



# ALGUNS PARÂMETROS E PADRÕES ÁUDIO-ACÚSTICOS

---

CSM – QS2020 – Prof. Mário Minami

# SUMÁRIO



Visão



Escalas, MIDI e Envelopes



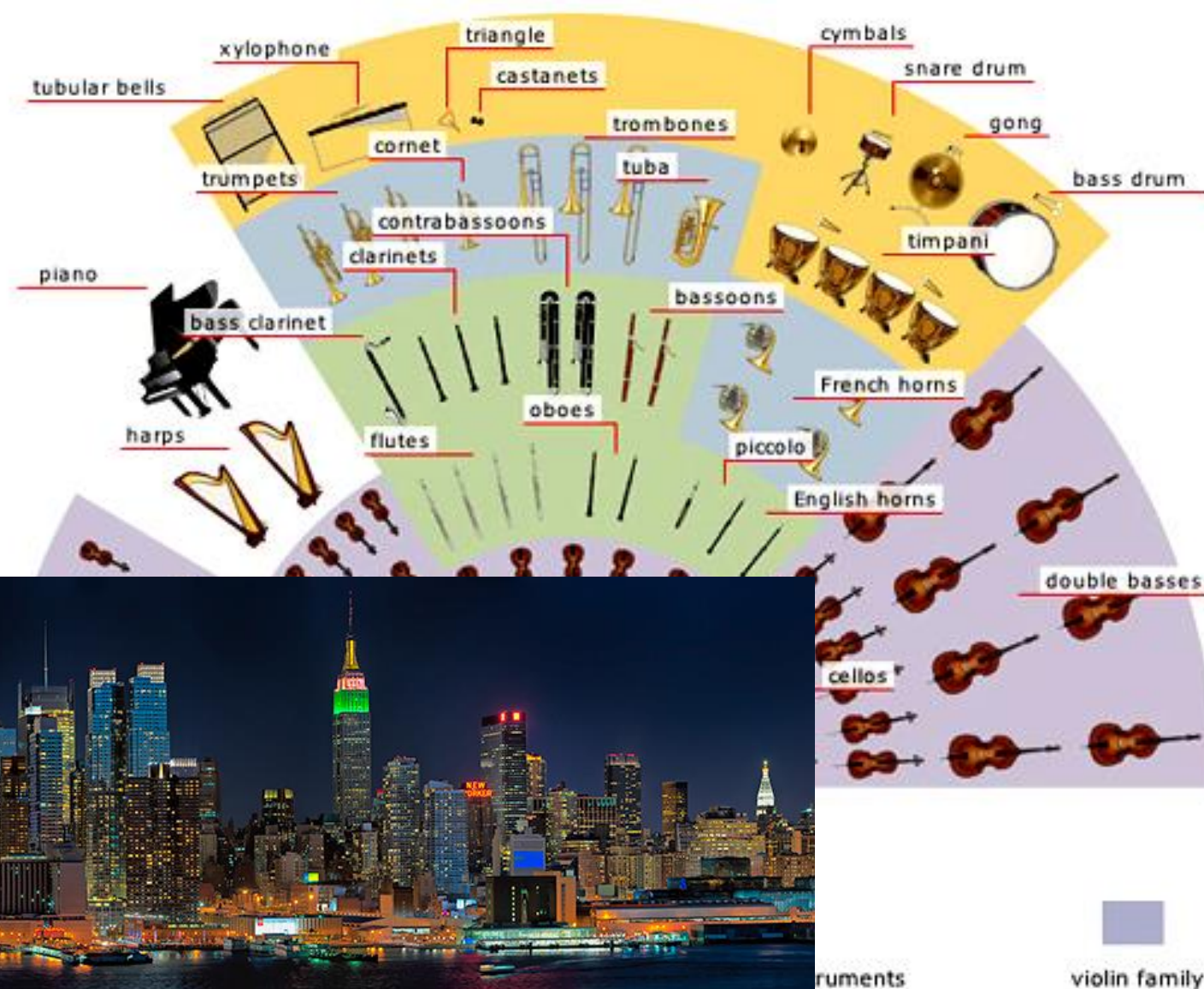
Algumas Quantidades Acústicas



Medidas de Qualidade de Fala

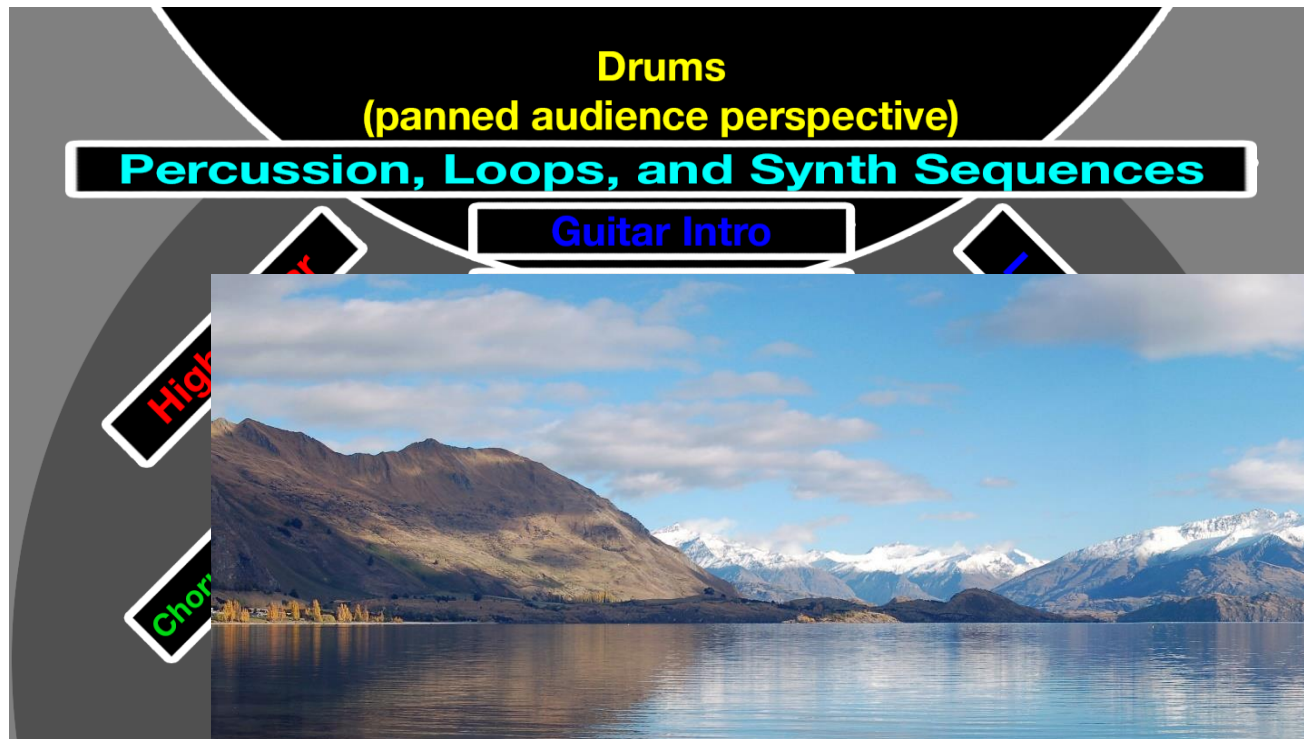
# SOM PANORÂMICO: OUVIR “VENDO” TUDO

“Pan” ou “Panning”

