2021

Speech Recognition (H02A6)

Dirk Van Compernolle

CHAPTER 1

Human Ear

1

Anatomy of

The Human Auditory System

2

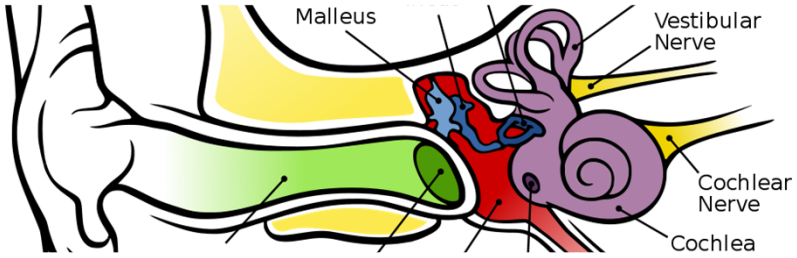
For animations, and much more on the ear …

- **Human Ear (intro for kids)**
  - https://www.youtube.com/watch?v=RiVx5Lih_44&list=TLPQMjMwOTIwMjCNyUqyTipIkw&index=5
- **Auditory Transduction (Brandon Pletsch)**
  - https://www.youtube.com/watch?v=PeTriGTENoc
  - https://www.youtube.com/watch?v=PeTriGTENoc&t=353 **(finale)**
- **Human Ear (.org) (online textbook)**
  - http://www.cochlea.eu/en/ear
- **Human Ear (Kahn Academy)**
  - https://www.youtube.com/watch?v=98-6WfdumZY

3

3

Anatomy of the Ear



Outer Ear	Middle Ear Ossicles	Inner Ear Cochlea	Cochlear Nerve, Brainstem, …
• microphone	• air-to-fluid impedance matching • protection against overload	• transduction from mechano-acoustic to electrical	• processing of electrical signals
	• non-linear compression of high intensity signals	• spectrogram like frequency analysis	• feature extraction • pattern matching

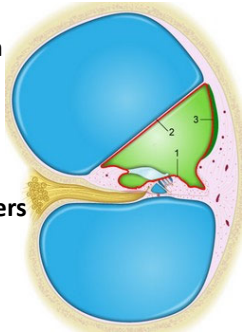
4

4

Inside the Cochlea

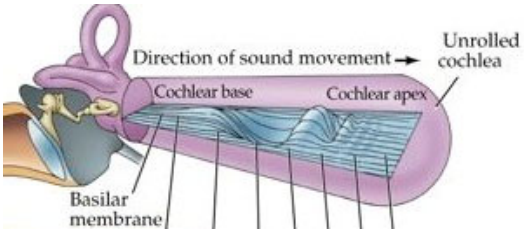
snail shaped cochlea

cross-section of a single turn



nerve fibers

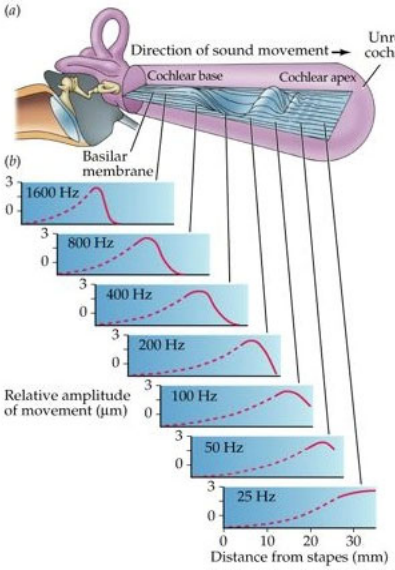
"unwrapped" as a longitudinal structure

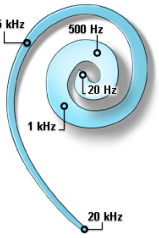


5

5

Spectral Analysis and Tonotopic Organization in the Cochlea





distance from base (cm)

0	10	20	30
20000	5000	500	50

characteristic frequency (Hz)

- basilar membrane + auditory nerve fibers behave as a filterbank with 30.000 channels
- center frequencies decrease with distance from the base = " tonotopic organization"
- high degree of redundancy (typical of all biological systems) allows for aging and trauma

6

6

Functional Processing in Inner Ear & Central Pathways

- **Inner Ear (Cochlea)**
  - **critical element in the whole processing chain**
  - **processing:**
    - a single time-domain signal is decomposed
    - in thousands of parallel channels (nerve fibers)
    - information carried by each channel is frequency dependent
    - this frequency analysis is performed by basilar membrane and hair cells
- **Higher Pathways & Central System**
  - **processing:**
    - the multi-channel input (signal on auditory nerve) goes through several stages of feature extraction
    - these features are used as input to the final recognition process

