Lecture 3: Auditory Perception

Part 2 of 3

Physiology Underlying Perception

Relevance (as for part 01)

Cheung, S., Han, E., Kushki, A., Anagnostou, E. and Biddiss, E., 2016. Biomusic: An auditory interface for detecting physiological indicators of anxiety in children. *Frontiers in neuroscience*, *10*, p.401.

Wu, J., Liu, Q., Zhang, M., Pan, Z., Li, H. and Tan, K.C., 2021. HuRAI: A brain-inspired computational model for human-robot auditory interface. *Neurocomputing*, *465*, pp.103-113.

Serafin, S., Geronazzo, M., Erkut, C., Nilsson, N.C. and Nordahl, R., 2018. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE computer graphics and applications*, *38*(2), pp.31-43

Zhao, Y., Bennett, C.L., Benko, H., Cutrell, E., Holz, C., Morris, M.R. and Sinclair, M., 2018, April. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-14).

Geronazzo, M., Bedin, A., Brayda, L., Campus, C. and Avanzini, F., 2016. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies*, *85*, pp.4-15.

Betlehem, T., Zhang, W., Poletti, M.A. and Abhayapala, T.D., 2015. Personal sound zones: Delivering interface-free audio to multiple listeners. *IEEE Signal Processing Magazine*, *32*(2), pp.81-91.

Learning Objectives

Provide an introduction to the structure and physiology of the auditory system.

Provide an overview of the concepts such as pitch, loudness, localisation and their relevance to auditory processing, including speech processing.

To generate an appreciation of the relevance of auditory perception research to interface design and evaluation.

Learning Outcomes

To be able to provide a description of the main areas of the auditory system relevant for the processing of auditory information.

To be able to provide a description of the way in which the auditory system codes pitch, loudness, and achieves sound localisation.

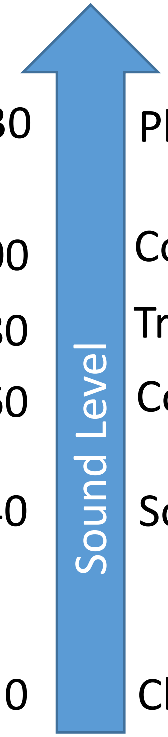
In Part 1:

“The human auditory system is both complex and remarkable.

To understand how our auditory system works, let’s pose some statements and questions and then investigate each element in more detail.”

Sound Level

Typical Sound Levels (dBA)



130	Plane taking off (200 ft away)
100	Construction site
80	Train (100 ft away)
60	Conversation (3 ft away)
40	Soft whisper
0	Close to threshold of hearing

<u>Sound level</u>

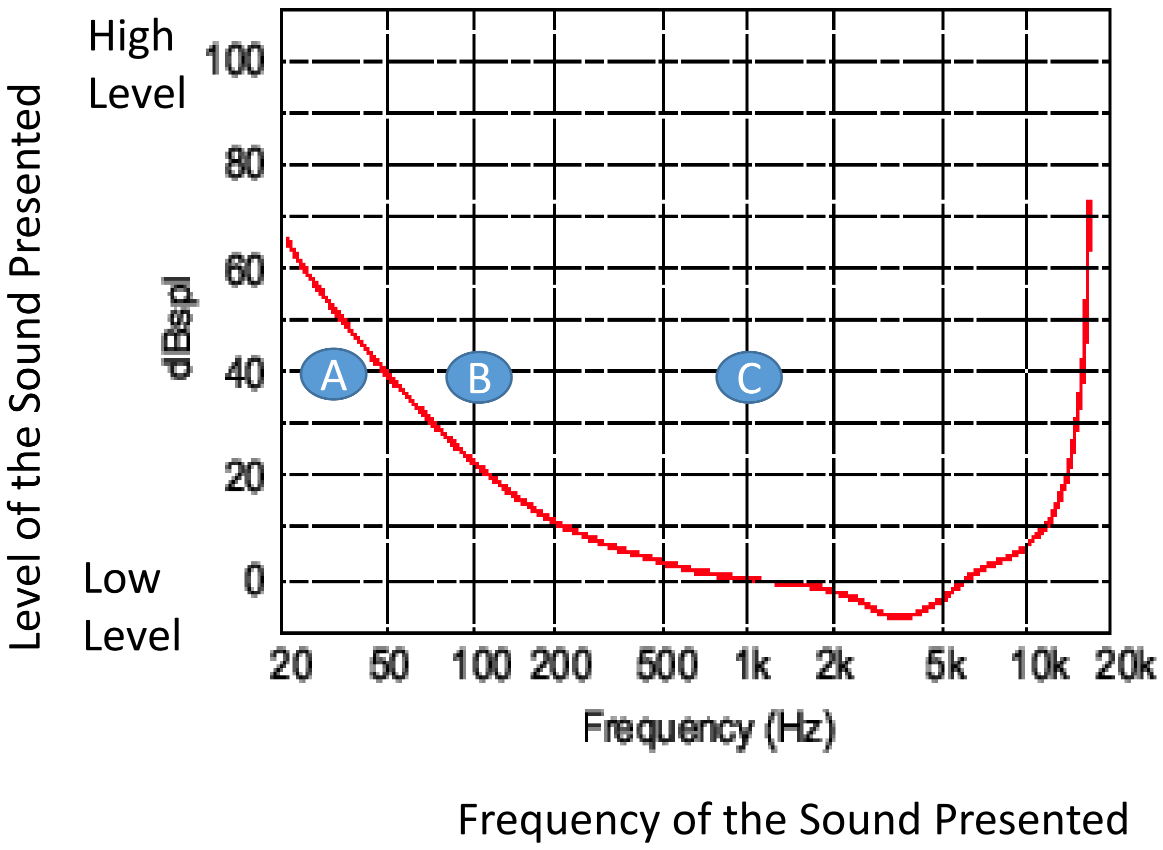
Statement: We can hear a wide variety of sounds from the quietest to the loudest.

Question: How is our auditory system able to process sound level information (which we subjectively describe as *loudness*) across such a large range of sound levels?

Loudness

Loudness is a perceived quality of sound, related to the amplitude of a sound.

Sensitivity of the Ear



The sensitivity of the ear varies with sound level and frequency.

Threshold curve is presented in red.

Sounds with combinations of level and frequency above the red line are audible.

Sounds with combinations of level and frequency below the red line are inaudible.

Much more sensitive to mid-range sound frequencies ~500 Hz to 5 kHz.

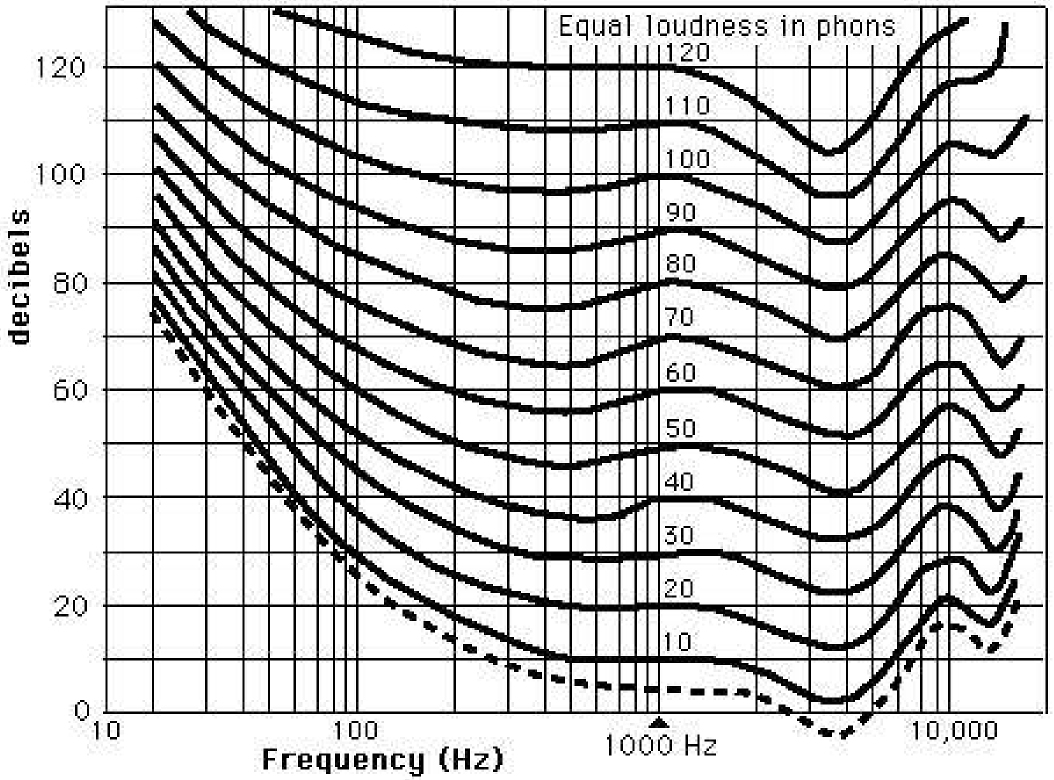
Consider 3 tones all presented at 40 dB SPL level, but with different frequencies.

