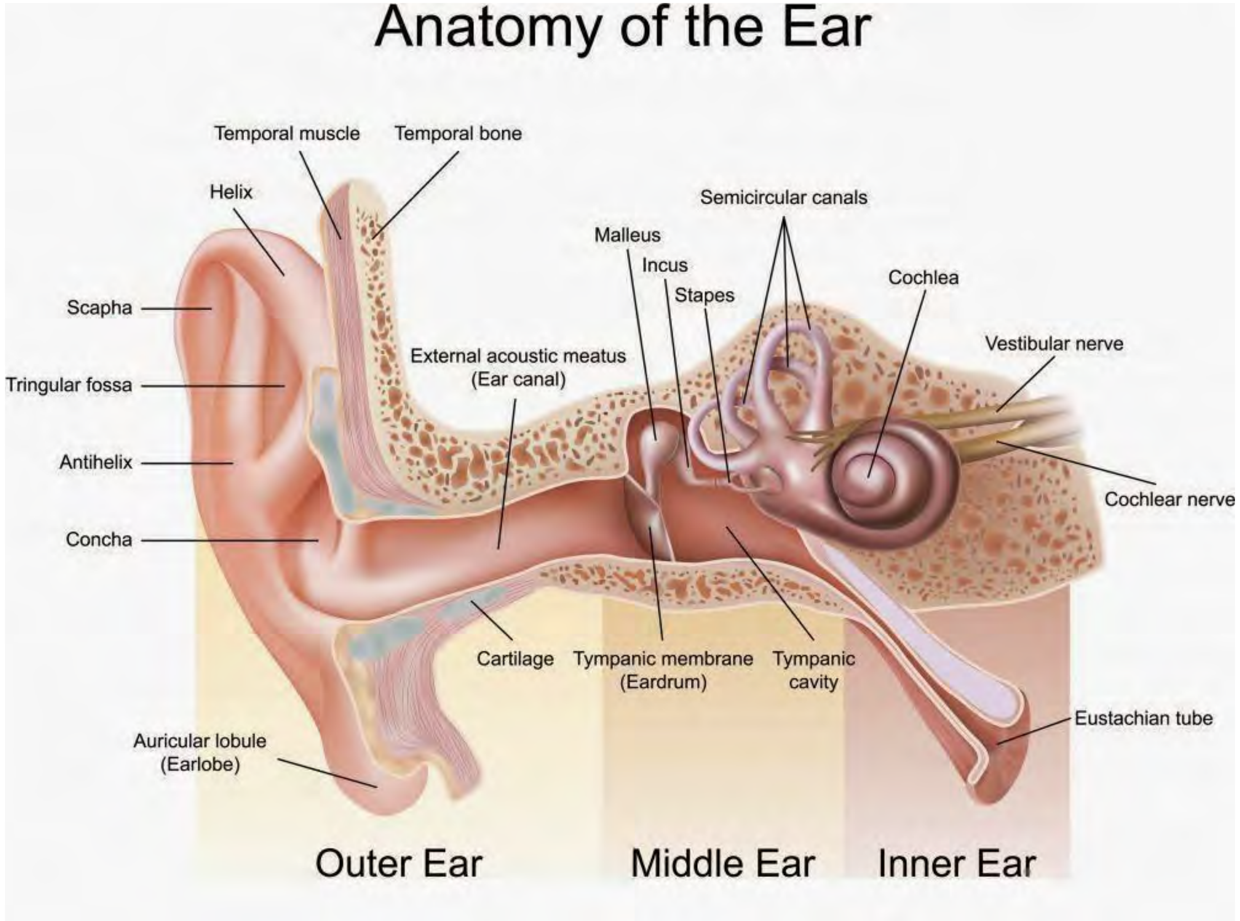
What do we know about hearing and brain function related to voice?

Mette Pedersen, Ear Nose specialist, FRSM PhD et h.c. research prof. IBC Cambridge, UK.

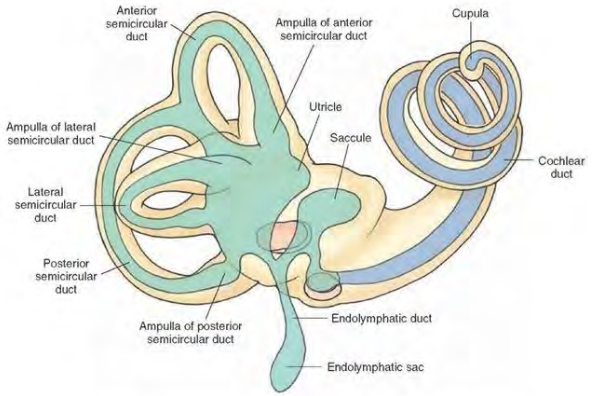
Sanila Mahmood, Anders Jønsson, M. Shezad Mahmood, Bilal Hussain, Anne A. Agersted – students at University of Copenhagen

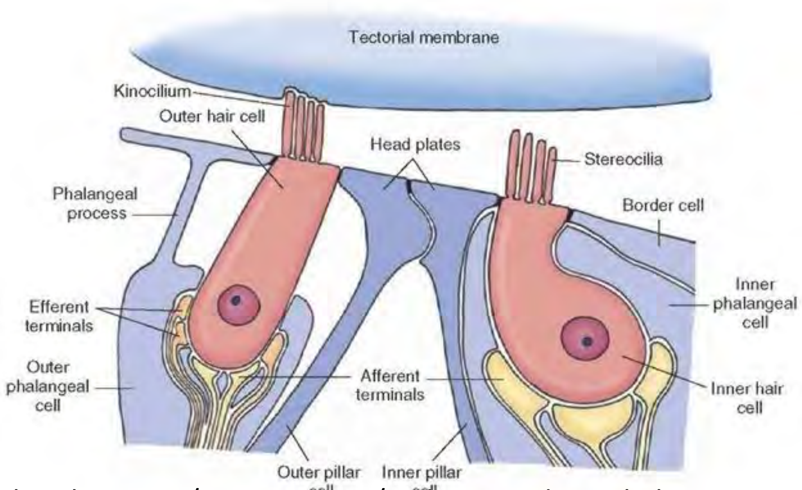
The diagram of the ear



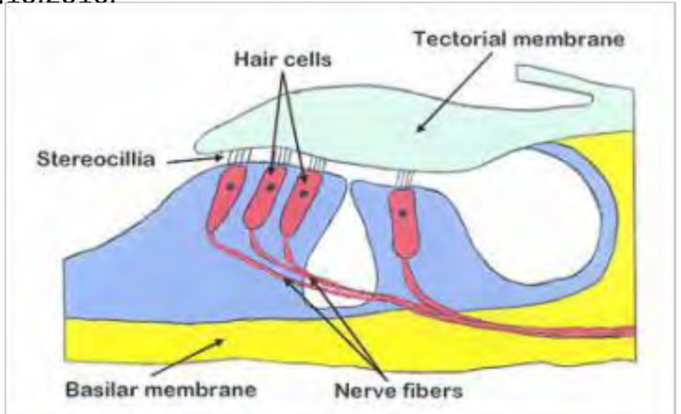
Medicinske illustrationer: Specialist- Biologisk, Videnskablig Anatomisk tegninger/illustrationer. http://www.ehealthideas.com/2013/10/external-ear-anatomy.html. 19-10-2016.

