



Processamento Audiovisual

Universidade de Coimbra

Licenciatura em Engenharia e Ciência de Dados



Professor



Luís Cruz

DEEC, Gab 3A.6 ou Lab S3.1 (LPM – Lab. Proc. Multimedia)

Email: lcruz@deec.uc.pt

Programa

Parte I - Processamento de sinais áudio

- Captura de sinais de áudio
- Representação de sinais de áudio, com e sem compressão
- Descritores (features) de sinais de áudio para indexagem e busca em repositórios de larga escala.
- Avaliação de qualidade de áudio e inteligibilidade de voz

Parte II - Processamento de imagem e vídeo

- Captura de imagem e vídeo
- Representações de imagem e vídeo com e sem compressão, seguindo formatos normalizados (JPEG, MPEG)
- Descritores (features) de imagem e vídeo para aplicação em indexagem e busca em repositórios de larga escala.
- Avaliação de qualidade de imagem e vídeo, métodos e normas

Parte III - Armazenamento e transmissão de áudio e vídeo em larga escala

Parte IV - Aplicações

Bibliografia

- Ian McLoughlin, "Speech and Audio Processing - A MATLAB-based Approach", Cambridge University Press 2016
- Thierry Dutoit, Ferran Marqués, "Applied Signal Processing A MATLAB-Based Proof of Concept", Springer 2009
- Yue-Ling Wong, "Digital Media Primer - Digital Audio, Video, Imaging and Multimedia Programming 2nd ed", Pearson 2013
- Salembier, Manjunath, Sikora, "Introduction to MPEG 7: Multimedia Content Description Language", Wiley, 2002
- Kim, Moreau, Sikora, "MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval", Wiley 2005
- Ian McLoughlin, "Applied Speech and Audio Processing with MATLAB Examples", Cambridge University Press 2009
- Jens-Rainer Ohm, "Multimedia Communication Technology: Representation, Transmission and Identification of Multimedia Signals", Springer 2004

Aulas Práticas

- Baseadas em MATLAB e aplicações específicas como compressores MP3, JPEG, MPEG, etc.
- Trabalhos práticos com duração de uma ou duas aulas
- Relatório preparado e entregue durante a aula

Avaliação

- Avaliação Contínua: $N_{AC}=11$ valores (mínimo: 3.5 valores)
 - Aulas Práticas Laboratoriais (assiduidade e relatórios)
 - Presença em pelo menos 10 das 14 aulas práticas (~71%)
- Teste Intercalar: 3 valores (TI_3) (dia 29 de Março)
- Exame: 6 valores (E_6)

Nota final: $N_{Final} = E_6 + TI_3 + N_{AC}$. (mínimo $TI_3 + E_6$: 3 valores)
- Exame de Recurso/Exame Especial: 9 valores (ER_9) (mínimo: 3 valores)

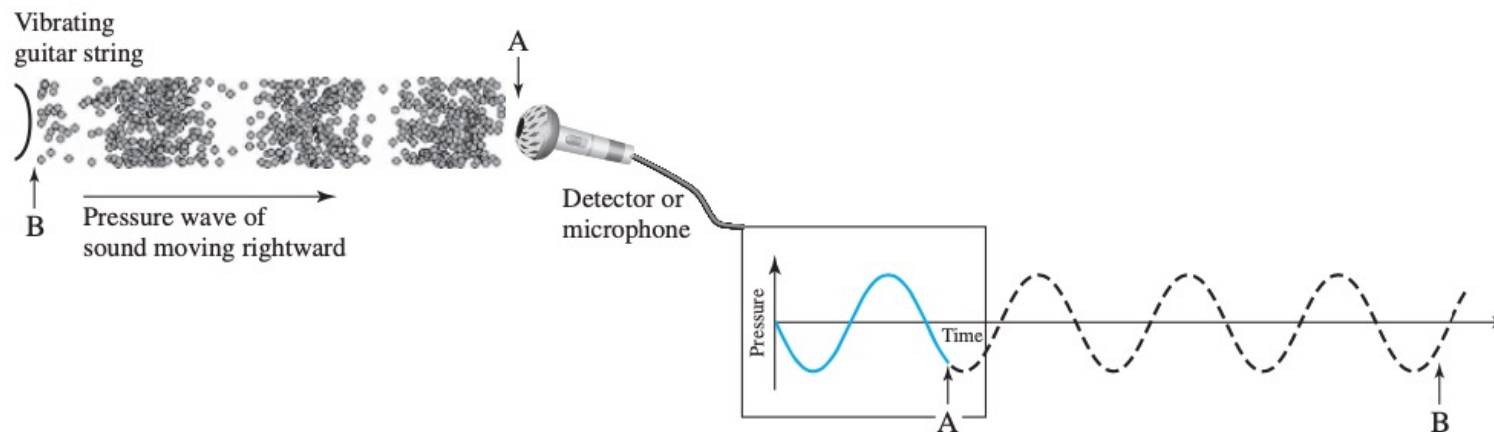
Nota final recurso: $N_{FinalRec} = ER_9 + N_{AC}$

