo 2) seria adequado para a função de um sinal de referência psicoacústica para mixagem, pois sua energia é distribuída no espectro de modo a que cada banda espectral tem a mesma quantidade de energia, resolvendo o problema da predominância dos agudos no *grey noise*.

É importante ressaltar que o uso do *pink noise* neste método de mixagem automatizada não é obrigatório. Pode-se, por exemplo, escolher uma gravação de referência, e fazer a *DTFT* da sua sessão com maior *loudness* e obter um espectro de referência válido. Para este estudo inicial fixamos o parâmetro do espectro de referência com o espectro do *pink noise* para todos os testes realizados, facilitando o exame dos dados e também seguindo as vantagens psicoacústicas deste ruído, descritas no capítulo 2.

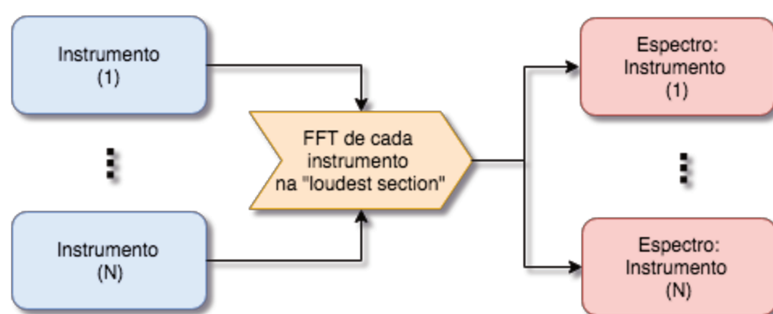
4.1.4 Comparação Espectral

Agora que temos o espectro da “Mix Inicial” e o espectro do sinal de referência (*pink noise* ou gravação de referência) podemos fazer uma comparação entre os dois espectros calculando a diferença de energia em cada uma das bandas de terço de oitava. Essa diferença nos mostra os primeiros problemas de equilíbrio espectral da “Mix Inicial”. Esta informação será utilizada nas próximas etapas do método.

4.1.5 Obtenção do Espectro de cada Instrumento

Assim como foi feito com o áudio da “Mix Inicial”, calculamos a *DTFT* (1/3 de oitava) da seção de maior *loudness* de cada instrumento a ser mixado e obtemos o espectro de cada instrumento musical. O vetor, resultado de cada *DTFT* é de dimensão 31×1 , mas como estamos trabalhando com N instrumentos, o resultado final deste processo é uma matriz de espectros de dimensão $31 \times N$. O fluxograma abaixo mostra a sequência desses passos.

Figura 30. Obtenção do Espectro de cada Instrumento



Essa matriz de espectros será usada em processos posteriores para a obtenção do equilíbrio do volume e balanço espectral.

4.2 REMOÇÃO DE FREQUÊNCIAS INDESEJÁVEIS

A grande maioria dos sinais provindos de captação sonora de instrumentos musicais contém em seu conteúdo espectral frequências indesejáveis do ponto de vista psicoacústico. Essas frequências são geralmente relacionadas a perda de clareza e definição da sonoridade acústica do instrumento e são causadas em sua maioria por problemas acústicos da sala de gravação (modos de sala), má confecção física ou eletrônica do instrumento ou aplicação de técnica de microfonação não apropriada para o instrumento em questão.

A remoção das frequências indesejáveis é importante para o processo da mixagem, pois, uma vez realizado este processo, é possível ter uma audição mais bem definida dos instrumentos gravados pela redução do mascaramento de certas frequências sobre outras. Abaixo discutimos dois tipos de frequências indesejáveis: frequências graves e frequências ressonantes.

4.2.1 Definição de Relevância por Banda

Para sabermos quais frequências podem ter suas amplitudes manipuladas, criamos uma matriz de “Relevância por Banda” de dimensão $31 \text{ bandas} \times N \text{ instrumentos}$. Essa matriz normalizada de 0 a 1 nos dá uma ideia da influência/peso de uma determinada banda de frequência sob a sonoridade geral do instrumento. O cálculo do peso de uma determinada banda de um instrumento é dado por:

$$P(n, b) = \frac{A(n, b)}{\sum_{b=1}^{31} A(n, b)} \quad (7)$$

Onde:

P é o peso da banda de terço de oitava normalizada entre 0 e 1

A é o vetor que armazena as amplitudes de cada terço de oitava

b é o índice da banda de terço de oitava entre 1 e 31

n é o índice do instrumento musical

4.2.2 Frequências Graves: Abaixo de 100Hz

A remoção do excesso de frequências graves é feita por um filtro passa-altas digital (HPF⁵⁰), o qual remove as frequências graves abaixo de uma frequência de corte dada. A frequência de corte deve ser próxima da frequência fundamental do instrumento, pois assim retiramos toda a sobra de grave “não musical” (como interferência do vento no microfone, sopro do instrumentista, ruídos do ar condicionado, ressonâncias da sala de gravação, etc.) e não retiramos qualquer conteúdo espectral relevante do som gerado pelo instrumento musical.

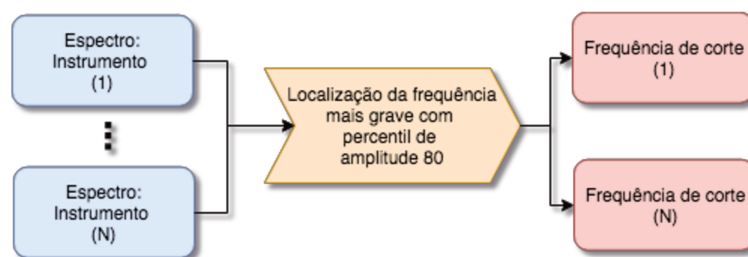
Para determinar a frequência de corte para cada instrumento da mixagem, ordenamos o vetor de espectro de cada instrumento e escolhemos o percentil de amplitude próximo ao 100 (valor da frequência fundamental). Nos experimentos realizados foi encontrado empiricamente que o percentil 80 seria adequado para o nosso estudo, já que esse corte não exige alta precisão.

⁵⁰ High Pass Filter

Como o decaimento aplicado do filtro passa-altas é suave (-12 dB por oitava⁵¹), mesmo que a frequência de corte fique próxima (acima ou abaixo) daquela da nota fundamental, esse corte removerá somente a sobra de graves sem afetar a sonoridade geral do instrumento.

O resultado desse processo é um conjunto de valores de frequências de corte que são armazenados num vetor de dimensão $1 \times N$, onde N é o número de instrumentos musicais. A Figura 31 mostra esse processo de obtenção das frequências de corte por instrumento musical.

Figura 31. Localização da Frequência de Corte



4.2.3 Frequências Ressonantes

Em áudio, as denominadas “Frequências Ressonantes” (*resonant/ringing frequencies*) são falhas específicas e pontuais do conteúdo espectral que aparecem em alguns estreitos intervalos de frequência. Tais frequências possuem grande amplitude e elevado destaque sonoro, prejudicando a sonoridade geral e o equilíbrio psicoacústico do conteúdo espectral do instrumento musical.

Em áudio, o nome de “frequências ressonantes” não está ligado a ressonância física do som gravado com o ambiente externo no momento da reprodução. Usualmente, engenheiros de som usam esta nomenclatura para este tipo de falha espectral gerada no momento da gravação/captação do instrumento. Essa falha espectral tem uma sonoridade característica parecida com uma ressonância acústica e pode ser causada por vários motivos como: mal posicionamento do microfone, sala de gravação inadequada, má qualidade do instrumento musical e outros.

De acordo com os aspectos físicos da onda sonora, frequências graves possuem comprimento de onda largo e por isso são mais propensas a geração de modos/nós de ressonância em salas pequenas. Podemos ver a relação entre a frequência e o comprimento de onda na equação abaixo:

⁵¹ Decaimento de energia de -12 dB para cada oitava abaixo da frequência de corte escolhida

$$\lambda = \frac{v}{f} \quad (8)$$

Onde:

λ é o comprimento de onda em metros

v é a velocidade do som no meio. Para o ar em 25°C o valor de $v \approx 346 \text{ m/s}$

f é a frequência em hertz

Nos experimentos realizados, encontramos empiricamente que o número máximo de frequências ressonantes que podem ser encontradas em um sinal de um instrumento musical é próximo de 10, porém, dentro da percepção humana, algumas não são psicoacusticamente relevantes, portanto não são prejudiciais à sonoridade do instrumento.

A práxis atual deste processo é baseada na localização/remoção somente das frequências ressonantes audíveis pelo engenheiro de som durante a mixagem. É importante lembrar que a construção da sala, qualidade das caixas de som, bem como o treinamento auditivo do engenheiro são aspectos essenciais para a identificação de problemas relativos ao conteúdo espectral dos sons de um instrumento musical.

No método que apresentamos abaixo, criamos uma forma de identificação desses pequenos problemas do conteúdo espectral de modo independente da sala de mixagem, dos equipamentos usados ou até mesmo da escuta humana. O método é implementado como um processo que deve ser repetido para todos os instrumentos musicais presentes na mixagem e consiste no seguinte:

- a) Localização da sessão com maior *loudness*
- b) Aplica-se à esta seção uma equalização temporária com os seguintes parâmetros
 - a. Frequência: *Sweep*⁵² de 50Hz a 2 kHz (onde há predominância do fenômeno)
 - b. Fator $Q = 10$ (aproximadamente 1/8 de banda de frequência)⁵³

$$Q = \frac{\sqrt{2BW}}{2BW - 1} \quad (9)$$

Onde:

Q é o fator Q adimensional. Mais detalhes sobre o cálculo do Fator Q em Bohn (2008)

⁵² Varredura espectral

⁵³ Largura de banda. Na prática de estúdio, o valor 10 é usual para a configuração de uma banda muito estreita

BW é a largura de banda (*bandwidth*) em oitavas

- c. Ganho = +12 dB (de acordo com a práxis de estúdio atual)
- c) Para cada valor de frequência instanciado verificamos a razão entre a amplitude máxima do sinal equalizado (*max peak value*) e a amplitude máxima do sinal original.
- d) Se o resultado dessa razão é maior ou igual a 2, ou seja +6 dB, então assumimos que a frequência instanciada é ressonante.

