

## Prova P2D - Codificação de sinais multimídia

### Integrantes:

Lucas Turano Machado RA 11201721752;

Bruno Sanches Rodrigues RA 11201721076;

Gabriel Ferreira Prodi RA 21017414;

Jonas Greco Onias RA 21021215;

Matheus Moreira Francisco RA 11201810390.

### Questões:

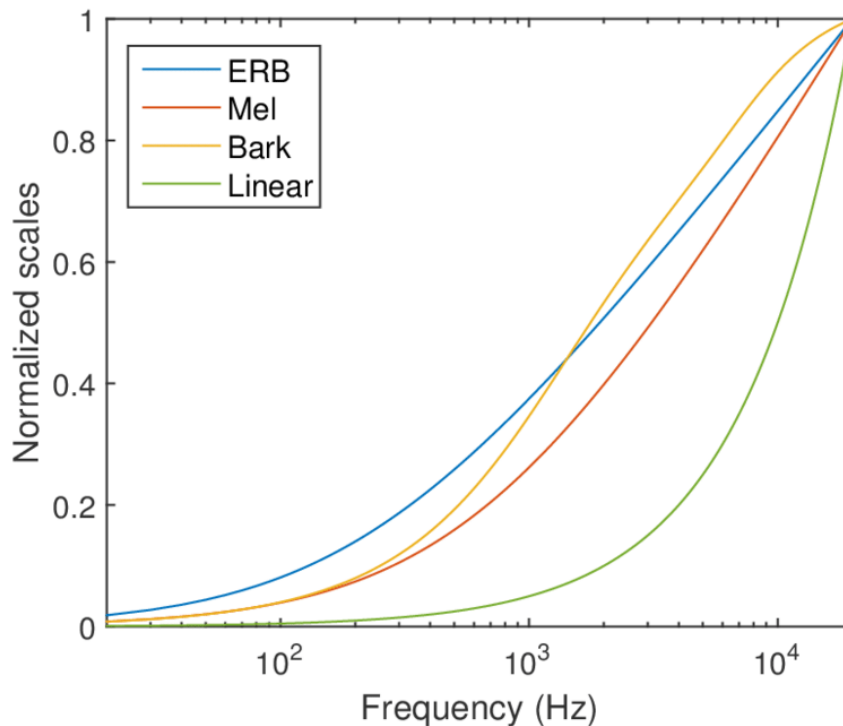
**1) (0,5) Defina as três escalas perceptivas e descreva-as. Compare-as fazendo um plot num gráfico sabendo que a escala Mel (m) é 100 vezes maior que a Bark (Z) ou a ERB (estas duas últimas de mesma ordem de grandeza):**

A escala mel é uma escala perceptiva de tons julgados pelos ouvintes como iguais em distância uns dos outros. O ponto de referência entre esta escala e a medição de frequência normal é definido atribuindo-se um tom perceptivo de 1000 mels a um tom de 1000 Hz, 40 dB acima do limite do ouvinte. Acima de cerca de 500 Hz, intervalos cada vez maiores são julgados pelos ouvintes para produzir incrementos de tom iguais. Como resultado, quatro oitavas na escala hertz acima de 500 Hz são consideradas como constituindo cerca de duas oitavas na escala mel. O nome mel vem da palavra melody para indicar que a escala é baseada em comparações de tons.

A escala de Bark é uma escala psicoacústica. Uma definição do termo é: uma escala de frequência em que distâncias iguais correspondem a distâncias perceptualmente iguais. Acima de cerca de 500 Hz, esta escala é mais ou menos igual a um eixo de frequência logarítmico. Abaixo de 500 Hz, a escala de Bark torna-se cada vez mais linear. A escala varia de 1 a 24 e corresponde às primeiras 24 faixas críticas de audição. É relacionado, mas um pouco menos popular do que a escala mel, uma escala perceptual de tons julgados pelos ouvintes como iguais em distância um do outro.

Conceito relacionado com o filtro auditivo é a largura de banda retangular equivalente (ERB). A ERB mostra a relação entre o filtro auditivo, frequência, e a largura de banda crítica. Um ERB passa a mesma quantidade de energia como o filtro auditivo que corresponde a e mostra como ela muda com a frequência de entrada. Em baixos níveis de ruído, o ERB é aproximado pela seguinte equação de acordo com Glasberg e Moore:  $ERB(f) = 24,7 * (4,37 f / 1000 + 1)$ , onde o ERB é em Hz e f representa a frequência central em Hz. banda crítica. Pensa-se que cada ERB é o equivalente a cerca de 0,9 milímetros na membrana basilar. A ERB pode ser convertida em uma escala que se refere à frequência e mostra a posição do filtro

auditivo ao longo da membrana basilar. Por exemplo, um número de ERB de 3,36 corresponde a uma frequência na extremidade apical da membrana basilar ao passo que um número de ERB 38.9 corresponde à base e um valor de 19,5 desce a meio caminho entre os dois. banda crítica.