Objectivos

- Perceber como são criados os sinais digitais de audio, imagem e vídeo (sinais audiovisuais)
- Perceber as características dos sinais audiovisuais nos domínios espectrais (frequência), espaciais e temporais
- Perceber como se reduz a quantidade de informação de sinais digitais audiovisuais minimizando a perda de qualidade (codificação)
- Perceber como se mede a qualidade subjectivamente relevante de sinais audiovisuais após codificação ou outro tipo de processamento
- Perceber como definir e extrair "features" para uso em tarefas de pesquisa de conteúdos audiovisuais
- Estudar as aplicações mais importantes dos sinais audiovisuais codificados ou não e suas features

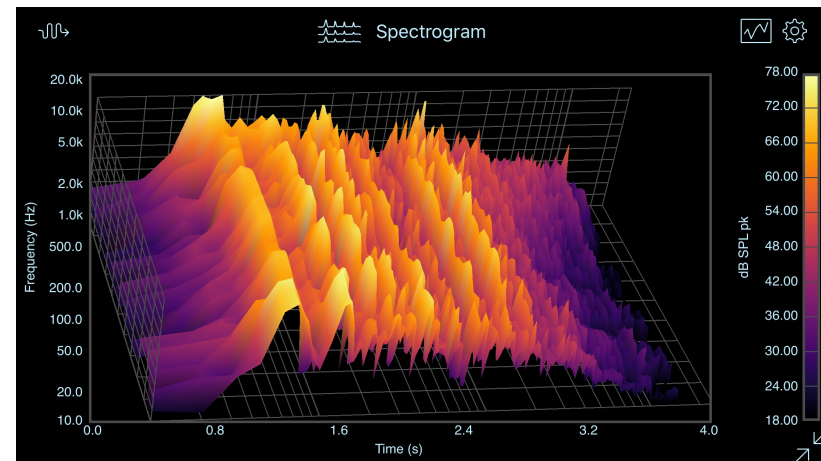
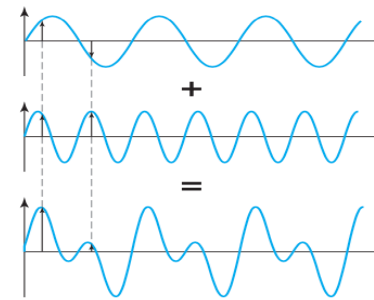
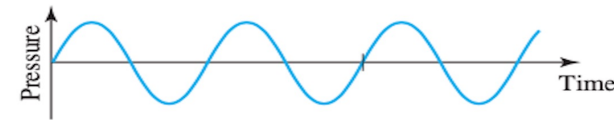
Som

- Vibrações mecânicas das moléculas do ar, que criam variações de pressão no tempo e no espaço