Assim, podemos descrever a ocorrência do fenômeno da localização de uma frequência ressonante seguindo o critério abaixo:

$$\frac{\max|A_{eq}(f)|}{\max|A_{or}|} \geq 2 \quad \text{para} \quad 50Hz \leq f \leq 2000Hz \quad (10)$$

Onde:

f é a frequência atual usada na equalização do instrumento musical

$A_{eq}(f)$ é o sinal do instrumento equalizado usando a frequência f

A_{or} é o sinal do instrumento original sem equalização⁵⁴

O resultado desse processo é um vetor de 1950 entradas (de 50 Hz a 2000 Hz). Em nosso processo computacional tomamos os 10 valores de amplitude mais proeminentes que ultrapassam o valor 2 e calculamos a redução de ganho a ser aplicado de modo a diminuir a ressonância naquela área espectral, como podemos ver na Eq. (11).

$$G(f) = \begin{cases} 0, & \text{se } \frac{\max|A_{eq}(f)|}{\max|A_{or}|} < 2 \\ -20 * \log\left(\frac{|A_{eq}(f)|}{|A_{or}|}\right), & \text{se } \frac{\max|A_{eq}(f)|}{\max|A_{or}|} \geq 2 \end{cases} \quad \text{para } 50 \leq f \leq 2000 \quad (11)$$

Onde:

G é o vetor que armazena os valores de redução de ganho em dB

⁵⁴ Note que Or não tem o índice f , pois é o sinal original e não varia com a frequência de equalização

Na subseção abaixo descrevemos mais detalhadamente o processo de remoção das frequências ressonantes.

4.2.4 Remoção das Frequências por Instrumento

Depois de obtido os valores das frequências ressonantes armazenadas em G e das frequências de corte de sobra de grave, instanciamos um equalizador paramétrico para cada instrumento musical e neste conjunto de equalizadores aplicamos as configurações previamente calculadas para a remoção das frequências indesejáveis. Experimentalmente, os novos vetores de áudio resultantes desse processo possuem uma sonoridade com um pouco mais de definição, clareza e limpidez quando comparados com os arquivos originais.

Essa etapa de pré-processamento espectral é importante para a próxima etapa (pré-processamento da dinâmica), pois o conjunto dos novos sinais usados nos cálculos de dinâmica não conterão *samples* com amplitudes desproporcionais à real sonoridade dos instrumentos musicais.

4.3 CONTROLE DA DINÂMICA

A dinâmica musical é um aspecto importante na interpretação e execução musical, sendo essencial o seu controle no momento da gravação de um instrumento.

Ao gravar um instrumento musical acústico, as variações de dinâmica executadas pelo músico têm uma implicação direta na amplitude do sinal gravado. Digitalmente falando, podemos dizer que quanto mais forte for a dinâmica executada pelo músico, maior será o valor de amplitude de cada amostra gravada.

Por mais que a variação de dinâmica seja algo intencional e de controle do músico executante, quando observamos cada amostra do sinal gravado percebemos que o controle da dinâmica feito pelo músico no momento da gravação é insuficiente para obtermos valores de amplitude de cada amostra dentro de um intervalo de amplitude rigorosamente calculado. Para exemplificar isso basta lembramos que a taxa de amostragem (*sample rate*) usada em áudio, que varia de 44.1 KHz até 192 KHz e a resolução usada que é de 16 ou 24 bits, implicam que, mesmo na menor taxa de amostragem e com baixa resolução, o músico teria que ter um controle “manual” sobre a amplitude de cada uma das 44.100 amostras por segundo, dentro de um intervalo de dinâmica entre 0 (*pianissíssimo*) e $2^{16} = 65.536$ (*fortissíssimo*), o que torna esse processo “manual” humanamente impossível.

Na prática atual os engenheiros de som usam compressores de áudio (detalhado no capítulo 3) para controlar a dinâmica musical e obter um intervalo de amplitudes mais controlado. É importante ressaltar que a normalização ou a compressão moderada de um sinal de áudio não afeta a nossa percepção de variação de dinâmica musical (*p*, *mf*, *ff*⁵⁵). A normalização de áudio é um processo que apenas multiplica todas as amostras por um único fator, de modo que o módulo da amostra com maior amplitude do vetor original seja igual a +1 depois da normalização. A compressão de um sinal de áudio (detalhada no capítulo 2) reduz os picos de amplitude do sinal, diminuindo a sua variação no tempo, porém isso não implica na mudança de dinâmica musical perceptual.

Exemplo de normalização: podemos gravar o som de um violino tocando em dinâmica *ppp* (*pianissíssimo*) sendo que o módulo da maior amplitude desta gravação seja igual a +0.1 dentro de um intervalo de -1 a +1, ou seja, 10 vezes menor que o limite do conversor A/D. Ao normalizarmos este vetor, este sinal será multiplicado por um fator (neste caso igual a 10) de modo que o módulo da amostra com amplitude +0.1 se torne +1, elevando todos os valores de amplitude do sinal do mesmo fator. Ao ouvirmos esse sinal normalizado, não haverá mudança na nossa percepção de dinâmica musical, ou seja, continuaremos a ouvir o som de um violino com as características sonoras do instrumento em *ppp*, porém com um maior *loudness*.

Exemplo de compressão: mesmo depois do sinal normalizado, podemos ainda observar que os picos do sinal estão muito mais altos que a média das amostras, impedindo a maximização do *loudness* por causa dos limites de amplitudes digitais (-1 e +1). O uso do compressor e a devida configuração de seus parâmetros (detalhados no capítulo 2) possibilita a redução da amplitude dos picos do sinal, de modo que seja possível aplicar mais ganho a todo o sinal, maximizando o *loudness* sem que haja distorção digital.

Por mais que haja um estudo formal dos parâmetros do compressor, a grande complexidade dos sinais musicais faz com que o uso e o ajuste dos parâmetros deste equipamento não seja algo estritamente matemático, deixando a regulagem de cada parâmetro à experiência e subjetividade dos critérios perceptivos intuitivamente adotados pelo engenheiro de som.

Sendo o foco deste trabalho a mixagem moderna de música comercial, levando em consideração prática moderna de estúdio e a redução de variação dinâmica das gravações de música comercial ao longo dos últimos 50 anos (*Loudness War* discutida no capítulo 2), iremos trabalhar com os valores de amplitude máxima (*Peak value*) próximos ao limite do conversor

⁵⁵ *p* = piano, *mf* = mezzo-forte, *ff* = fortíssimo

A/D (*Peak value* $\cong -0.1$ dBFS). A partir desta consideração desenvolvemos um método de compressão automática que é usado no pré-processamento de cada arquivo de áudio a ser usado na mixagem.

Para cada sinal de um instrumento musical, analisamos o seu valor eficaz (RMS^{56}) e a sua amplitude máxima (*Peak value*). O cálculo do RMS (σ) do n ésimo instrumento é dado por:

$$\sigma(n) = \sqrt{\frac{1}{S} \sum_{k=1}^S A_n^2(k)} \quad \text{para } n = 1, \dots, N \quad (12)$$

Onde:

$\sigma(n)$ é o valor eficaz (RMS) do n ésimo instrumento

k é o índice da amostra (*sample*)

S é o número total de amostras

$A_n(k)$ é a amplitude da amostra k do instrumento n

O cálculo do valor de pico é dado por:

$$P(n) = \max_{1 \leq k \leq S} |A_n(k)| \quad (13)$$

Onde:

P é o valor de pico do instrumento n , ou seja, é a amplitude máxima do sinal

k é o índice da amostra (*sample*)

S é o número total de amostras

$A_n(k)$ é a amplitude da amostra k do instrumento n

Com os valores de RMS e de pico de cada instrumento e com base na pequena variação de dinâmica previamente relatada e de testes realizados em estúdio para música comercial, estimamos empiricamente que a relação entre o valor de RMS e o valor de pico de cada instrumento deve obedecer uma razão de 2:1, ou seja, o valor de pico deve ser no máximo duas vezes maior (+6 dB) que o valor RMS . Auditivamente, ou seja, dentro das práticas de estúdio, percebemos que poderíamos reduzir a variação dinâmica de sinais de instrumentos musicais

⁵⁶ Root Mean Square

sem grande perda perceptual até o limite de aproximadamente $+6 \text{ dB}$ entre o valor RMS e o valor de pico (*Peak Value*).

De acordo com Audio Engineering Society (2009), o nível de amplitude de sinal digital usando a unidade dBFS pode ser descrito como:

$$L(n, k) = 20 * \log(|A_n(k)|) \quad \text{para } k = 1, \dots, N \quad (14)$$

Onde:

$L(n, k)$ é o nível da amostra k do instrumento n em *decibel full-scale* (dBFS)

A é o valor de amplitude da amostra (*sample*)

n é o índice do instrumento

k é o índice da amostra

N é o total de amostras

No nosso caso, quando o valor de pico $P(n)$ é duas vezes maior que o valor RMS $\sigma(n)$, ou seja, $P(n) = 2 * \sigma(n)$, temos a relação desejada de $+6 \text{ dB}$, como pode ser visto na equação abaixo:

$$\Delta L = 20 * \log\left(\frac{2 * \sigma(n)}{\sigma(n)}\right) \approx +6 \text{ dB} \quad (15)$$

Onde:

ΔL é a diferença de nível entre os valores de pico $P(n)$ e de RMS $\sigma(n)$

$\sigma(n)$ é o valor RMS do sinal do instrumento n

n é o índice do instrumento

Deste modo, não só mantemos a sensação psicoacústica da variação de dinâmica musical (*p*, *mf*, *ff*), mas também controlamos com mais rigor as amplitudes de cada amostra gravada. Com esta compressão, podemos mixar um maior número de instrumentos sem distorção do sinal final. Além disso ela também nos possibilita aumentar os níveis de *loudness* nos processamentos futuros.

No nosso caso, como já temos os valores de RMS de cada instrumento, fixamos os valores de *Threshold* de cada compressor com os valores RMS obtidos anteriormente e calculamos os valores de *Ratio* usando a equação abaixo.

$$R(n) = \frac{P(n) - \sigma(n)}{r * \sigma(n)} \quad (16)$$

Onde:

n é o índice do instrumento

$R(n)$ é o valor de *Ratio* do instrumento n

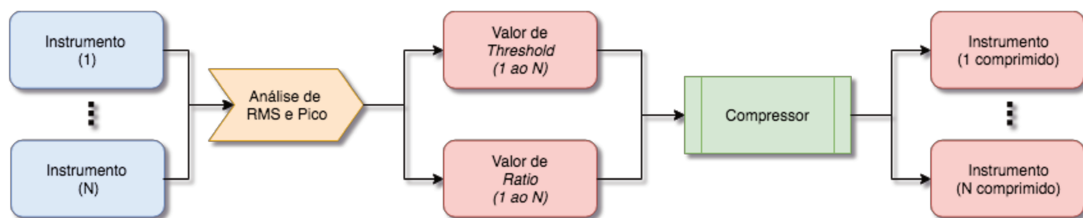
$P(n)$ é o valor de pico do instrumento n

$\sigma(n)$ é o valor de *RMS* do instrumento n

r é a razão desejada entre o novo valor de pico e RMS. Neste caso $r = 2$, ou seja, +6dB

Os valores de *Threshold* e *Ratio* são armazenados numa matriz de dimensão $2 \times N$ onde posteriormente serão utilizados nos N compressores, os quais irão processar os sinais de cada instrumento musical individualmente. Detalhes do funcionamento do compressor podem ser vistos no capítulo 3. A figura abaixo mostra a sequência deste processo de controle de dinâmica por instrumento musical, o qual finaliza o pré-processamento dos arquivos de áudio.