Structure of the inner ear



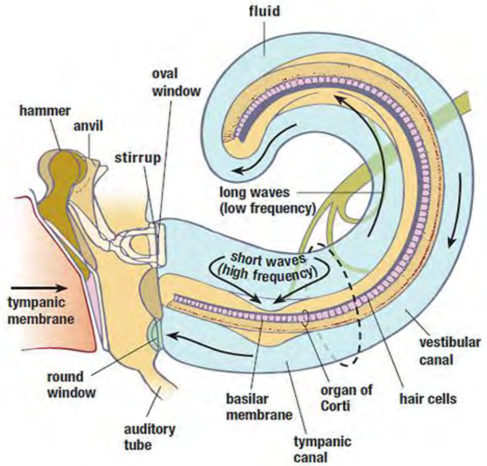


Auditory and Vestibular Systems (sensory System) Part 1. http://what-when-how.com/neuroscience/auditory-and-vestibular-systems-sensory-system-part-1/. 19.10.2016.



The inner ear, including prominent structures: basilar membrane, Organ of Corti, hair cells (3 rows of outer, 1 row of inner), tectorial membrane, and the auditory nerve fibers

http://sphsc461ohcloss.weebly.com/anatomy-of-the-ear.html

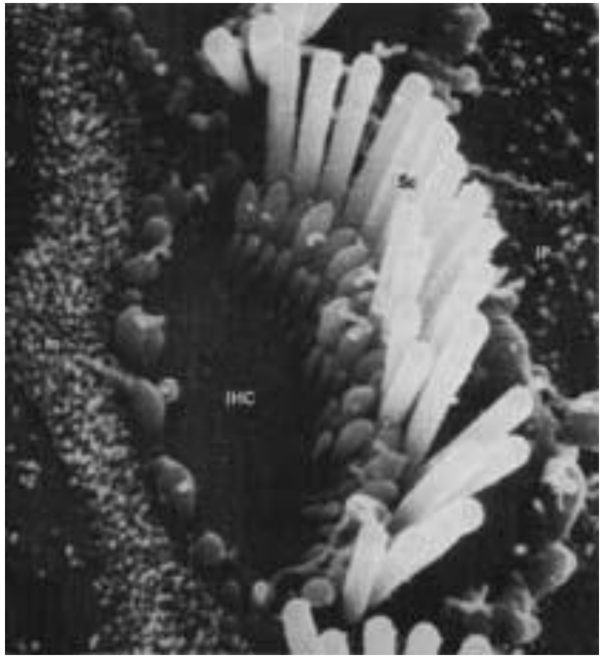


http://www.prosoundweb.com/article/perception_is_reality_psychoacoustics_from_an_audio_engineers_perspective/P2/

Sansory cells – hair cells

Internal (left) and outer hair cells of the cochlea. Inner hair cells is about 35µm, about outer hair cells 25

In total there are approximately 2500 IHC (inner hair cells) and 12500 OHC (outer hair cells) in the cochlea.

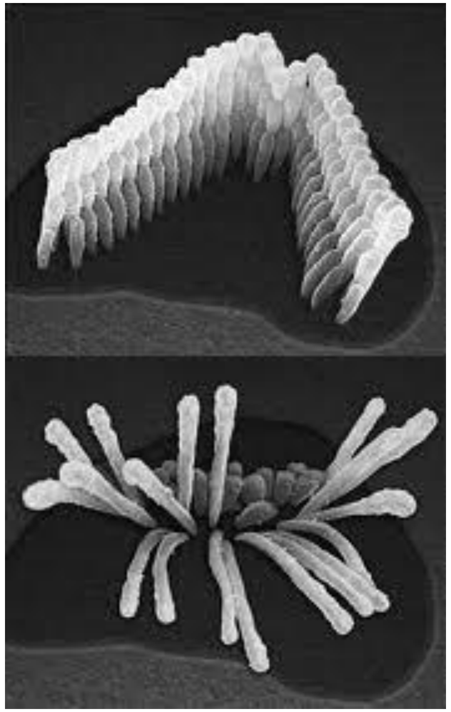


Øret og hørelsen-fra lyd til sansning. 2015. Jakob Christensen-Dalsgaard, Biologisk Institut, SDU. jcd@biology.sdu.dk

Intact and undamaged hair cells

- Especially the outer hair cells are damaged by intense sounds (lower image)

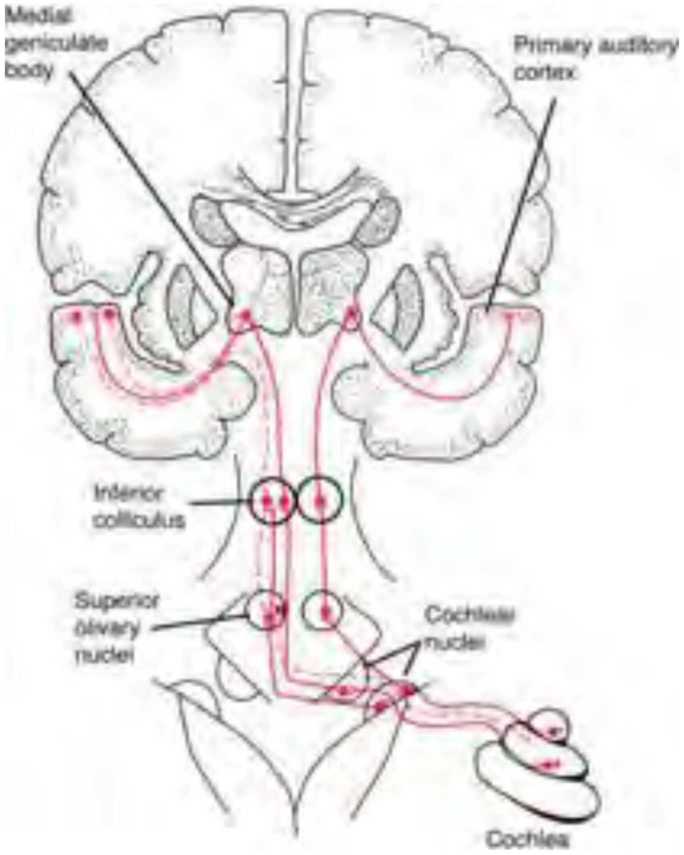
- The outer hair cells provide a feed-back reinforcement of vibration of the basilar membrane and make more cochlear sensitive - thresholds down to 0.4 nm (threshold has been largely limited by internal noise in the ear)



Active mechanisms in cochlea

- The active mechanisms are due to external hair cells. Apparently there is some form of feedback between the wide, passive stimulation of basilar membrane in the migrant wave - and active processes, because of the contraction of the outer hair cells in the stimulated region.