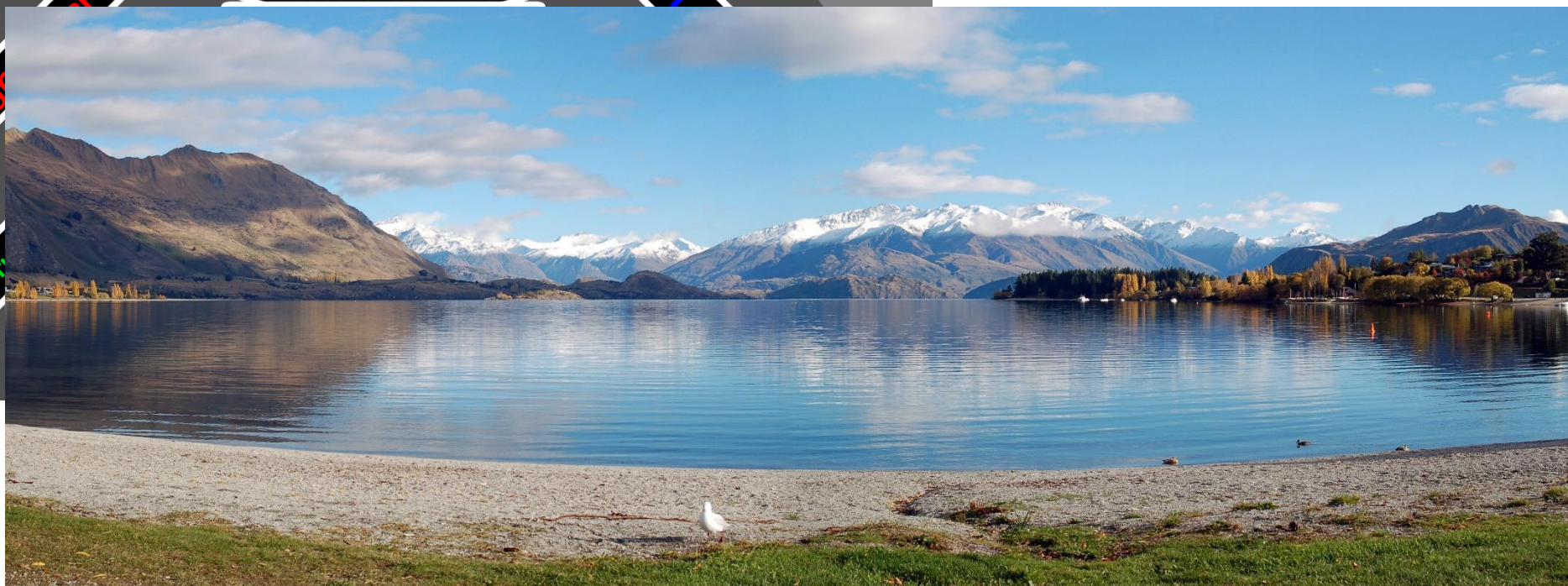




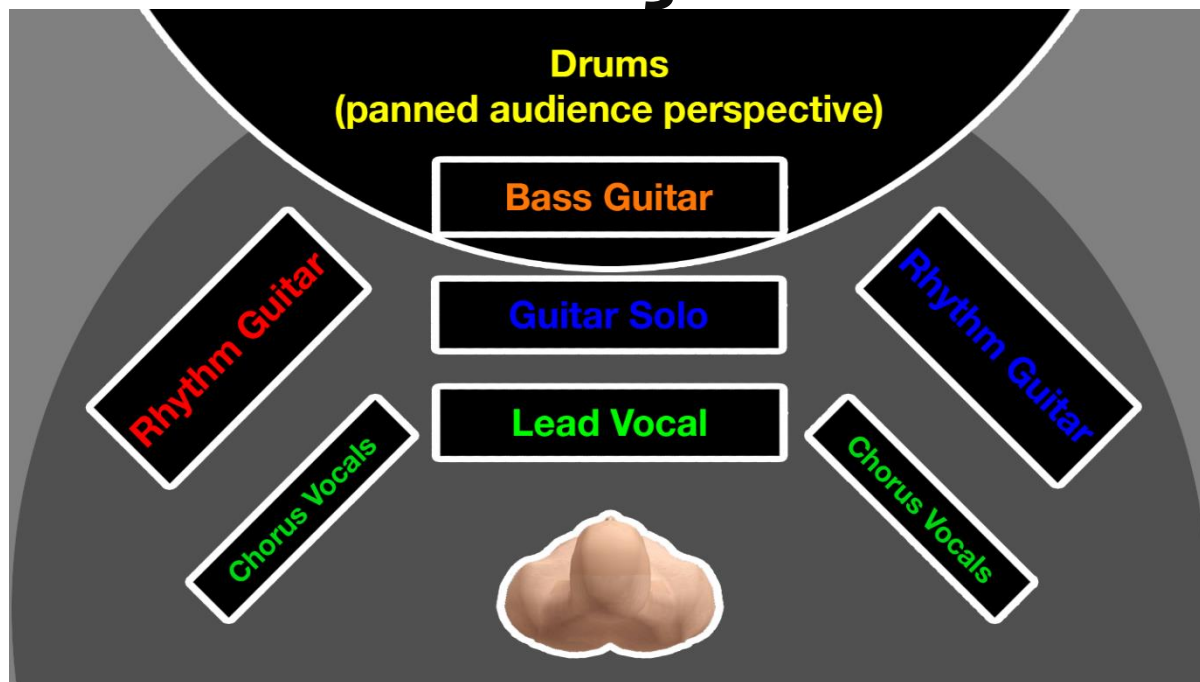
# “Visão” também depende da Paisagem



Pan Jazz, cool



# MENOS VARIEDADE, MAIS “DISTINÇÃO”

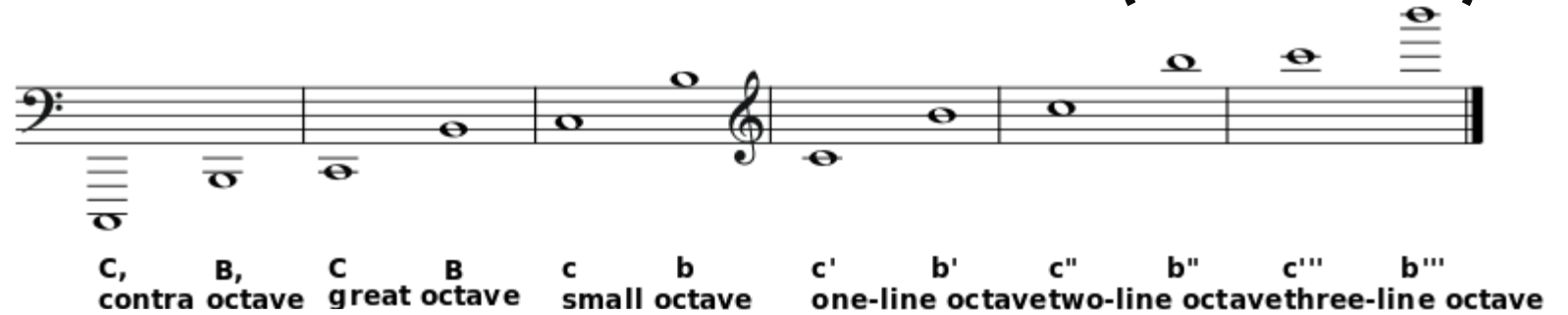


Pan Pop&Rock

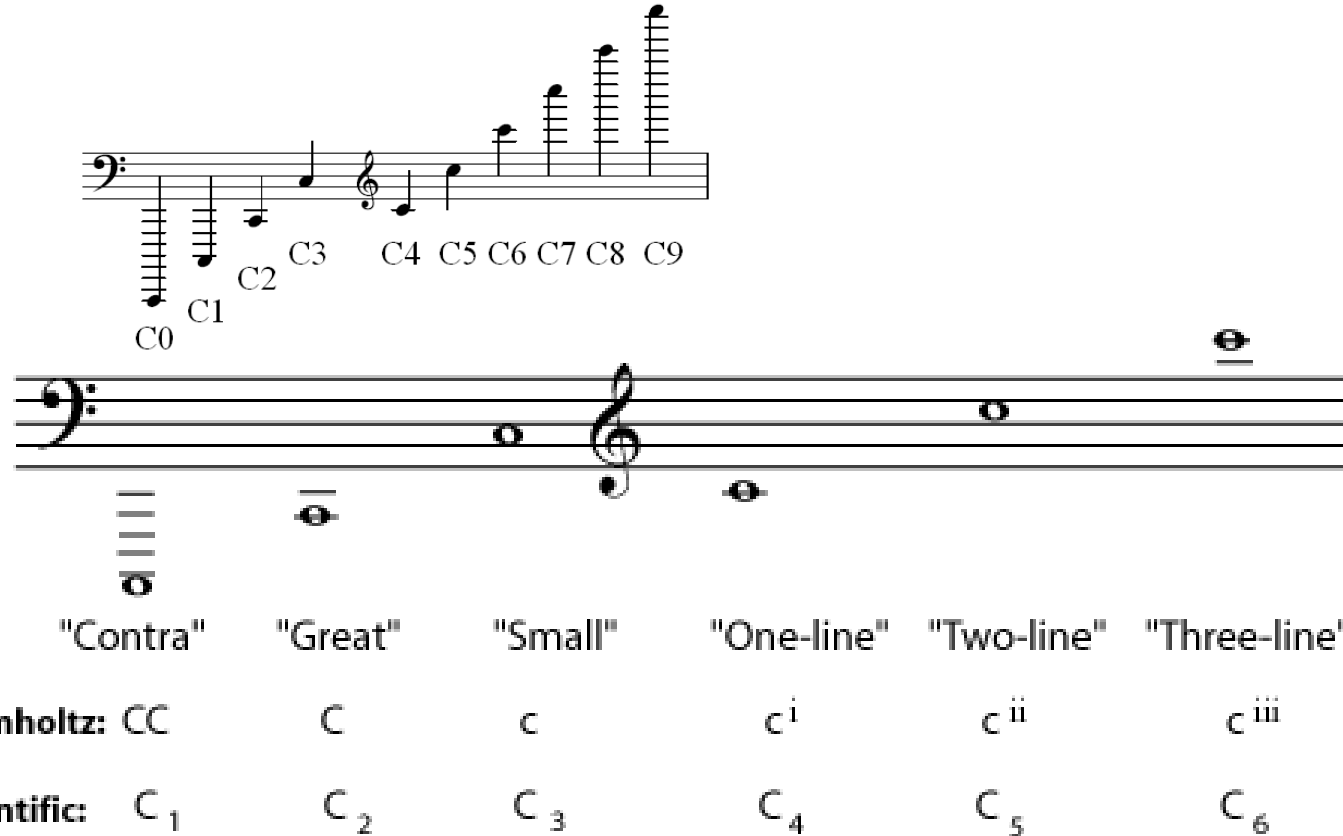


# PITCH - SENSAÇÃO AUDITIVA DOS TONS MUSICAIS (“NOTAS”)

## Sistema de Helmholtz



## Sistema Científico



Say:	"Contra"	"Great"	"Small"	"One-line"	"Two-line"	"Three-line"
Helmholtz:	CC	C	c	c <sup>i</sup>	c <sup>ii</sup>	c <sup>iii</sup>
Scientific:	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>

# PITCH DE NOTAS ESCRITAS: CLAVE

*Bass Clef*

48 50 52 53 55 57 59 60 62 64 65 67 69 71 72 74 76 77 79 81 83 84 86 88 89 91 93 95 96

C2 C3 C4 C5 C6 C7

c d e f g a b c' d' e' f' g' a' b' c'' d'' e'' f'' g'' a'' b'' c''' d''' e''' f''' g''' a''' b''' c''''

24 26 28 29 31 33 35 36 38 40 41 43 45 47

C0 C1 C2

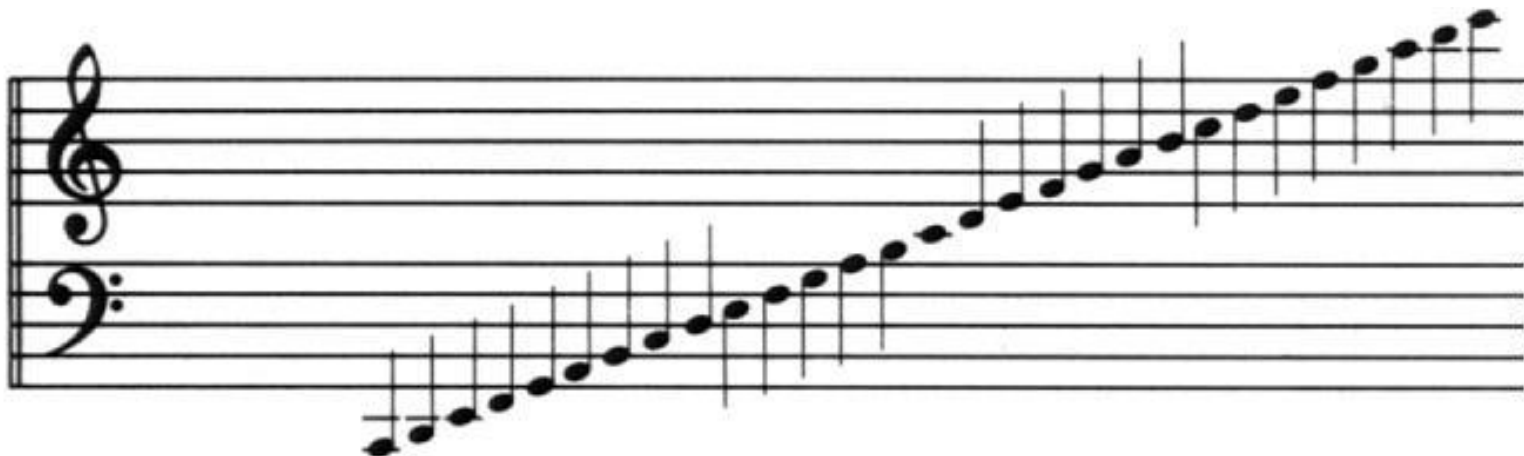
C, D, E, F, G, A, B, C D E F G A B

Middle C

A 440



# FREQUÊNCIAS



Frequency (Hz) →	65	73	82	87	98	110	123	131	147	165	175	196	220	247	262	294	330	349	392	440	494	523	587	659	698	784	880	987	1047
Note →	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C
Keyboard →																													