Figura 32. Controle da Dinâmica por Instrumento



4.4 EQUILÍBRIO ENTRE OS VOLUMES

No pré-processamento corrigimos os sinais em termos espectrais e de variação de amplitude (dinâmica), tornando-os mais próximos de “sinais ideais” no âmbito da psicoacústica musical. Nos processos seguintes nosso foco estará no sinal resultante da soma temporal de todos os sinais dos instrumentos musicais a serem mixados. É este sinal resultante que será o objeto final a ser escutado pelo engenheiro de som e que estará sendo avaliado em termos de seus parâmetros psicoacústicos de equilíbrio tonal, espectral e dinâmico, conforme descritos no capítulo 3. Chamaremos esse sinal temporal de “Mix Intermediária”.

Para obtermos uma sensação de equilíbrio tonal na Mix Intermediária precisamos determinar a relação de volume entre os instrumentos musicais. É da escuta deste equilíbrio tonal inicial que o engenheiro de som parte para ajustes de segunda ordem, ou seja, ajustes mais

finos como quantidade de *reverb* por instrumento, efeitos como *delay*, *flanger*, *echo*, alterações no espectro ou na dinâmica de um instrumento, ou até mesmo leves desequilíbrios de volume entre os instrumentos para um fim estético ou de estilo musical. Na prática atual este ajuste de volume é feito pelo engenheiro de som através dos *faders*, que são os controles deslizantes de volume de uma mesa de som, como ilustrado na figura a seguir.

Figura 33. Faders das mesas Yamaha CL Series



Fonte: Yamaha⁵⁷

Cada *fader* atua como controle de ganho do sinal do instrumento musical gravado naquele canal de áudio. A escala dessas marcações é padronizada em dB para facilitar o uso e age num intervalo de +12 dB⁵⁸ até $-\infty$, ou seja, quando o *fader* tem um valor positivo ele aumenta a amplitude do sinal e quando ele tem um valor negativo ele diminui a amplitude do sinal. É importante ressaltar que devido a função logarítmica na definição de decibel, não podemos fazer operações lineares diretamente em unidade dB, por exemplo $10\text{ dB} + 10\text{ dB} \neq 20\text{ dB}$. Assim, todos os nossos cálculos estão em unidades de amplitude (lineares), podendo eventualmente ser apresentados resultados em dB apenas para facilitar a compreensão do fenômeno sonoro nesta unidade usual.

⁵⁷ Disponível em: < <http://www.yamahaproaudio.com/global/en/about/interview/knobs/> >

⁵⁸ Em algumas mesas de som o máximo valor de um *fader* é +6 dB

A equação que descreve um sinal musical passando por um *fader* de volume é dada pela relação linear:

$$A_F(n, k) = F(n) * A(n, k) \quad \text{para } n = 1, \dots, N \quad (17)$$

Onde:

n é o índice do instrumento

$A_F(n)$ é o sinal temporal do instrumento n obtido após o *fader* de volume

$F(n)$ é o valor do *fader* de volume para o instrumento n . Veja Eq. (18) abaixo.

$A(n)$ é o sinal temporal original do instrumento n

k é o índice da amostra (*sample*)

O nosso método de ajuste automático de volume entre os instrumentos pode ser dividido nos seguintes passos:

- a) Cálculo da taxa temporária de volume a ser aplicada em cada instrumento musical:

$$F(n) = \min_{1 \leq b \leq 31} \left(\frac{W_{ref}(b)}{W(n, b)} \right) \quad \text{para } n = 1, \dots, N \quad (18)$$

Onde:

$F(n)$ é a taxa temporária de volume do instrumento n

W é a matriz de espectro dos N instrumentos

W_r é o espectro do sinal de referência

b é o índice da banda de terço de oitava num intervalo de 1 até 31

- b) Aplicação da taxa temporária de volume $F(n)$ em cada instrumento musical é dado pela Eq. (17).
- c) “Mixagem temporária” dos novos sinais modificados é dada pela Eq. (6)
- d) Obtenção do espectro da “mixagem temporária” é dado pela Transformada de Fourier de Tempo Discreto (*DTFT*) descrita no capítulo 2. Chamaremos esse vetor espectral de W_{mt}

- e) Encontro da banda espectral com maior diferença de amplitude entre a “mixagem temporária” W_{mt} e o espectro de referência (*pink noise*):

$$\Delta W_{max} = \max_{1 \leq b \leq 31} \{W_{mt}(b) - W_{ref}(b)\} \quad (19)$$

Onde:

ΔW_{max} é a máxima diferença de amplitude entre os espectros

b é o índice da banda de terço de oitava no intervalo de 1 a 31

$W_{mt}(b)$ é o espectro em terço de oitava da “mixagem temporária”

$W_{ref}(b)$ é o espectro em terço de oitava do sinal de referência

- f) Definimos um novo ajuste de *fader* (de segunda ordem) o qual será aplicado no sinal obtido no passo b. Este ajuste $H(n)$ é definido pela razão entre os valores dos espectros na banda b_{max} :

$$H(n) = 1 - \frac{\Delta W_{max}}{W_{mt}(b_{max})} \quad \text{para } n = 1, \dots, N \quad (20)$$

Onde:

b_{max} é o índice da banda onde ΔW_{max} é realizado

$H(n)$ é o novo ajuste de *fader* para o instrumento n

ΔW_{max} é a máxima diferença de amplitude entre os espectros

$W_{mt}(b_{max})$ é o valor da banda b_{max} da mixagem temporária

- g) Aplicação da taxa de volume final $H(n)$ em cada um dos N instrumentos musicais como descrita na Eq. (17), trocando $F(n)$ pelo seu valor atualizado $H(n)$.
- h) Atualização do vetor “Mix Intermediária” com os novos sinais como indicada na Eq. (6).

Em resumo, ao final desse conjunto de pequenas etapas temos um vetor $H(n)$ de dimensão $1 \times N$ o qual armazena o valor da taxa de volume final aplicado em cada instrumento musical. Além disso atualizamos o nosso vetor temporal “Mix Intermediária” de acordo com o equilíbrio

tonal proporcionado por $H(n)$, como descrito no item f). Os gráficos com os resultados dessas operações estão sendo mostrados em conjunto na sessão de conclusão deste trabalho.

4.5 CORREÇÃO ESPECTRAL

Depois de obtido o equilíbrio tonal dos instrumentos musicais através do ajuste dos *faders* de volume, passamos para a etapa da busca pelo equilíbrio espectral de cada instrumento dentro da mixagem. É importante lembrar que o resultado sonoro que buscamos otimizar não se encontra nos arquivos dos instrumentos musicais individuais, mas sim na somatória de suas sonoridades individuais na mixagem. Sendo assim, é possível que haja instrumentos que, quando escutados individualmente, perceba-se um desequilíbrio espectral, porém ao inseri-lo no conjunto de todos os instrumentos da mixagem, o efeito da soma espectral de todos os instrumentos tocando simultaneamente pode mudar a interpretação psicoacústica do som geral para o ouvinte, gerando uma sensação de equilíbrio espectral e consequentemente uma sonoridade agradável.

Exemplo: Podemos pensar num cenário de mixagem de uma música de *rock*, na qual é fundamental a sensação de peso nas baixas frequências (grave). O erro ocorre quando, na busca desta sonoridade específica, ressaltarmos o grave de todos os instrumentos gravados, o que gera um desequilíbrio sonoro da mixagem. A prática de estúdio nos mostra que o aumento do grave de apenas 1 ou 2 instrumentos colabora para a reinterpretação psicoacústica da sonoridade geral, nos dando a sensação de que todos os instrumentos tiveram suas baixas frequências ressaltadas e que atingimos a nossa sonoridade alvo, mesmo sabendo que o som de cada instrumento separadamente não tenha todas as componentes espectrais esperadas.

Dito isso precisamos saber como cada instrumento musical gravado está influenciando a sonoridade da mixagem específica e depois modificar suas qualidades espectrais de modo a atingir um espectro da mixagem próximo ao espectro de referência.

4.5.1 Contribuição Espectral de cada Instrumento

O cálculo da contribuição espectral nos informa o quanto cada banda de um determinado instrumento musical influencia na banda do espectro da mixagem total. Esse resultado é armazenado numa matriz $31 \times N$ e está normalizado entre 0 e 1.

$$C(n, b) = \frac{W(n, b)}{W_{mix}(b)}, \quad \text{para } n = 1, \dots, N \quad (21)$$

Onde:

$C(n, b)$ é a contribuição de amplitude por banda normalizado entre 0 e 1

n é o índice do instrumento

b é o índice da banda de terço de oitava no intervalo de 1 até 31

$W(n, b)$ é o vetor de amplitudes espectrais do instrumento n

$W_{mix}(b)$ é o vetor de amplitudes espectrais da “Mix Intermediária”

4.5.2 Comparação entre a Mix Intermediária e a Referência

Paralelamente, verificamos a razão as amplitudes dos espectros da “Mix Intermediária” e da referência e armazenamos esse resultado em um vetor 31×1 . De acordo com o equilíbrio de volumes realizado anteriormente, temos que o valor de amplitude por banda da curva espectral do sinal de referência nunca será inferior aos valores de amplitude da curva espectral da mixagem. Por esse motivo calculamos a distância entre cada banda usando a equação abaixo.

$$Q(b) = \frac{W_{ref}(b)}{W_{mix}(b)} \quad (22)$$

Onde:

Q é a razão entre as amplitudes dos espectros de referência e da “Mix Intermediária”

b é o índice da banda de terço de oitava no intervalo de 1 até 31

$W_{ref}(b)$ é o vetor de amplitudes do espectro de referência

$W_{mix}(b)$ é o vetor de amplitudes do espectro da “Mix Intermediária”

4.5.3 Definição das Magnitudes de Correção

Tendo calculado os valores de $C(n, b)$ e $Q(b)$ definimos a correção espectral do enésimo instrumento musical como:

$$Eq(n, b) = C(n, b) * Q(b) \quad (23)$$

Onde:

Eq é a matriz de correção espectral (equalização)

Q é a razão entre as amplitudes dos espectros de referência da “Mix Intermediária”

$C(n, b)$ é a contribuição de amplitude por banda normalizado entre 0 e 1

b é o índice da banda de terço de oitava no intervalo de 1 até 31

n é o índice do instrumento musical

Em outras palavras, a matriz Eq nos fornece os valores de equalização por banda de terço de oitava para cada instrumento musical. A dimensão de Eq é $31 \times N$.