The hearing anatomy

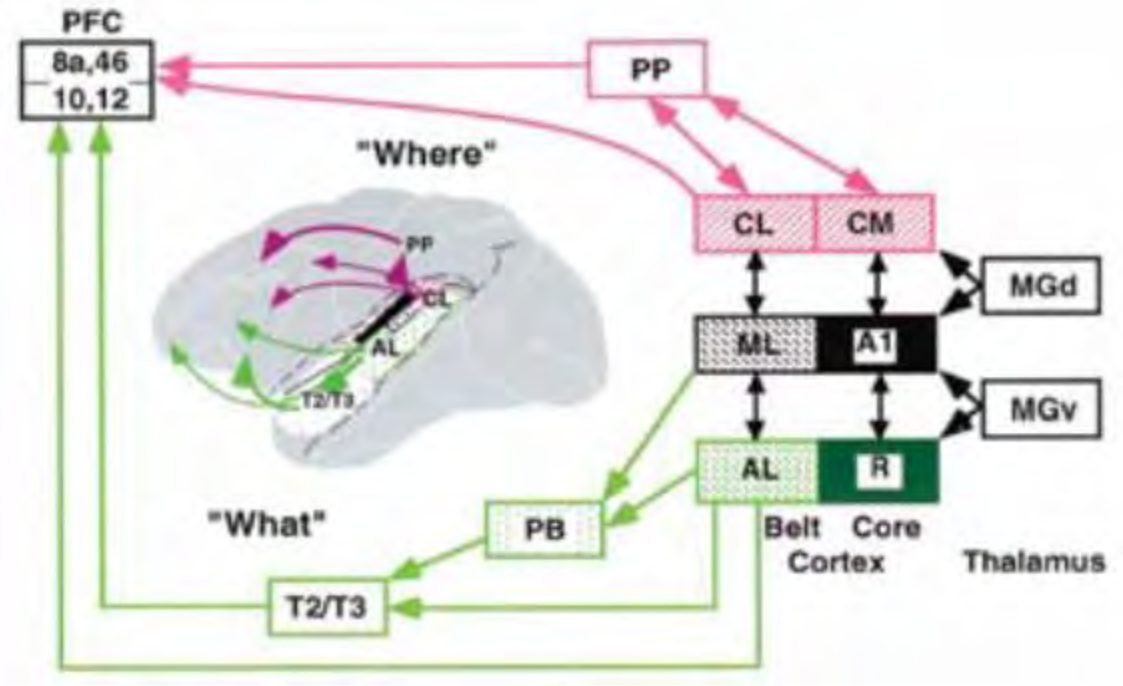


- Nerve signals propagate with a rate of approximately 10-30 m/s and will therefore be delayed up through the hearing pitch.
  In each synapse is a further delay of approximately 1 ms.

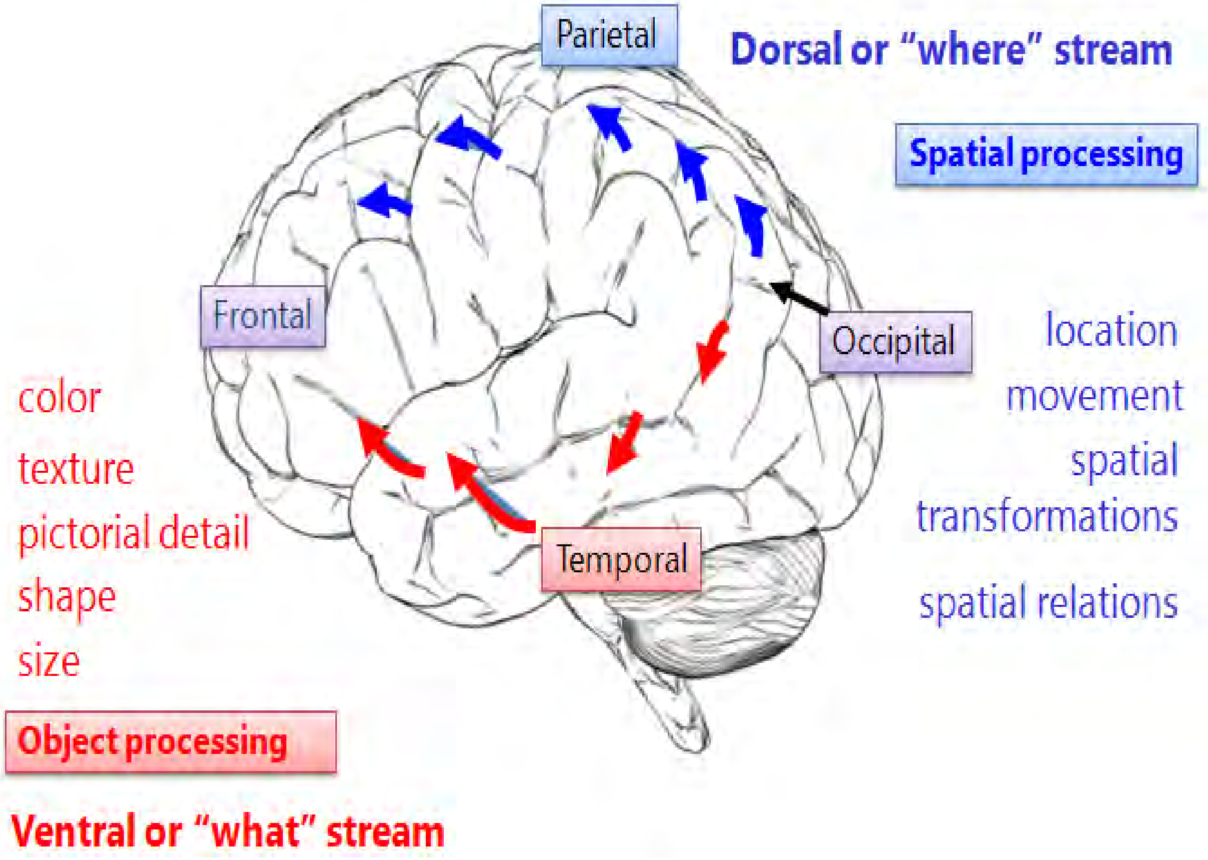
Øret og hørelsen-fra lyd til sansning. 2015. Jakob Christensen-Dalsgaard, Biologisk Institut, SDU. jcd@biology.sdu.dk

‘What’ and ‘where’

- Categorization (What) and Localization(Where) is treated probably in various streams in the cortex, known what and where pathways.



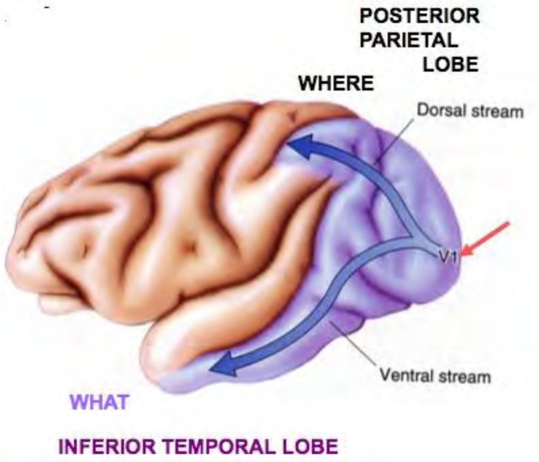
What (green) and where (violet)-pathways



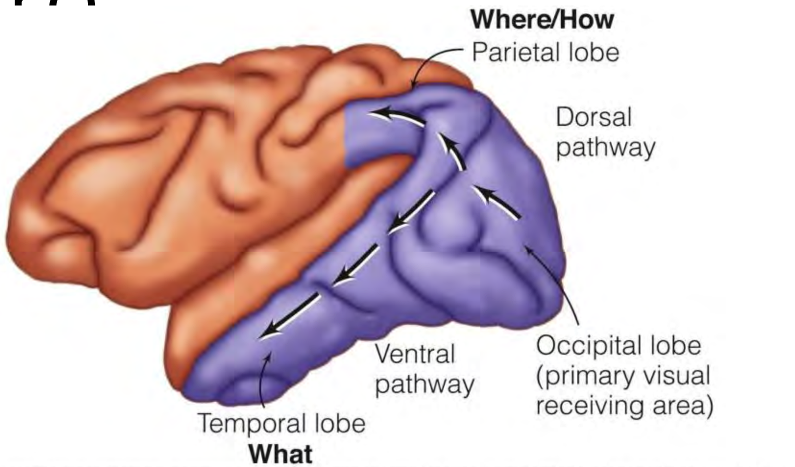
Øret og hørelsen-fra lyd til sansning. 2015. Jakob Christensen-Dalsgaard, Biologisk Institut, SDU. jcd@biology.sdu.dk

Imagery Lab. http://www.nmr.mgh.harvard.edu/mkozhevnlab/?page_id=663

Ekstra billeder af what and where



http://www.ssc.education.ed.ac.uk/courses/vision/vmay092ii.html



From "Object Vision and Spatial Vision: Two Central Pathways," by M. Mishkin, L. G. Ungerleider, & K. A. Macko, 1983, *Trends in Neuroscience, 6,* 414–417, figure 1. Copyright © 1983, with permission from Elsevier.