# MUSICAL INSTRUMENT DIGITAL INTERFACE



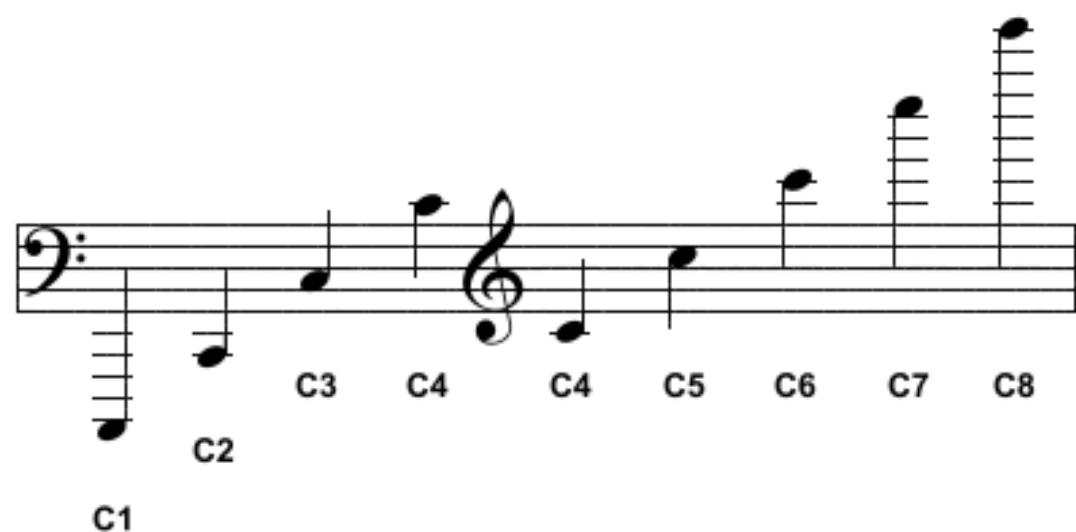
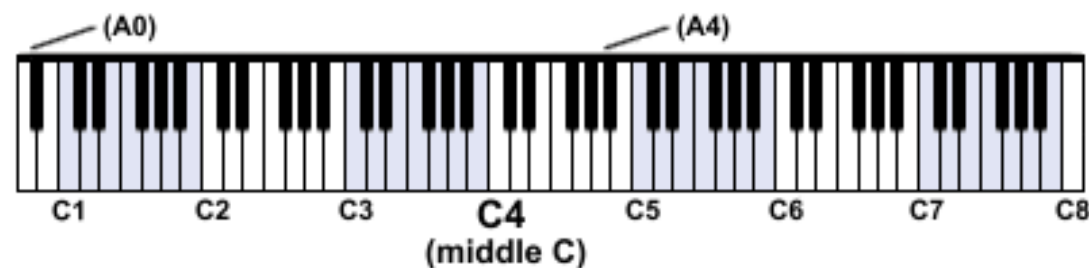
## MIDI Chunk

Chunk Header

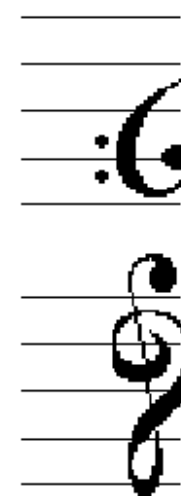
ID	Length	Data
----	--------	------

Variable length specified  
by the Length field.

# NOTAS E MIDI



MIDI number	Note name	Keyboard	Frequency Hz	Period ms
21	A0		27.500	36.36
23	B0		30.868	29.135
24	C1		32.703	30.58
26	D1		36.708	27.24
28	E1		41.203	24.27
29	F1		43.654	22.91
31	G1		48.999	20.41
33	A1		55.000	18.18
35	B1		61.735	16.20
36	C2		65.406	15.29
38	D2		73.416	13.62
40	E2		82.407	12.13
41	F2		87.307	11.45
43	G2		97.999	10.20
44	A2		110.00	9.091
45	B2		123.47	8.099
47	C3		130.81	7.645
48	D3		146.83	6.811
50	E3		164.81	6.068
52	F3		174.61	5.727
53	G3		196.00	5.102
55	A3		220.00	4.545
57	B3		246.94	4.050
59	<b>C4</b>		<b>261.63</b>	<b>3.822</b>
61	D4		293.67	3.405
62	E4		329.63	3.034
64	F4		349.23	2.863
65	G4		392.00	2.551
67	A4		440.00	2.273
69	B4		493.88	2.025
71	C5		523.25	1.910
72	D5		587.33	1.703
74	E5		659.26	1.517
76	F5		698.46	1.432
77	G5		783.99	1.276
79	A5		880.00	1.136
81	B5		987.77	1.012
83	C6		1046.5	0.9556
84	D6		1174.7	0.8513
86	E6		1318.5	0.7584
88	F6		1396.9	0.7159
89	G6		1568.0	0.6378
91	A6		1760.0	0.5682
93	B6		1975.5	0.5062
95	C7		2093.0	0.4778
96	D7		2349.3	0.4257
98	E7		2637.0	0.3792
100	F7		2793.0	0.3580
101	G7		3136.0	0.3189
103	A7		3520.0	0.2841
105	B7		3951.1	0.2531
107	C8		4186.0	0.2389



# TABELA COM ESCALA CROMÁTICA EXTENDIDA

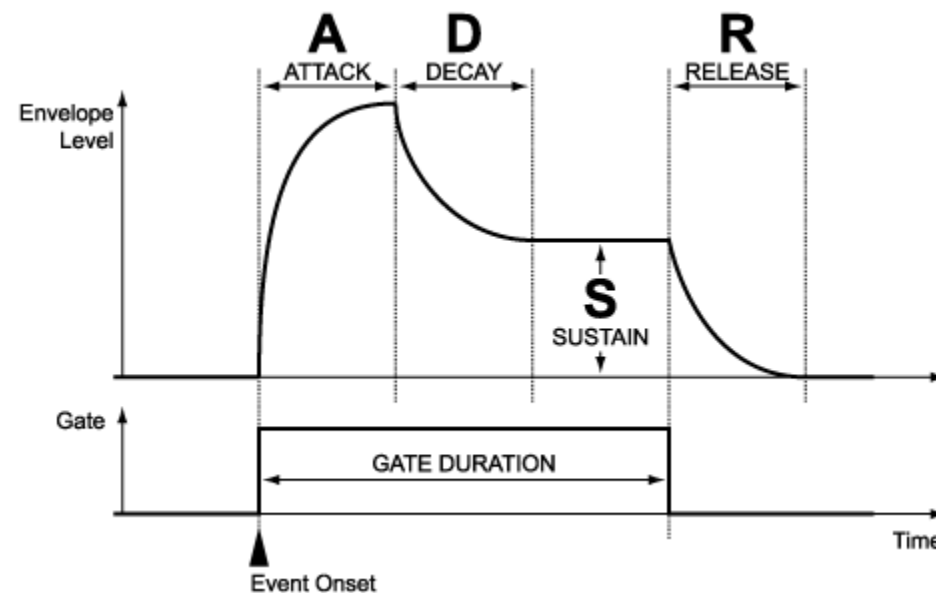
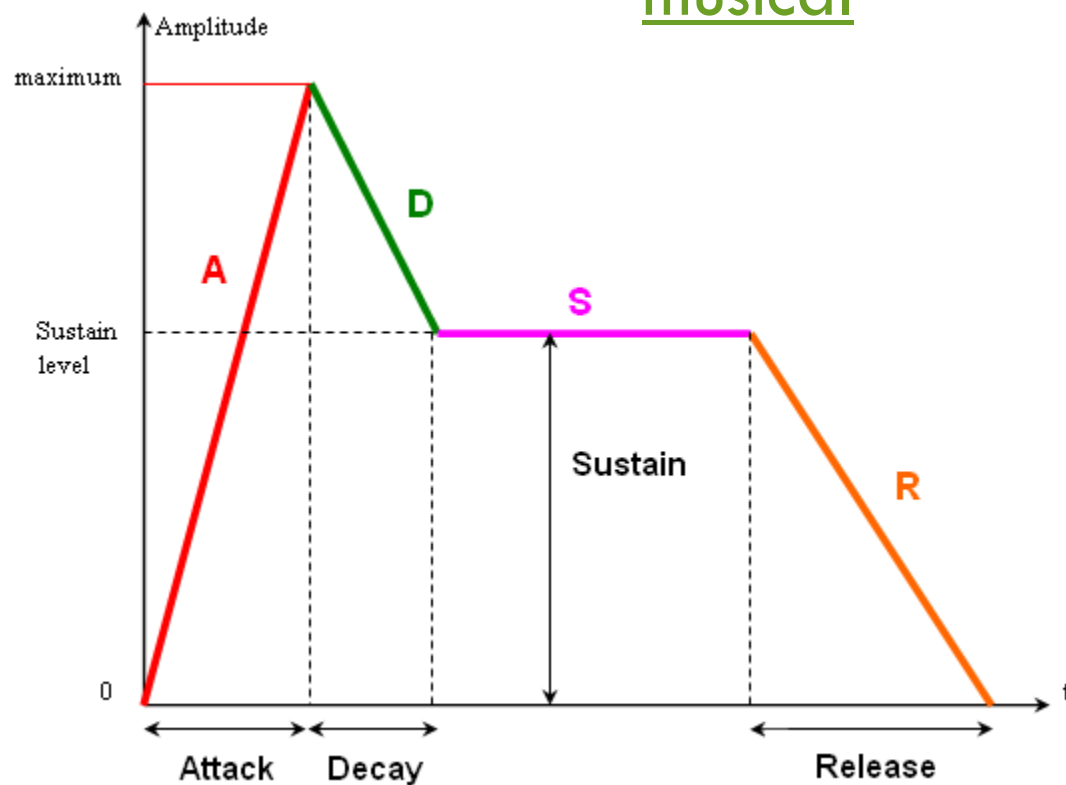
escala cromática é uma escala que contém 12 notas com intervalos de semitons entre elas