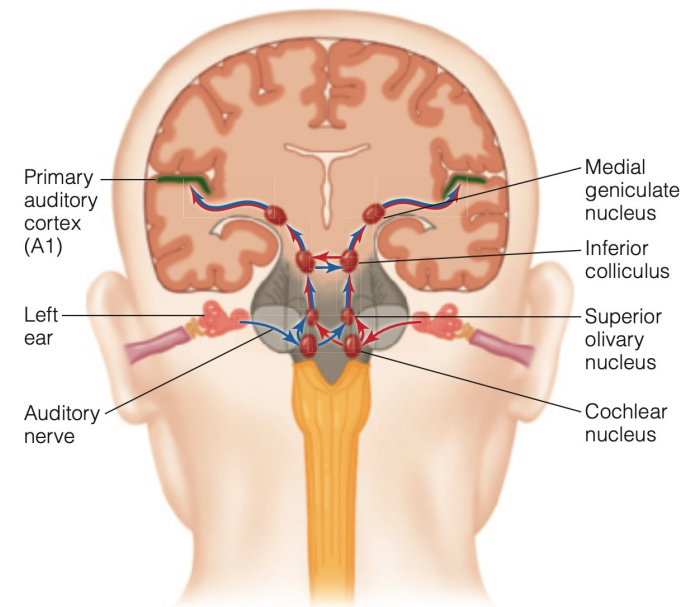
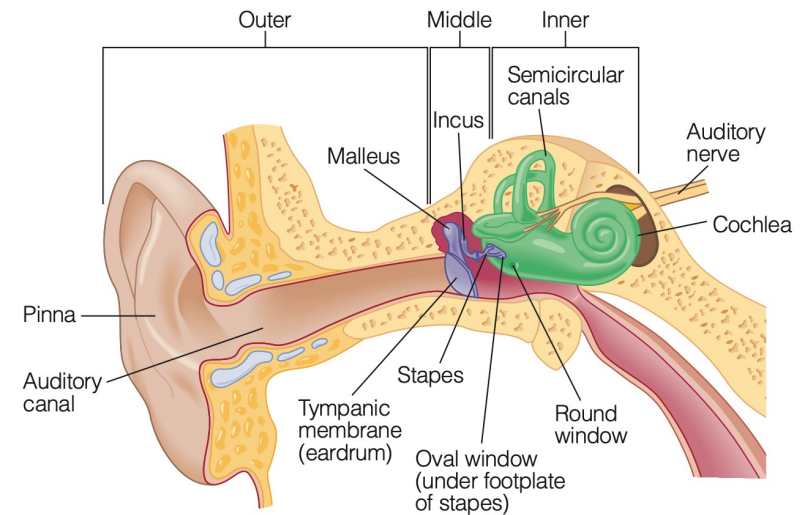
Som

- Sons ou tons puros: uma só frequência
- Sons complexos obtidos por combinação de tons puros com frequências, amplitudes e fases diferentes
- No caso geral um som pode ser representado por uma mistura/soma de um número muito grande de tons puros com amplitudes e fases distintas: decomposição espectral
- **Característica importante: largura de banda que é mede a extensão no domínio da frequência do sinal.**



Audio

- Som percebido pelos seres humanos
- Largura de banda específica, em geral de 20Hz a 20KHz
- Várias fontes:
 - Instrumentos musicais
 - Sistema vocal humano:
 - Canto
 - Fala
 - Vibrações mecânicas produzidas por árvores, vento, etc.
- Percepção humana resulta da ação do ouvido e processamento no cérebro