https://www.studyblue.com/notes/note/n/psych-test-2/deck/12335029

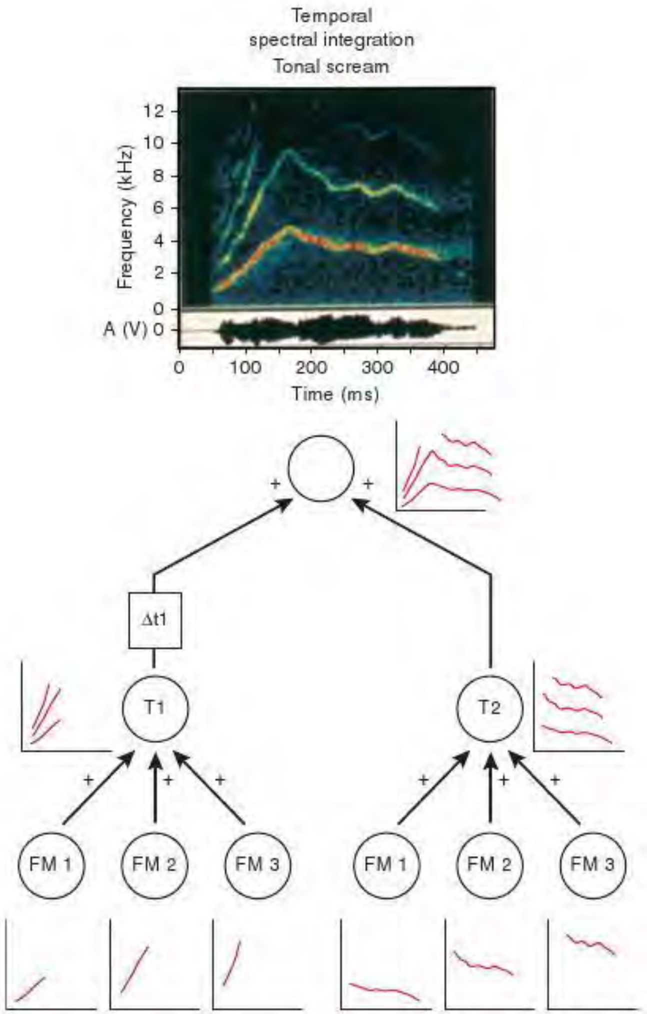
The what pathway sweeps forward from V1 and V2 along the lower edge of the brain on each side, processing information about the identity of an object.

The where pathway sweeps forward higher up on the brain and processes information about where objects in the world are located.

The what pathway is concerned with the identification of objects in the environment. It helps identify if a particular pattern of light and color represents a chicken, a Volkswagen, or Aunt Mabel. The where pathway is concerned with the location of information and with guiding actions in the world, such as reaching out and grabbing things, moving from place to place, or making eye movements. What makes an object easy to find is how

http://www.easynotecards.com/print_cards/10262

Model of neural processing in cortex



Communication calls consist of elementary features, such as bandpass noise bursts or **frequency-modulated** (FM) sweeps. Harmonic calls, such as the vocal scream from the rhesus monkey repertoire depicted here by its spectrogram and time signal amplitude (A, measured as output voltage of a sound meter), consist of fundamental frequencies and higher harmonics.

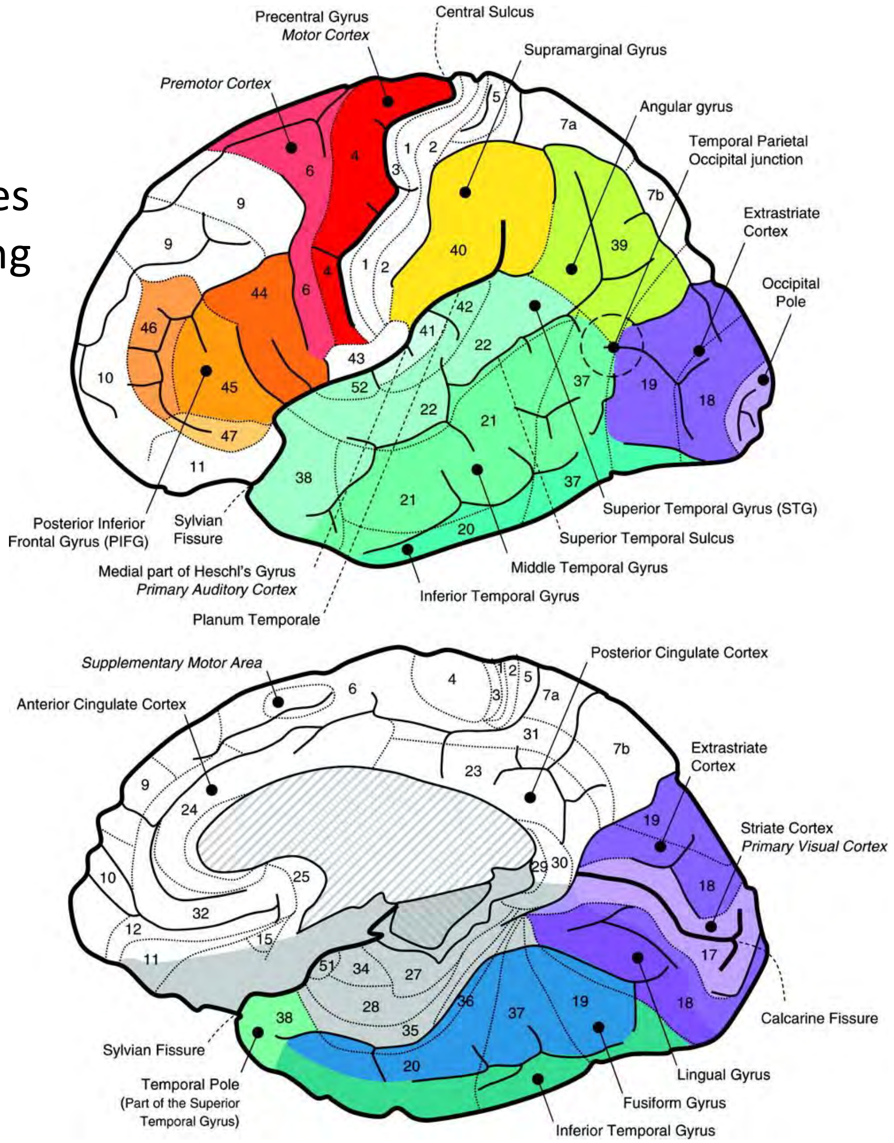
The neural circuitry for processing such calls is thought to consist of small hierarchical networks. At the lowest level, there are neurons serving as FM **detectors** tuned to the rate and direction of FM sweeps; these detectors extract each FM component (shown in cartoon spectrograms) in the upward and downward sweeps of the scream. **The output of these FM detectors is combined nonlinearly at the next level: the target neurons T1 and T2 possess a high threshold and fire only if all inputs are activated. At the final level, a 'tonal-scream detector' is created by again combining output from neurons T1 and T2 nonlinearly.** Temporal integration is accomplished by having the output of T1 pass through a delay line with a latency Δt1 sufficient to hold up the input to the top neuron long enough that all inputs arrive at the same time.

Early processing of human speech sounds in the antero-lateral auditory belt and parabelt cortex is thought to be accomplished in a similar way.

Rauschecker J, Scott S. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. 2009. Nat Neurosci. Vol. 12(6), pp. 718-724.