4.5.4 Remoção do ganho comum a todas as bandas

Em alguns casos pode ocorrer que a matriz de correção espectral Eq tenha um valor de mínimo comum a todas as bandas de um determinado instrumento. Podemos ver um exemplo disso na Tabela 5 onde temos um exemplo da matriz Eq . Nesta tabela, apresentamos os valores em dB apenas para facilitar a compreensão do estudo.

Tabela 5 - Exemplo de valor de correção comum a um instrumento

Instrumento	20 Hz	25 Hz	31.5 Hz	...	20 KHz
Violão	+ 2 dB	+ 8 dB	+ 7 dB	...	+ 10 dB
Vocal	+10 dB	+ 5 dB	+ 3 dB	...	+ 5 dB

Considerando que o mínimo valor Eq é de 2 dB para o violão e 3 dB para o vocal, reescrevemos a matriz Eq subtraindo 2 dB de todas as bandas do violão e 3 dB de todas as bandas do vocal. Este leve ajuste colabora para a mínima distorção possível do sinal de áudio mantendo a proporção dos ganhos por banda. O resultado dessa subtração é apresentado na Tabela 6 abaixo.

Tabela 6 - Valores de ganho atualizados

Instrumento	20 Hz	25 Hz	31.5 Hz	...	20 KHz
Violão	+ 0 dB	+ 6 dB	+ 5 dB	...	+ 8 dB
Vocal	+ 7 dB	+ 2 dB	+ 0 dB	...	+ 2 dB

A equação (24) nos mostra esse processo de remoção do ganho comum a todas as bandas para cada um dos N instrumentos.

$$Eq(n, b) = Eq(n, b) - \min_{1 \leq b \leq 31} (Eq(n, b)) \quad \text{para } n = 1, \dots, N \quad (24)$$

4.5.5 Limite de Distorção Espectral

A princípio poderíamos aplicar uma correção espectral a cada instrumento musical através da matriz $Eq(n, b)$ calculada previamente, porém algumas de suas entradas podem ter valores muito altos, o que causaria uma grande distorção espectral quando aplicada ao conjunto de instrumentos musicais.

Tais valores em são devidos à tentativa do cálculo de correção espectral em bandas espectrais que o devido instrumento não atua predominantemente. Por exemplo, de acordo com Eargle (1994), um contrabaixo elétrico possui predominância espectral nas baixas frequências (grave, em torno de 40Hz a 400 Hz), não fazendo sentido os valores de correção espectral nas bandas de alta frequência entre 5 KHz e 20KHz. Ao contrário, um prato de bateria possui predominância nas altas frequências (agudo, entre 5KHz a 20KHz), não fazendo sentido os valores de correção espectral nas bandas de baixa frequência entre 40 Hz a 400 Hz.

Para evitar grandes distorções espectrais impomos uma limitação nos valores contidos na Matriz de Correção Espectral Eq , possibilitando a correção espectral de uma maneira mais suave a musical. Definimos o Limite de Equalização por banda em ± 6 dB (dobro/metade da amplitude) de acordo com a prática atual dos engenheiros de som e através da experimentação sonora do algoritmo. Caso o valor de equalização do instrumento exceda este limite, é interpretado que o instrumento musical em questão não tem predominância sonora naquela banda de terço de oitava e por isso o seu valor de correção nesta banda deve ser zero.

Podemos resumir isto, matematicamente, como a definição de uma nova Matriz de Correção Espectral, digamos, renormalizada, Eq_{lim} cujos valores são dados por:

$$Eq^{lim}(n, b) = \begin{cases} Eq(n, b), & \text{se } |Eq(n, b)| \leq l_{eq} \\ 0, & \text{se } |Eq(n, b)| > l_{eq} \end{cases} \quad (25)$$

onde l_{eq} é o limite de equalização. No nosso caso $l_{eq} = 2$, ou seja, +6 dB.

4.5.6 Aplicação da Correção Espectral

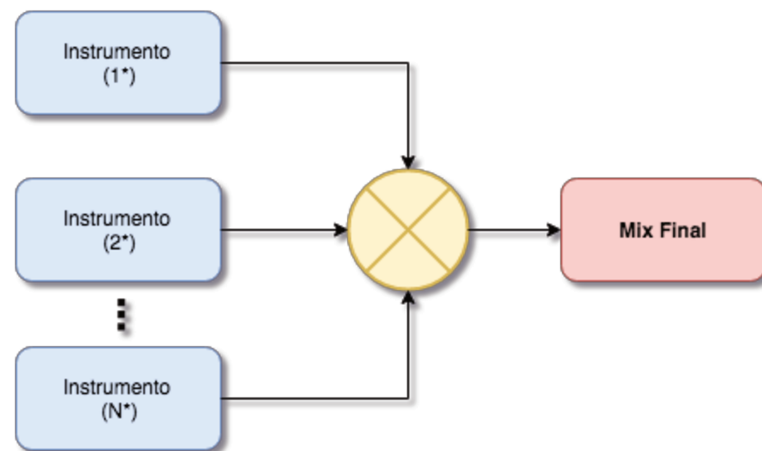
Com a matriz Eq devidamente atualizada, podemos aplicar a equalização por instrumento musical em cada terço de oitava. Para isso abrimos um equalizador paramétrico no MATLAB de 31 bandas e processamos cada sinal temporal de cada instrumento musical usando

os seus parâmetros de correção calculados previamente em *Eq* como descrito na Eq. (25). O resultado desse processo é um novo conjunto de sinais temporais corrigidos espectralmente de modo que o espectro da soma de todos os sinais (mixagem) é próximo ao espectro do sinal de referência, no nosso caso, *pink noise*.

4.6 MIXAGEM FINAL

Com os novos sinais corrigidos, podemos somá-los temporalmente para obter um único sinal temporal o qual chamamos de “Mix Final”. Esse processo de soma (mixagem) está definido pela Eq. (6). A principal característica da “Mix Final” é que o seu espectro é próximo ao espectro de referência, devido às correções de dinâmica, equilíbrio tonal e equilíbrio espectral, atingindo o nosso objetivo inicial. A figura abaixo mostra a soma dos sinais temporais de cada instrumento musical para formar a “Mix Final”.

Figura 34. Mixagem Final: Soma dos instrumentos modificados



4.7 MASTERIZAÇÃO

De acordo com Katz (2007), a masterização é o processo final de refinamento no controle da dinâmica e do conteúdo espectral da mixagem, ou seja, do resultado da soma dos sinais individuais em um canal (mono) ou 2 canais (estéreo). Na masterização já não temos acesso aos sinais dos instrumentos musicais individuais e por isso o nosso controle atua apenas no sinal da “Mix Final”.

Mesmo tendo feito a maior parte do controle dinâmico e espectral durante o processo de mixagem, o engenheiro de som pode otimizar ainda mais esses parâmetros sonoros no

processo de masterização, onde ele analisa a variação de dinâmica e conteúdo espectral da soma resultante, o que possibilita um ajuste de segunda ordem na otimização sonora da mixagem.

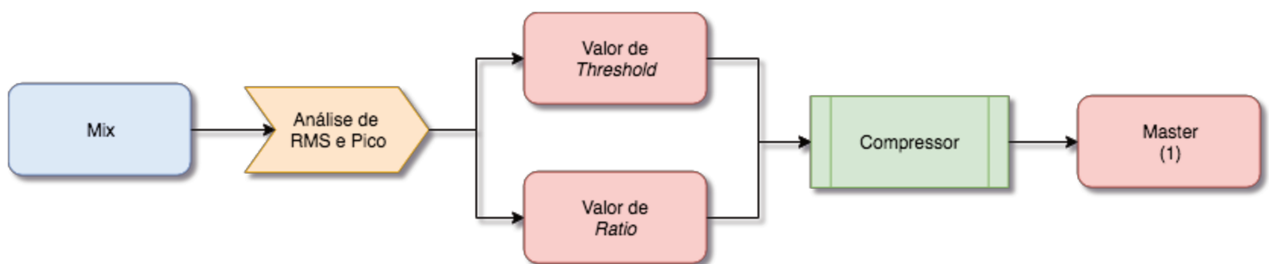
Como já dito anteriormente, o controle de dinâmica é feito pelo compressor de áudio e o controle espectral é feito por um equalizador de áudio. Na prática atual, a masterização de áudio musical serve para:

- a) Aumentar a sensação de volume (*loudness*)
- b) Remover frequências indesejáveis (sobra de grave e ressonâncias)
- c) Tentar homogeneizar o conteúdo espectral de todas as músicas do mesmo álbum

Como no nosso caso de estudo já removemos as frequências indesejáveis em processos anteriores e queremos masterizar uma única música por vez (e não uma coleção de músicas que precisam soar como um álbum), nos resta apenas definir e aplicar os processos de aumento de *loudness* para a música gravada. Na nossa implementação do processo de masterização aplicamos o mesmo método de análise e controle de dinâmica dos instrumentos individuais no sinal resultante da mixagem para remover as grandes diferenças de amplitudes do sinal. Depois disso fizemos um limite de amplitude máximo que delimita o nosso intervalo dinâmico e finaliza o nosso processo de masterização.

A figura abaixo mostra o processo de compressão tendo como entrada a “Mix Final”. Pode se ter um aprofundamento de cada processo desse fluxograma observado as equações (12), (13) e (16).

Figura 35. Análise e Controle da Dinâmica da Mixagem



4.7.1 Limite de Dinâmica da Mixagem

A última etapa do processo de masterização consiste em dar um ganho positivo no sinal temporal pré-masterizado, aumentado a amplitude de todos os seus *samples*, porém garantindo que nenhum deles ultrapasse o intervalo numérico entre -1 e +1. Esse intervalo entre -1 e +1 garante que o nosso sinal não foi distorcido e que não houve perda de informação no processamento do sinal.