**2) (0,5) Pesquise qual a diferença entre os formatos 0, 1 e 2 do MIDI e como se diferenciam os instrumentos, exemplificando.**

Arquivo MIDI, formalmente, denominado Standard MIDI Files, é utilizado para armazenar os dados MIDI. Este formato armazena as mensagens MIDI (status bytes com o apropriado data bytes) mais o tempo de cada mensagem (série de bits que representam a quantidade de pulsos que a mensagem deve esperar antes de "tocar"). Além disso, este formato permite armazenar informações sobre tempo, time, key signatures, entre outras informações necessárias a um equipamento que "toca" a música. O arquivo MIDI é composto de "chunks". Ele contém no cabeçalho um header chunk, que é seguido por um ou mais track chunks. O header chunk indica um de três tipos de arquivos:

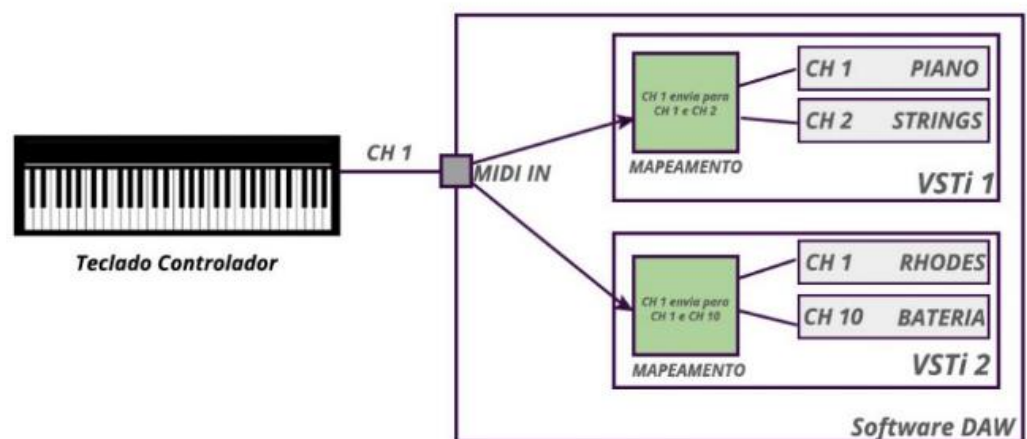
- Tipo 0: Single Track das informações MIDI com possibilidade de muitos canais de dados de nota MIDI;
- Tipo 1: Multitrack utilizado para sequenciadores de dados;
- Tipo 2: Dados para multisequencer ou programas baseados em padrões, para armazenar um número independentes de sequências ou padrões num único arquivo MIDI;

O arquivo de formato 0 tem um pedaço de cabeçalho seguido por um pedaço de trilha. É a representação de dados mais intercambiável. É muito útil para um simples reprodutor de faixa única em um programa que precisa fazer sintetizadores fazer sons, mas que se preocupa principalmente com outra coisa, como mixadores ou caixas de efeitos sonoros. É muito desejável ser capaz de produzir tal formato, mesmo que seu programa seja baseado em trilhas, para trabalhar com esses programas simples.

Um arquivo de Formato 1 ou 2 tem um bloco de cabeçalho seguido por um ou mais blocos de trilha. programas que suportam várias faixas simultâneas devem ser capazes de salvar e ler dados no formato 1, uma forma verticalmente unidimensional, ou seja, como uma coleção de faixas. Os programas que suportam vários padrões independentes devem ser capazes de salvar e ler dados no formato 2, uma forma horizontal unidimensional. Fornecer esses recursos mínimos garantirá a intercambiabilidade máxima.