A = A tone of 30 Hz at 40 dB SPL is under the red threshold curve so it is not audible.

B = A tone of 100 Hz at 40 dB SPL is just above the red threshold curve, so is just audible.

C = A tone of 100 Hz at 40 dB SPL is much higher above the red threshold curve, so would be louder than B.

Loudness



So, as we have seen- 3 tones all presented at 40 dB SPL level, but with different frequencies have different loudness.

Can generate equal loudness curves, determined experimentally:

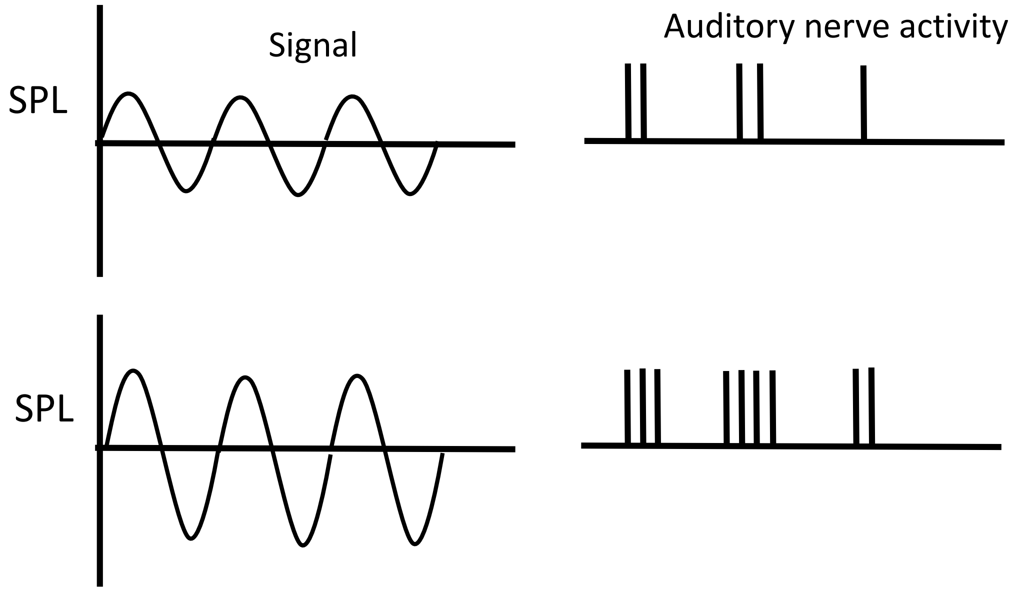
Measure the sound pressure level (SPL) required for different sound frequencies to sound equally as loud as a 1-kHz tone at a given level.

E.g., A listener is presented with a 1 kHz tone at a given level (the *standard*), and the listener adjusts the level of other tones at different frequencies to match the loudness of this *standard*.

E.g., Curve marked “40”: Was determined by matching the loudness of sound frequencies to the loudness of a 1-kHz tone presented at 40 dB SPL (the standard). “the 40-Phon curve”

Loudness

Firing Rate Hypothesis

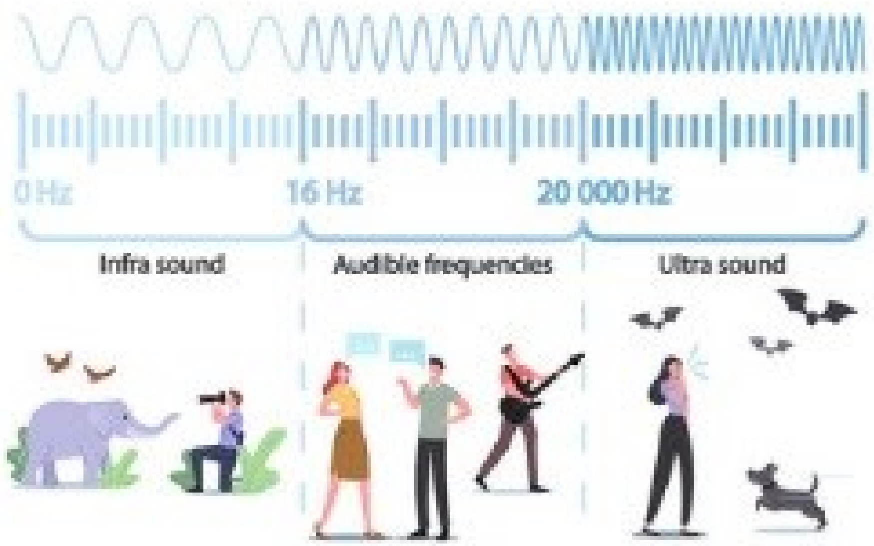


How is loudness coded physiologically?

Auditory nerve fibres increase their firing as the intensity of a sound increases.

But could also include mechanisms whereby the number of neurons activated is also increased (spread of excitation).

Sound Frequency



<u>Sound frequency</u>

Statement: We can hear a wide variety of sound frequencies, from low frequencies to high frequencies ~ 20 Hz to 20 kHz; including the frequencies important for speech processing.

Question: How does our auditory system process sound frequency information, and how do we perceive the *pitch* of sounds, which is important for speech processing as well as music perception?

Pitch

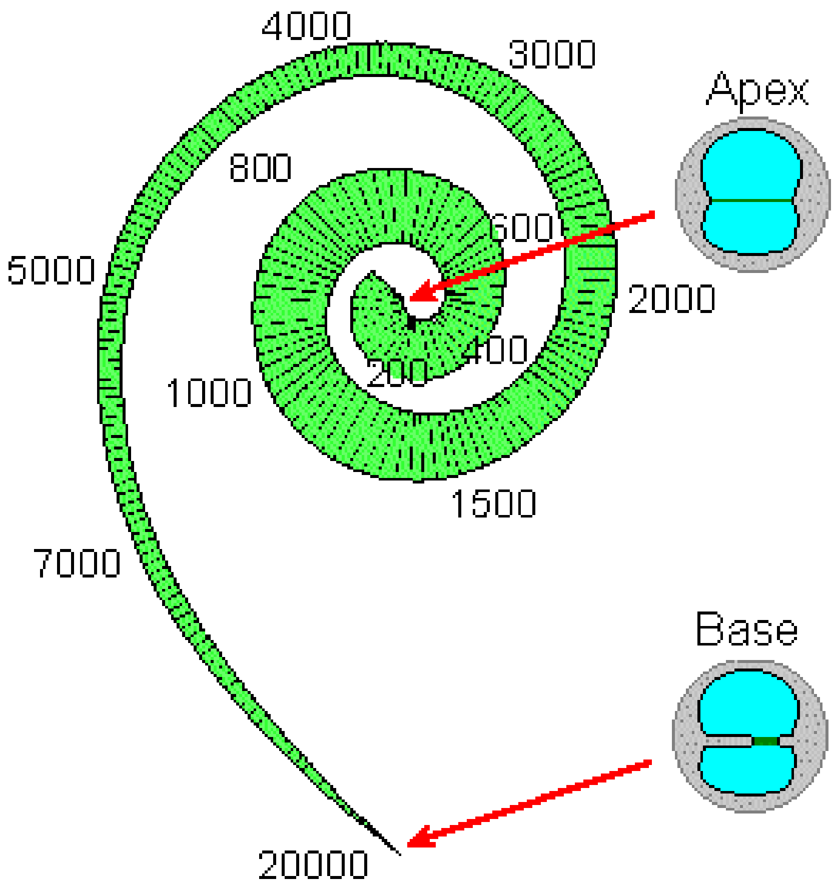
Pitch is…..

“…..*that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale*”

Closely related to the repetition rate of a sound.

So, Pitch is the *perceptual correlate* of a sound’s repetition rate.

Frequency analysis



Firstly, the auditory system performs a partial separation of the frequency components of the incoming sound. This occurs on the *basilar membrane* in the cochlea.

八度

Tone height, Chroma and Octaves

色度

Tone height = perceptual experience of increasing pitch accompanying increases in tone's fundamental frequency.