O equipamento que nos possibilita esse controle fino da dinâmica sonora é um tipo especial de compressor, o qual é chamado de *limiter*. A principal característica do *limiter* é que ele possui a opção de *ratio* igual a infinito, ou seja, os valores de amplitude de cada sample que passar do *threshold* serão iguais ao valor do próprio *threshold*. De acordo com GIANNOULIS et al. (2012), a equação que descreve o funcionamento de um compressor padrão pode ser observada abaixo.

$$A_y(n, s) = \begin{cases} A_x(n, s), & \text{para } A_x(n, s) \leq Th(n) \\ Th(n) + \frac{A_x(n, s) - Th(n)}{R(n)}, & \text{para } A_x(n, s) > Th(n) \end{cases} \quad (26)$$

Onde:

n é o índice do instrumento musical

s é o índice da amostra (*sample*)

A_y é o valor da amplitude do *sample* de saída após a compressão

A_x é o valor de amplitude do *sample* original

Th é o valor do *threshold*

R é o valor do *ratio*

Na Eq. (26) podemos observar que quando:

$$A_x(s) > Th \quad \text{e} \quad R \rightarrow \infty, \quad \text{então} \quad A_y(s) \rightarrow Th \quad (27)$$

No nosso método, para definir um valor de *threshold* usamos o cálculo do percentil $P(99)$ de modo a obter um valor bem próximo à amplitude máxima do sinal pré-masterizado como podemos ver na equação abaixo.

$$Th = P_{lim}(|A_{pm}|) \quad (28)$$

Onde:

Th é a amplitude máxima desejada (*threshold*)

P_{lim} é o percentil *lim*. No nosso caso $lim = 99$.

A_{pm} é o vetor de amplitudes pré-masterizado $A_{pm}(k)$ com $k = 1, \dots, S$

Depois disso instanciamos um compressor com função de *limiter* ($ratio \approx \infty$) e processamos o áudio pré-masterizado por ele. Para garantir que o nosso sinal final terá *samples* com grandes amplitudes, porém sem distorção, aplicamos um ganho positivo ao sinal de modo que o módulo da máxima amplitude do novo sinal seja +1 ou 0 dBFS. Esse processo de ganho é chamado de normalização e pode ser visualizado na figura abaixo.

Figura 36. Limite de Dinâmica da Mixagem



5 RESULTADOS

Para testar o algoritmo de mixagem automatizada desenvolvido neste trabalho utilizamos arquivos de áudio de gravações musicais disponíveis em repositórios de livre acesso na internet, os quais armazenam todos os arquivos das sessões de gravações (*multitracks*) de alguns grupos musicais. O engenheiro de gravação Mike Senior mantém um repositório de sessões intitulado “Cambridge Music Technology”⁵⁹, o qual se destaca de outros repositórios devido à quantidade e variedade de sessões de gravação lá armazenadas.

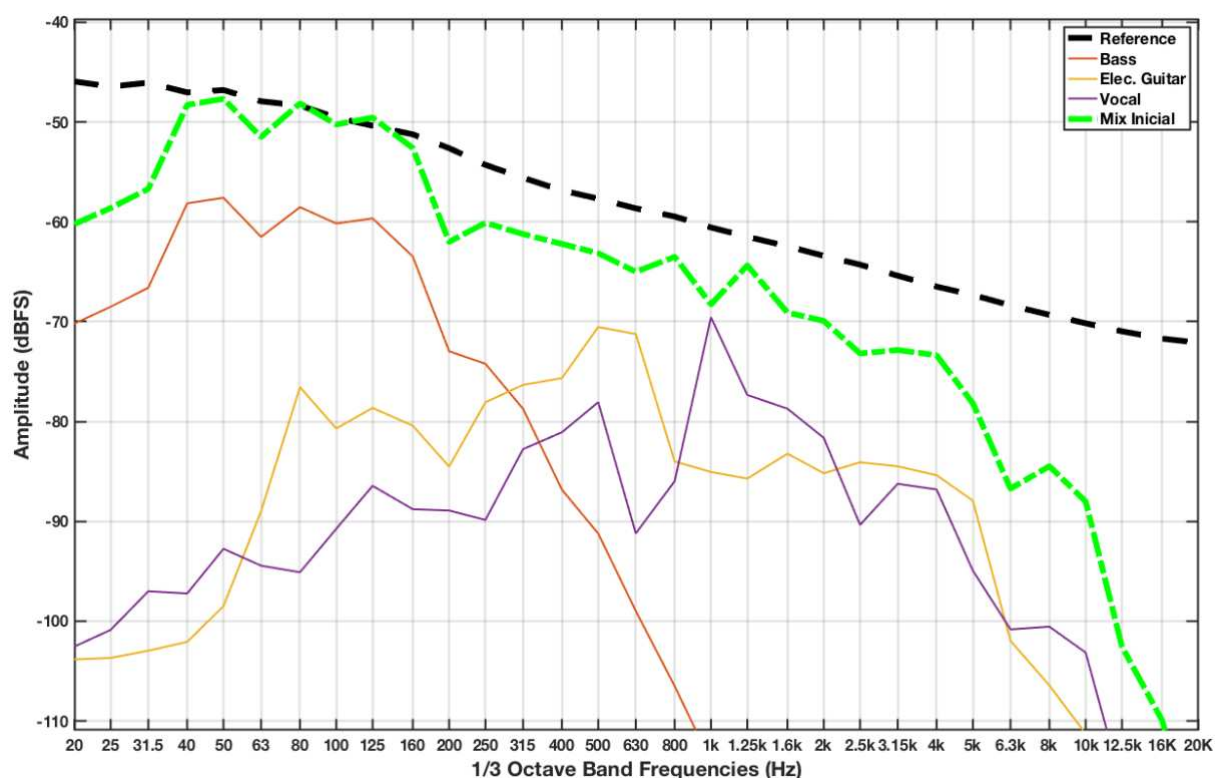
Dentre as músicas testadas, para a apresentação dos resultados deste trabalho, escolhemos a sessão da música comercial chamada “*Rescue Me*” (2015) do grupo “*Amy Helm & The Handsome Strangers*”, a qual se enquadra dentro dos estilos *country* e *blues* norte-americanos encontrada no “Cambridge Music Technology” e que representa bem os resultados encontrados em outras músicas testadas. A formação instrumental mixada é composta de contrabaixo elétrico, guitarra elétrica e vocal feminino.

Como já discutido nos capítulos 2 e 4, devido à fisiologia e à psicoacústica do ouvido humano, desenvolvemos nosso algoritmo com o objetivo de que a mixagem final tenha uma curva espectral próxima à curva espectral do sinal de referência, neste caso, o ruído rosa. A Figura 37 apresenta os espectros dos instrumentos musicais individuais (contrabaixo em vermelho, guitarra em amarelo e vocal em roxo) originais, o espectro do ruído rosa (em preto pontilhado) e o espectro da mixagem inicial (em verde pontilhado), ou seja, aquela em que não houve alteração alguma dos arquivos de áudio originais.

Nesta figura podemos ver que a região entre 40 Hz e 160 Hz está bem próxima a curva de referência com predominância sonora do contrabaixo nesta região, porém acima de 200 Hz o espectro da mixagem inicial vai se afastando do espectro de referência, o que proporciona a sensação psicoacústica de “falta de agudo” ou “excesso de grave”. Além disso não há uma relação de volume adequada entre os instrumentos musicais. Graficamente podemos observar essa relação desigual pela área abaixo da curva espectral de cada instrumento (vide Figura 37). Essa área nos dá a quantidade de energia percebida pelo nosso ouvido, impactando diretamente na nossa percepção de volume de um determinado instrumento.

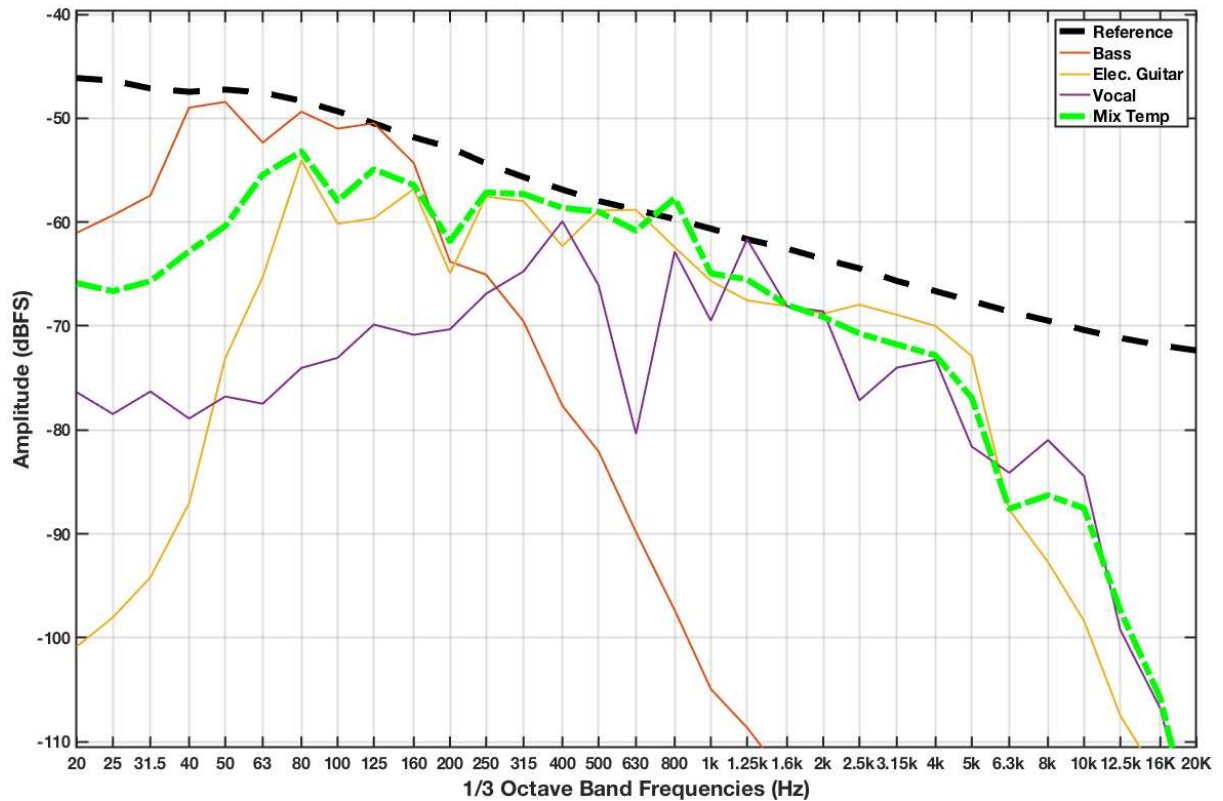
⁵⁹ Disponível em: <<https://www.cambridge-mt.com/ms-mtk.htm>>. Acesso em set. 2018.