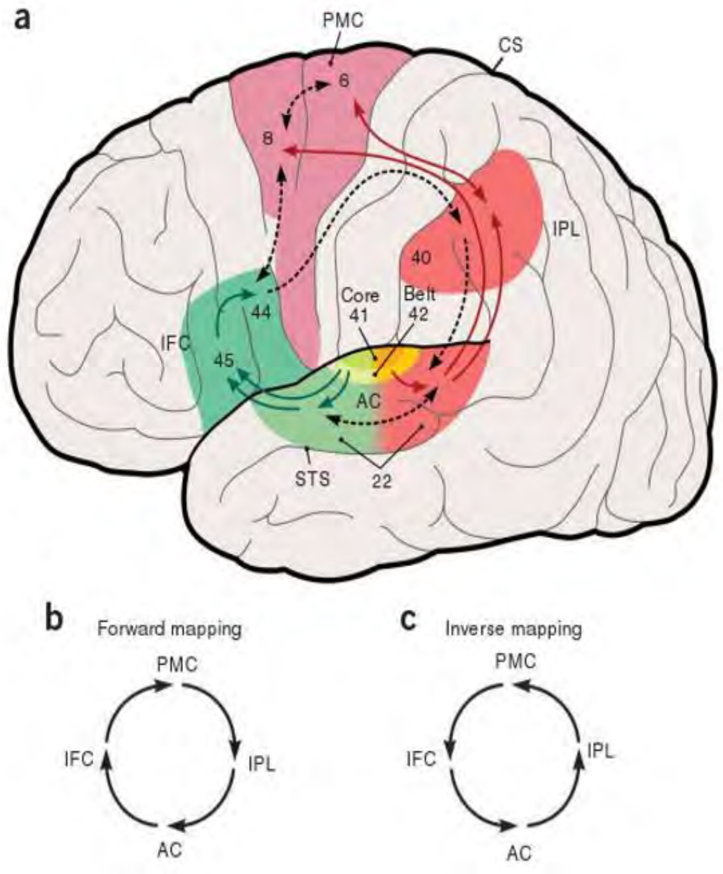
Main brain regions involved in language processing

The main anatomical structures involved in language processing are colored and labeled.



Démonet J, Thierry G, Cardebat D. Renewal of the Neurophysiology of Language: Functional Neuroimaging Physiol Rev 2005;85:49-95

Main brain regions involved in auditory processing



Dual auditory processing scheme of the human brain and the role of internal models in sensory systems.

This expanded scheme closes the loop between speech perception and production and proposes a common computational structure for space processing and speech control in the postero-dorsal auditory stream.

(a) Antero-ventral (green) and postero-dorsal (red) streams originating from the auditory belt. The postero-dorsal stream interfaces with premotor areas and pivots around inferior parietal cortex, where a quick sketch of sensory event information is compared with a predictive efference copy of motor plans.

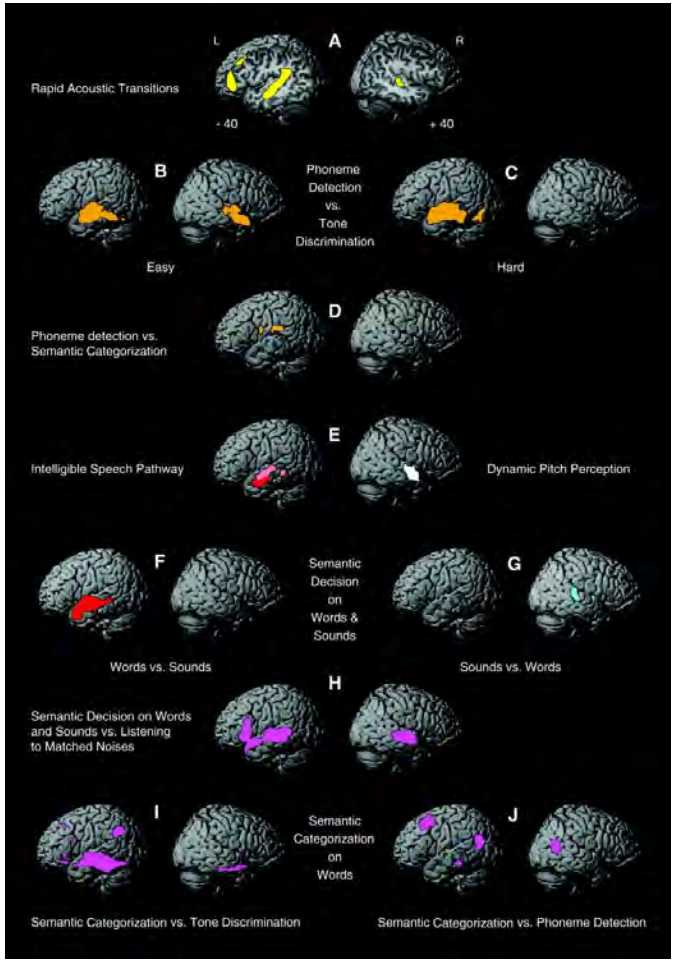
(b) In one direction, the model performs a forward mapping: object information, such as speech, is decoded in the antero-ventral stream all the way to category-invariant inferior frontal cortex (area 45), and is transformed into motor-articulatory representations (area 44 and ventral **PMC**), whose activation is transmitted to the IPL (and posterior superior temporal cortex) as an **efference copy**. **(c) In reverse direction, the model performs an inverse mapping, whereby attention- or intention-related changes in the IPL influence the selection of context-dependent action programs in PFC and PMC.**

Both types of dynamic model are testable using techniques with high temporal precision (for example, magnetoencephalography in humans or single-unit studies in monkeys) that allow determination of the order of events in the respective neural systems.

AC, auditory cortex; STS, superior temporal sulcus; **IFC**, inferior frontal cortex, PMC, premotor cortex; **IPL**, inferior parietal lobule; **CS**, central sulcus. Numbers correspond to Brodmann areas.

Rauschecker J, Scott S. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. 2009. Nat Neurosci. Vol. 12(6), pp. 718-724.

Synapsis of PET studies of auditory language processing



This figure depicts the impact of several factors on the distribution of language-related neural activities. Schematic activations are redrawn from the original studies using the significance threshold reported in each study.

- ***A*: Belin et al. showed that rapid acoustic transitions elicit more activity** in the left STG than contralateral right-sided regions.
- ***B* and *C*: Démonet et al. showed that increasing difficulty of phoneme monitoring tasks induces a left-sided asymmetry in the activity of the STG.**
- ***D*: Démonet et al. found increased activity in the dorsal pathway by contrasting phoneme monitoring** (the difficult variant) with semantic categorization (the reverse contrast is shown in *J*).
- ***E*: Scott et al. showed that intelligible speech samples activate a specific pathway along the anterior part of the left superior temporal sulcus**. Pale red depicts activations elicited by speechlike stimuli including **unintelligible** samples, and red depicts **intelligible** speech activations. The white cluster in the right hemisphere is interpreted as a correlate of dynamic pitch perception.
- ***F* and *G*: Thierry et al. found two regions specifically activated for accessing **semantic** contents from **spoken words vs. environmental sounds** (**anterior** part of the left STG) (*F*) and **environmental sounds vs. spoken words** (**posterior** part of the right STG) (*G*). Note the congruence of *E, left*, and *F*. *H*: Thierry et al. found common activations for verbal and nonverbal **meaningful** inputs relative to matched noises in **perisylvian regions.**

  I* and *J: Démonet et al. (103 and 105, respectively) showed a set of regions activated during **semantic processing compared with pure tone monitoring** (*I*) or phoneme detection (*J*). In addition to the activity in the anterior part of the left superior temporal sulcus (*E*), the “ventral” pathway involved in lexical semantic processing included an area at the **junction of the left** inferior and middle temporal gyri. The right angular gyrus appears activated in *J* and not in *I* because it was activated to the same level in semantic categorization and pure tone monitoring in *I*.
- **Why no study on singing?**

Démonet J, Thierry G, Cardebat D. Renewal of the Neurophysiology of Language: Functional Neuroimaging Physiol Rev 2005;85:49-95

Understanding voice perception

- Voices carry large amounts of socially relevant information on persons, much like 'auditory faces'. Following *Bruce and Young(1986)*'s seminal model of face perception, we propose that the cerebral processing of vocal information is organized in interacting but functionally dissociable pathways for processing the three main types of vocal information: **speech, identity**, and **affect**. The predictions of the 'auditory face' model of voice perception are reviewed in the light of recent **clinical, psychological**, and **neuroimaging evidence**.

