COMP47700
Speech and Audio

Dr Alessandro Ragano

Office: 3rd Floor, Science East, Insight Centre
School of Computer Science
University College Dublin

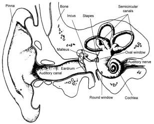
alessandro.ragano@ucd.ie

4.1.2
How We Hear:
Speech Perception

Human Auditory Perception

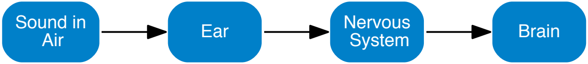
Physical Systems

- Process of Hearing
- The ear
- Frequency selectivity
- Brain interpretation



Perception

- Psychoacoustics
- Equal Loudness
- Combination tones
- Phase Locking
- Signal Processing
- Temporal Integration
- Masking
- Precedence

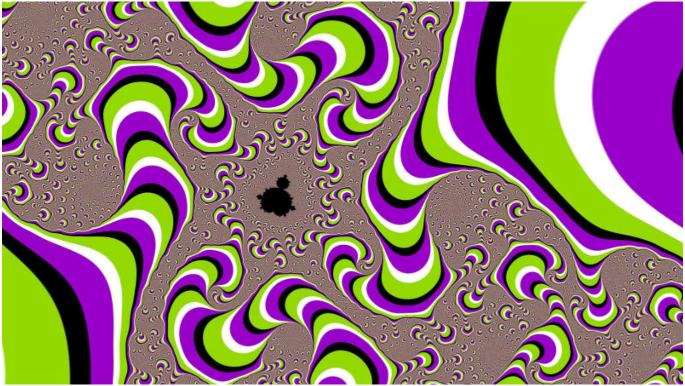


Perception

- Psychoacoustics
- Equal Loudness
- Combination tones
- Phase Locking
- Signal Processing
- Temporal Integration
- Masking
- Precedence

Psychoacoustics

What you hear may not be what you think you should hear!



Psychoacoustics

Signal + Brain

Perception is a mixture of signal and inference

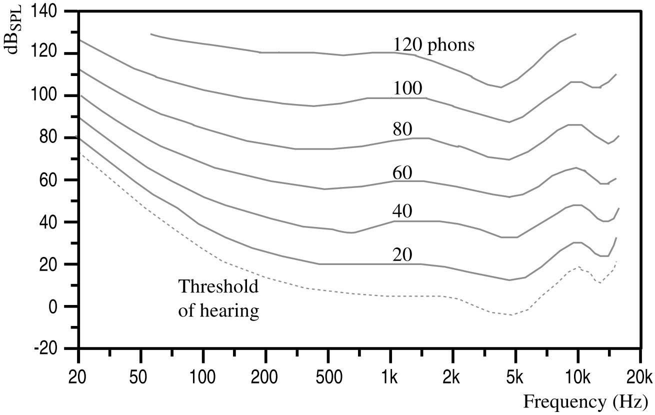
> *"I know that you believe you understand what you think I said, but I'm not sure you realise that what you heard is not what I meant"*
>
> – Robert McCloskey, 1984.

- Psychology of acoustics
- Auditory illusions occur like optical illusions
- Psychoacoustics looks at how we perceive sound
- **Why is it useful to learn?** e.g. Leads to useful analysis, compression and transformation tricks for speech and audio

Equal Loudness

- Tones of **same amplitude** but **different frequency** judged as **different loudness**
- Phons of a sound are the dB SPL of a sound at a frequency of 1 kHz that sounds just as loud
- 0 phon is the limit of perception (negative phon levels are inaudible)
- Curves are as a result of human hearing system factors (e.g. *orthotelephonic gain* from the pinna)

Equal Loudness



Combination tones

Two pure tones with frequencies f_1 and f_2 ($f_2 > f_1$).
Non-linear active processing by the auditory system.
Perceived additional tone at $f_{ct} = 2f_2 - f_1$.

Example: $2f_1 - f_2$ Tone Induction

This effect is used by musicians to create harmonies.
We can experience this by listening to a tone ($f_2 = 1800\ Hz$) and a sinusoidal sweep (quadratic chirp signal, $f_1 = 2000 \rightarrow 2200$).
We will perceive an additional downward sweeping tone ($f_{ct} = 1,600 \rightarrow 1,400\ Hz$).

Combination Tones Notebook: Demo should be done with loudspeakers to show mixing occurs in ears

Phase Locking

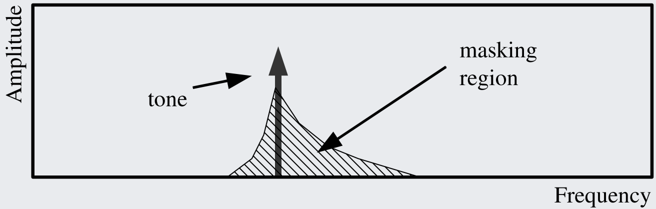
- Hair cells at any particular point on the basilar membrane have a tuned average firing frequency
- The may not fire for every cycle of a signal but the might every $2^{nd}, 4^{th}, 6^{th}$ without altering the overall firing rate (remember pushing a swing metaphor?)
- They have a **recovery period**, i.e. how long before they can fire again meaning sounds can be missed
- Also, a higher amplitude tone may **suppress** a similar frequency at a lower amplitude

Temporal Integration

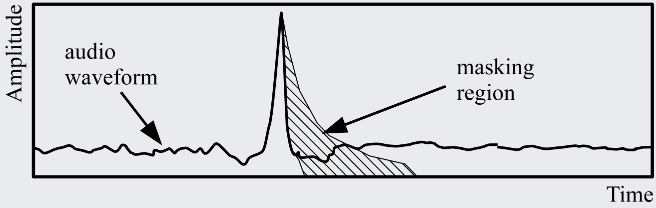
- Tones with a duration greater than approximately 500 ms are detected irrespective of duration, complexity or pattern
- The shorter the tones duration the higher the intensity required

Masking

Frequency Masking



Temporal Masking



Masking Demo: `https://www.coursera.org/lecture/internet-of-things-multimedia/audio-processing-abelQ`

Precedence Effect

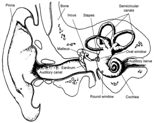
Haas or precedence effect

- Multiple similar versions of a sound reach an listener with delayed arrival times
- Listener hears the first signal but **suppresses** the subsequent versions (even if delayed versions are up to 10 dB louder)
- Only when arrival times are < 50 ms apart
- Explains how speech remains intelligible in small reverberant room

Summary

Physical Systems

- Process of Hearing
- The ear
- Frequency selectivity
- Brain interpretation



Perception

- Psychoacoustics
- Equal Loudness
- Combination tones
- Phase Locking
- Signal Processing
- Temporal Integration
- Masking
- Precedence

Sound in Air → Ear → Nervous System → Brain