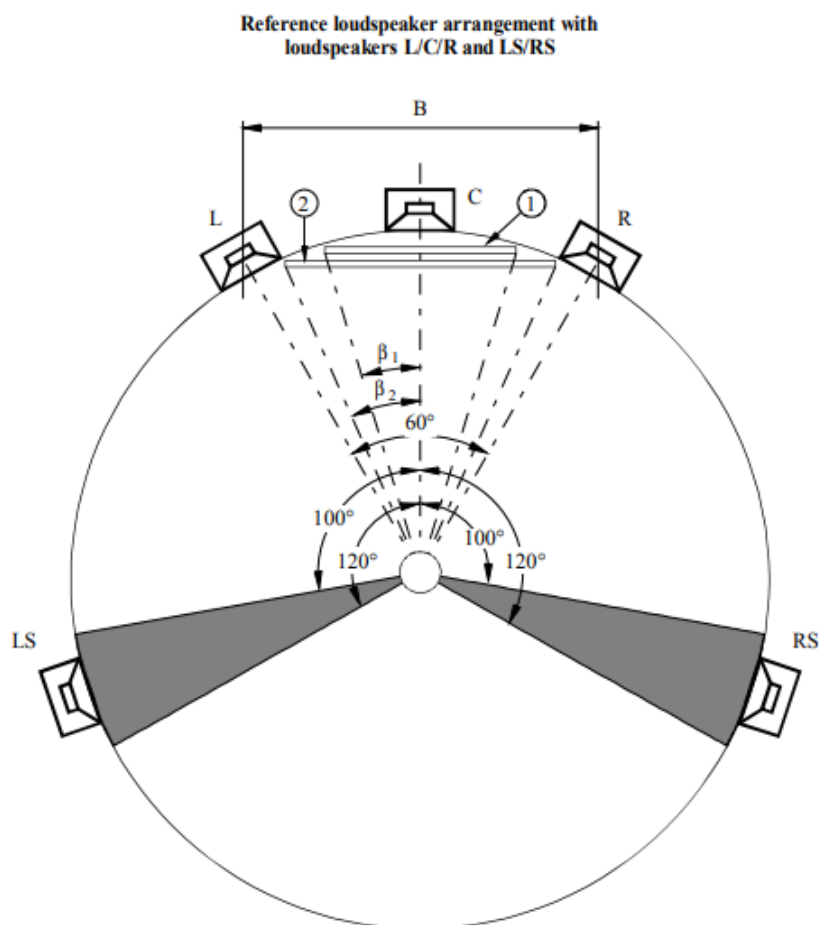
Os canais permitem enviar mensagens MIDI para diferentes instrumentos de forma simultânea. Você pode atribuir, por exemplo, o piano no canal 1, strings no canal 2 e bateria no canal 10. Se você tem um teclado controlador que está ligado a outro instrumento, você deve conhecer o canal de saída do teclado controlador, bem como, o canal de entrada no outro instrumento. Se você configurar o controlador para enviar no canal 1, você deve garantir que o outro instrumento esteja configurado para receber no canal 1. Os *softwares* sequenciadores, bem como, as *DAWs*, possuem pistas (*tracks*) MIDI. Normalmente, você configura o canal de entrada para cada pista.

As DAW's específicas para performances ao vivo costumam oferecer a opção de mapear os canais. Dessa forma, é possível personalizar a distribuição dos canais entre os instrumentos virtuais dentro do *software*. Se, por exemplo, o teclado controlador enviar mensagens apenas para o canal 1, você pode configurar a DAW para que um instrumento específico reproduza o canal 1 nos canais 1, 2 e 3.



**3) (0,5) Pesquise e ilustre quais as diferenças e aplicações das instalações 3+2 e 3/4 na recomendação ITU-R BS.775-1, exemplificando qual seria a sua escolha na sua própria sala de TV. Anexar uma foto e as dimensões da sala, e o layout da instalação pretendida**

A instalação 3/2 tem 3 caixas de som frontais, (esquerda, central e direita), e 2 caixas de som traseiras (esquerda e direita). Seguindo a configuração abaixo:



Screen 1 HDTV – Reference distance =  $3 H$  ( $2\beta_1 = 33^\circ$ )

Screen 2 =  $2 H$  ( $2\beta_2 = 48^\circ$ )

H: height of screen

B: loudspeaker base width

Loudspeaker	Horizontal angle from centre (degrees)	Height (m)	Inclination (degrees)
C	0	1.2	0
L, R	30	1.2	0
LS, RS	100 ... 120	$\geq 1.2$	0 ... 15 down

D01

Para um maior envolvimento acústico, pode-se ter 4 ou mais caixas de som traseiras, configurando a instalação 3/4. As caixas de som então deverão ser dispostas da seguinte forma:

The diagram illustrates the experimental setup for the study. It features a circular arrangement of six speakers. At the top center is speaker C. Below it are speakers L and R. At the bottom left are speakers LS<sub>1</sub> and LS<sub>2</sub>. At the bottom right are speakers RS<sub>1</sub> and RS<sub>2</sub>. The speakers are positioned around a central point, with angles of 60° between adjacent speakers and 150° between speakers on opposite sides. A horizontal dimension B is indicated at the top.