Belin P, Bestelmeyer P. E. G, Latinus M, Watson R. Understanding voice perception. 2011. British Journal of Psychology.

Human voice perception

- We are all voice experts. First and foremost, we can produce and understand speech, and this makes us a unique species. But **in addition to speech perception, we routinely extract from voices a wealth of socially-relevant information** in what constitutes a more primitive, and probably more universal, non-linguistic mode of communication.
- Consider the following example: you are sitting in a plane, and you can hear a conversation in a foreign language in the row behind you.
- You do not see the speakers' faces, and you cannot understand the speech content because you do not know the language. Yet, an amazing amount of information is available to you.
- You can **evaluate the physical characteristics** of the different protagonists, including their **gender**, approximate **age and size,** and associate an identity to the different voices.
- You can form a good idea of the different **speaker's mood** and **affective state**, as well as more subtle cues as the perceived attractiveness or dominance of the protagonists.
- In brief, you can form a fairly detailed picture of the type of social interaction unfolding, which a brief glance backwards can on the occasion help refine — sometimes surprisingly so.
- **What are the acoustical cues** that carry these different types of vocal information? How does our brain process and analyse this information? Here we briefly review an emerging field and the main tools used in voice perception research.

Latinus M, Belin P. Human voice perception. 2011. Curr Biol. Vol.21(4), pp.143-5.

How do we recognize who is speaking?

- **The human brain effortlessly extracts a wealth of information from natural speech**, which allows the listener to both **understand** the speech message and **recognize** who is speaking.
- This article reviews behavioural and neuroscientific work that has attempted to characterise how listeners achieve speaker recognition. Behavioral studies suggest that the action of a **speaker's glottal folds** and the overall **length of their vocal tract** carry important voice-quality information. Although these cues are useful for discriminating and recognising speakers under certain circumstances, **listeners may use virtually any systematic** feature for recognition. Neuroscientific studies have revealed that speaker recognition relies upon a predominantly **right-lateralised** network of brain regions.
- Specifically, the **posterior parts of superior temporal sulcus** appear to perform some of the acoustical analyses necessary for the perception of speaker and message, whilst anterior portions may play a more abstract role in perceiving speaker identity.
- This voice-processing network is supported by direct, early connections to non-auditory regions, such as the **visual face-sensitive** area in the fusiform gyrus, which may serve to optimize person recognition.

Mathias SR, Von Kriegstein K. How do the recognize who is speaking? 2014. Front Biosci. Vol. 1(6), pp. 92-109.

How do auditory cortex neurons represent communication sounds?

- The present review aims at investigating **the role of auditory cortex in the processing of speech**, bird songs and other vocalizations, which all are spectrally and temporally highly structured sounds.
- Whereas earlier studies have simply looked for neurons exhibiting **higher firing rates** to particular conspecific vocalizations over their modified, **artificially** synthesized versions, more recent studies determined the coding capacity of temporal **spike patterns**, which are prominent in primary and non-primary areas **(and also in non-auditory cortical areas).**
- In several cases, this information seems to be correlated with **the behavioral performance** of human or animal subjects, suggesting that **spike-timing based coding strategies might set the foundations of our perceptive abilities**. Also, it is now clear that the responses of auditory cortex neurons are highly **nonlinear** and that their responses to **natural stimuli** cannot be predicted from their responses to **artificial stimuli** such as moving ripples and broadband noises.
- Since auditory cortex neurons **cannot follow rapid fluctuations** of the vocalizations envelope, they only respond at specific time points during communication sounds, which can serve as temporal markers **for integrating the temporal and spectral processing taking place at subcortical relays.**
- Thus, the **temporal sparse code** of auditory cortex neurons can be considered as a first step for generating high level representations of communication sounds **independent** of the acoustic characteristic of these sounds.

Gaucher Q, Huetz C, Gourévitch B, Laudanski J, Occelli F, Edeline JM. How do auditory cortex neurons represent communication sounds? 2013. Hear res. Vol.305, pp. 102-12.

Representation of speech in human auditory cortex: is it special?

- Successful categorization of **phonemes** in speech requires that the brain analyze the **acoustic signal** along both **spectral** and **temporal** dimensions.
- Neural encoding of the stimulus amplitude envelope is critical for parsing the speech stream into **syllabic units.**
- Encoding of voice onset time **(VOT)** and place of articulation **(POA)**, cues necessary for determining phonemic identity, occurs within shorter time frames. An unresolved question is whether the neural representation of speech is based on processing mechanisms that are unique to **humans** and shaped by learning and experience, **or** is based on rules governing general auditory processing that are also present in **non-human animals.** This question was examined by comparing the neural activity elicited by speech and other complex vocalizations in primary auditory cortex of macaques, who are limited vocal learners, with that in **Heschl's gyrus, the putative location of primary auditory cortex in humans.**
- Entrainment to the amplitude envelope is neither specific to humans nor to human speech.
- VOT is represented by responses time-locked to consonant release and voicing onset in both humans and monkeys. Temporal representation of VOT is observed both for isolated syllables and for syllables embedded in the more naturalistic context of running speech.
- **The fundamental frequency** of male speakers is represented by more rapid neural activity phase-locked to the glottal pulsation rate **in both humans and monkeys**.
  In both species, the differential representation of **stop consonants** varying in their POA can be predicted by the relationship between the frequency **selectivity of neurons** and the onset spectra of the speech sounds.
- These findings indicate that the **neurophysiology of primary auditory cortex is similar in monkeys and humans** despite their vastly different experience with human speech, and that **Heschl's gyrus is engaged in general auditory, and not language-specific, processing.**

Steinschneider M, Nourski KV, Fishman YI. Representation of speech in human auditory cortex: is it special? 2013. Hear Res. Vol. 305, pp. 57-73.

Auditory signal processing in communication: perception and performance of vocal sounds

- **Learning** and **maintaining** the sounds we use in **vocal communication** require accurate perception of the sounds we hear performed by others and **feedback-dependent imitation** of those sounds to produce **our own vocalizations.**
- **Understanding how the central nervous system integrates auditory and vocal-motor information to enable communication is a fundamental goal of systems neuroscience**, and insights into the mechanisms of those processes **will profoundly enhance clinical therapies for communication disorders.**
- Gaining the high-resolution insight necessary to define the circuits and **cellular mechanisms** underlying human vocal communication is presently impractical. **Songbirds** are the best animal model of human speech, and this review highlights recent insights into the neural basis of auditory perception and feedback-dependent imitation in those animals. **Neural correlates of song perception** are present in auditory areas, and those correlates are preserved in the auditory responses of **downstream neurons** that are also active when the bird sings. Initial tests indicate that singing-related activity in those downstream neurons is associated with **vocal-motor performance** as opposed to the bird simply hearing itself sing.
- Therefore, action potentials related to auditory perception and **action potentials related to vocal performance are co-localized in individual neurons.**
- **Conceptual models of song learning involve comparison of vocal commands and the associated auditory feedback** to compute an error signal that is used to guide refinement of subsequent song performances, **yet the sites of that comparison remain unknown. Convergence of sensory and motor activity** onto individual neurons points to a possible mechanism through which auditory and vocal-motor signals may be linked to enable learning and maintenance of the sounds used in vocal communication.

Prather Jf. Auditory signal processing in communication: perception and performances of vocal sounds. 2013. Hear Res. Vol. 305, pp.144-55.

Pathology

Hearing loss in singers: a preliminary study

- **OBJECTIVE**

  Singers need good hearing; however, they may be exposed to loud noises during their musical activities. The objectives of this study were to describe the incidence and type of **hearing loss (HL)** in singers.

