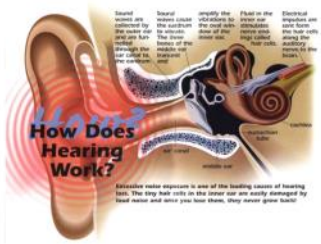
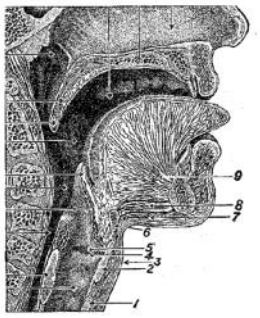


Infocommunication
Sound, hearing and speech

Tamás Csapó
<csapot@tmit.bme.hu>



Copyright

- This lecture material was created by Tamás Gábor CSAPÓ from the Budapest University of Technology and Economics. Using the materials without explicit permission is considered copyright infringement.

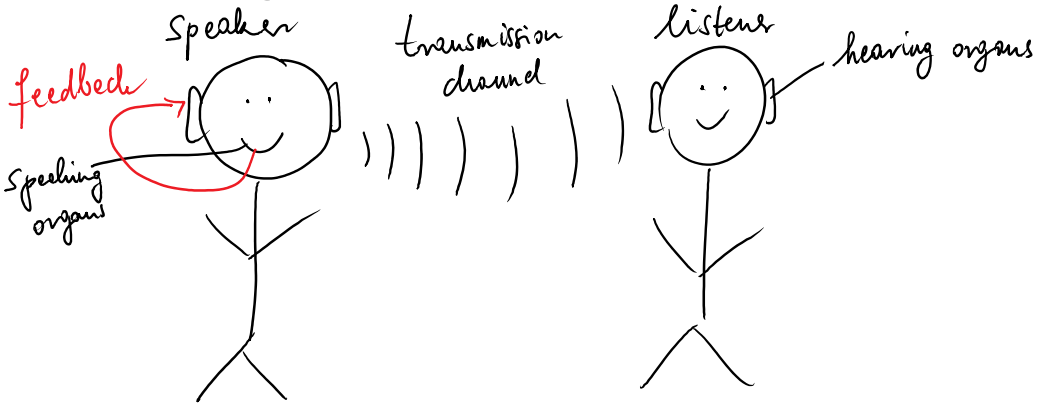
2

Topics

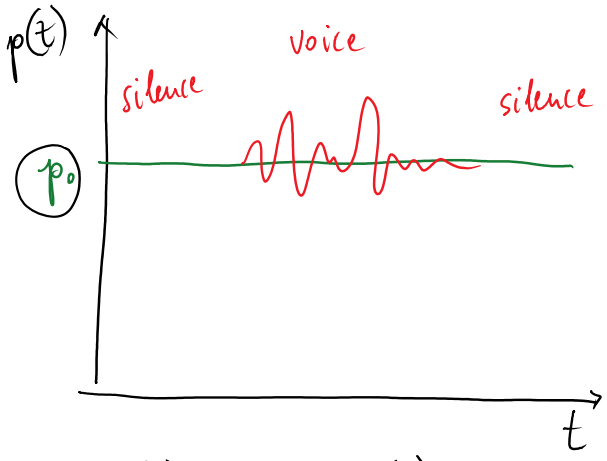
- Basic signal processing
- Sampling and quantization
- Analog modulation
- Digital baseband modulation
- Digital carrier modulation
- Error Detection Coding
- Error Correction Coding
- Radio, guided waves

- **Sound, hearing and speech**
- Light and vision
- Radio Communication
- Video Broadcasting
- GSM, Mobile communication

Natural speech communication chain



Physical modelling of sounds



Sound . mechanical vibration of an elastic medium

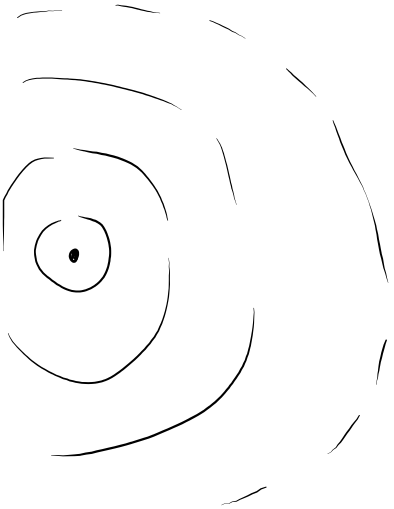
Solid — liquid — gaseous air

$P(t) = p_0 + p(t)$

Sound pressure change

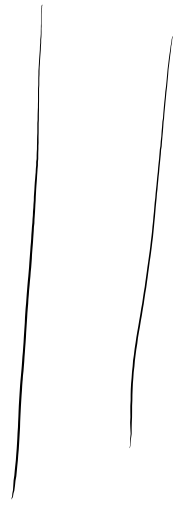
$p = \frac{F}{A}$ $\left[\frac{N}{m^2}\right] = [Pa]$ $\sim \frac{1}{r}$ decreasing