7

7

Psychoacoustics Concepts

Loudness, Pitch & Timbre

8

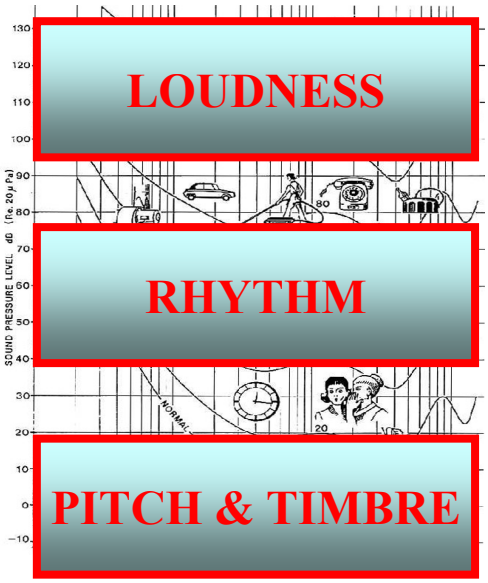
Psychoacoustics

- The theory of how the brain interprets audio signals
- The study of subjective human perception of sounds
- The study of the relationship between physical measures of sound (e.g., amplitude and frequency) and the perception of them

9

9

Perception and Intuition



- audible
- loud vs. soft
- bursty, repetitive, rhythmic
- sudden, constant
- melodical, not tonal
- high vs. low tones

10

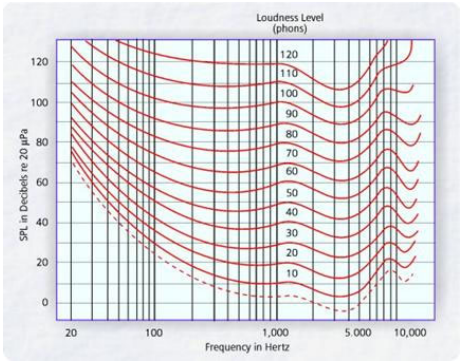
10

Perception of pure tones (sine waves)
Pitch & Loudness

- **Pitch = tonal percept**
  - **ability to rank on a tonal scale**
  - **directly linked to the physical frequency**
  - **range: 20Hz – 20kHz**
  - **frequency discrimination abilities drop logarithmically above 1kHz**
- **Loudness = intensity percept**
  - **ability to rank on a loudness scale**
  - **physical measure: sound pressure level = SPL (expressed in dB)**
  - **a frequency dependent mapping from SPL to the perceptual scale**
  - **intensity range (in mid-frequency range): > 100dB SPL**

$$I = 20\log\left(\frac{A}{A_0}\right)$$

Equal Loudness Curves



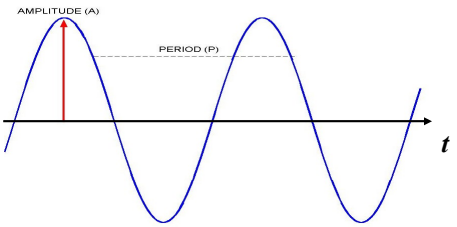
11

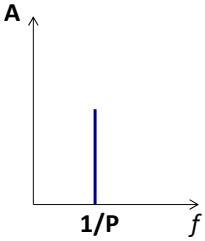
11

Time and Frequency Domains
sine wave = single frequency

$$x(t) = A\sin(2\pi f_0 t)$$

Time Domain *Waveform*	⟺ Fourier Transform	*Frequency Domain* *Spectrum*





f_0 = frequency = 1/P (expressed in Hz = number of cycles per second)

A = Amplitude

12

12

Fourier Series of Periodic (Harmonic) Signals

Any periodic signal with period *P* can be written as a sum of harmonics with fundamental frequency $f_0=1/P$

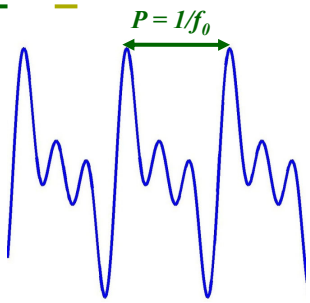
$$x(t)=\sum_k A_k \sin(2\pi k f_0 t+\Phi_k)$$

f_0 = fundamental frequency *(=1/P)*

k = harmonic index

$\{A_k\}$ = amplitude spectrum

$\{\Phi_k\}$ = phase spectrum



The FOURIER SERIES is a form of the Fourier Transform that applies to harmonic signals and allows for intuitive interpretations