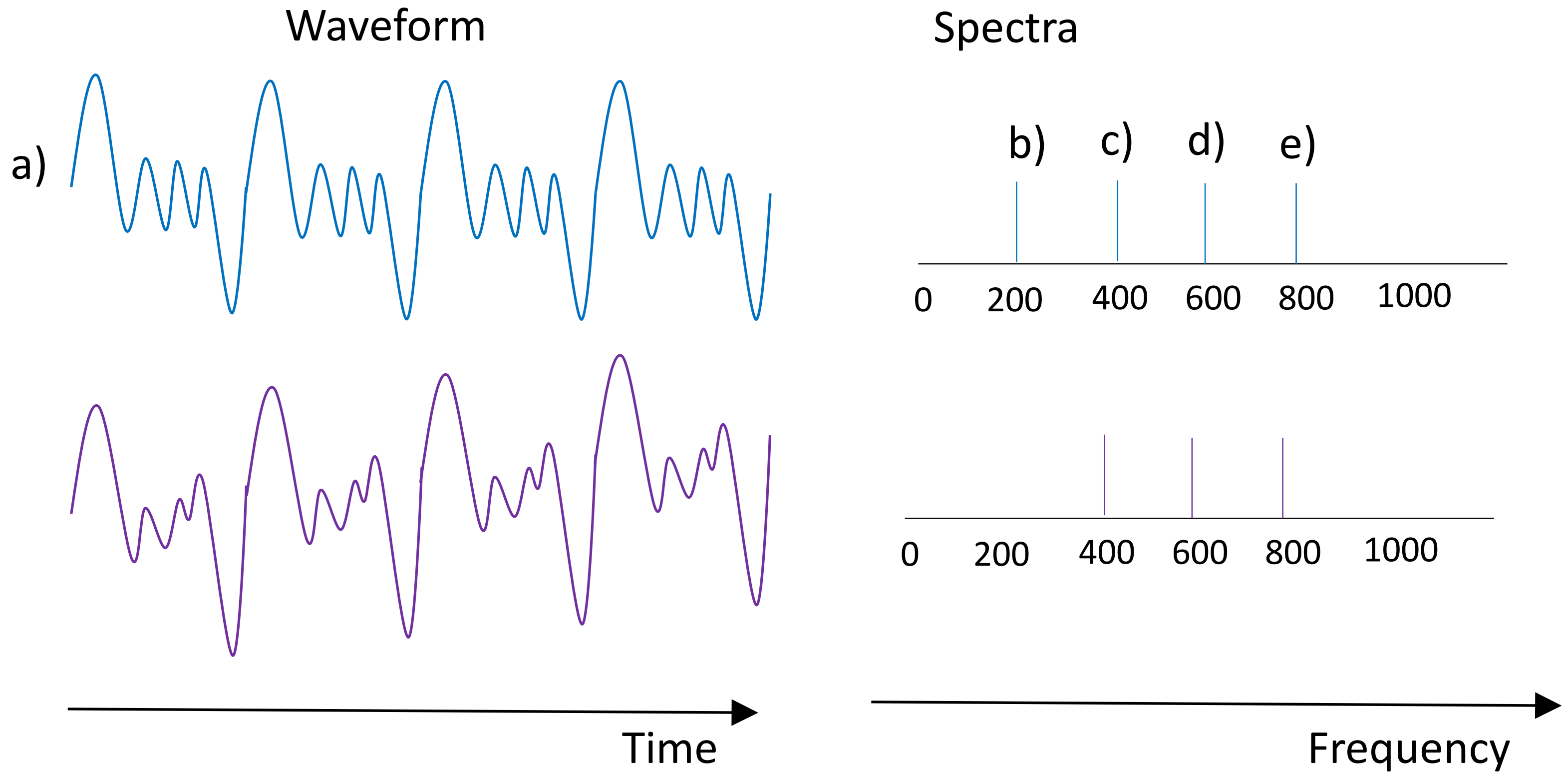
Piano- Lowest note on left = A then moving to highest note – perceived as increasing tone height.

Piano- letters of the notes A, B, C, D, E, F, G repeat. Notes with same letter sound similar. Notes with the same letter have the same tone chroma.

Every time move from one letter to the same letter across the keyboard- increase in an octave.

Octave is a doubling of frequency.

The Mystery of the Missing fundamental. . .



Periodic Complex Waveform

a) = b) + c) + d) + e)

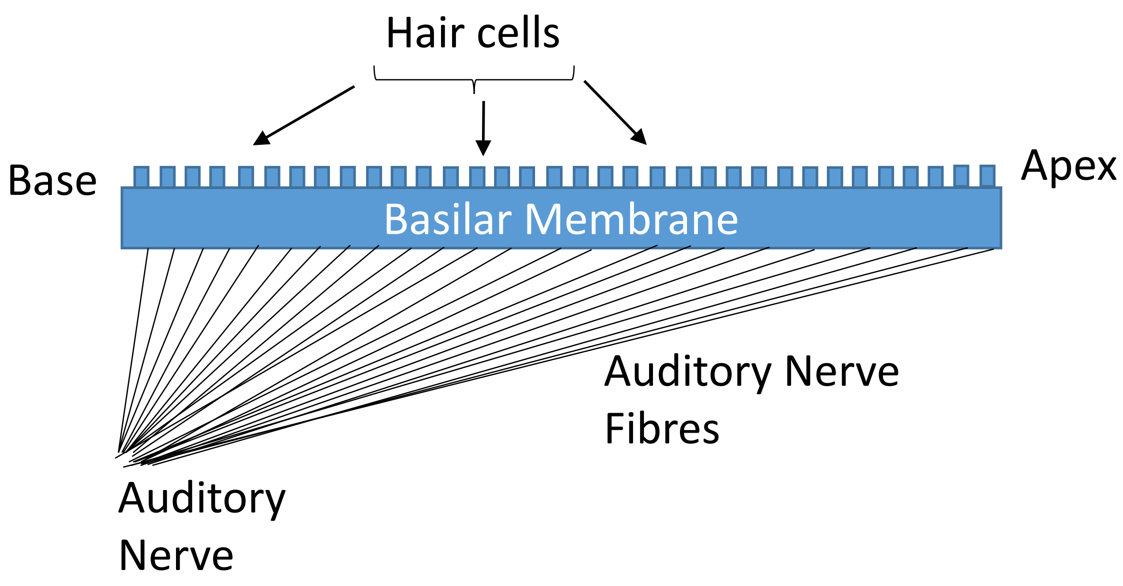
What happens to the waveform if the fundamental frequency is removed?
…the waveform's repetition rate remains unchanged (although it may change aspects of its waveform).
Therefore the pitch of the complex sound remains the same.

What does this mean? – Pitch in the human brain can be calculated without the fundamental frequency; information about the spacing between harmonics and the repetition rate (which indicates the fundamental frequency) is sufficient for the brain to compute pitch.

How is Pitch information Conveyed by the Auditory Neurons?

Regions of the membrane that respond best to sounds of high frequencies

Regions of the membrane that respond best to sounds of low frequencies



Different regions of the *basilar membrane* respond to different sound frequencies.

Hair cells and nerve fibres are also associated with each place along the basilar membrane.

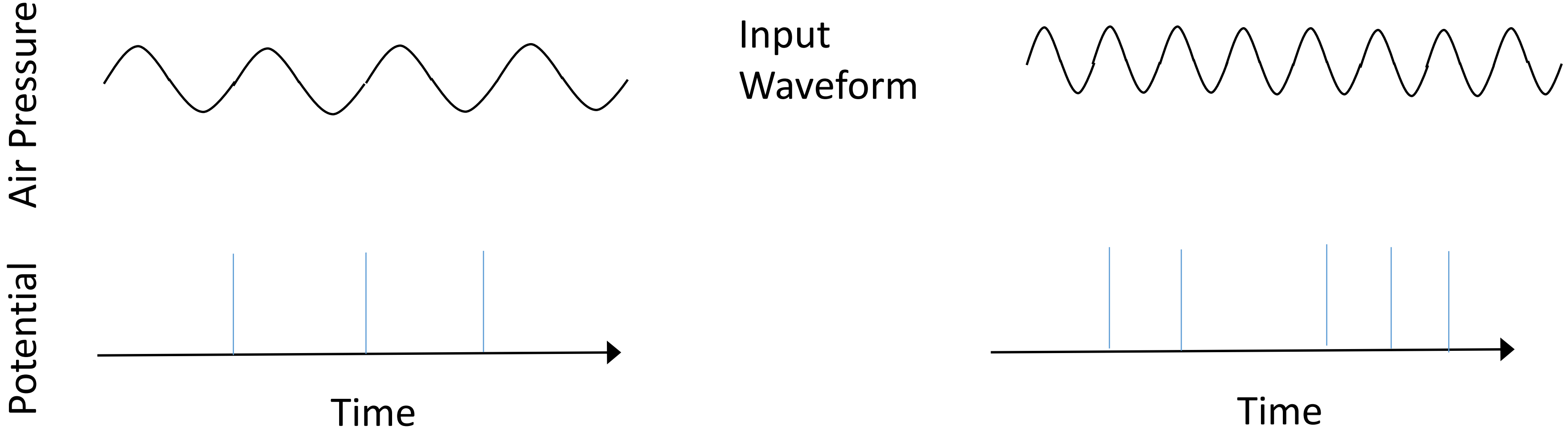
Early pitch theory- place theory. Frequency of a tone is indicated by the place along the basilar membrane at which the auditory nerve fibres are activated.