Frequency in hertz (semitones above or below middle C)											
Octave → Note ↓	0	1	2	3	4	5	6	7	8	9	10
C	16.352 (-48)	32.703 (-36)	65.406 (-24)	130.81 (-12)	261.63 (±0)	523.25 (+12)	1046.5 (+24)	2093.0 (+36)	4186.0 (+48)	8372.0 (+60)	16744.0 (+72)
C#/D♭	17.324 (-47)	34.648 (-35)	69.296 (-23)	138.59 (-11)	277.18 (+1)	554.37 (+13)	1108.7 (+25)	2217.5 (+37)	4434.9 (+49)	8869.8 (+61)	17739.7 (+73)
D	18.354 (-46)	36.708 (-34)	73.416 (-22)	146.83 (-10)	293.66 (+2)	587.33 (+14)	1174.7 (+26)	2349.3 (+38)	4698.6 (+50)	9397.3 (+62)	18794.5 (+74)
E♭/D♯	19.445 (-45)	38.891 (-33)	77.782 (-21)	155.56 (-9)	311.13 (+3)	622.25 (+15)	1244.5 (+27)	2489.0 (+39)	4978.0 (+51)	9956.1 (+63)	19912.1 (+75)
E	20.602 (-44)	41.203 (-32)	82.407 (-20)	164.81 (-8)	329.63 (+4)	659.26 (+16)	1318.5 (+28)	2637.0 (+40)	5274.0 (+52)	10548.1 (+64)	21096.2 (+76)
F	21.827 (-43)	43.654 (-31)	87.307 (-19)	174.61 (-7)	349.23 (+5)	698.46 (+17)	1396.9 (+29)	2793.8 (+41)	5587.7 (+53)	11175.3 (+65)	22350.6 (+77)
F♯/G♭	23.125 (-42)	46.249 (-30)	92.499 (-18)	185.00 (-6)	369.99 (+6)	739.99 (+18)	1480.0 (+30)	2960.0 (+42)	5919.9 (+54)	11839.8 (+66)	23679.6 (+78)
G	24.500 (-41)	48.999 (-29)	97.999 (-17)	196.00 (-5)	392.00 (+7)	783.99 (+19)	1568.0 (+31)	3136.0 (+43)	6271.9 (+55)	12543.9 (+67)	25087.7 (+79)
A♭/G♯	25.957 (-40)	51.913 (-28)	103.83 (-16)	207.65 (-4)	415.30 (+8)	830.61 (+20)	1661.2 (+32)	3322.4 (+44)	6644.9 (+56)	13289.8 (+68)	26579.5 (+80)
A	27.500 (-39)	55.000 (-27)	110.00 (-15)	220.00 (-3)	440.00 (+9)	880.00 (+21)	1760.0 (+33)	3520.0 (+45)	7040.0 (+57)	14080.0 (+69)	28160.0 (+81)
B♭/A♯	29.135 (-38)	58.270 (-26)	116.54 (-14)	233.08 (-2)	466.16 (+10)	932.33 (+22)	1864.7 (+34)	3729.3 (+46)	7458.6 (+58)	14917.2 (+70)	29834.5 (+82)
B	30.868 (-37)	61.735 (-25)	123.47 (-13)	246.94 (-1)	493.88 (+11)	987.77 (+23)	1975.5 (+35)	3951.1 (+47)	7902.1 (+59)	15804.3 (+71)	31608.5 (+83)

# ENVELOPES

## ADSR

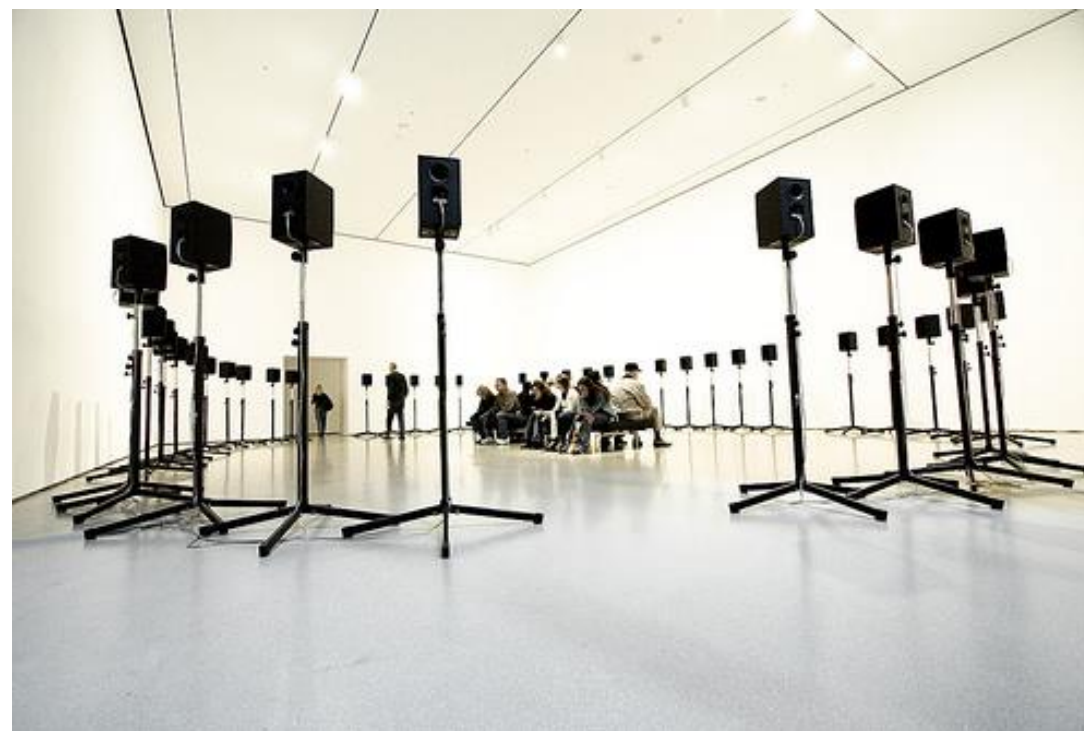
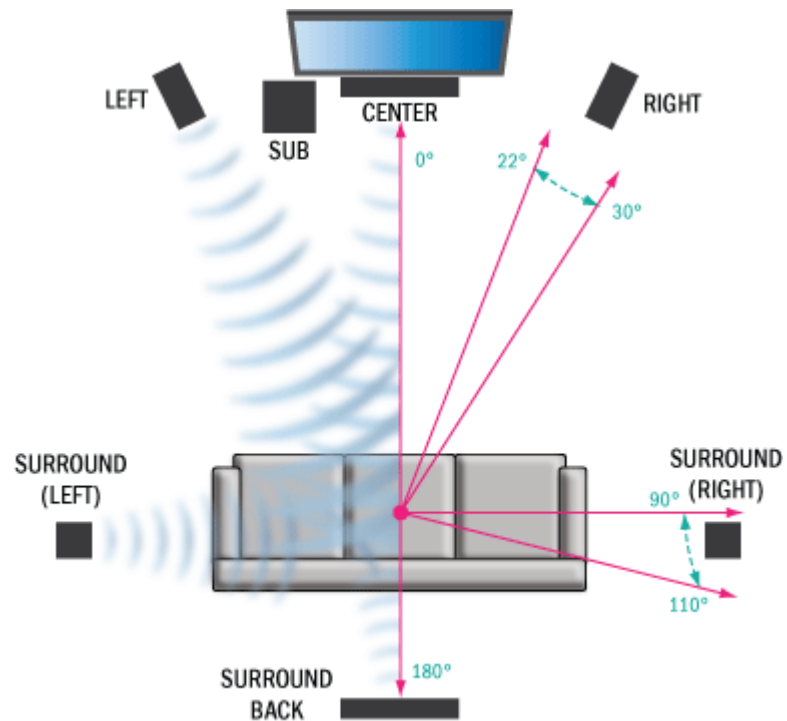
É uma das formas mais comumente utilizadas para aplicar um envelope de amplitude a um som para produzir um timbre característico de um instrumento musical



**Onset detection:** detecção de Ataque da nota musical

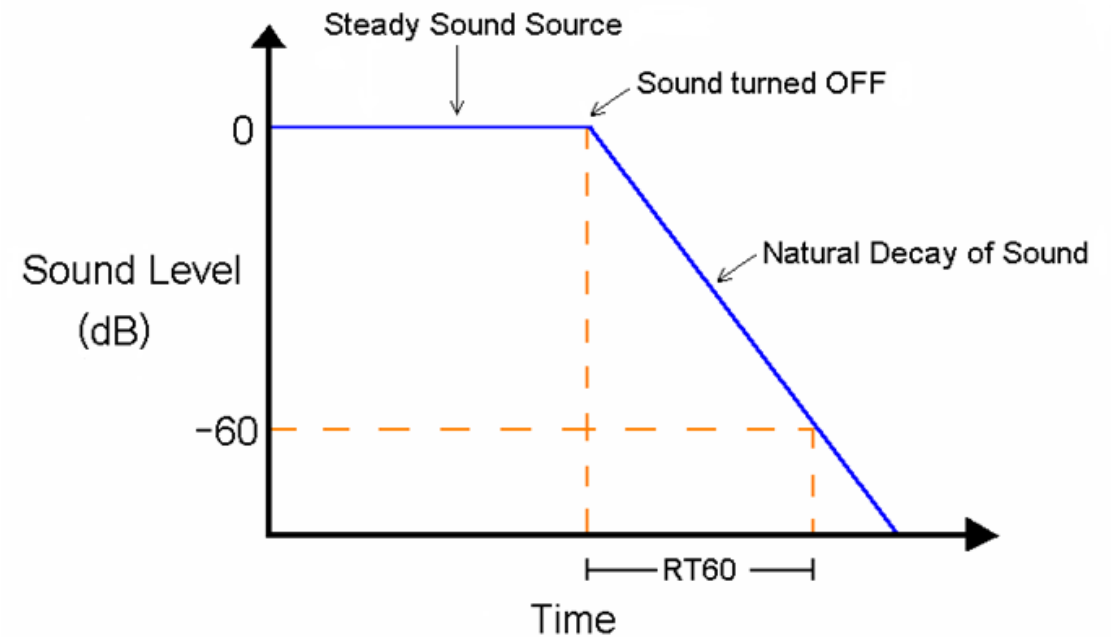
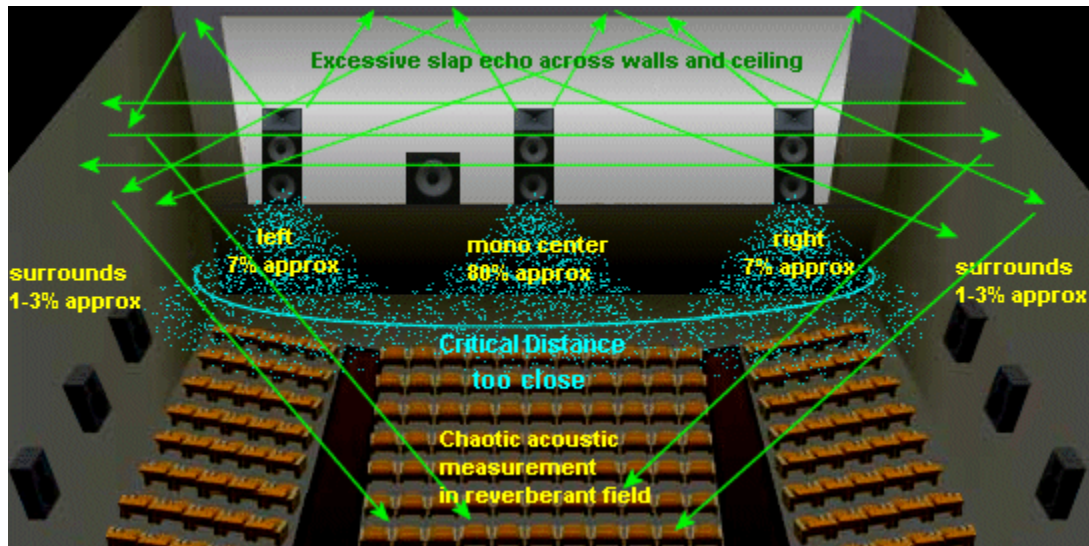
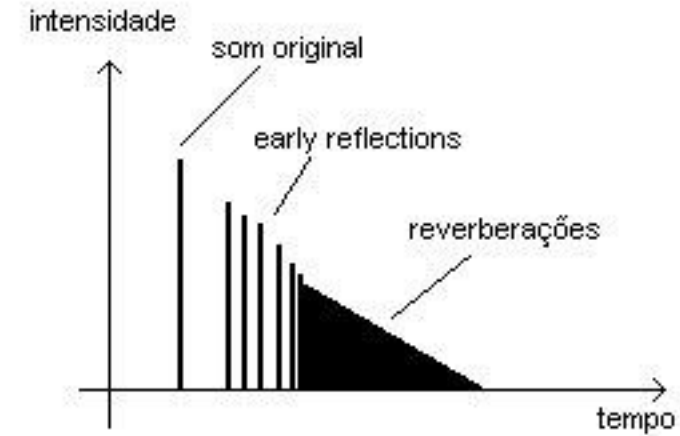
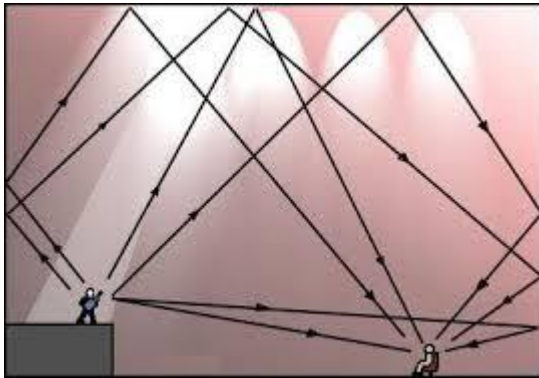