Figura 37. Mix Inicial da música “Rescue Me”



Dado este cenário inicial apresentado na Figura 37, damos início ao nosso método de mixagem automatizada. A Figura 38 mostra os espectros dos sinais processados até a etapa da geração da mixagem temporária, na qual podemos observar uma curva espectral mais próxima da curva de referência, principalmente na região média entre 250 Hz e 4 KHz (região de maior sensibilidade do ouvido humano). Com a região média com mais energia e uma leve queda da região grave o ouvinte começa a ter uma sensação psicoacústica de equilíbrio espectral ou equilíbrio tonal.

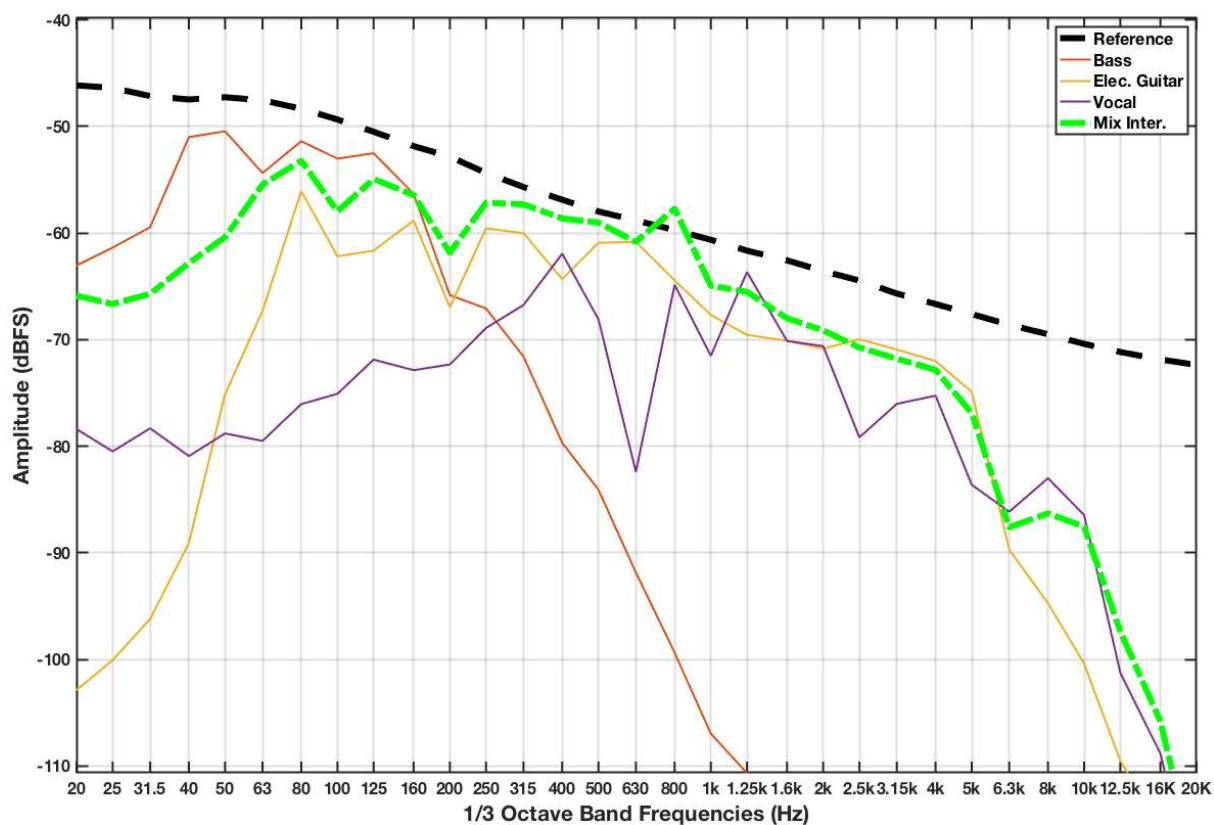
Figura 38. Mix Temporária da música “Rescue Me”



Um detalhe importante é a redução de amplitude não esperada do espectro da mixagem temporária na região grave (entre 40 Hz e 160 Hz). Como o contrabaixo está com amplitude maior do que a mixagem nesta região, acreditamos que pode ser um caso de cancelamento de fase entre este instrumento e os demais, já que o nosso método não verifica a fase dos sinais musicais importados. Em trabalhos futuros, esperamos implementar este tipo de verificação.

A Figura 39 mostra os sinais processados até a etapa da mixagem intermediária, na qual é feito um ajuste de segunda ordem, definido pela Eq. (20), nos volumes de cada instrumento visando uma melhor aproximação entre o espectro da mixagem e o espectro de referência.

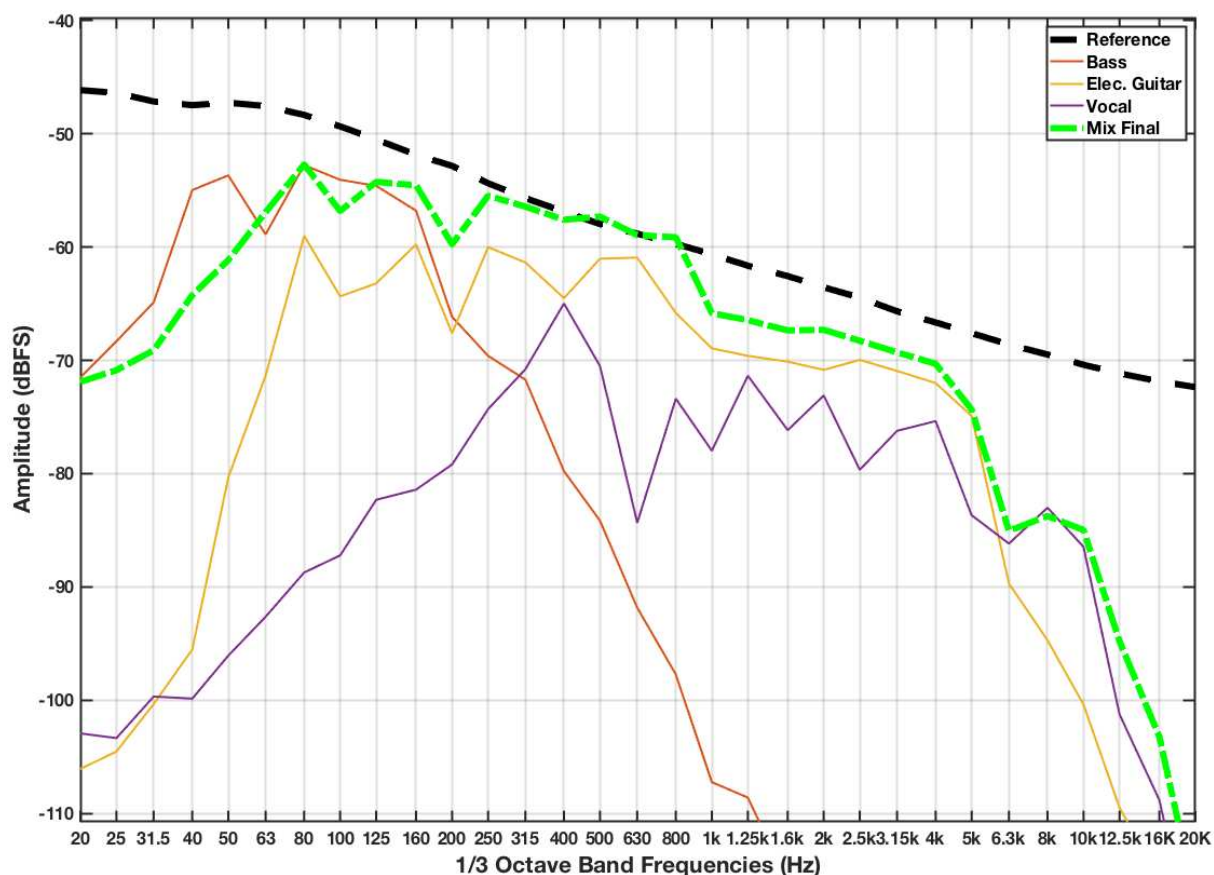
Figura 39. Mix Intermediária da música "Rescue Me"



Nos testes realizados, esse ajuste de segunda ordem é da ordem de $\pm 2\text{dB}$, sendo assim de difícil percepção auditiva, mas interessante do ponto de vista formal e matemático para este modelo.

Por fim, a Figura 40 mostra o final do processo da mixagem automatizada, onde aplicamos a correção espectral em cada um dos instrumentos musicais de modo a aproximar a nova mixagem à curva do ruído rosa.

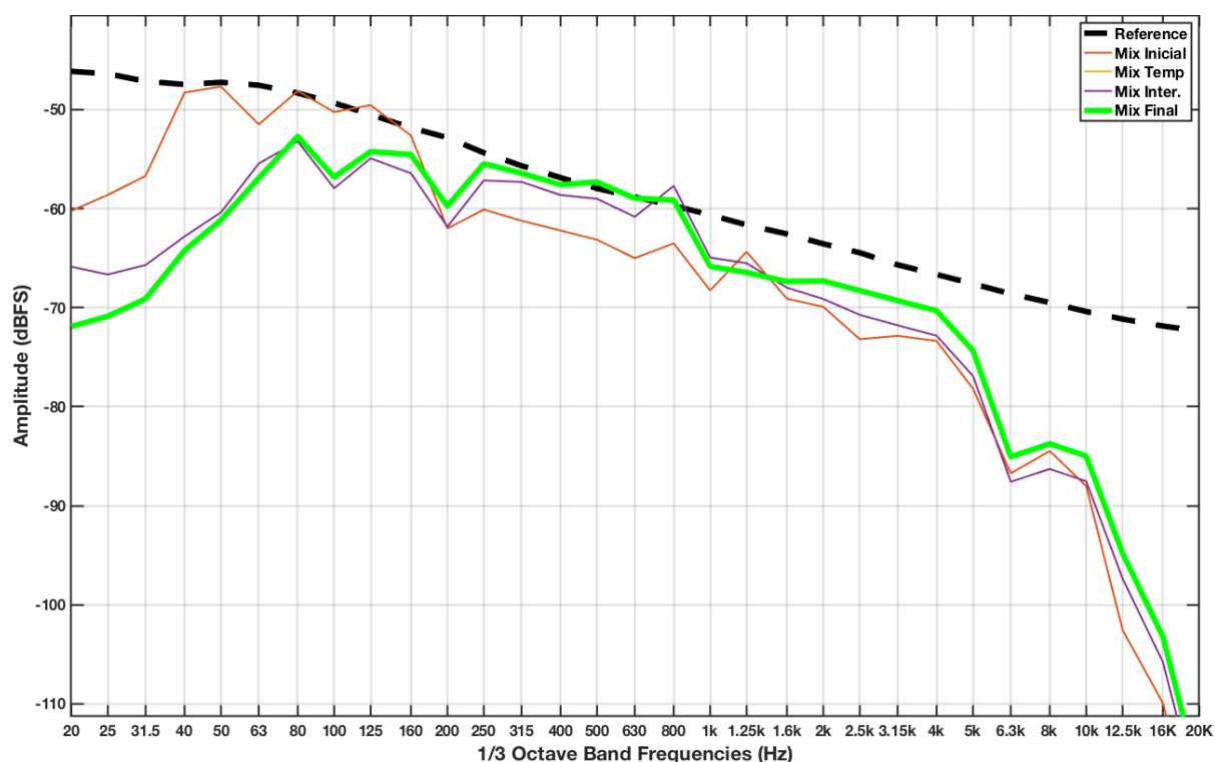
Figura 40. Mix Final da música "Rescue Me"



Nesta imagem podemos observar que a curva da mixagem está mais próxima da curva de referência, mantendo uma relação mais adequada entre grave, médios e agudos. Há uma redução, conforme nossa hipótese, do cancelamento de fase entre 50 Hz e 160 Hz. A esse resultado atribuímos aos filtros de passa altas aplicados em todos os instrumentos de acordo com a frequência de corte particular de cada um deles previamente calculada, diminuindo assim o cancelamento de fase nesta região. Houve também uma melhora na resposta região média entre 250 Hz e 800 Hz, onde foi removido o pico acima da referência em 800 Hz e aproximado toda a região para perto do espectro de referência.