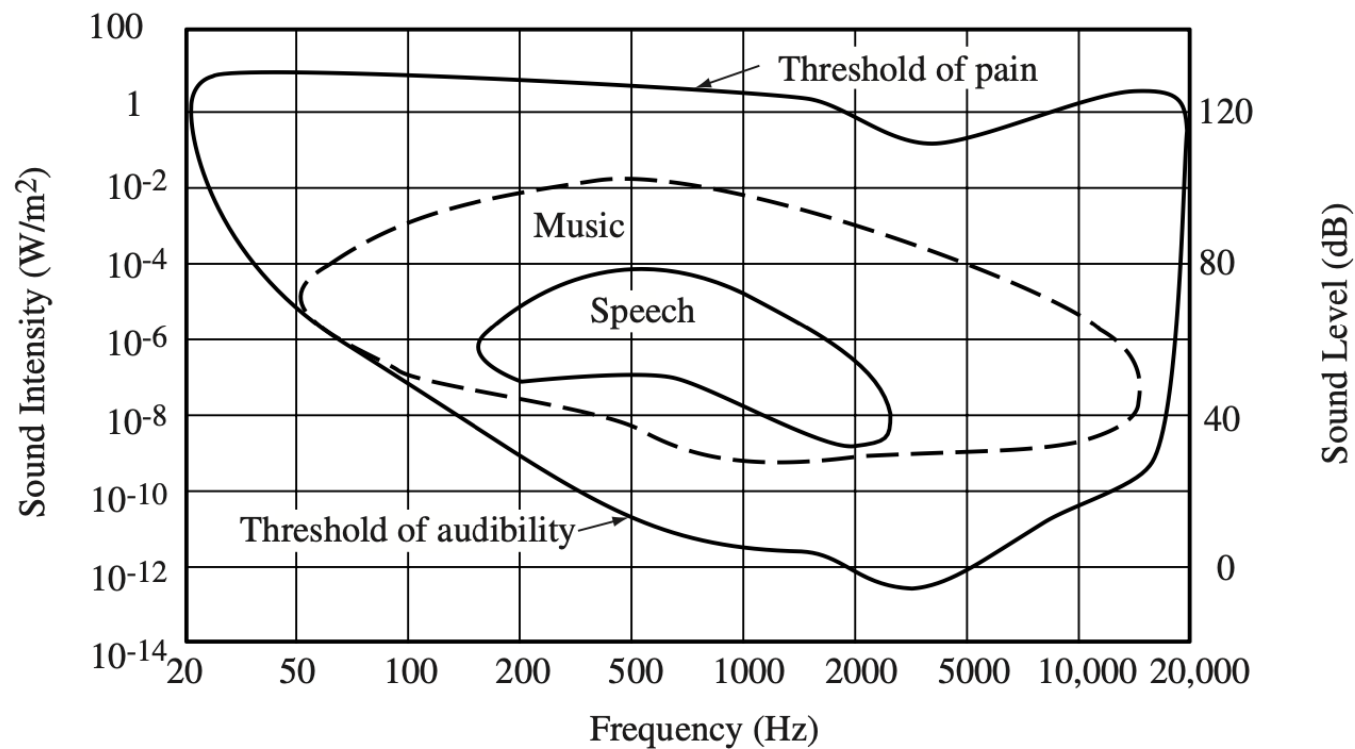


Pressão acústica

- A intensidade de um som é medida usando uma escala relativa, logarítmica expressa em decibels (dB)
- O decibel acústico é definido por $dB_{SPL} = 10 \log \left(\frac{P}{P_{ref}} \right)$ em que P_{ref} é a pressão acústica correspondente a 20 μPa (micro Pascal, Pascal=N/m²)
- A gama de dBs suportáveis por seres humanos vai de cerca de -5 dB a 80-90 dB, com a sensibilidade a depender da frequência e variando de pessoa para pessoa

Sensibilidade auditiva humana

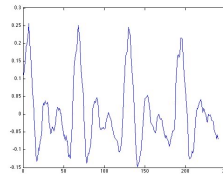
- A audição humana tem os limites representados na figura



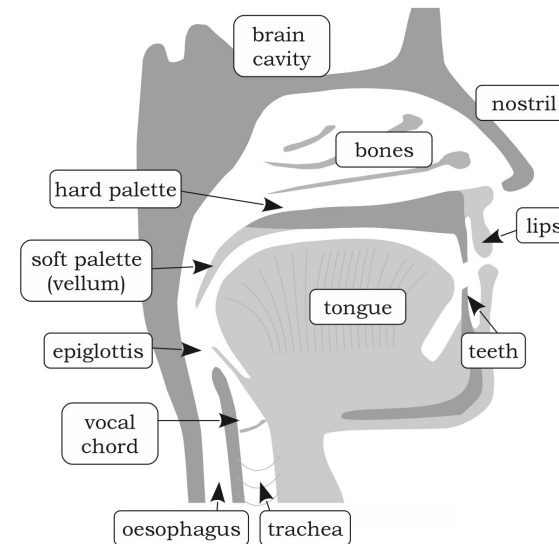
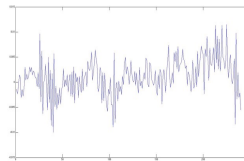
Voz

- A voz humana é produzida pela acção coordenada de vários órgãos do corpo:
 - Diafragma e Pulmões
 - Cordas Vocais
 - Cavidades como a boca, nariz, traqueia
- Os tipos de som produzidos quando falamos são em geral divididos em:

- Sons vozeados, quasi-periódicos

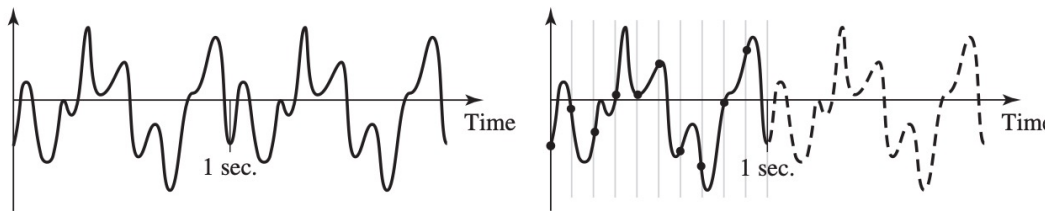


- Sons não-vozeados, sem estrutura periódica



Aquisição- amostragem

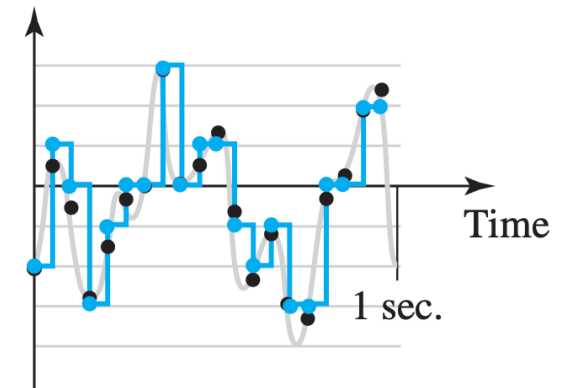
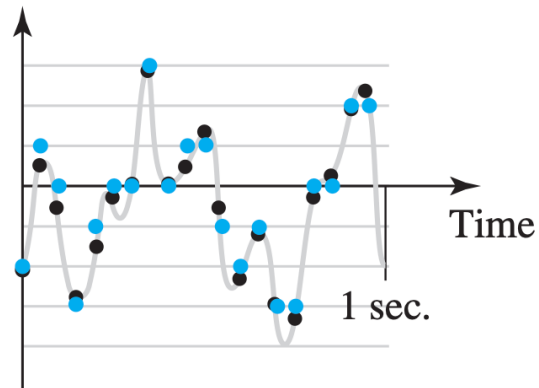
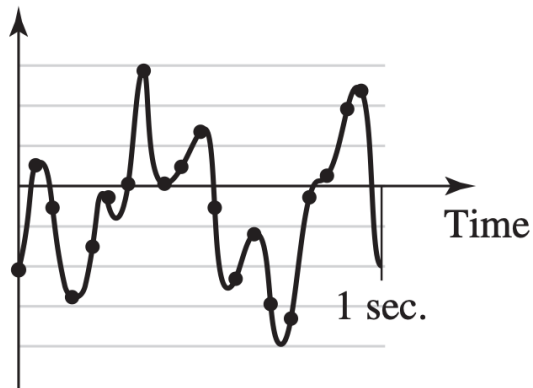
- Sons de origem natural são representados por sinais contínuos no tempo e na amplitude
- Conversão para sinal discreto no tempo por amostragem



- Para não haver distorção na amostragem (aliasing), necessário:
 - Sinal contínuo tem de ter banda limitada, B_{\max} Hz e
 - Amostragem tem de ser feita com ritmo superior a $2 \times B_{\max}$ (Critério de Nyquist), isto é $f_{\text{amost}} > 2B_{\max}$

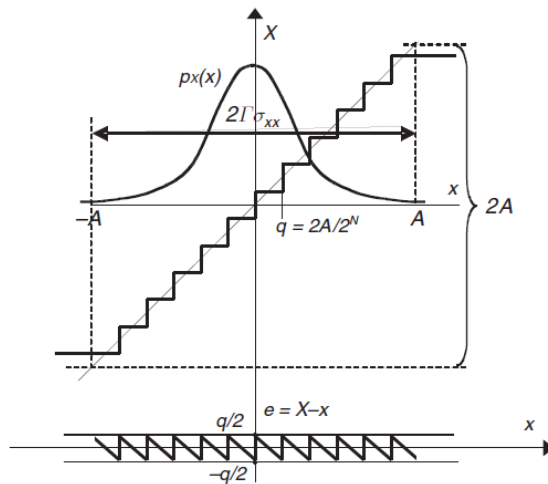
Aquisição- quantização

- Após amostragem, a amplitude das amostras ainda pode assumir um contínuo de valores.
- Para se obter uma representação digital, estas amplitudes têm de ser convertidas para valores discretos (quantização)



Aquisição – quantização com N bits (análise)

- Sinal de áudio $x(n)$ com amplitudes podendo tomar valores entre $-A$ e A
- Quantização uniforme com N bits e passo de quantização q dado por $q = 2A/2^N$
- Erro de quantização: $e(n) = X(n) - x(n)$ $-q/2 \leq e(n) \leq q/2$



Energia do erro

$$\sigma_{ee}^2 = \frac{q^2}{12}$$

- SNR dada por $SNR(dB) = 6.02N + 4.77 + 20\log_{10}(\Gamma)$ ($\Gamma = A/\sigma_{xx}$)
- Ruído com **distribuição espectral aproximadamente uniforme** afeta igualmente todas as frequências.
- Para amplitudes baixas o erro é correlacionado com o sinal e audível!