How is Pitch information Conveyed by Auditory Nerve Fibres?

Stimulus pitch information can also be provided in the time pattern of neural activation. Phase locking operates for sounds with frequencies up to around 5000 Hz. Phase locking means that regularity in a sound waveform leads to a regular pattern of neural activity.

Low frequency

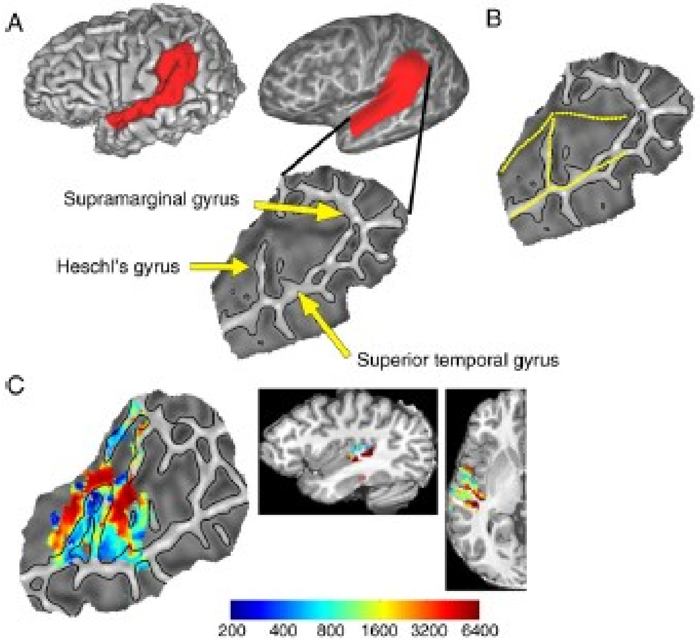
Higher frequency



Neural spikes (phase locked to the input waveform).

A nerve fibre does not necessarily fire on every cycle of the stimulus, but when they do occur , they occur at roughly the same phase of the waveform each time.

Pitch information is Maintained at Different Levels of the Auditory System

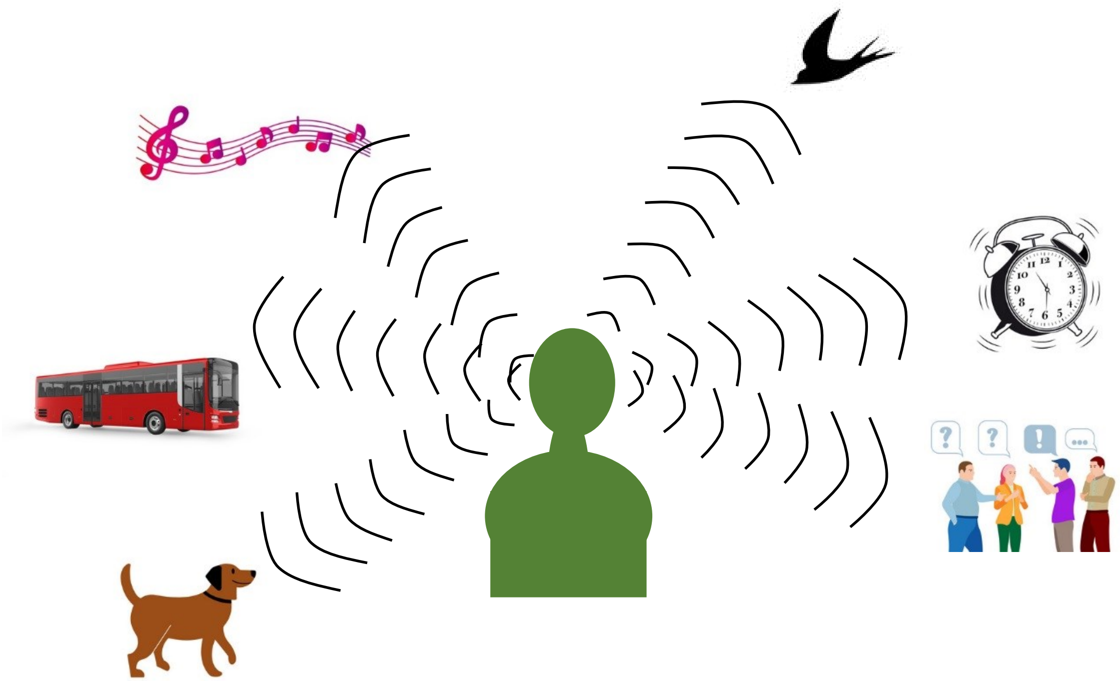


Pitch information appears to be very important and this information is preserved at many levels of the auditory system from ear to brain.

The research area of auditory pitch perception is very extensive, with different theories regarding pitch, perfect pitch, with a variety of computational models of pitch perception.

Humphries, C. et al. (2010). NeuroImage, 50, 1202-1211

Sound Localisation



Sound localisation

Statement: We can detect sounds from behind, in front, above, and below.

Question: How is the information from each ear and both ears combined in order for us to accurately judge where a sound is located in space?

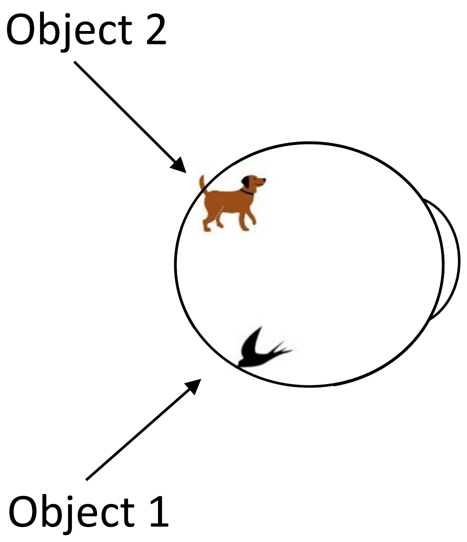
Sound Localisation

Ability to localise sound sources is very important – provides information about the direction of objects to seek, or objects to avoid, and also indicates appropriate direction to direct visual attention.

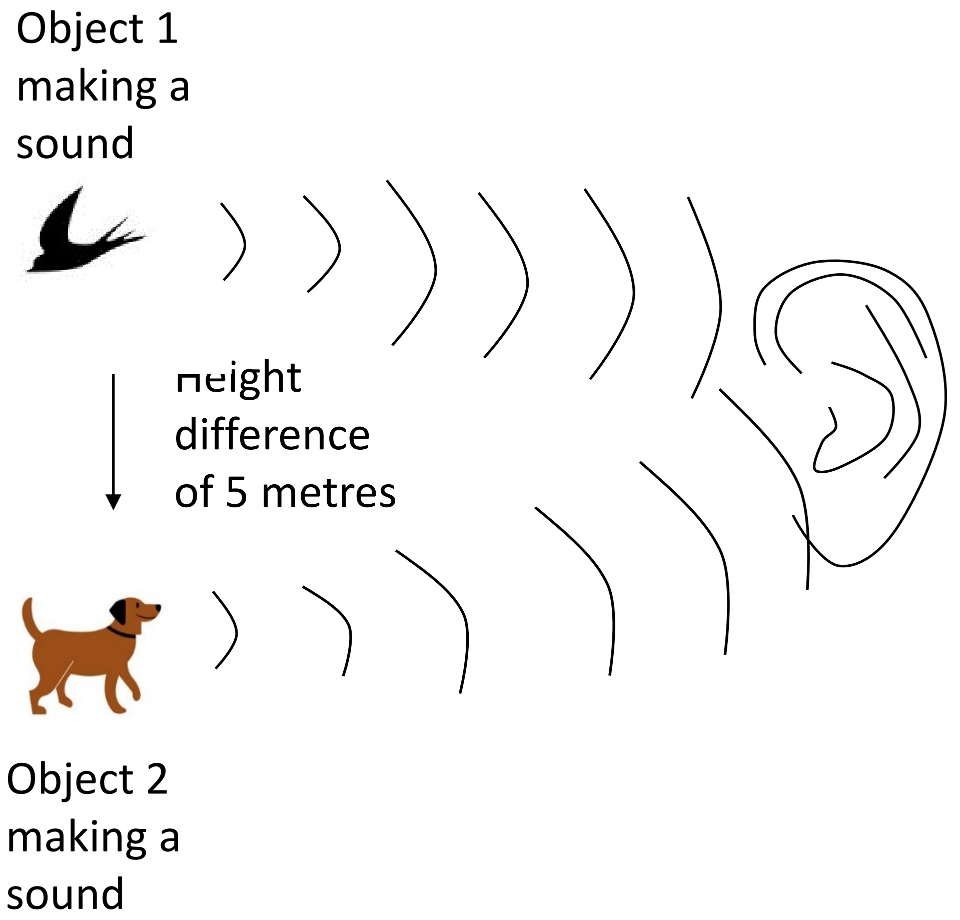
Localisation of sounds can depend on a comparison of sound information from both ears as well as the sound information processed by a single ear.