Minha sala possui uma porta de vidro na parte esquerda (encoberta pelas cortinas na foto), e uma passagem aberta na parte direita. A configuração 3/4 não seria viável, pois uma caixa de som teria que ficar no meio do caminho, e outra caixa de som teria que ficar em frente a porta.



A configuração 3/2 ficaria no limite da viabilidade em minha sala, com as caixas de som LS e RS dispostas da seguinte forma:



**4) (0,5) Faça um resumo da norma speech intelligibility index, ANSI S3.5-1997, descrevendo uma aplicação da mesma**

Trata-se de uma medida entre 0.0 e 1.0 de inteligibilidade de voz, que quantifica a porção de voz que é tanto audível quanto inteligível ao ouvinte. É definida como um método de calcular uma medida física que tem alta correlação com inteligibilidade de voz. Uma SII de 0 significa que nenhuma informação foi útil para o entendimento da fala. Esta medida é calculada ao longo de diferentes faixas de frequência.

Seu cálculo segue a fórmula geral abaixo:

$$SII = \sum_{i=1}^n I_i A_i$$

Onde:  $n$  é o número de bandas de frequências utilizado, que pode variar dependendo da precisão desejada;  $I_i$  é um fator de importância da banda de frequência para a inteligibilidade da voz; e  $A_i$  é a audibilidade da banda, que assume valores entre 0 e 1, e é calculado a partir da proporção entre nível de voz e nível de ruído na banda. [Fonte.](#)

Na área de audiolgia, a norma é utilizada na verificação do desempenho de aparelhos de amplificação sonora individual. [Fonte.](#)

**5) (0,5) Pesquise sobre a importância do tempo de reverberação na construção de grandes salas de concerto e cite alguns exemplos no mundo (três ou mais).**

O tempo de reverberação em ambientes fechados, em que há a propagação de som de um emissor até um receptor, é muito importante devido ao entendimento desse receptor.

Aqui falaremos de grandes salas de concerto, imaginemos um espetáculo acontecendo nesse ambiente, quando algum ator ou músico canta ou faz qualquer atividade que emite som, ele emite ondas. As ondas que vão diretamente ao espectador são chamadas de som direto, já as ondas que vão em direção a parede, ao teto e ao chão são refletidas até chegar no espectador, nesse caso são chamadas de som refletivo.

O tempo de reverberação entra justamente nesse som refletido, assim ele tem uma reverberação, que nada mais é que a persistência do som no ambiente, parametrizado pelo tempo de reverberação.

Assim se o tempo de reverberação for baixo, as ondas do som direto e do som refletidos, chegam ao mesmo tempo ao receptor, assim aumentando a potência sonora e causando uma melhor sensação. Agora quando o tempo de reverberação for alto, fica perceptível ao receptor a diferença do momento que chegam as ondas em seu ouvido, causando assim uma sensação de ouvir dois sons dessincronizados.

Portanto, o tempo de reverberação em um ambiente de salas de concerto é importante, pois, a sua principal função é entreter o público e causar uma boa sensação em relação ao som.

Um dos parâmetros para verificar esse gap é o critério de clareza, que é calculada pela fórmula:

$$C_{80} = 10 \log \frac{\text{energia\_sonora\_0} - 80ms}{\text{energia\_sonora\_após\_80ms}}$$

Existe um intervalo, caso o resultado de menor igual à -5dB, a acústica da sala é confusa, caso o resultado de maior que +6dB, a acústica é muito seca.

Citando alguns exemplos de grandes salas de concerto:

- **Grosser Musikvereinsaal (Viena)**  
Tempo de reverberação de 2s e clareza de -2,8dB
- **Concertgebouw (Amsterdam)**  
Tempo de reverberação de 2s e clareza de -3,3dB
- **London Royal Festival Hall (Londres)**  
Tempo de reverberação de 1,5s e clareza de 0,33dB
- **Berlin Philharmonic (Berlin)**  
Tempo de reverberação de 1,9s e clareza de -0,6dB.