- **STUDY DESIGN**

  Retrospective case cohort.

- **METHODS**

  Billing records identified patients who had undergone videostroboscopy and audiogram during the same visit over a 3 year period. A singer was defined as anyone who self-identified as a singer (professional or avocational). Age and gender matched nonsingers were used as controls. Patients with otologic diagnoses, surgery, or complaints were excluded. Retrospective chart review was conducted for the presence of HL, type of HL, and pure tones audiogram results. Statistical analysis included descriptive statistics, Students t test, chi-square test, and Fisher exact test.

- **RESULTS**

  Of 172 singers (44.7 years, 37.8% male), 31 (17.5%) had HL. Pure tone thresholds for the singers with HL subgroup at 3, 4, and 6 kHz were 21.0, 26.5, and 34.4 dB in the right, and 22.8, 30.3, and 38.8 dB in the left ear, respectively. Older age ($P = 0.000000000000001$), male gender ($P < 0.001$), longer number of years of singing ($P = 0.0000000003$), and baritone voice ($P < 0.001$) were associated with HL. **There was no association with genre of music**. When compared with controls, the incidence of HL (19.8%) was not significantly different ($\chi 2 = 0.300$, $P = 0.58$). Pure tones at 3, 4, and 6 kHz were not significantly different than controls with HL. Most common type of HL in singers was bilateral sensorineural (83.9%), which was significantly higher than controls (39.0%, $\chi 2 = 14.6$, $P < 0.001$).

- **CONCLUSIONS**

  The incidence of HL in singers was 17.5%, which was not significantly different from controls. Bilateral sensorineural HL was most common.

Hu A, Hofmann E, Davis J, Capo J, Krane N, Sataloff RT. Hearing loss in singers: a preliminary study. 2015. J Voice. Vol.29(1), pp.120-4.

Variability in voice fundamental frequency of sustained vowels in speakers with sensorineural hearing loss

- **In a previous study,** the low-frequency modulation extent (LFP) of the vocal fundamental frequency (F(0)) showed a significant increase in the presence of binaural noise masking for the **healthy individuals.**
- This study was to investigate the F(0) of subjects with sensorineural hearing loss (SNHL) using **sustained phonations** to explore the changes of F(0) modulations in SNHL. Twenty-three SNHL subjects and **14 age**-matched subjects without hearing loss were enrolled in the study. Sustained vocalizations of vowel /a/ for more than 5 seconds were digitally recorded. **The F(0) contour** of each phonation was acquired using digital signal processing. The modulation extent at different frequencies **was obtained using Fourier transformation** of F(0) contour. The LFP of F(0) (<3Hz) was significantly greater for the SNHL subjects ($P<0.001$, independent samples t test). Although the correlation analysis was limited to the **auditory-evoked brainstem response (ABR) thresholds because of their disagreement with the pure-tone thresholds** in some subjects with functional hearing disorder, the correlation between LFP and ABR thresholds was significant ($\rho=0.45$, $P=0.03$, Spearman's correlation analysis).
- The LFPs of F(0) were significantly greater for the SNHL subjects and the changes of F(0) modulations could be detected using power spectral analysis of F(0). **The method may be used for evaluation of audio-vocal feedback in SNHL.**

Lee GS. Variability in voice fundamental frequency of sustained vowels in speakers with sensorineural hearing loss. 2012. J Voice. Vol.26(1), pp. 24-9.

Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech

- In cochlear implants (CIs), acoustic speech cues, especially for **pitch, are delivered in a degraded form**. This study's aim is to **assess** whether due to degraded pitch cues, normal-hearing listeners and CI users employ **different perceptual strategies to recognize vocal emotions**, and, **if so**, how these differ.
- Voice actors were recorded pronouncing a nonce word in four different emotions: **anger, sadness, joy, and relief**.
- These recordings' pitch cues were **phonetically analyzed**. The recordings were used to test 20 normal-hearing listeners' and 20 CI users' emotion recognition.
- In congruence with previous studies, **high-arousal emotions had a higher mean pitch, wider pitch range, and more dominant pitches than low-arousal emotions.** Regarding pitch, speakers did not differentiate emotions based on valence **but on arousal**.
- **Normal-hearing listeners outperformed CI users in emotion recognition**, even when presented with CI simulated stimuli. However, only normal-hearing listeners recognized one particular actor's emotions worse than the other actors'.
- The groups behaved differently when presented with similar input, showing that they had to employ **differing strategies**.
- Considering the respective speaker's deviating pronunciation, it appears that for normal-hearing listeners, **mean pitch** is a more salient cue than pitch range, whereas CI users are biased toward **pitch range cues.**

Gilbers S, Fuller C, Gilbers D, Broersma M, Goudbeek M, Free R, Baskent D. Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. 2016. i-Perception.Vol. 6(5).

Systemic analysis of the benefits of cochlear implants on voice production

- **PURPOSE**

  To perform a systematic analysis of the research regarding **vocal characteristics of hearing impaired** children or adults with cochlear implants.

- **RESEARCH STRATEGY**

  A literature search was conducted in the databases Web of Science, Bireme, and Universidade de **São Paulo's** and CAPES' thesis and dissertations databases using the keywords voice, **voice quality**, and **cochlear implantation**, and their respective correspondents in **Brazilian** Portuguese.

- **SELECTION CRITERIA**

  The selection criteria included: title consistent with the purpose of this review; participants necessarily being children or adults with severe to profound pre-lingual or post-lingual hearing loss using cochlear implants; and data regarding participants' performance on perception and/or acoustic analysis of the voice.

- **RESULTS**

  Twenty seven papers were classified according to the levels of evidence and quality indicators recommended by the American Speech-Language-Hearing Association (ASHA). The designs of the studies were **considered of low and medium levels of evidence**. Six papers were classified as IIb, 20 as III, and one as IV.

- **CONCLUSION**

  The voice of hearing impaired children and adults with cochlear implants **has been little studied**. There is not an effective number of studies with **high evidence levels** which precisely show the effects of the cochlear implantation on the quality of voice of these individuals.

Coelho AC, Brasolotto AG, Bevilacqua MC. Systemic analysis of the benefits of cochlear implants on voice production. 2012. J Soc Bras Fonoaudiol. Vol. 24(4), pp.395-402.

Perception of music timbre by cochlear implant listeners: a multidimensional scaling study

- Several studies have shown that the ability to identify the timbre of musical instruments is reduced in cochlear implant (CI) users compared with **normal-hearing (NH)** listeners. However, most of these studies have focused on tasks that **require specific musical knowledge**.
- In contrast, the present study investigates the perception of timbre by CI subjects using a multidimensional scaling (MDS) paradigm. The main objective was to investigate whether CI subjects use the same cues as NH listeners do to differentiate the timbre of musical instruments.
- **DESIGN**

  Three groups of 10 NH subjects and one group of 10 CI subjects were asked to make dissimilarity judgments between **pairs of instrumental sounds**. The stimuli were 16 synthetic instrument tones spanning a wide range of instrument families. All sounds had the same fundamental frequency (261 Hz) and were balanced in loudness and in perceived duration before the experiment. One group of NH subjects listened to **unprocessed stimuli**. The other two groups of NH subjects listened to the same stimuli passed through a **four-channel or an eight-channel noise vocoder**, designed to simulate the signal processing performed by a **real CI**. Subjects were presented with all possible combinations of pairs of instruments and had to estimate, for each pair, the amount of dissimilarity between the two sounds. These estimates were used to construct dissimilarity matrices, which were further analyzed using an MDS model. The model output gave, for each subject group, an optimal graphical representation of the perceptual distances between stimuli (the so-called "timbre space").
- **RESULTS**