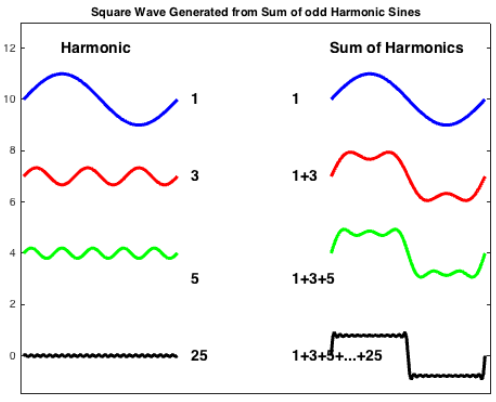
13

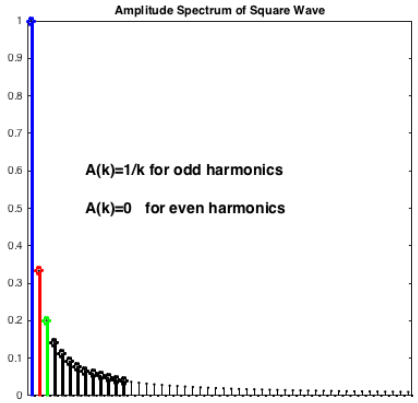
13

Example: Square Wave

$$x(t)=\sum_{k=1,3,5,\ldots}(1/k)\sin(2\pi k f_0 t)$$

$$P=1/f_0 \qquad A_k=1/k \quad \text{k = odd}$$

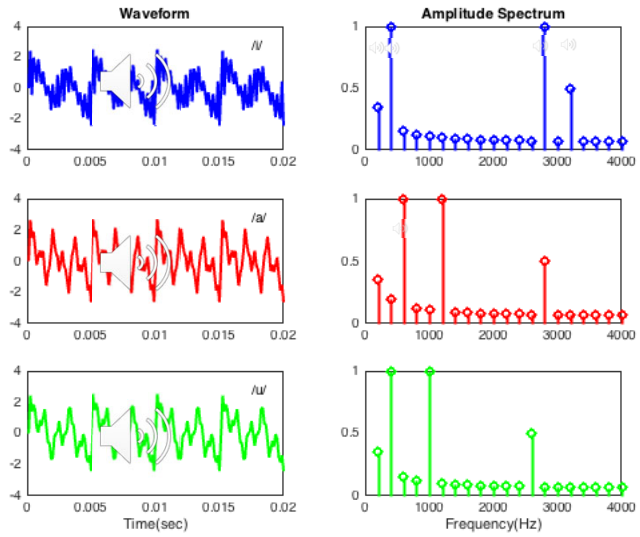




14

14

Perception of Harmonic Complexes
Pitch, Timbre



These sounds have

the same fundamental: on a musical scale they are perceived as the same note (=periodicity =pitch =200Hz)

the same harmonics, but they come with very different amplitudes: so they are sounds with a very different timbre (related to spectral envelope)

15

15

Perception of Harmonic Signals

Perception[Sum(Harmonics)]

≠

Sum(Perceptions[Harmonic])

- **We do not hear the harmonics in the complex in an analytic/independent way**
- **Perception of the complex is based on 'group' properties**

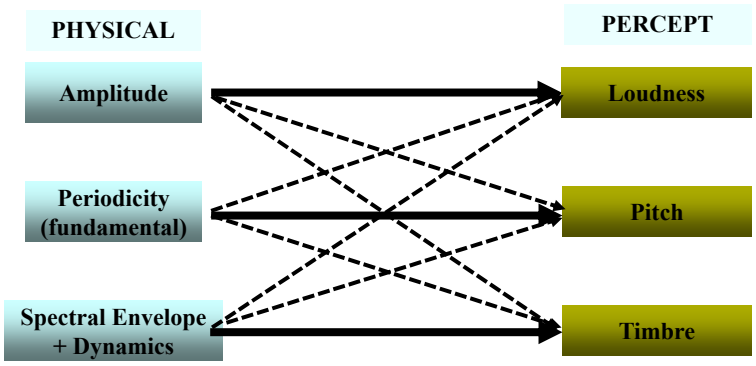
PHYSICAL PROPERTIES	PERCEPTION
• Fundamental frequency	• Pitch = "tonal" percept
• Amplitude spectrum – integrated energy – shape	• Loudness • Timbre, sound quality