What Information is Required for Localisation?

<u>EYE</u>



<u>EAR</u>

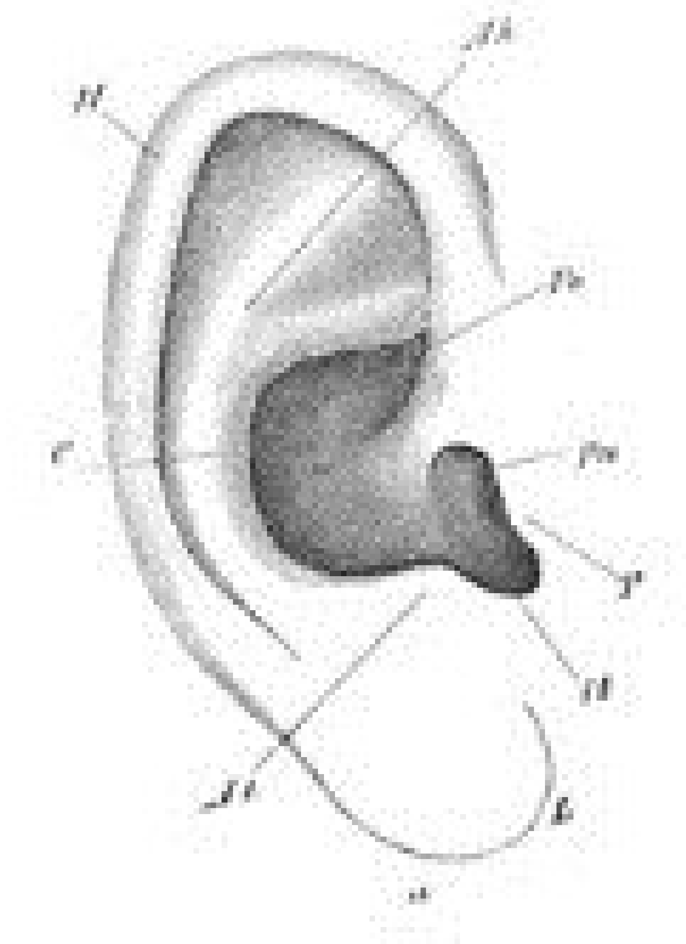


Comparing location information for vision and hearing.

Vision: Objects 1 and 2 are imaged on *different places* on the retina - - > provides location information

Hearing: Sound frequencies from Object 1 and 2 create a spread of *vibrations across the cochlea*- how can the location information be extracted/coded?

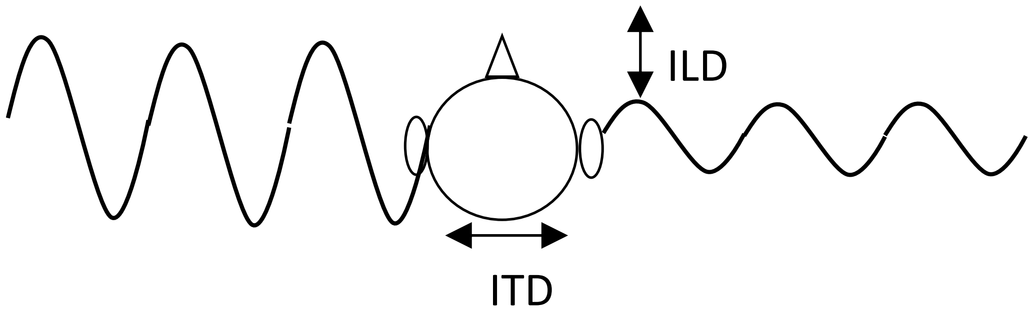
Monaural (single ear) Sound Localisation Cues



folds and curves of the outer ear (pinna) are unique to the individual, and help introduce spectral changes into the sound that arrives at the ear. This spectral modification contains important information for the localisation of sounds.

Can obtain information about how sound spectra are altered by the pinna by placing microphones inside the ear and recording changes in the sound reaching the inside of the ear.

Binaural Sound Localisation Cues



There are two binaural cues for sound localisation -both based on a comparison of the sound signal reaching the left and right ears.

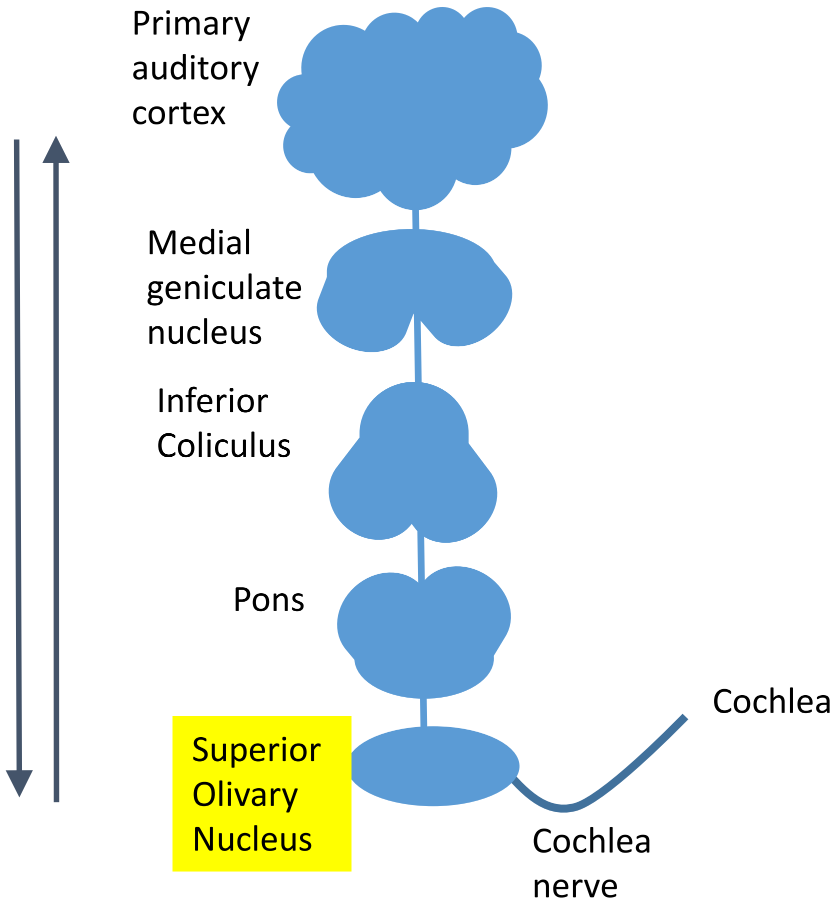
Difference between time of arrival of sound at two ears - Interaural Time Difference (ITD)

ITD is useful at low frequencies

Difference between loudness of sounds at two ears - Interaural Level Difference (ILD)

ILD is useful at high frequencies

Where is the Information From Two Ears Combined?