Propagation in short distance / long distance



spherical sound waves

$c_{sound} = f \cdot \lambda = 340\ m/s$

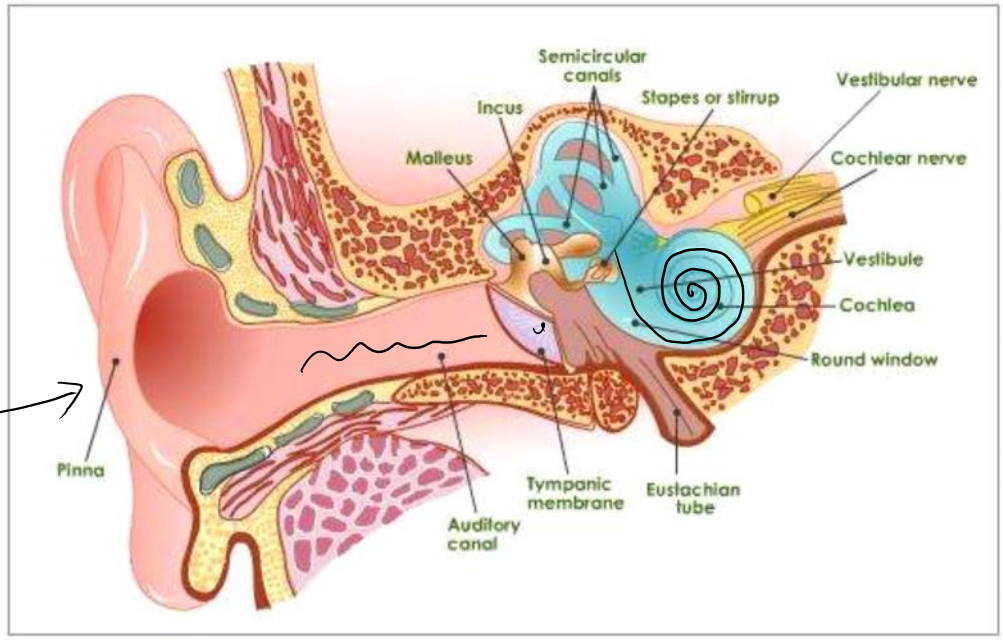


plain waves

$\frac{p}{v} = \rho_0 \cdot c =$

$410\ \frac{kg}{m^2 s}$

Structure of the human ear



Source: http://www.tutorvista.com/content/science/science-i/sound/structure-ear.php

Békésy György / Georg von Békésy

Nobel prize in 1961 (function of the cochlea)

Source: http://braintour.harvard.edu/archives/portfolio-items/von-bekesy-experiments-hearing

- Intensity: $I = \frac{P}{A} \left[\frac{W}{m^2}\right] \sim \frac{1}{r^2}$ decreasing
- Volume: (sound pressure level)

$$L = 20 \cdot \lg \frac{p}{p_0} = 10 \cdot \lg \frac{I}{I_0} \; [dB]$$

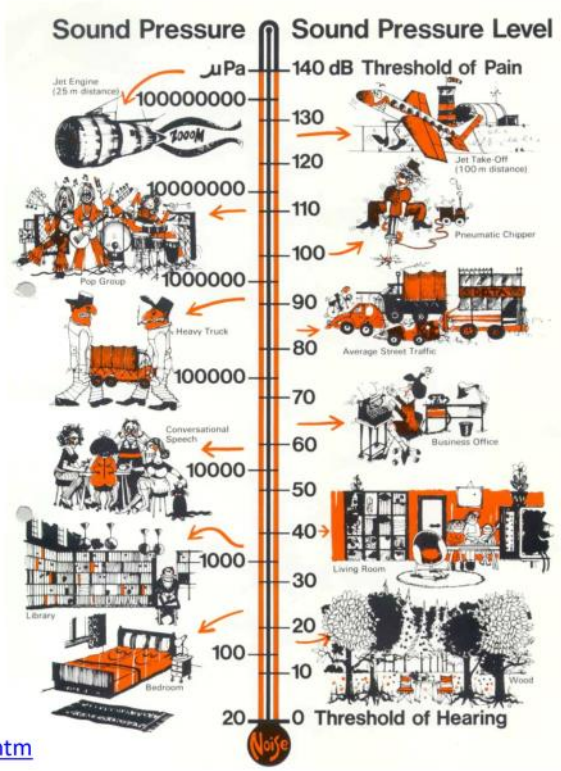
$p_0 = 2 \cdot 10^{-5}$ Pa reference $I_0 = 10^{-12}$ W/m²