Codificação PCM

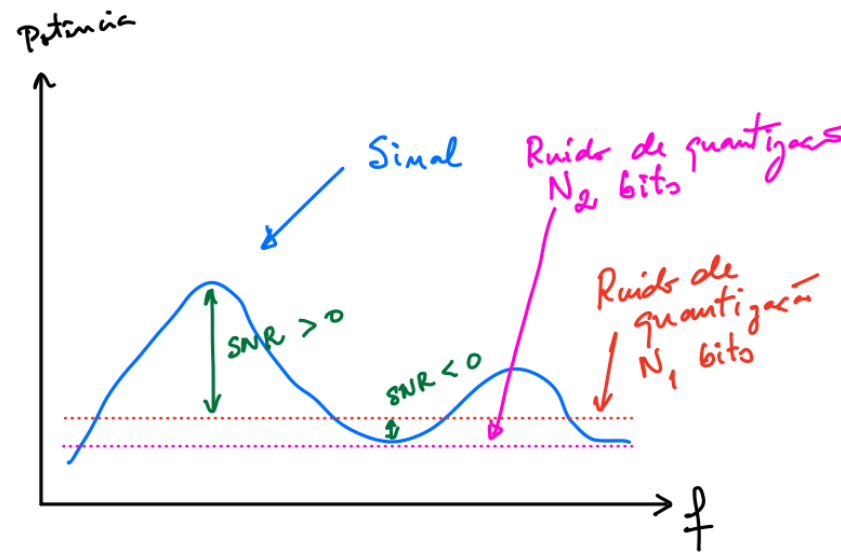
- A quantização com N bits fornece o primeiro formato de representação da audio: PCM (pulse code modulation)
- É o formato de base para muitas aplicações, como sistemas telefónicos, CD audio e outros
- **Exemplos de aplicações:**

Table 1.3. Sampling and raw bit rates for audio signals

	Frequency range [Hz]	Sampling rate [KHz]	PCM resolution [bits/sample]	PCM rate [kb/s]
Telephony speech	300 – 3,400	8	8-16	64-128
Broadband speech	50 – 7,000	16	8-16	128-256
CD audio	10 – 20,000	44.1	2 x 16 (stereo)	1410
DVD multichannel	10 – 22,000	48	(5+1) x 16	$4.6 \cdot 10^3$
DVD audio	10 – 44,000	96	2 x 24	$4.6 \cdot 10^3$
Multichannel audio	10 – 44,000	96	(7+1) x 24	$18.4 \cdot 10^3$

Codificação PCM - Problema

- O problema principal de PCM é que o ruído de quantização afeta igualmente todas as componentes espectrais, causando uma qualidade (SNR) variável na frequência



- Na figura mostra-se que se pode garantir $SNR > 0$ para todas as frequências usando um número de bits $N_2 > N_1$
- Porém, se o perfil de potência do sinal tiver diferenças muito grandes, essa solução implica utilizar um número de bits muito elevado, conduzindo a uma representação PCM pouco económica.
- Há outra solução que permita uma SNR acima de um valor mínimo sem utilizar um bitrate muito elevado ?

Leitura sobre PCM

Estude os seguintes capítulos e secções de [1]; Cap. 1, Cap. 2: 2.1 a 2.3, Cap 6: 6.1 (Preâmbulo), 6.1.1. Nesses capítulos e secções encontrará informação sobre as funções Matlab usadas para ler, escrever e reproduzir áudio, assim como sobre a quantização de sinais.

Estude também o preâmbulo do Cap. 2 de [2], assim como as secções 2.1 (introdução), 2.1.1, 2.2 (introdução). De seguida estude e reproduza a demonstração apresentada na secção 2.2.1. Por fim leia a secção 2.4.

[1]-Ian McLoughlin, “Speech and Audio Processing - A MATLAB-based Approach”, Cambridge University Press 2016

[2]-Thierry Dutoit, Ferran Marqués, “Applied Signal Processing A MATLAB-Based Proof of Concept” , Springer 2009