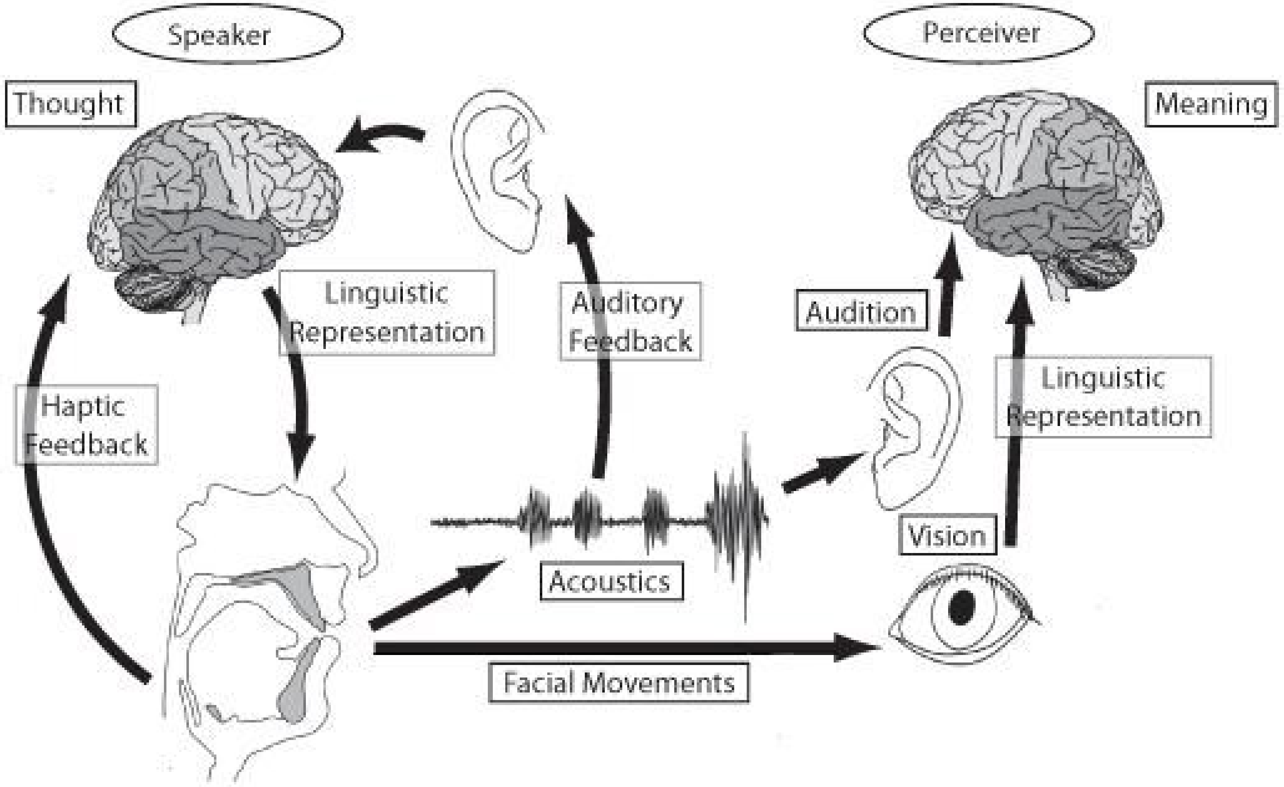
Superior Olivary Complex (SOC): 1st place where signals from two ears come together.

Cells in lateral superior olive detect differences in sound intensity

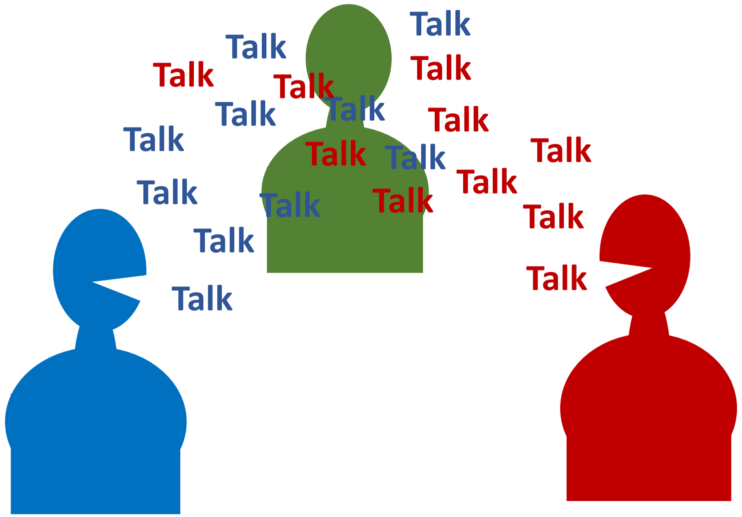
Cells in medial superior olive detect differences in timing

Speech Processing

The speech chain:



Speech Processing



Speech is a very complex signal- sound levels, location information, frequencies.

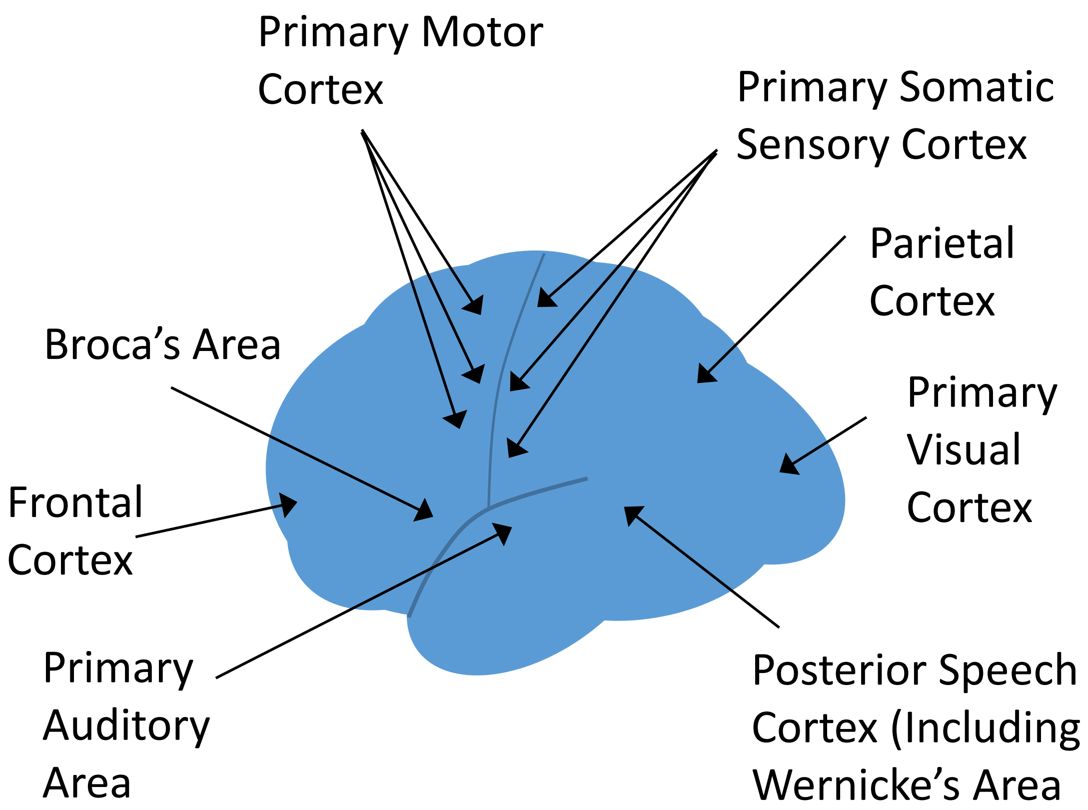
Language- linguistic unit, how we perceive strings of phonemes that make up phrases, sentences.

Context – Prior knowledge is important in processing meaning.

Speech perception depends on processing of the auditory signals by the ear (bottom-up processing) and prior knowledge held in the brain about meaning, grammatical rules, social situation etc. (top-down processing)

Auditory Cortex of the Brain and Speech

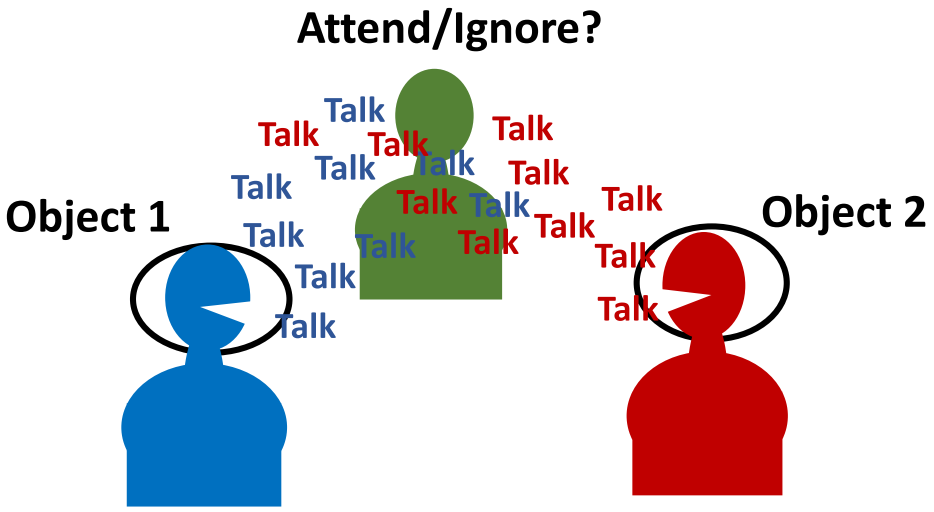
皮层



Wernicke's Area in temporal lobe of the brain: Controls comprehension of speech and other complex sounds. 1871 Carl Wernicke described patients who could speak fluently but had difficulty comprehending speech. They had lesions in left temporal lobe also parts of inferior/superior parietal lobe.