- SPL:

↓

Sound pressure relative to the reference in acoustic decibels

Sound pressure level



Source:
http://personal.cityu.edu.hk/~bsapplec/sound.htm

Sound pressure level

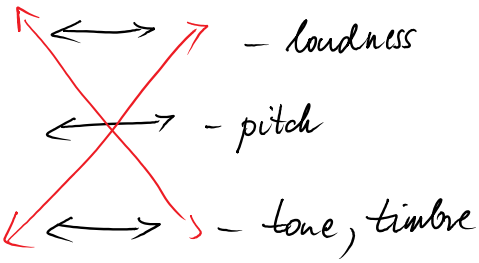
- 440 Hz tone (A4 on musical scale)
  - reduced in 1 dB steps
  - reduced in 3 dB steps
  - reduced in 5 dB steps

Source: http://www.ece.uvic.ca/~elec499/2003a/group09/p/demos.htm

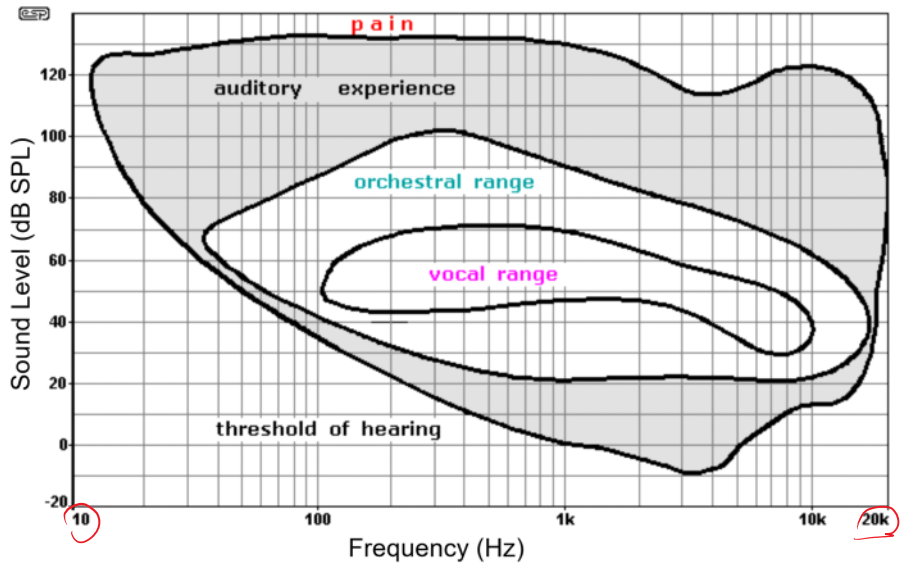
Physiological & psychoacoustical properties of hearing

- Objective parameters
  [can be measured]
  - pressure / intensity level
  - fundamental frequency / period
  - spectrum

Subjective parameters
[can be perceived]



Limits of human hearing



Source: https://people.ece.cornell.edu/land/courses/ece5030/FinalProjects/s2014/kkp37_rjs483/kkp37_rjs483/AudioGram.html

Equal loudness level contours

FM contours

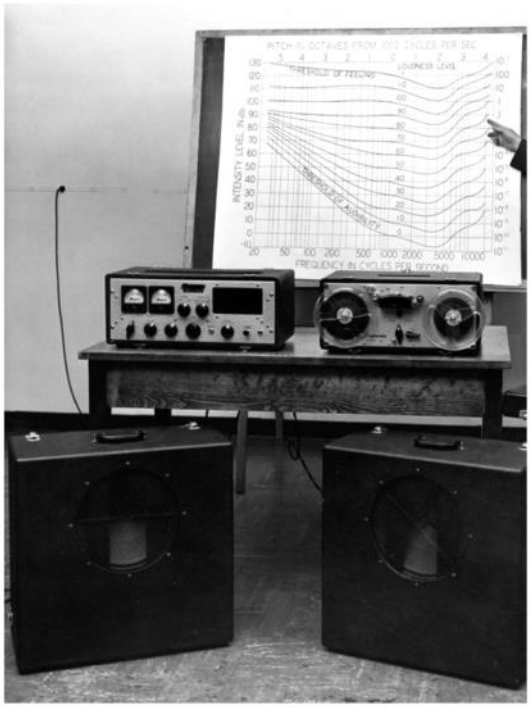
- Def: loudness level of an arbitrary sound is as many phons as many dB SPL of the 1 kHz sinusoidal sounds