16

16

Psychoacoustics of Complex Signals

- **Short Term Properties**



- **Long Term Properties**
  - **Duration**
  - **Rhythm**

17

17

Frequency Perception

- **Frequency Range**
  - **Full range of human hearing: 20Hz-20kHz**
  - **Essential for day to day voice communication: 300-3400Hz (Telephone!)**
- **Pitch**
  - **tonal percept of melodic sounds**
  - **highly adapted to the human voice range**
    - 50-150Hz for male, 200-300Hz for female, 400Hz for children
- **Rhythm**
  - **< 20Hz**
- **Timbre**
  - **typically we do NOT hear individual frequency components**
  - **the overall shape of the amplitude spectrum (full range) is a major contributor to the 'timbre' percept**
  - **temporal properties play an important role as well**
  - **" ... that attribute of auditory sensation in terms of which a listener can judge that two sounds with the same loudness and pitch are different .. "**

18

18

Rhythm

- What happens to the frequencies below 20Hz ?
- Frequencies >20Hz
  - contribute to frequency perception (pitch, timbre)
- Frequencies <10Hz
  - contribute to temporal perception (rhythm, isolated events, ..)
- Questions:
  - How many notes can a musician play per second
  - How many separate notes per second can you hear, before everything blurs together (+- 10) ?

19

What happened to the phase ?

- Our ears are 'phase-deaf' -- (almost)
- Frequency and Amplitude almost completely dominate the perception
- Phase has only a minimal impact on speech perception
- Reverberation has great impact on phase
  - this is primarily perceived as an impact on "sound quality"
  - limited reverberation has no impact on speech understanding
  - strong reverberation (reverberation times > duration of single phonemes) can have a detrimental impact on speech understanding as consecutive sounds may now be heard simultaneously

20

Towards the perception of Time-varying Signals: Thinking "Time-Frequency"

- Observations
  - acoustic signals vary over time
  - speech is a sequence of sounds
- A single spectrum
  - captures the properties of stationary sounds
  - can not capture the transient nature of most sounds or represent sound sequences
- Perception of time-varying signals
  - ~ a complex combination of frequency domain and time domain properties
  - ~ a sequence of short-time spectra

➔ Short-Time Fourier Analysis, Spectrogram
➔ Speech Perception
➔ Speech Recognition

21

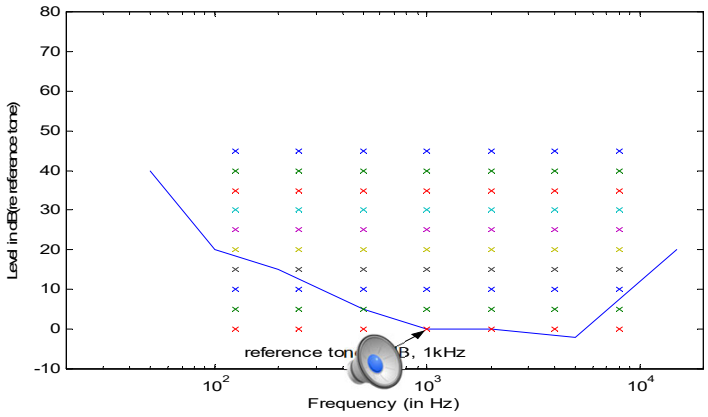
Auditory Demo 2
Loudness of Broadband signals

- **Broadband noise played at various levels of intensity**
  - 10 * 6dB steps (IPO-CD track8))
  - 20 * 1dB steps (IPO-CD track 10))
- **Speech at various distances from a microphone**
  - **distances: 25cm, 50cm, 100cm, 200cm (IPO-cd track 11)**
  - **REMARK: this is with an omni-directional microphone in an anechoic room !!**
- **As long as the energy is well distributed over the whole auditory spectral range, intensity and loudness are correlated well in a very similar manner as holds for simple tones**

23

Auditory Demo 1
Loudness Perception of Pure Tones

Pure Tones played in 10 * 5dB decreasing levels



24

Auditory Demo 3
Timbre

- **PERCEPT:**
  - **“ ... that attribute of auditory sensation in terms of which a listener can judge that two sounds with the same loudness and pitch are different .. “**
  - **complex sound quality, … difficult to describe**

- **Effect of SPECTRUM on Timbre**
  - **Strike note of an instrument = +- pitch**
  - **The timbre is largely dominated by the spectral envelope, i.e. by how much of which harmonic**
  - **Examples: add harmonics 1, 2, 3, 4, 5+6, 7+8, 9+10+11, 12+ for**
    - Carillon Bell
    - Guitar:

25