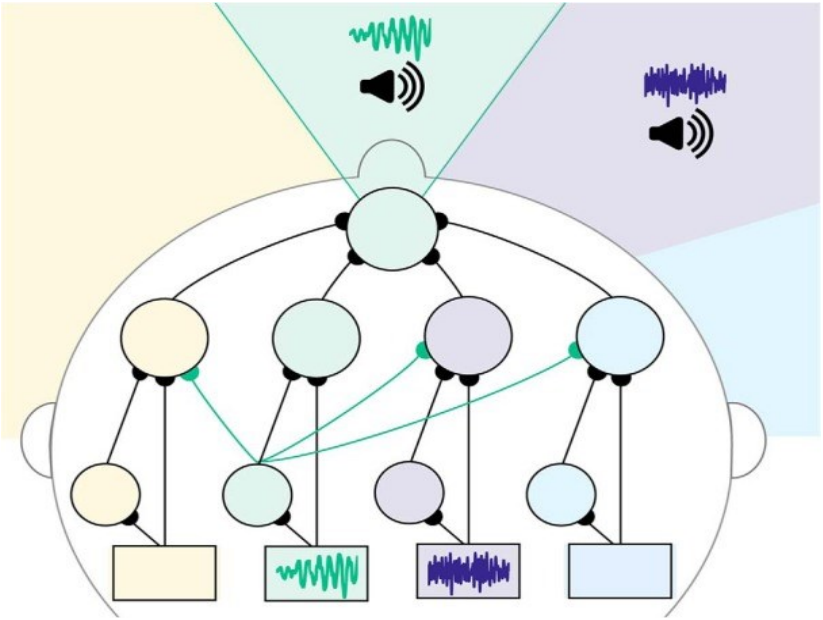
Broca's area in lower left frontal lobe of brain: Controls speech production including motor movements of speech. 1861 Paul Broca described a patient who could not speak but could understand spoken language. Large lesion in lower left frontal lobe; confirmed in others.

Cocktail Party Problem



Cocktail Party Effect (Cherry 1953)

In a cocktail party environment- How do we manage to follow speech from one person amongst other talkers? Can use loudness, pitch and location cues to help segregate the speech of interest.



Prominent area of research: Algorithms to separate out voice mixtures.

Applications: Speech Recognition Systems, assistive devices.

Computational neural network models to separate out mixtures of speech based on physiological processes. e.,g Dong et al., 2016.

Hearing Impairment and Assistive Devices

Hearing Impairment

The speech chain:

Reduced auditory feedback

Reduced auditory input

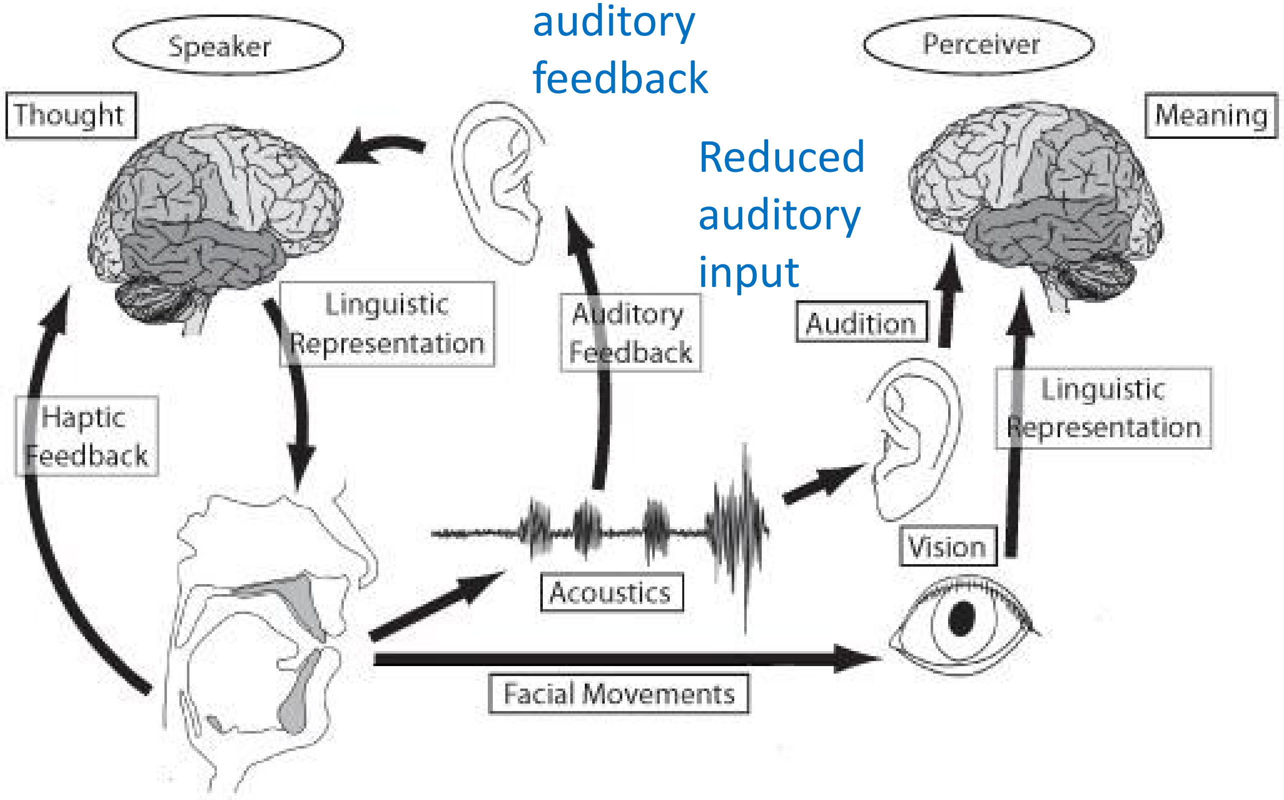


Image accessed (28/2/17): http://www.phon.ucl.ac.uk/courses/spsci/iss/week1.php

History: Variety/miniaturisation

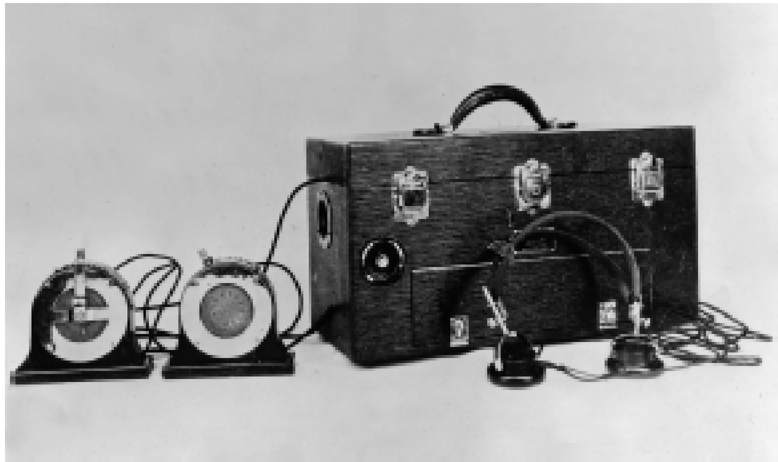


Figure 4. The WE 24-A Audiphone (1924). One of Western Electric's first portable models. (Courtesy of Kenneth Berger Hearing Aid Museum and Archives)

Mills, M., 2011. Hearing aids and the history of electronics miniaturization. *IEEE Annals of the History of Computing*, *33*(2), pp.24-45

Assistive Devices - Cochlear Implants

Electrodes are inserted by a surgeon into the cochlea to stimulate hearing by electrically stimulating regions of the cochlea.

The cochlear implant electrically stimulates the auditory nerve fibres directly.

Sound signals are picked up by the microphone outside the ear. This signal is transformed into electrical pulses.

Electrical pulses are sent to a magnetic coil behind the ear, and then transmitted across the skin to the implant.

Implant sends a pattern of electrical pulses to the electrodes in the cochlea.

Auditory nerve picks up these electrical pulses and sends them to the brain.
The brain recognises these signals as sound.

Questions [10mins]

Can the electrodes of a cochlear implant be matched across ears, and will this affect sound localisation?

What kinds of new developments, e.g., machine learning, could be applied to a cochlear implant?

What other factors may contribute to the audio separation of multiple speakers ?