1933, Fletcher & Munson

1956, Robinson & Dadson

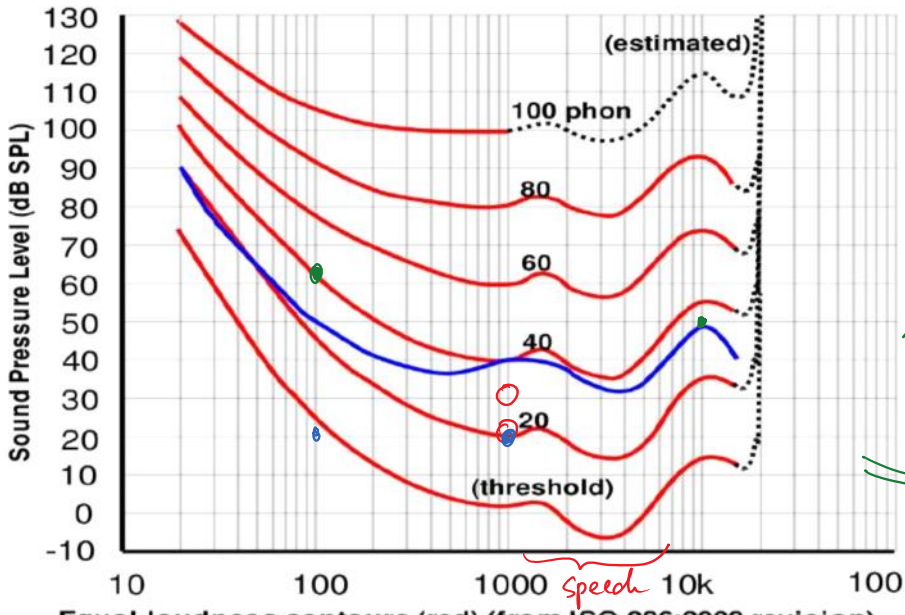
1961, ISO standard

Fletcher & Munson 1933 experiment



Source: http://www.effectrode.com/wp-content/uploads/fletcher_munson_chart.jpg

Equal loudness contours



Equal-loudness contours (red) (from ISO 226:2003 revision)
Original ISO standard shown (blue) for 40-phons

Source: OP

Examples which is louder

1 kHz, 20 dB < 1kHz, 30dB

1kHz, 20dB (20 phon) > 100Hz, 20dB (0phon)

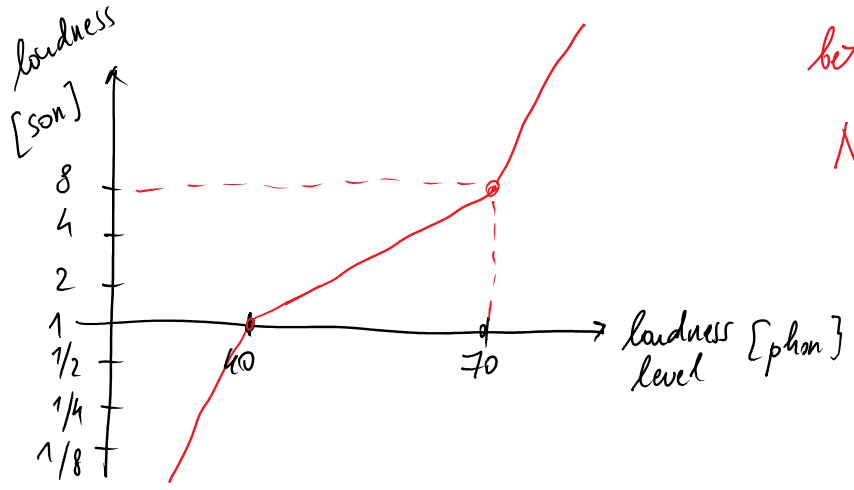
100Hz, 40phon > 10kHz 50dB 37phon

⟹ dB ⟷ phon conversion

f [Hz]