Para uma melhor visualização das etapas deste processo, a Figura 41 compara todas as mixagens obtidas em uma única imagem.

Figura 41. Comparação entre todas as mixagens do método



Nesta imagem podemos observar que, mesmo que a Mixagem Inicial (em vermelho) esteja bem próxima a curva de referência nos graves, ela vai se distanciando do espectro do ruído rosa no sentido dos médios e agudos. É possível ver um equilíbrio melhor na Mixagem Temporária (em amarelo) e Intermediária (em roxo), sobrepostas neste gráfico, porém apresenta um pico em 800 Hz acima da curva de referência e pouca amplitude na região dos agudos (acima de 2 KHz). Por fim, o espectro da Mixagem Final apresenta uma melhor aproximação à curva de referência, sem regiões acima da referência e com amplitudes levemente maiores nas regiões agudas.

Auditivamente, podemos perceber que a Mixagem Final possui um maior equilíbrio tonal, proporcionando uma maior sensação de equilíbrio espectral. Também percebemos uma melhor relação de volume entre os instrumentos de modo que não haja grandes disparidades de volume entre eles. Justamente por este método proporcionar a mesma sensação de volume entre os instrumentos, sua limitação está em não ser possível criar camadas sonoras (*layers*) com volumes diferentes, como é o caso quando temos canais de *back-vocals*⁶⁰, os quais não devem soar de maneira igual ao vocal principal.

⁶⁰ Vocais secundários ou “Vocais de Fundo”

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho propõe um método de mixagem automática baseado numa aproximação do espectro da mixagem com o espectro do ruído rosa realizada computacionalmente por processos imitativos à prática atual de estúdio. Inicialmente, apresentamos uma breve história da gravação musical, sua evolução no tempo e como as tecnologias de áudio criadas influenciaram tanto aspectos técnicos quanto artístico-musicais. Discutimos também alguns fundamentos teóricos básicos para o estudo do áudio como a busca do equilíbrio estético na música desde os tempos gregos até a atualidade, o equilíbrio notacional na música erudita e a perda da referência sonora no âmbito da música comercial gravada. Apresentamos os detalhes da fisiologia auditiva e sensações psicoacústicas derivadas desta fisiologia das quais surge o cerne do nosso método e o motivo da escolha do ruído rosa como sinal de referência de mixagem musical. Também mostramos as dificuldades técnicas de se obter uma sala adequada para o trabalho de mixagem musical. No Capítulo 3 apresentamos os principais métodos de mixagem usuais e quais os equipamentos de áudio utilizados para este fim e no Capítulo 4 expomos o nosso método de mixagem automatizada como contribuição original deste trabalho.

Este método foi implementado em linguagem MATLAB, cujo código fonte disponibilizamos para *download* no repositório do Núcleo Interdisciplinar de Comunicação Sonora⁶¹ – NICS/UNICAMP, e seus resultados preliminares mostram bons indicativos no âmbito da mixagem automatizada. Dados os resultados obtidos neste trabalho, acreditamos que o estudo aprofundado da fisiologia auditiva e psicoacústica podem guiar a melhoria deste método em diversos campos a fim de se tornar uma ferramenta de uso geral dos músicos e operadores de áudio como técnicos e engenheiros de som.

Como continuação desta pesquisa pretendemos desenvolver métodos para evitar cancelamentos de fase na soma de sinais, adicionar espacialização sonora (panorama), incluir reverberação artificial através de convolução de resposta de impulso, adicionar a possibilidade de ajustar diferentes parâmetros de mixagem para diferentes estilos musicais, suporte a *surround* 5.1 e 7.1 e adição de efeitos de áudio como *delay*, *chorus* e *flanger*. Um outro possível trabalho seria usar computação evolutiva para a análise e mixagem dos arquivos de áudio, possibilitando melhores ajustes de todos os parâmetros através da mutação e comparação do áudio mixado com um áudio de referência.

⁶¹ Disponível em: < <https://www.nics.unicamp.br/> >. Acesso em set. De 2018.

REFERÊNCIAS

ADAMS, S. B.; BUTLER, O. R. **Manufacturing the Future: A History of Western Electric**. First Edition edition ed. Philadelphia: Cambridge University Press, 1999.

ANDREWS, T. Computation Time Comparison Between Matlab and C++ Using Launch Windows. , p. 6, 2012.

ARAÚJO, D. V. G. DE. Uma breve história da mixagem: Origem, Técnicas, Percepção e Futuros Avanços. , 2015. Acesso em: 5/9/2018.

AUDIO ENGINEERING SOCIETY. **AES17-1998 - Measurement of digital audio equipment**. Audio Engineering Society, 2009.

BAZIL, E. **Sound Mixing Tips and Tricks**. 1 edition ed. Merton, Thetford, Norfolk, UK: PC Publishing, 2008.

BAZIL, E. Sound on Sound. **Mixing To A Pink Noise Reference**, dez. 2014. Disponível em: <<https://www.soundonsound.com/techniques/mixing-pink-noise-reference>>. Acesso em: 24/5/2018.

BENESTY, J.; SONDHI, M. M.; HUANG, Y. **Springer Handbook of Speech Processing**. Springer Science & Business Media, 2007.

BENTO, R. F.; MARTINS, G. DE S. Q.; PINNA, M. H. **Tratado de Otologia**. 2º ed. Atheneu, 2013.

BHATARA, A.; TIROVOLAS, A. K.; DUAN, L. M.; LEVY, B.; LEVITIN, D. J. Perception of emotional expression in musical performance. **Journal of Experimental Psychology: Human Perception and Performance**, v. 37, n. 3, p. 921–934, 2011.

BISTAFA, S. R. **Acústica Aplicada ao Controle do Ruído**. Edição: 2 ed. São Paulo: Blucher, 2011.

BOFF, K. R.; KAUFMAN, L.; THOMAS, J. P. **Handbook of Perception and Human Performance, Vol. 1: Sensory Processes and Perception**. Volume 1 edition ed. New York: Wiley-Interscience, 1986.

BOHN, D. Bandwidth in Octaves Versus Q in Bandpass Filters. , p. 3, 2008.

CABRAL, Á. **Dicionário técnico de psicologia**. Editora Cultrix, 1996.

CARTER, R. **The Human Brain Book: An Illustrated Guide to its Structure, Function, and Disorders**. Expanded, Illustrated, Updated edition ed. New York, New York: DK, 2014.

EARGLE, J. M. **Frequency Ranges of Musical Instruments and the Human Voice**. Springer, Boston, MA, 1994.

EUROPEAN BROADCASTING UNION. R68-2000 Alignment level in digital audio production equipment. , 2000.

EUROPEAN BROADCASTING UNION. R128 Loudness Parameters for Short-Form Content. , jan. 2016. Disponível em: <<https://tech.ebu.ch/docs/r/r128s1.pdf>>. Acesso em: 15/10/2018.

EVEREST, F. A. **Master Handbook of Acoustics**. 4th edition ed. New York: McGraw-Hill/TAB Electronics, 2001.

FASTL, H.; ZWICKER, E. **Psychoacoustics: Facts and Models**. 3º ed. Berlin Heidelberg: Springer-Verlag, 2007.

FECHNER, G. T. **Elementos de Psicofísica**. Leipzig: DE Breitkopf & HÄTEL, 1860.

FERNANDES, J. C. Acústica, Ruídos e Perda de Audição. , , n. 9th Brazilian Conference on Dynamics, Control and their Applications, p. 55, 2010.

FLETCHER, H.; MUNSON, W. A. Loudness, Its Definition, Measurement and Calculation. **Bell System Technical Journal**, v. 12, n. 4, p. 377–430, 1933.

FRANCELIN, M. A. S.; MOTTI, T. F. G.; MORITA, I. As implicações sociais da deficiência auditiva adquirida em adultos. **Saúde e Sociedade**, v. 19, n. 1, p. 180–192, 2010.

GAZZANIGA, M. S.; IVRY, R. B.; MANGUN, G. R. **Cognitive Neuroscience: The Biology of the Mind, 4th Edition**. 4th edition ed. New York, N.Y: W. W. Norton & Company, 2013.

GESCHIEDER, G. A. **Psychophysics: the fundamentals**. 3º ed. United States: Mahwah, N.J. : L. Erlbaum Associates, 1997.

GIANNOULIS, D.; MASSBERG, M.; REISS, J. D. Digital Dynamic Range Compressor Design— A Tutorial and Analysis. **J. Audio Eng. Soc.**, v. 60, n. 6, p. 10, 2012.

GIBSON, D.; PETERSEN, G. **The art of mixing: a visual guide to recording, engineering, and production**. Emeryville, CA (6400 Hollis St., Emeryville, 94608): MixBooks, 1997.

GLINSKY, A. **Theremin: Ether Music and Espionage**. University of Illinois Press, 2000.

GROUT, D. J.; PALISCA, C. V. **História da Música Ocidental**. Edição: 6ª ed. Lisabon: Gradiva, 2014.

HAROLD RICHARDSON, J. The Spotify Paradox: How the Creation of a Compulsory License Scheme for Streaming On-Demand Music Services Can Save the Music Industry. **SSRN Electronic Journal**, 2014.