Audio Applications

Application

Serafin, S., Geronazzo, M., Erkut, C., Nilsson, N.C. and Nordahl, R., 2018. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE computer graphics and applications*, *38*(2), pp.31-43

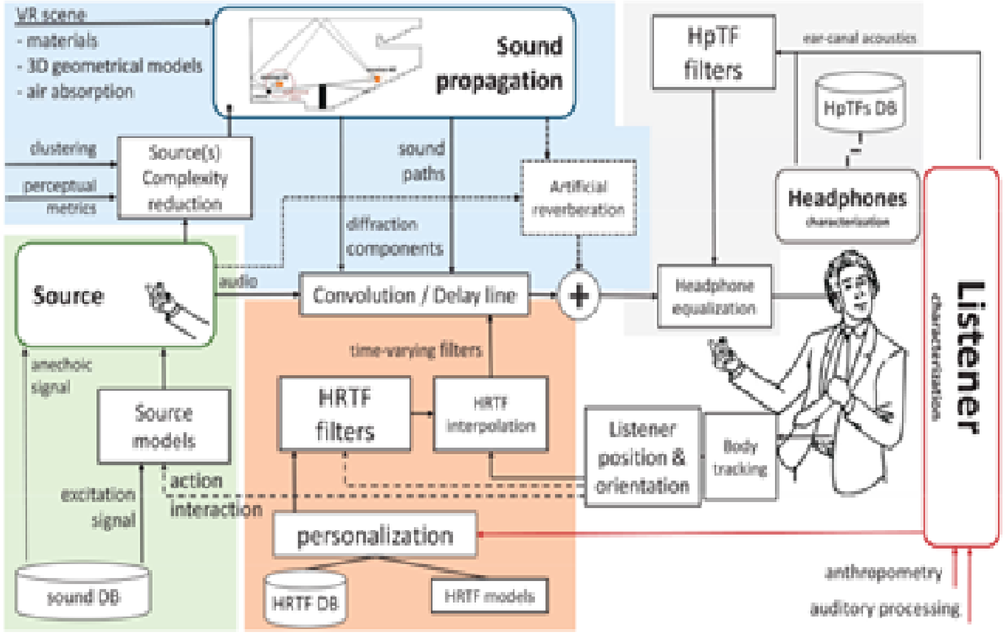


Figure 2. Block diagram of a typical system for binaural rendering and auralization.

Advances in hardware and software technologies have facilitated the improvement in the development of immersive interactive sound-rendering experiences in virtual environments.

Different elements are involved that, combined, provide a complete interactive sonic experience (e.g., physics-based simulation of sound effects, propagation in space, binaural rendering to simulate the position of sound sources).

Headphone-based sound rendering is often used as it makes it possible to completely control the sound arriving at each ear.

Current challenges: Personalisation, adaptation.

Application

Zhao, Y., Bennett, C.L., Benko, H., Cutrell, E., Holz, C., Morris, M.R. and Sinclair, M., 2018, April. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-14).

Canecontroller

Overlay of virtual scene and real scene

Navigating a virtual road crossing using Canecontroller

Canetroller, a haptic cane controller that simulates white cane interactions, enabling people with visual impairments to navigate a virtual environment by transferring their cane skills into the virtual world.

Canetroller provides:

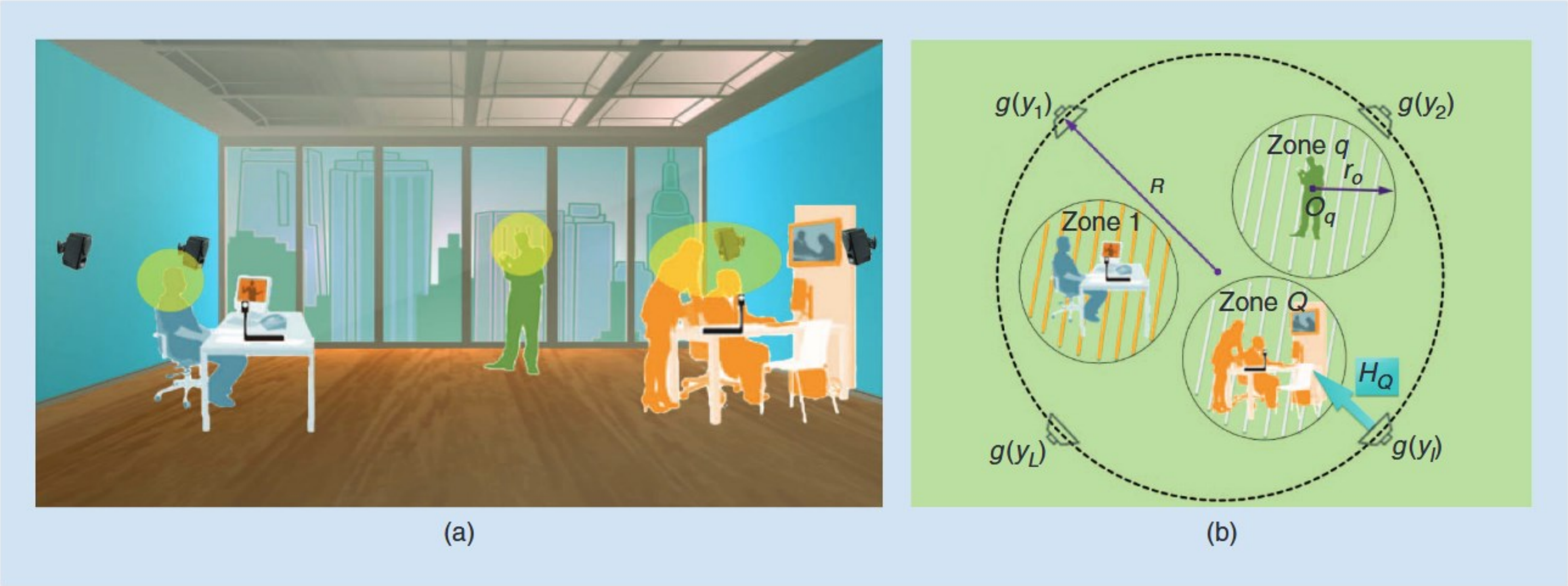
(1) physical resistance when the virtual cane comes in contact with a virtual object;

(2) vibrotactile feedback simulating cane vibrations when it hits an object or touches a surface.

(3) spatial 3D auditory feedback simulating the sound of real-world cane interactions.

Applications

Betlehem, T., Zhang, W., Poletti, M.A. and Abhayapala, T.D., 2015. Personal sound zones: Delivering interface-free audio to multiple listeners. *IEEE Signal Processing Magazine*, *32*(2), pp.81-91.



[FIG1] (a) An illustration of personal sound zones in an office environment. (b) A loudspeaker array is used to create multiple sound zones for multiple listeners.

Sound rendering is increasingly being required to extend over certain regions of space for multiple listeners, known as personal sound zones, with minimum interference to listeners in other regions. Could apply to all kinds of enclosures (offices, cars, exhibition areas) and generating quiet zones in noisy environments.

Betlehem et al. (2015). Presented a design of directional loudspeakers for sound reproduction, to control sound fields over wide areas (reduces total number of loudspeaker units), making it suitable for establishing personal sound zones.

Overall Summary

Covered an introduction to how the auditory system processes loudness, pitch, and localisation of sounds.

Provided examples of application of audio research to interface design/evaluation.

Resources

<u>Essential</u>:
Sensation and Perception 8th edition (book), p259-289; 291-309; 311-327.

<u>Supplementary</u>:

Cheung, S., Han, E., Kushki, A., Anagnostou, E. and Biddiss, E., 2016. Biomusic: An auditory interface for detecting physiological indicators of anxiety in children. *Frontiers in neuroscience*, *10*, p.401.

Wu, J., Liu, Q., Zhang, M., Pan, Z., Li, H. and Tan, K.C., 2021. HuRAI: A brain-inspired computational model for human-robot auditory interface. *Neurocomputing*, *465*, pp.103-113.

Serafin, S., Geronazzo, M., Erkut, C., Nilsson, N.C. and Nordahl, R., 2018. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE computer graphics and applications*, *38*(2), pp.31-43

Zhao, Y., Bennett, C.L., Benko, H., Cutrell, E., Holz, C., Morris, M.R. and Sinclair, M., 2018, April. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-14).

Geronazzo, M., Bedin, A., Brayda, L., Campus, C. and Avanzini, F., 2016. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies*, *85*, pp.4-15.

Betlehem, T., Zhang, W., Poletti, M.A. and Abhayapala, T.D., 2015. Personal sound zones: Delivering interface-free audio to multiple listeners. *IEEE Signal Processing Magazine*, *32*(2), pp.81-91.

<u>Not necessary, but if interested in reading further about pitch perception:</u>

Oxenham, A.J., 2012. Pitch perception. Journal of Neuroscience, 32(39), pp.13335-13338.

Erfanian Saeedi, N., Blamey, P.J., Burkitt, A.N. and Grayden, D.B., 2016. Learning pitch with STDP: A computational model of place and temporal pitch perception using spiking neural networks. *PLoS computational biology*, *12*(4), p.e1004860.