Loudness

perceptual correlate of intensity



between 40–70 phon

$N = 2^{\frac{L_N - 40}{10}}$

relative loudness, loudness growth

40 phon → 1 son

70 phon → 8 son

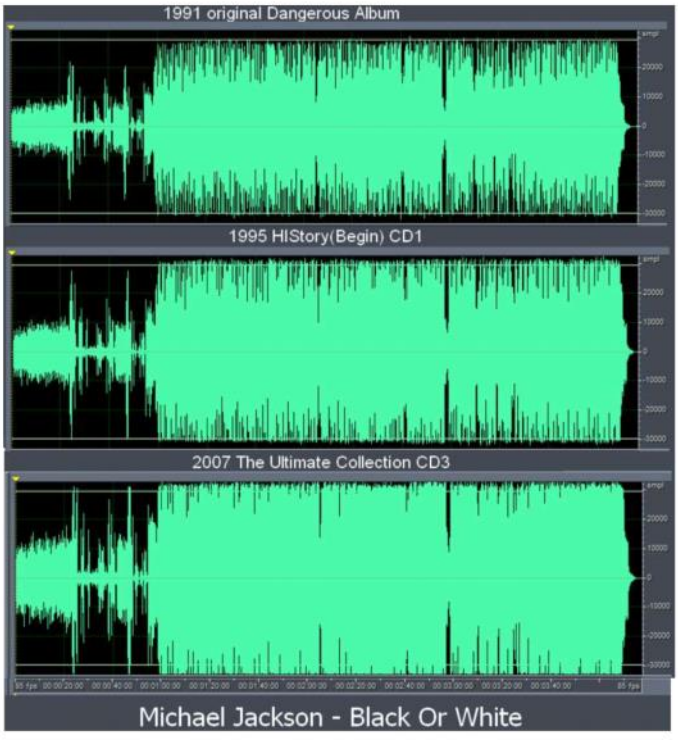
Loudness

- various frequencies at a constant SPL
  (the perceived loudness of tones varies at equal sound intensity)

- which tone sounds twice as loud as the reference tone?
  - reference tone + same tone 5 dB higher
  - reference tone + same tone 8 dB higher
  - reference tone + same tone 10 dB higher

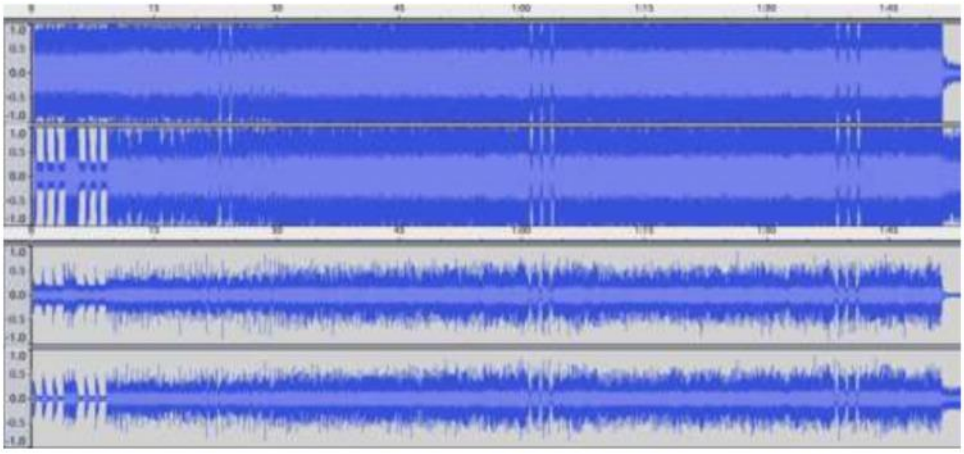
Source: http://www.ece.uvic.ca/~elec499/2003a/group09/p/demos.htm

Loudness war



Source:
https://en.wikipedia.org/wiki/Loudness_war

Loudness war
Metallica: Death Magnetic



Source:
https://en.wikipedia.org/wiki/Loudness_war

Loudness war

Source: https://www.youtube.com/watch?v=3Gmex_4hreQ

Loudness war

Some of the albums that have been criticized for their sound quality include the following:

Artist	Album
Arctic Monkeys	*Whatever People Say I Am, That's What I'm Not*[6]
Black Sabbath	*13*[57]
Bob Dylan	*Modern Times*[40]
	Together Through Life[40]
Christina Aguilera	*Back to Basics*[3]
The Cure	*4:13 Dream*[58]
Depeche Mode	*Playing the Angel*[59]
The Flaming Lips	*At War with the Mystics*[6][note 3]
Led Zeppelin	*Mothership*[60]
Lily Allen	*Alright, Still*[61]
Los Lonely Boys	*Sacred*[3]
Nine Inch Nails	*Pretty Hate Machine (2010 Remaster)*[62]
Metallica	*Death Magnetic*[63][note 4]
Miranda Lambert	*Revolution*[64]
Oasis	*(What's the Story) Morning Glory?*[6]
Paul McCartney	*Memory Almost Full*[65]
Paul Simon	*Surprise*[66]
Pearl Jam	*Ten* (2009 remaster)[67][68][69]
Queens of the Stone Age	*Songs for the Deaf*[6]
Red Hot Chili Peppers	*Californication*[3][6]
Ghost	*Infestissumam*[70]
Rush	*Vapor Trails*[71]
The Stooges	*Raw Power* (1997 remaster)[66]

Source: https://en.wikipedia.org/wiki/Loudness_war

Dynamic range compression artistic effect

Listen at around 0:43 for the bass drum; you'll heard the rest of the track's volume drop.