# SURROUND SOUND



Existe uma Recomendação ITU-R  
BS.775-1

# REVERBERAÇÕES, GRANDES SALAS

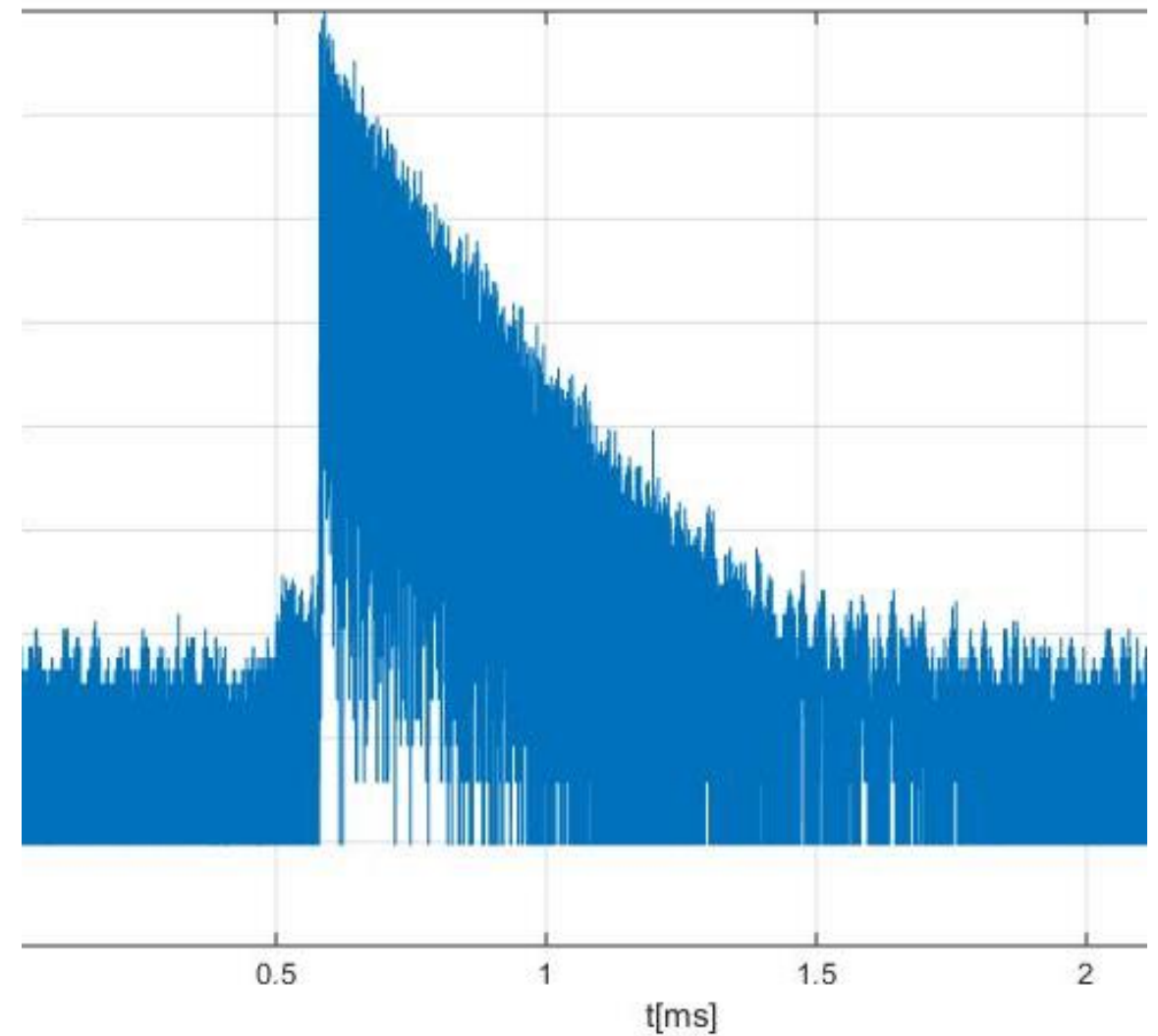


# QUANTIDADES PSICOACÚSTICAS BÁSICAS

Loudness

Pitch

Análise Temporal e Percepção da  
Modulação



# DIMENSÕES DE TRANSFORMAÇÕES NA PERCEPÇÃO AUDITIVA HUMANA

Intensidade sonora no loudness percebido subjetivamente

Componentes de frequência predominantes em pitch subjetivo

Padrões temporais e ritmos nas flutuações percebidas subjetivamente

Conteúdo espectro-temporal dos sinais acústicos em timbre subjetivo

Disparidades intra-aurais entre os dois ouvidos e conteúdos espectro-temporais de sinais acústicos na localização espacial e da extensão espacial de um objeto

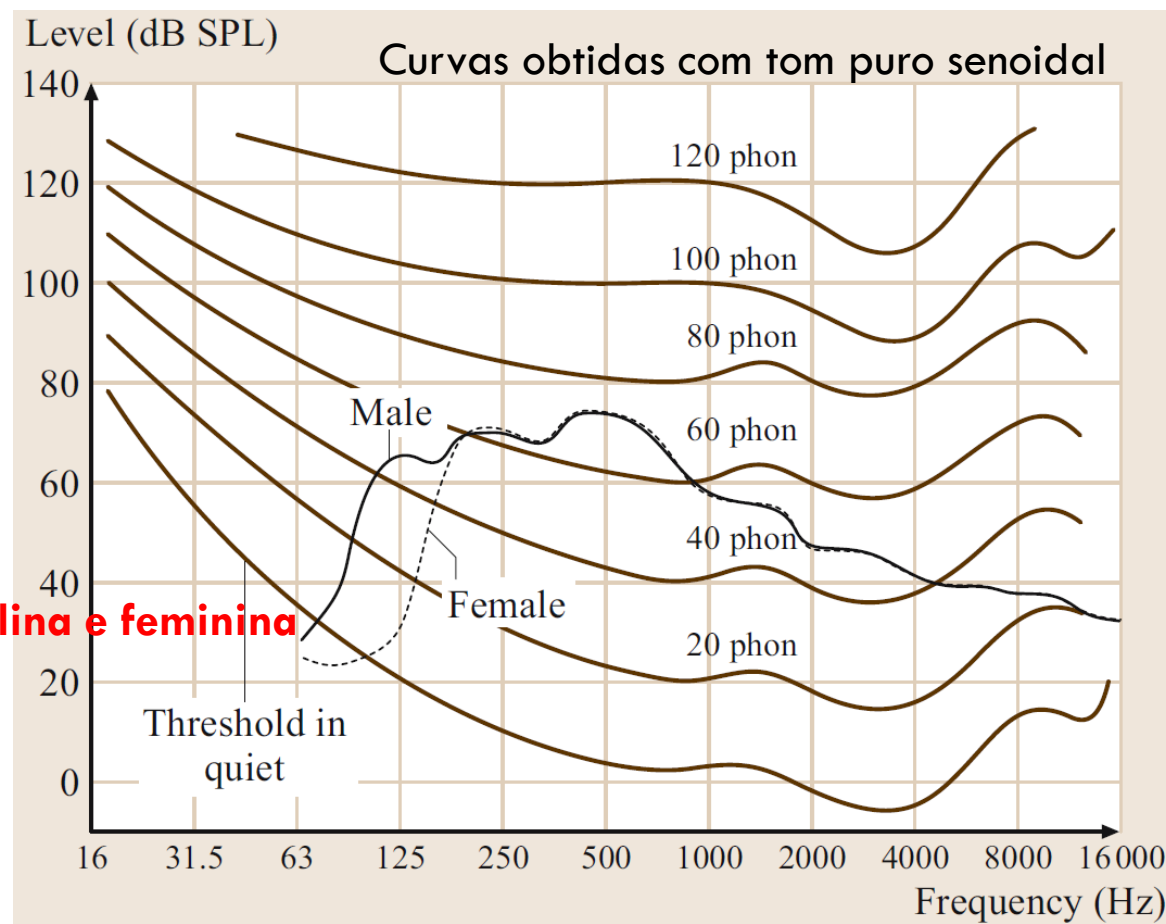


# Mapeamento da Intensidade de Loudness

- Curvas de percepção de tom igual - isophons
- A percepção da voz humana é diferente, pois ela não é composta de um único tom senoidal (vide curva isophons)

**Espectros médios da voz masculina e feminina**

[Intensidade sonora] = phon



Maior sensibilidade perto de 1kHz

# Como expressar as bandas críticas?

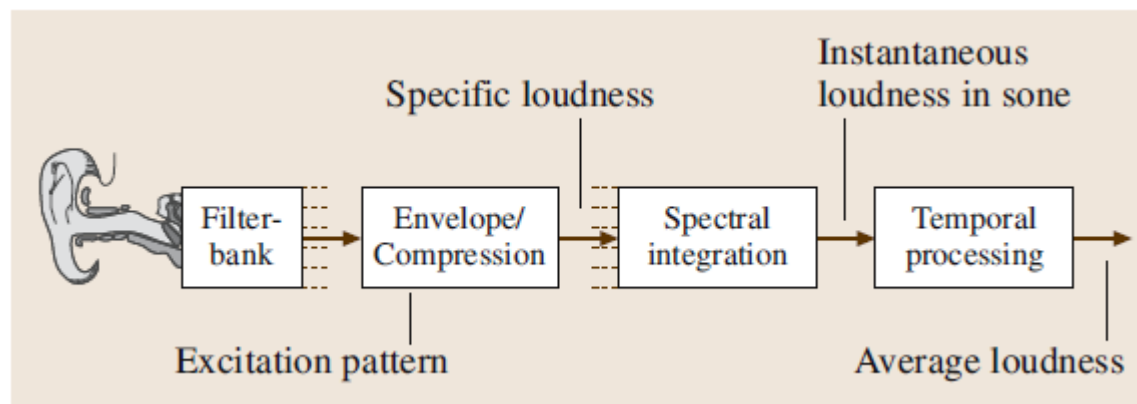
(sons de maior banda são percebidos como de menor loudness que os de menor banda)