HOLMES, T. **Electronic and Experimental Music: Technology, Music, and Culture**. Routledge, 2012.

IMMINK, K. A. S. Shannon, Beethoven, and the Compact Disc. , p. 5, 2007.

IMMINK, K. S. The future of digital audio recording. **Journal of the Audio Engineering Society. Audio Engineering Society**, v. 47, p. 171–172, 1999.

IMMINK, K. S. Any Song, Anytime, Anywhere. **Journal of the Audio Engineering Society**, v. 58, 2010.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 226:2003 - Normal equal-loudness-level contours**. International Organization for Standardization, 2003.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 532:2017 - Methods for calculating loudness**. International Organization for Standardization, 2017.

IZHAKI, R. **Mixing Audio, Second Edition: Concepts, Practices and Tools**. 2 edition ed. Amsterdam ; Boston: Focal Press, 2011.

JENKINS, J. S. The Mozart effect. **Journal of the Royal Society of Medicine**, v. 94, n. 4, p. 170–172, 2001.

KATZ, B. **Mastering Audio: The Art and the Science**. Edição: 2 ed. New York: Focal Press, 2007.

LOUDEN, M. M. Dimension-Ratios of Rectangular Rooms with Good Distribution of Eigentones. , v. 24, p. 101–104, 1971.

LYON, R. F. **Human and Machine Hearing: Extracting Meaning from Sound**. 1 edition ed. Cambridge, United Kingdom ; New York, NY: Cambridge University Press, 2017.

MACEDO, F. A. B. O processo de produção musical na indústria fonográfica: questões técnicas e musicais envolvidas no processo de produção musical em estúdio. , v. XI, p. 7, 2007.

MAIA, P. A. **Estimativa de Exposições não Contínuas a Ruídos**. São Paulo: FUNDACENTRO, 2002.

MAKIBARA, R. R.; FUKUNAGA, J. Y.; GIL, D. Eustachian tube function in adults with intact tympanic membrane. **Brazilian Journal of Otorhinolaryngology**, v. 76, n. 3, p. 340–346, 2010.

MANSBRIDGE, S.; FINN, S.; REISS, J. D. Implementation and evaluation of autonomous multitrack fader control. in 132nd Convention of the Audio Engineering Society. **Anais...** , 2012.

MAULER, J. P. **[Anatomia] Orelha**. 2015.

MIDDLEBROOKS, J. C.; GREEN, D. M. Sound Localization by Human Listeners. **Annual Review of Psychology**, v. 42, n. 1, p. 135–159, 1991.

MIRANDA-VILELA, A. L. Anatomia e Fisiologia Humanas. **AFH**, 2018. Disponível em: <<https://afh.bio.br/>>. .

MORRIS, J. **Manhattan '45**. 2014.

MOSS, J. Top-Down Mixing – Members Area. **Behind the Speakers**, 2018a. Disponível em: <<https://members.behindthespeakers.com/module-3/top-down-mixing/>>. Acesso em: 21/8/2018.

MOSS, J. How To Easily Mix Guitars Like A Pro (3 Powerful Tips). **Behind The Speakers**, 29. abr. 2018b. Disponível em: <<https://behindthespeakers.com/mixing-guitars/>>. Acesso em: 21/8/2018.

MOYLAN, W. **Understanding and Crafting the Mix: The Art of Recording**. 2 edition ed. Amsterdam ; Boston: Focal Press, 2006.

NETTER, F. H. **Atlas de Anatomia Humana**. 6º ed. Elsevier, 2015.

NINO, R.; JAHIR, B.; HERRERA MARTINEZ, M. Design of an Algorithm for VST Audio Mixing Based on Gibson Diagrams. , 2017. Audio Engineering Society. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=18719>>. Acesso em: 8/12/2017.

NISHIDA, S. M.; WEBER, S. A. T.; OLIVEIRA, F. A.; TROLL, J. Como ouvimos o mundo? O aparelho da audição. , set. 2011. Universidade Estadual Paulista. Disponível em: <http://www2.ibb.unesp.br/Museu_Escola/2_qualidade_vida_humana/Museu2_qualidade_corpo_sensorial_audicao1.htm>. Acesso em: 11/10/2018.

OLIVEIRA, J. A. A. Tratado de Otorrinolaringologia. . p.510–530, 1994. São Paulo: Roca.

OLIVEIRA, J. A. A. Fisiologia Clínica da Audição – Cóclea Ativa. **Perda Auditiva Induzida pelo Ruído**. p.101–121, 1997. Porto Alegre: Bagagem Comunicação.

OLSON, H. F. **Music, Physics and Engineering**. Dover Publications, 1967.

OPPENHEIM, A. V.; SCHAFER, R. W.; BUCK, J. R. **Discrete-time Signal Processing (2Nd Ed.)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1999.

OSMAN, R. Designing small music practice rooms for sound quality. , p. 8, 2010.

PALUDO, T. R. Reconfigurações Musicais: Os novos caminhos da música na era da comunicação digital. , mar. 2010. Disponível em: <<repositorio.pucrs.br/dspace/bitstream/10923/.../000422751-Texto%2BCompleto-0.pdf>>. Acesso em: 27/5/2018.

PERETZ, I. CARTA: Evolutionary Origins of Art and Aesthetics: The Emotional Power of Music. , 2009. Disponível em: <<https://www.linggars.com/video/79lxJxUMpzQ/carta-evolutionary-origins-of-art-and-aesthetics-the-emotional-power-of-music-isabelle-peretz.html>>. Acesso em: 8/12/2017.

PISTON, W. **Orchestration**. 1st edition ed. New York: W. W. Norton & Company, 1955.

REISS, J. A Meta-Analysis of High Resolution Audio Perceptual Evaluation. **Journal of the Audio Engineering Society**, v. 64, n. 6, p. 364–379, 2016.

- REISS, J. D.; PEREZ-GONZALEZ, E. Automatic gain and fader control for live mixing. , 2009. Disponível em: <https://www.researchgate.net/publication/224086775_Automatic_gain_and_fader_control_for_live_mixing>. Acesso em: 6/9/2018.
- RICHTER, B. F. **389 Choralgesänge für vierstimmigen gemischten Chor**. Breitkopf & Härtel, 1966.
- RIMSKY-KORSAKOV, N. **Principles of Orchestration**. S. l.: CreateSpace Independent Publishing Platform, 1873.
- ROSSING, T. D.; MOORE, F. R.; WHEELER, P. A. **The Science of Sound, 3rd Edition**. 3rd edition ed. San Francisco: Pearson, 2001.
- RUMSEY, F. Mixing and Artificial Intelligence. **Journal of the Audio Engineering Society**, v. 61, n. 10, p. 806–809, 2013.
- RUSSO, I. C. P. **Acústica e Psicoacústica Aplicadas À Fonoaudiologia**. São Paulo, 1999.
- SCOTT, J.; PROCKUP, M.; SCHMIDT, E. M.; KIM, Y. E. AUTOMATIC MULTI-TRACK MIXING USING LINEAR DYNAMICAL SYSTEMS. , p. 6, 2011.
- SEPMeyer, L. W. Computed Frequency and Angular Distribution of the Normal Modes of Vibration in Rectangular Rooms. **The Journal of the Acoustical Society of America**, v. 37, n. 3, p. 413–423, 1965.
- SILVA, J. A. DA. Qual A Verdadeira Função Psicofísica: de Potência ou Exponencial? **Arquivos Brasileiros de Psicologia**, v. 34, p. 75–79, 1982.
- SILVA, J. A. DA; MACEDO, L. DE. A função-potência na percepção: significado e procedimentos de cálculo do expoente. **Arquivos Brasileiros de Psicologia**, v. 34, n. 4, p. 27–45, 1982.
- SILVA, J. A. DA; ROZESTRATEN, R. J. A. Manual Prático de Psicofísica. , 2014. Disponível em: <<https://pge3.wikispaces.com/file/view/manual+de+psicofisica.pdf>>. Acesso em: 24/5/2018.
- SILVERSTEIN, D. How to mix faster with the pink noise mixing trick. **Audio Hertz**, 15. mar. 2018. Disponível em: <<https://audiohertz.com/2018/03/15/how-to-mix-faster-with-the-secret-pink-noise-mixing-trick/>>. Acesso em: 24/5/2018.
- SKOVENBORG, E.; NIELSEN, S. H. Evaluation of Different Loudness Models with Music and Speech Material. , p. 34, 2004.
- STEVENS, S. S. Cross-Modality Validation of Subjective Scales for Loudness, Vibration, and Electric Shock. **Journal of experimental psychology**, v. 57, p. 201–9, 1959.
- STEVENS, S. S. To Honor Fechner and Repeal His Law. **Science**, v. 133, n. 3446, p. 80–86, 1961.

STEVENS, S. S. **Psychophysics**. Transaction Publishers, 1975.

STEVENS, S. S.; VOLKMANN, J. The Relation of Pitch to Frequency: A Revised Scale. **The American Journal of Psychology**, v. 53, n. 3, p. 329–353, 1940.

STEVENS, S. S.; WARSHOFISKY, F. **Som e Audição**. Rio de Janeiro: José Olympo Editora, 1968.

SUZUKI, Y.; TAKESHIMA, H. Equal-loudness-level contours for pure tones. **The Journal of the Acoustical Society of America**, v. 116, n. 2, p. 918–933, 2004.

SWANSON, K. A Case Study on Spotify: Exploring Perceptions of the Music Streaming Service. **MEIEA Journal; Nashville**, v. 13, n. 1, p. 207–230, 2013.

TILSON THOMAS, M. Music and emotion through time. , fev. 2012. Disponível em: <https://www.ted.com/talks/michael_tilson_thomas_music_and_emotion_through_time>. Acesso em: 12/8/2017.

WEBER, E. H. **De Pulsu, Resorptione, Auditu et Tactu. (Annotationes Anatomicae et Physiologicae. Programmata Collecta. Fasciculi Tres)**. Lipsiae Leipzig: C.F. Koehler, 1834.

WOSZCZYK, W.; ENGEL, J.; USHER, J.; AARTS, R.; REEFMAN, D. Which of the Two Digital Audio Systems Matches the Quality of the Analog System? **Journal of the Audio Engineering Society**, , n. 57, 2009. Disponível em: <https://www.researchgate.net/publication/239575695_WHICH_OF_THE_TWO_DIGITAL_AUDIO_SYSTEMS_BEST_MATCHES_THE_QUALITY_OF_THE_ANALOG_SYSTEM>. Acesso em: 15/10/2018.

ZORZETTO, N. L. Anatomia da Orelha. . 2º ed, p.1216, 2006. Porto Alegre: Artmed.