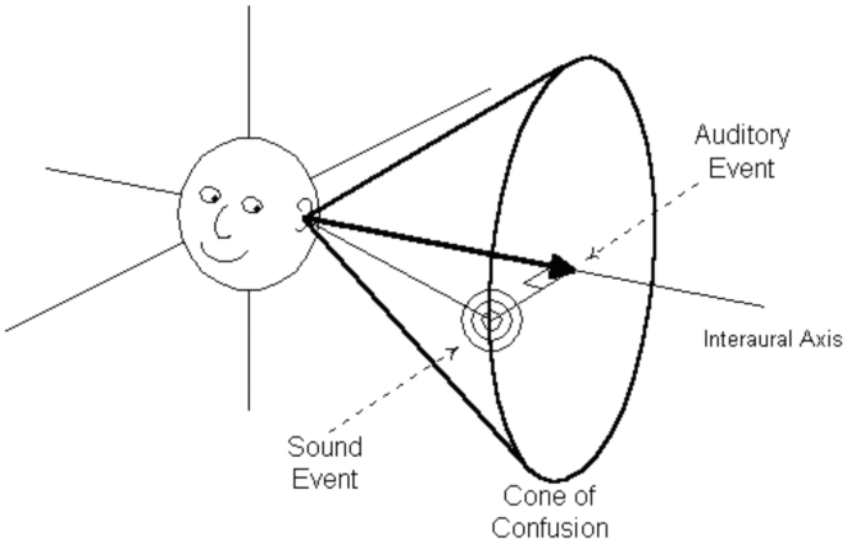
Source: http://www.howtogeek.com/57903/htg-explains-how-does-dynamic-range-compression-work/
https://www.youtube.com/watch?v=RIZdjT1472Y

Spatial parameters, direction of sound

– horizontal plane
  - low freq. (wavelength < size of head) 50–1600Hz
    phase difference
  - high freq. (w.l. > size of head) 1600Hz –...)
    intensity difference

– vertical plane

Cone of confusion



Source: http://www.music.miami.edu/programs/mue/Research/jwest/Chap_2/Chap_2_Spatial_Hearing.html

Spatial hearing
„hearing throne”

Source: https://auditoryneuroscience.com/book/export/html/15

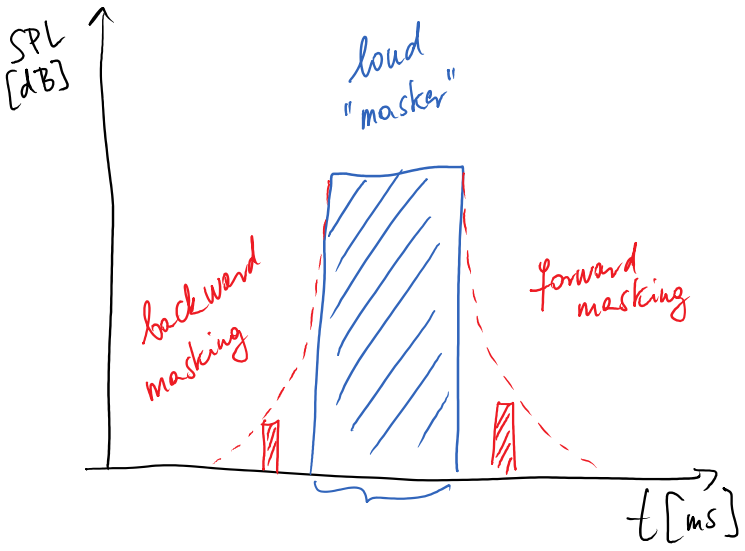
~~Critical bands /1~~

Critical bands /2

Masking

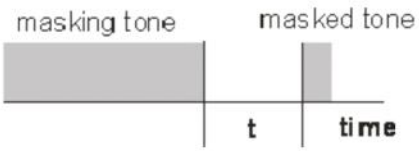
- Time domain masking
- Frequency domain masking
- Directional masking

Time domain masking



a) forward: several 10ms
b) backward: several ms

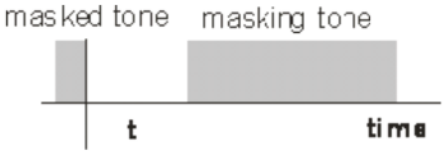
Time domain masking - Forward



- masking tone + tone that is semitone down
  - with a 100 ms delay in between
  - with a 10 ms delay in between