Nível RMS não ponderado, em dB SPL, ou  $10\log(I/I_0)$ , sendo  $I_0$  um nível de referência padronizado

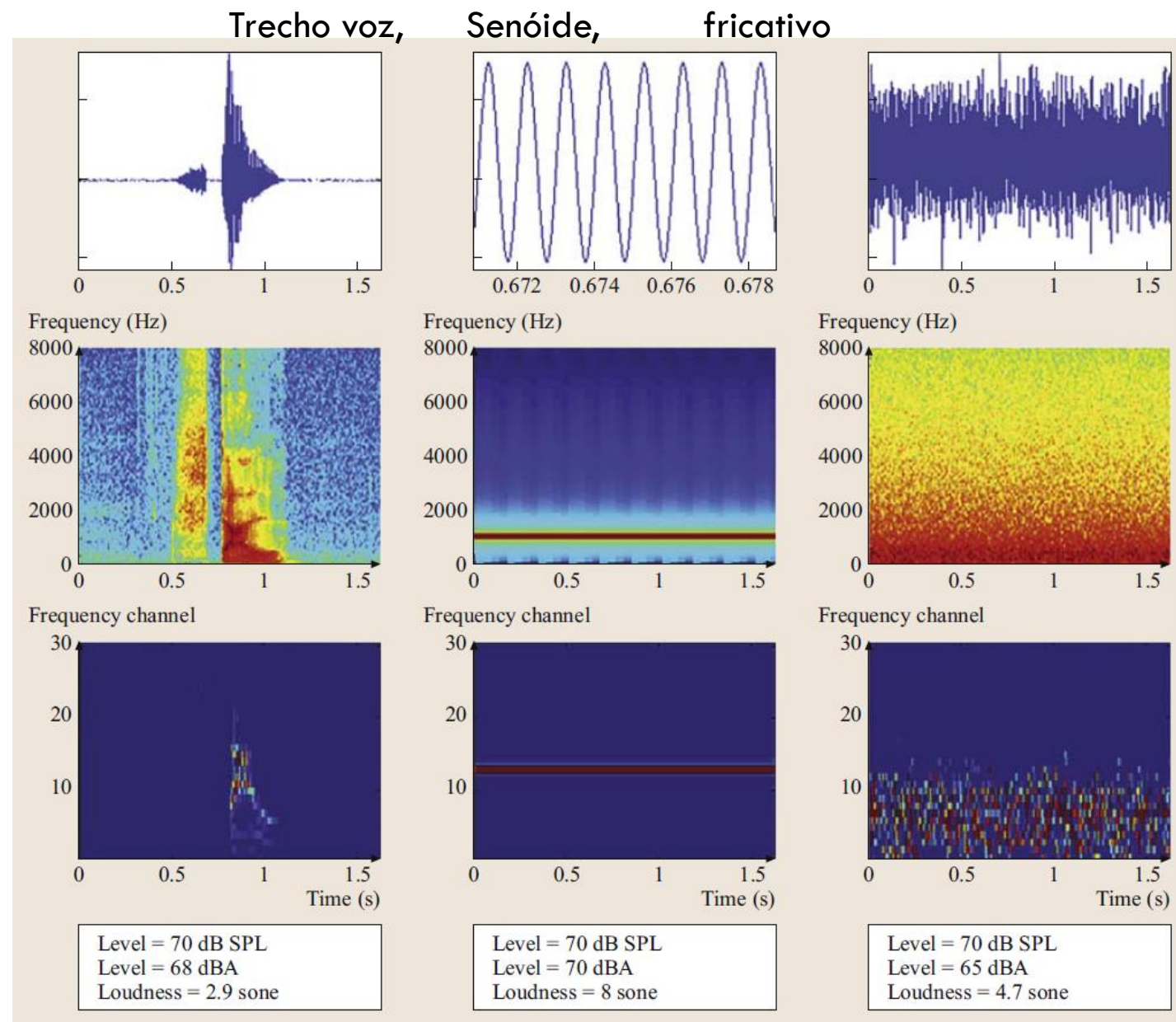
Nível de sinal A-ponderado – uma função de ponderação aproximada entre 20 e 30 phons (A-weighting) é aplicada em todo o espectro sendo a soma o nível total sonoro.

Loudness em sone.

$$N[sone] = (I/I_0)^\alpha, \alpha \approx 0,3 \text{ para NB e } I_0 = 40 \text{ SPL}$$



# DIFERENÇA ENTRE DB SPL, DBA E SONE



# PITCH — ESCALA MEL

Se um sinal senoidal de baixa frequência vai aumentando gradativamente até 500 Hz, a percepção de cada tom, ou tonalidade sonora - Pitch, incrementa também linearmente com a frequência.

Nas frequências acima de 1 kHz a percepção do pitch muda quase que de forma logarítmica.

A combinação destas duas regiões de percepção geraram a escala psicofísica mel:

$$m = \frac{1000}{\log 2} \log \left( 1 + \frac{f}{1000} \right)$$

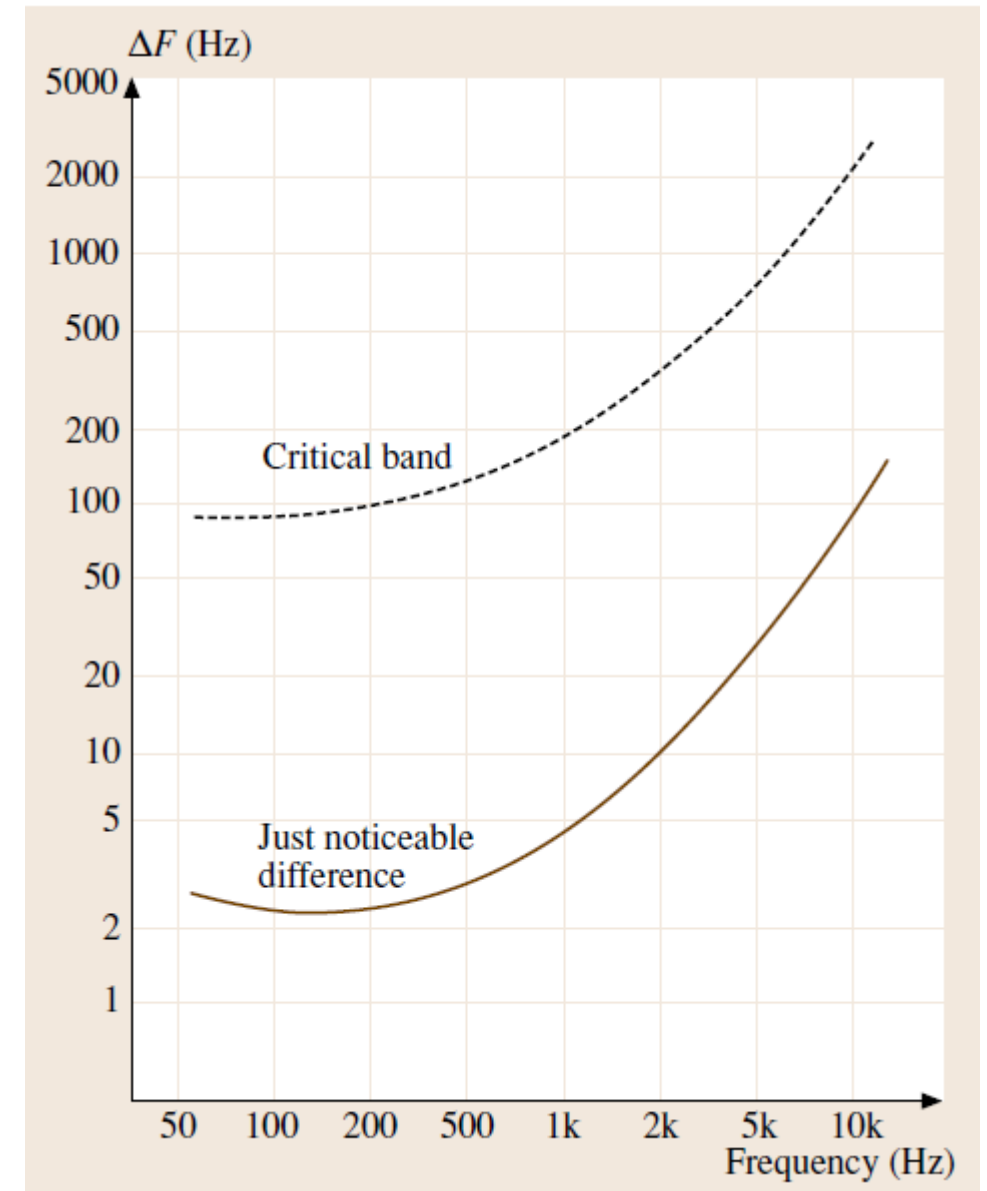


# PITCH — ESCALA BARK

O mapeamento das frequências ocorre na membrana basilar, onde as frequências até 2 kHz vão até a metade da membrana e de 2 kHz a 20 kHz a outra metade.

A inclinação do mapeamento resulta na JND (just noticeable difference): cerca de 3 Hz de frequência JND abaixo de 500 Hz e cerca de 0.6% para frequências acima de 1 kHz, que são aproximadamente 3 mel.

O valor de  $1/30$  da largura de banda da banda crítica (CB) é relevante tanto na somatória de loudness quanto no mascaramento espectral



# BARKS E ESCALA BARK (Z)

Toda energia espectral numa CB é somada e mascara (ou desabilita) a detecção de tons senoidais centrados naquela CB, desde que estejam abaixo do nível de disparo de mascaramento (*masked threshold*).