Source: http://www.ece.uvic.ca/~elec499/2003a/group09/p/demos.htm

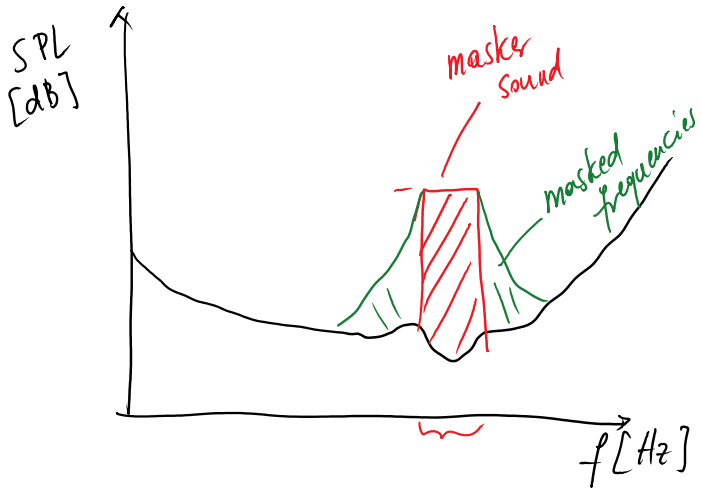
Time domain masking - Backward



- initial tone is going to be masked by the tone that follows
  - delay: 100 ms
  - delay: above 10 ms
  - delay: below 10 ms

Source: http://www.ece.uvic.ca/~elec499/2003a/group09/p/demos.htm

Frequency domain masking

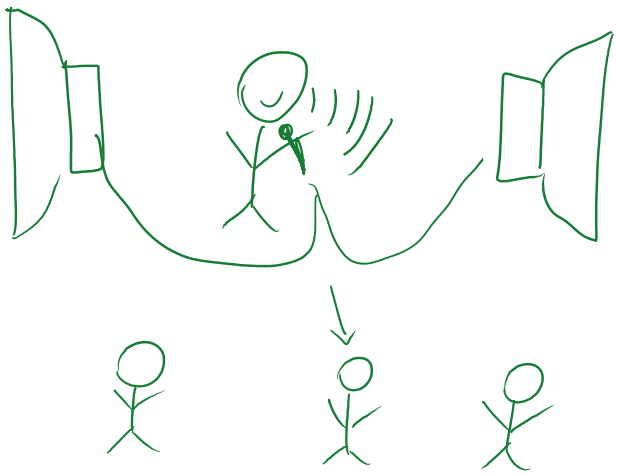


Frequency domain masking

- Pure tones mask higher frequencies better than lower frequencies
  - Mask high freqs
  - Mask low freqs

Source: http://www.ece.uvic.ca/~elec499/2003a/group09/p/demos.htm

Directional masking

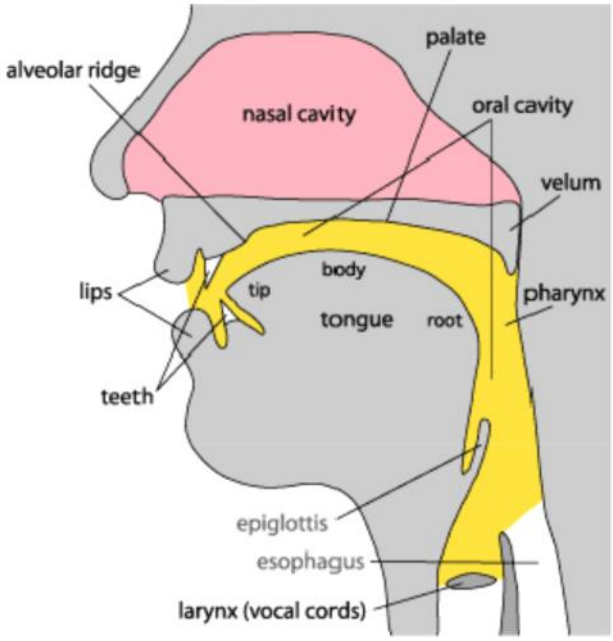


37

~~Sum volume of sinusoidal sounds~~

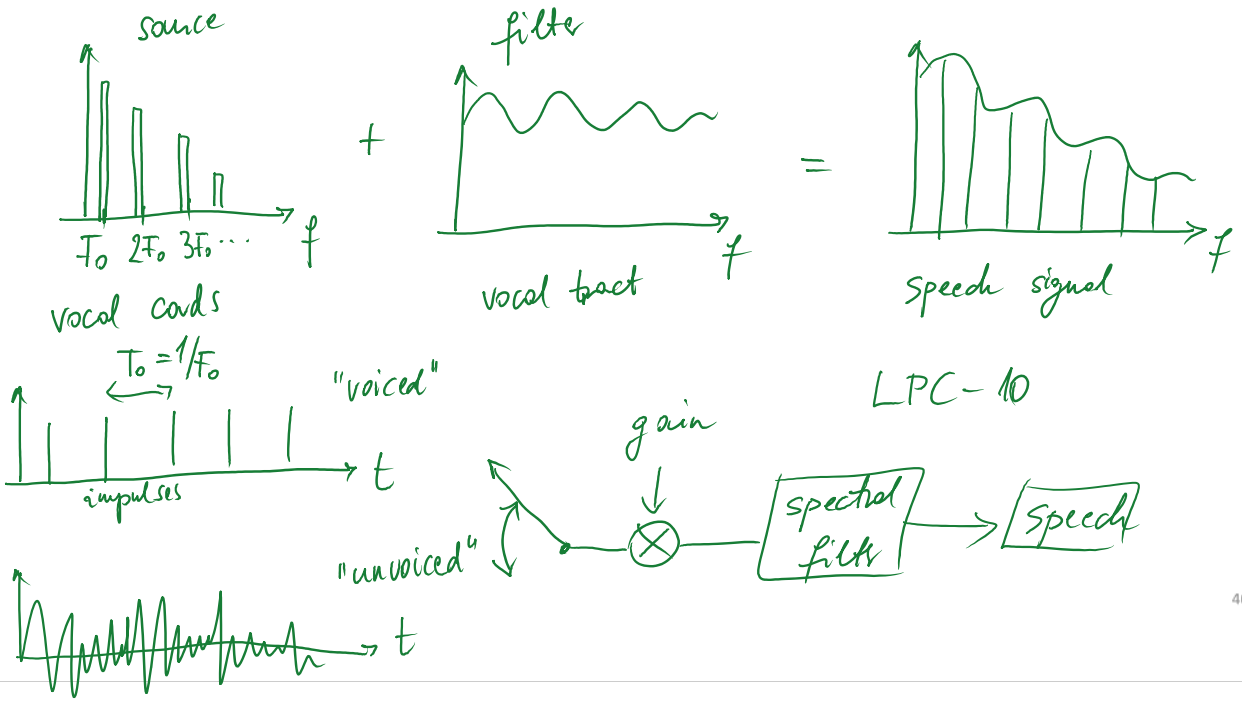
38

The organs of speech



Source: http://www.englishbaby.com/lessons/3201/member_submitted/vocal_organs_of_speech

Source-filter model of speech production



Waveform coding vs Speech coding

- sampling & quant.
- any kind of signal
- store every sample
- e.g. f_s = 8 kHz) & bit quantization

64 kbit/s

20–30× compression

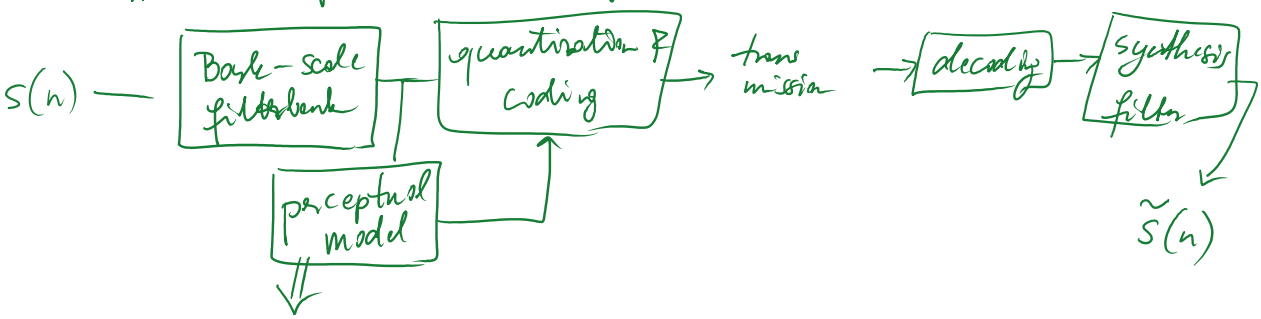
LPC-10

GSM

- analyse speech to spectral parameters
- use knowledge about speech production
- high compression as low as 2 kbit/s

Perceptual / subband coding

- masked components can be left out

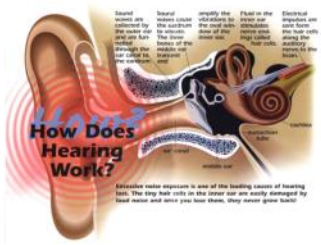


e.g. MP3, MPEG1-Layer 3

LPC

- Waveform coding
  - Original (64 kbps)
  - ADPCM (32 kbps)
- Linear Predictive Coding
  - CELP (4800 bps)
  - LPC-10 (2400 bps)

Source: http://www.data-compression.com/speech.html



The END

Infocommunication
Sound, hearing and speech

Tamás Csapó
<csapot@tmit.bme.hu>