Esta largura de banda crítica pode ser expressa em bark (homenagem ao físico alemão Barkhausen) como função da frequência  $f_0$  como:

$$\begin{aligned} 1 \text{ bark} &= 100 \text{ mel} \\ &\approx 100 \text{ Hz para } f \leq 500 \text{ Hz} \\ &\approx \frac{1}{5f_0} \text{ para } f > 500 \text{ Hz} \end{aligned}$$

A escala psicofísica resultante da integração das CBs nas frequências é a escala bark Z, aproximada por um arco seno-hiperbólico:

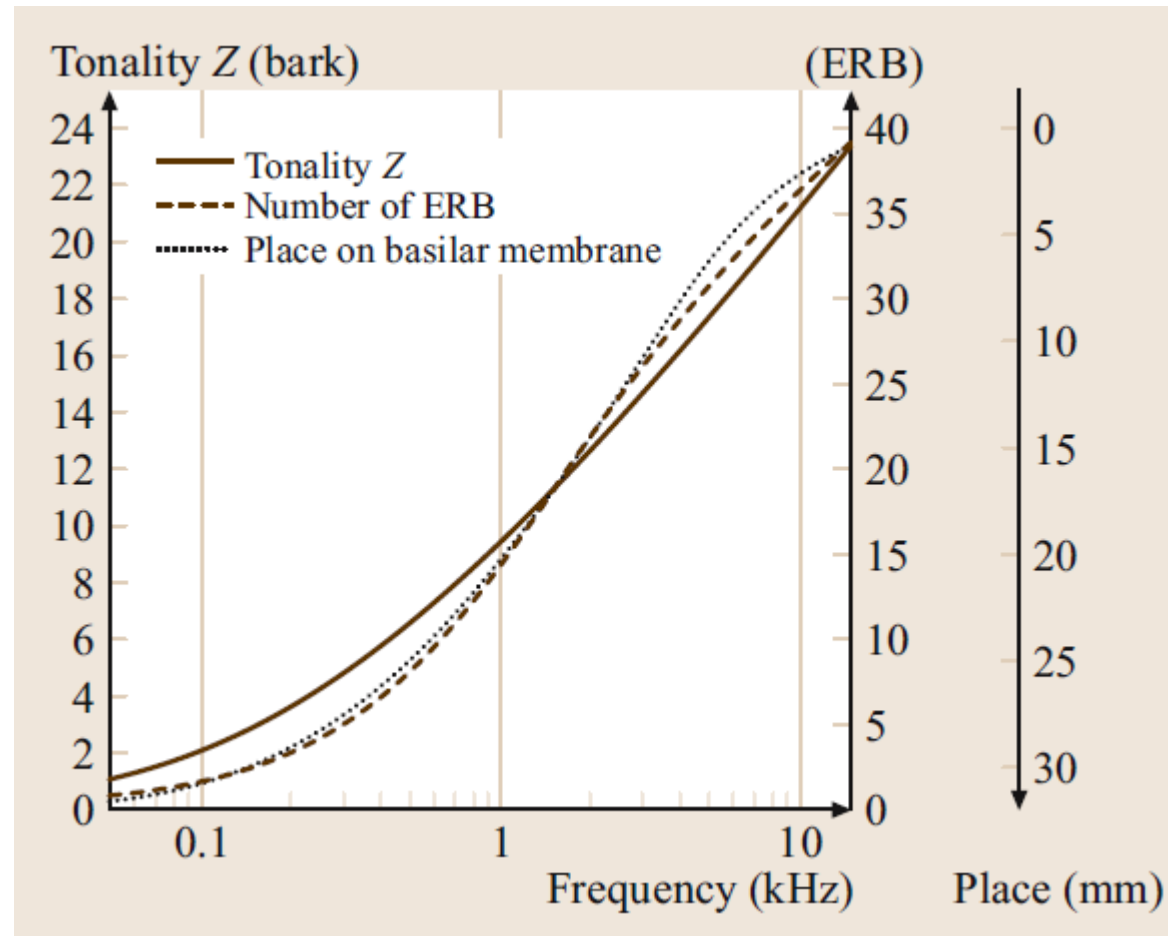
$$Z[\text{bark}] = 7 \operatorname{arcsinh} \left( \frac{f_0}{650} \right)$$

# LARGURA DE BANDA RETANGULAR EQUIVALENTE (ERB — *EQUIVALENT RECTANGULAR BANDWIDTH*)

Na determinação mais precisa das CB's o trabalho de Moore e Patterson<sup>1</sup> cunhou o termo de Largura de Banda Retangular Equivalente, ERB na sigla em inglês.

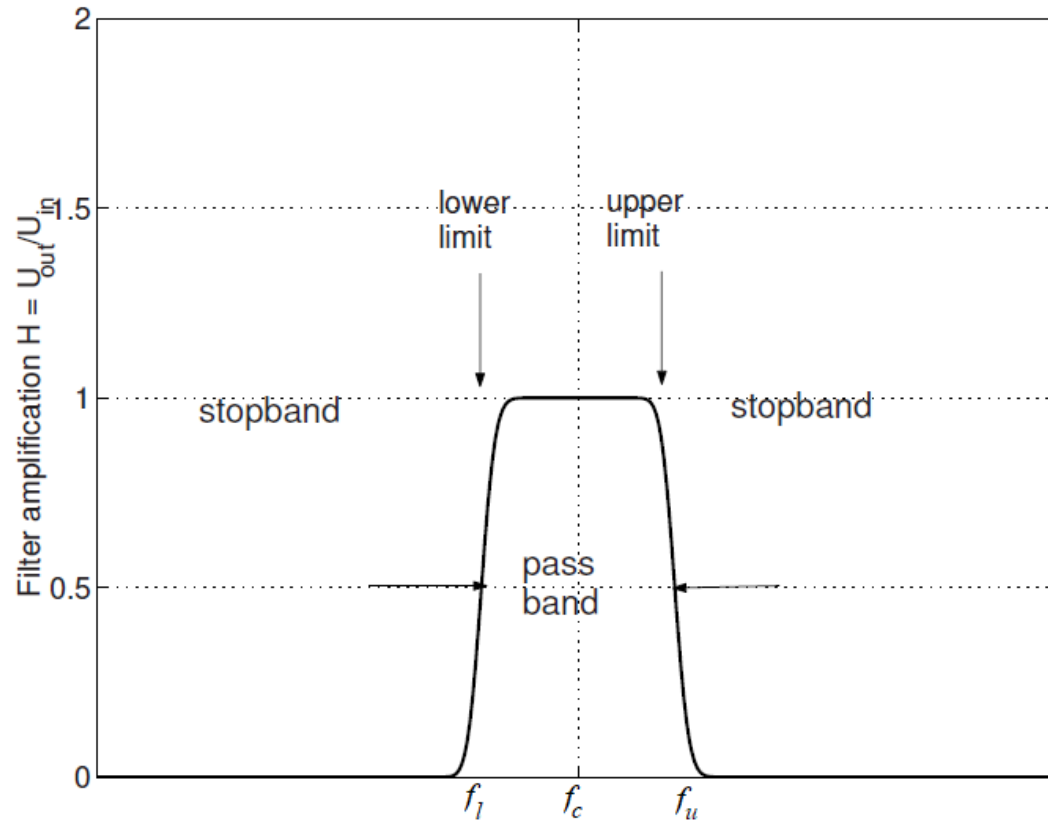
Na figura temos a representação da escala Z em barks, as ERBs e a posição na membrana basilar, sendo que 1 bark = 100 mel

$$ERB = 21.4 \log(0.00437 f_0 + 1)$$



[1] B.C.J. Moore, R.D. Patterson: *Auditory Frequency Selectivity* (Plenum, New York 1986)

# FILTROS PASSA BANDA\* EM OITAVA E 1/3 DE OITAVA



$$f_c = \sqrt{f_l f_u}$$

**Largura de Banda em Oitava:**  $f_u = 2f_l$

Logo:  $f_c = 2f_l$ ,

$$\Delta f = f_u - f_l = f_l = f_c / \sqrt{2}$$

**Largura de Banda em 1/3 Oitava:**

$$f_u = \sqrt[3]{2} f_l = 1.26 f_l$$

$$f_c = \sqrt[6]{2} f_l = 1.12 f_l$$

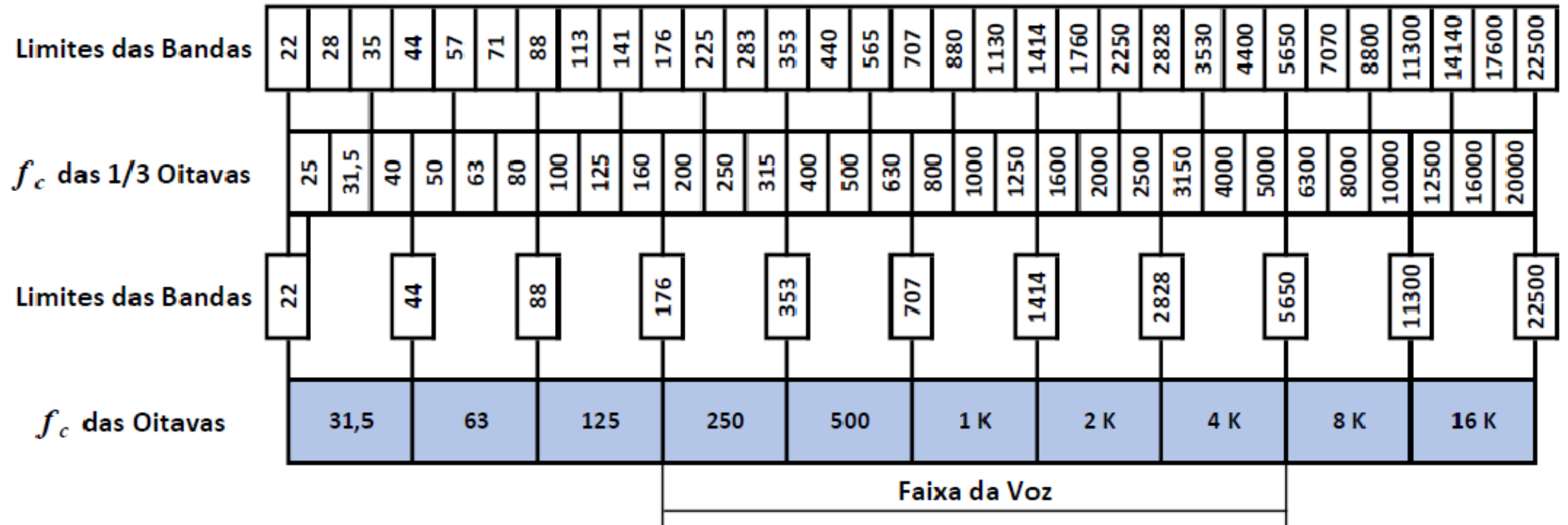
$$\Delta f = 0.26 f_l$$

(\*) **N = 11** frequências definidas na IEC 61672:

$f_c = 16, 31.5, 63, 125, 250, 500, 1k, 2k, 4k, 8k, 16kHz$

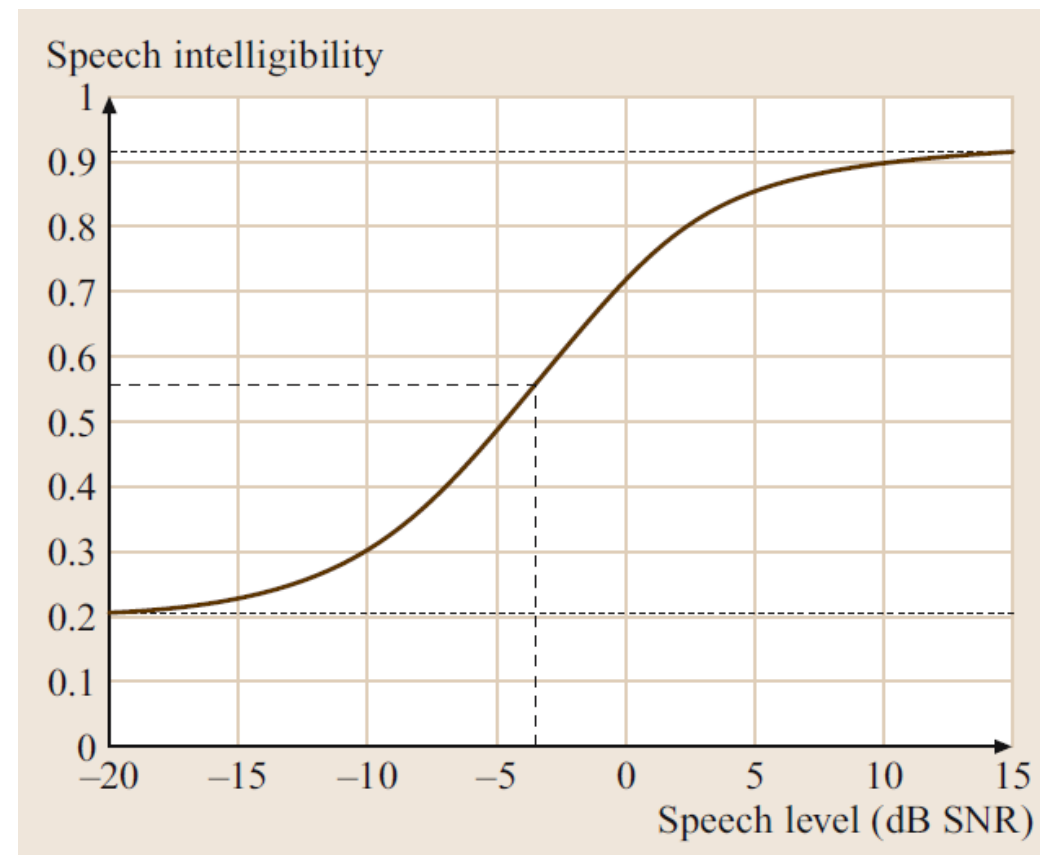


# FAIXA AUDÍVEL EM OITAVAS E 1/3 OITAVAS



# INFORMAÇÕES ACÚSTICAS PARA PERCEPÇÃO DE VOZ

**Inteligibilidade da Voz – SI (Speech Intelligibility):** é a proporção de itens de voz (p.ex. sílabas, palavras ou sentenças) corretamente repetidas por um (ou vários) ouvinte(s) para um dado teste de inteligibilidade da voz.



Exemplo de função **SI** (*linha sólida*) para teste de inteligibilidade de palavras (com cinco alternativas). *Linhas tracejadas*  $L_{\text{med}}$ . As *linhas pontilhadas* são o limite inferior ( $1/A = 0.2$ ) e o máximo assintótico  $SI_{\text{assintot}}$  da função **SI**. Parâmetros:  $L_{\text{med}} = -3.5$  dB **SNR**,  $SI_{\text{max}} = 0.9$  ( $SI_{\text{assintot}} = 0.92$ ),  $A = 5$ , inclinação =  $0.05/\text{dB}$  ( $s = 3.6$  dB)

# AJUSTE DO FUNÇÃO LOGÍSTICA DE SI(L) PARA UM CONJUNTO DE DADOS EMPÍRICOS

$$SI(L) = \frac{1}{A} \left( 1 + SI_{max} \frac{A - 1}{1 + \exp\left(-\frac{L - L_{med}}{s}\right)} \right)$$

Onde:

$L_{med}$ : nível de voz do ponto médio da função de inteligibilidade

$s$  parâmetro de inclinação, sendo  $s = \frac{SI_{max}(A-1)}{4A}$

$SI_{max}$ : parâmetro para máxima inteligibilidade ( $<1$ )

O máximo assintótico de SI será  $SI_{max} + (1-SI_{max})/A$ .

$A$  é o número de alternativas da resposta. P.ex. para dígitos 0 a 9,  $A = 10$ .

$\sigma_{SI} = \sqrt{\frac{SI(1-SI)}{n}}$ , pois SI tem distribuição binomial

# FATORES QUE INTERFEREM NA INTELIGIBILIDADE DA VOZ

## ➤ Método de medição

- AI

ANSI: *Methods for the calculation of the **articulation index***, ANSI S3.5-1969 (American National Standards Institute, New York 1969)

- SII

ANSI: *Methods for calculation of the **speech intelligibility index***, ANSI S3.5-1997 (American National Standards Institute, New York 1997)

- STI

IEC: Sound system equipment – Part 16: Objective rating of speech intelligibility by **speech transmission index**. INTERNATIONAL STANDARD 60268-16.

## ➤ Acústica da Sala e ruído

## ➤ Fenômeno Cocktail Party

- A. Bronkhorst: **The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions**, *Acustica* **86**, 117–128 (2000)

# SPEECH INTELLIGIBILITY INDEX (STI)

In auditoria used for speech, such as lecture halls or theaters, the influence of the acoustics on intelligibility is a major issue

A common way to assess objectively *speech intelligibility* in rooms is by measurement of the Speech Transmission Index STI:

- This measure is based on the idea that speech can be regarded as an amplitude-modulated signal in which the degree of modulation carries the speech information. If the transmission path adds noise or reverberation to the signal, the degree of modulation in the signal will be reduced, resulting in reduced intelligibility.
- The modulation transfer is tested by emitting noise in seven octave bands, each modulated with 14 different modulation frequencies as listed in the table of next slide and then calculating the ratio between the original and the received degree of modulation, the modulation reduction factor, in each of these 98 combinations.

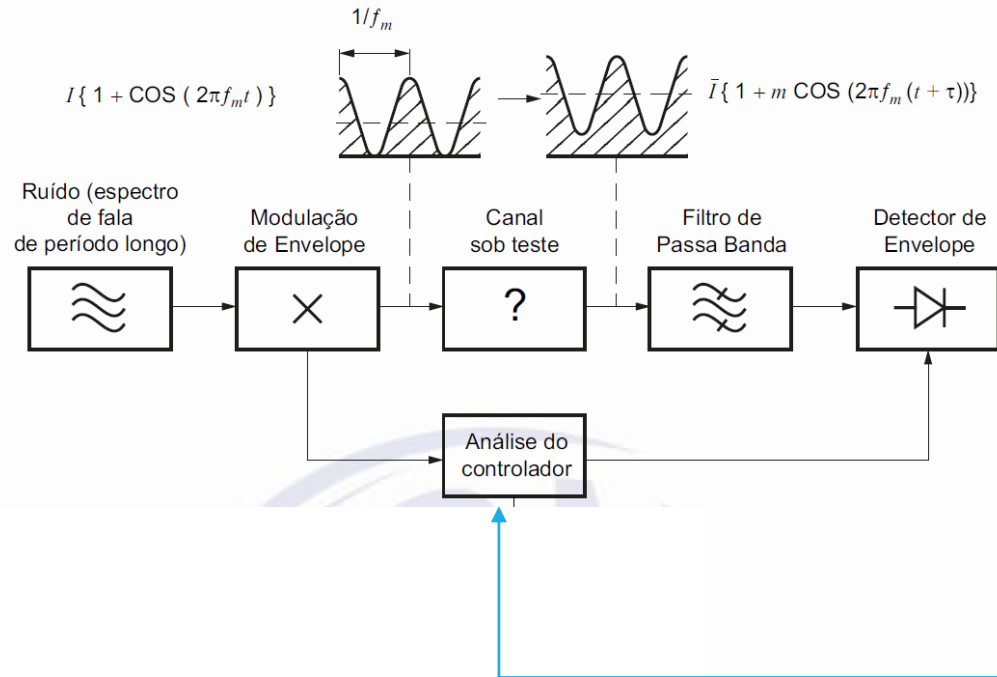
A weighted average of the modulation reduction factor then results in a number between 0 and 1, corresponding to very poor and excellent conditions respectively.

PS.: Other (and easier) measure is the ‘centre time’  $t_s$ .



# STI COMPLETE METHOD

(Ref.: ABNT IEC 60286-16: 2018)



Banda de oitava Hz	125	250	500	1k	2k	4k	8k
$f_1 = 0,63 \text{ Hz}$	$m$						
$f_2 = 0,8 \text{ Hz}$							
$f_3 = 1,0 \text{ Hz}$							
$f_4 = 1,25 \text{ Hz}$							
$f_5 = 1,6 \text{ Hz}$							
$f_6 = 2,0 \text{ Hz}$							
$f_7 = 2,5 \text{ Hz}$							
$f_8 = 3,15 \text{ Hz}$							
$f_9 = 4,0 \text{ Hz}$							
$f_{10} = 5,0 \text{ Hz}$							
$f_{11} = 6,3 \text{ Hz}$							
$f_{12} = 8,0 \text{ Hz}$							
$f_{13} = 10 \text{ Hz}$							
$f_{14} = 12,5 \text{ Hz}$							
$\bar{I}_k$							

NOTA O valor da função de transferência de modulação ( $m$ ) é determinado para todas as células da matriz de sete bandas de frequências de uma oitava e 14 frequências de modulação. Além disso, os níveis de intensidade de oitava ( $I_k$ , igual ao quadrado dos níveis de pressão sonora) são obtidos para uso no cálculo dos efeitos de mascaramento auditivos.

# QUESTÕES PARA A PROVA P2D

***OBS.: PODE SER FEITA EM GRUPO DE ATÉ 5 ALUNOS, MAS CADA UM DEVE SUBMETER INDIVIDUALMENTE, CITANDO OS COLEGAS QUE COLABORARAM.***

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:  $m = \frac{1000}{\log 2} \log \left( 1 + \frac{f_0}{1000} \right)$

b) Escala Bark:  $Z[bark] = 7 \operatorname{arc} \sinh \left( \frac{f_0}{650} \right)$

c) Escala ERB:  $ERB = 21.4 \log(0.00437 f_0 + 1)$

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

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.

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

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).

# REFERÊNCIAS

- EVEREST, F. A. **Master Handbook of Acoustics**, 4th Ed., McGrawHill, 2001.
- TOOLE, F., **Sound Reproduction, Loudspeakers and Rooms**, Focal Press, 2009.
- Benesty, J., Sondhi, M. M. and Huang, Y., **Springer Handbook of Speech Processing**, ch. 4, Springer, 2007.
- Harrington, J. and Cassidy, S., **Techniques in Speech Acoustics, Springer-Science-Business Media**, ch.2, 1999.
- Allen, J. Anthony. **Music Theory for Electronic Music Producers: The producer's guide to harmony, chord progressions, and song structure in the MIDI grid**. Slam Academy, 2018.