  For all groups, **the first two dimensions of the timbre space were strikingly similar** and correlated strongly with the logarithm of the attack time and with the center of gravity of the spectral envelope, respectively. The acoustic correlate of the third dimension differed across groups but only accounted for a small proportion of the variance explained by the MDS solution. Surprisingly, CI subjects and NH subjects listening to noise-vocoded simulations gave relatively more weight to the **spectral envelope** dimension and **less weight to the attack-time dimension** when making their judgments than NH subjects **listening to unprocessed stimuli.** One possible reason for the relatively higher salience of spectral envelope cues in real and simulated CIs may be that the degradation of local fine spectral details produced a **more stable spectral** envelope across the stimulus duration.
- **CONCLUSIONS**

  The internal representation of musical timbre for isolated musical instrument sounds was found **to be similar in NH and in CI listeners**. This suggests that training procedures designed to improve timbre recognition in CIs **will indeed train CI subjects to use the same cues as NH listeners**. Furthermore, NH subjects listening to noise-vocoded sounds appear to be a good model of CI timbre perception as they show the same first two perceptual dimensions as CI subjects do and also exhibit a similar change in perceptual weights applied to these two dimensions. This **last finding validates the use of simulations to evaluate and compare training procedures to improve timbre perception in CIs.**

Macherey O, Delpierre A. Perception of musical timbre by cochlear implant listerners: a multidimensional scaling study. 2013. Ear Hear. Vol.34(4), pp. 426-36.

Auditory cortex activation to natural speech and stimulated cochlear implant speech measured with functional near-infrared spectroscopy

- **The primary goal of most cochlear implant procedures is to improve a patient's ability to discriminate speech.** To accomplish this, cochlear implants are programmed so as to maximize speech understanding. However, programming a cochlear implant can be an iterative, labor-intensive process that takes place over months.
- **In this study, we sought to determine whether functional near-infrared spectroscopy (fNIRS), a non-invasive neuroimaging method which is safe to use repeatedly and for extended periods of time, can provide an objective measure of whether a subject is hearing normal speech or distorted speech.**
- We used a **140 channel fNIRS** system to measure activation within the auditory cortex in 19 normal hearing subjects while they listed to speech with different levels of intelligibility. Custom software was developed to analyze the data and compute **topographic maps** from the measured changes in oxyhemoglobin and deoxyhemoglobin concentration.
- Normal speech reliably evoked the **strongest responses within the auditory cortex**.
- Distorted speech produced **less region-specific cortical activation.**
- **Environmental sounds** were used as a control, and they produced the least cortical activation. **These data collected using fNIRS are consistent with the fMRI literature and thus demonstrate the feasibility of using this technique to objectively detect differences in cortical responses to speech of different intelligibility.**

Pollonini L, Olds C, Abaya H, Bortfeld H, Beauchamp M. S, Oghalai J. S. Auditory cortex activation to natural speech and stimulated cochlear implant speech measured with functional near-infrared spectroscopy. 2013. Hearing Research.

The use of acoustic cues for phonetic identification: effects of spectral degradation and electric hearing

- Although some cochlear implant (CI) listeners can show good **word recognition accuracy, it is not clear** how they perceive and use the various acoustic cues that contribute to phonetic perceptions.
- In this study, the use of acoustic cues was assessed for normal-hearing (NH) listeners in optimal and spectrally degraded conditions, and also for CI listeners. Two experiments tested the **tense/lax vowel** contrast (varying in **formant structure, vowel-inherent spectral change, and vowel duration**) and the **word-final** fricative voicing contrast (varying in **F1 transition, vowel duration, consonant duration**, and consonant voicing).
- Identification results were modeled using **mixed-effects logistic regression**.
- These experiments suggested that under spectrally-degraded conditions, NH listeners decrease their use of formant cues and increase their use of **durational cues**. Compared to NH listeners, CI listeners showed decreased use of spectral cues like **formant structure** and **formant change** and **consonant voicing**, and showed use of **durational cues** (especially for the fricative contrast).
- The results suggest that although NH and CI listeners may show similar accuracy on basic tests of word, phoneme or feature recognition, they may be using **different perceptual strategies** in the process.

Winn MB, Chatterjee M, Idsardi WJ. The use of acoustic cues for phonetic identification: effects of spectral degradation and electric hearing. 2012. J Acoust Soc Am. Vol.131(2), pp. 1465-79.

The role of fundamental frequency and formants in voice gender identification

- **Purpose**

  The present study examined how **fundamental frequency** (F0) and **formant frequencies contribute to the identification of gender.**

- **Method**

  Speech stimuli were synthesized from recorded voices of men and women using a formant scaling factor of 1.2 and F0 range of 100–250 Hz. Listeners who were native speakers of Cantonese were instructed to judge the perceived gender of the voice stimuli. Percent-correct gender identification of male and female stimuli at different F0–formant combinations was obtained.

- **Results**

  F0 was found to be the primary cue for gender perception and listeners showed a higher accuracy in identifying male's than female's voices.

- **Conclusions**

  The findings are consistent with results previously reported, although other acoustic cues such as voice quality may also affect gender perception.

Morris S. F. P, Manwa L. N. The role of fundamental frequency and formants in voice gender identification. 2014. Taylor&Francis. Pp.161-165

Does knowledge of medical diagnosis bias auditory-perceptual judgments of dysphonia

- **OBJECTIVE/HYPOTHESIS**

  To determine whether knowledge of medical diagnosis biases listeners with varied experience levels in their **judgments of dysphonia.**

- **STUDY DESIGN**

  Prospective, mixed experimental and comparative design.

- **METHODS**

  Twenty-six speakers with **dysphonia** and four **normal controls** provided speech recordings. Twenty novice and eight experienced clinicians **evaluated** speech samples for roughness and breathiness using 100-mm visual analog scales. **In one condition,** the speech samples were presented without diagnostic information; **in the second condition**, samples were presented in conjunction with the medical diagnosis.

- **RESULTS**

  Regardless of experience level, listeners judged the samples as significantly **more severe when the speakers' diagnoses were known**. Specifically, novice listeners (NLs) significantly increased the severity of judgments for speakers who were mildly breathy or mildly or moderately rough when diagnostic information was known. In addition, listeners in both groups judged speakers with mass lesions to be significantly rougher when diagnosis was known; this bias was not observed for speakers with other diagnoses. NLs also trended toward increasing the severity of breathiness judgments for individuals with known vocal fold paralysis but not other diagnoses.

- **CONCLUSIONS**

  Sources of bias such as **knowledge of medical diagnoses should be considered** when listeners with varied experience levels use auditory-perceptual measures to evaluate dysphonia.

Eadie T, Sroka A, Wright DR, Merati A. Does knowledge of medical diagnosis bias auditory-perceptual judgments of dysphonia? 2011. J Voice. Vol.25(4), pp. 420-9.

Functional magnetic resonance imaging study of brain activity associated with pitch adaptation during phonation in healthy women without voice disorders

- **Objectives**

  This functional magnetic resonance imaging (fMRI) study investigated the brain activity associated with pitch adaptation during phonation in healthy women without voice disorders.

- **Study Design**

  This is an **interventional prospective study.**

- **Methods**

  Sixteen healthy women (mean age: 24.3 years) participated in a blocked design fMRI experiment involving two phonation (**comfortable phonation** and **high-pitched phonation**) and **exhalation (prolonged exhalation) tasks**. ***BrainVoyager*** *QX* Version 2.4 software was used for group-level general linear model analysis ($q[FDR] < 0.05$).

- **Results**

  Analyses showed a significant **main effect of phonation with pitch adaptation** compared with rest period in the **bilateral precentral gyrus, superior frontal gyrus, posterior cingulate gyrus, superior and middle temporal gyrus, insula and cerebellum, left middle and inferior frontal gyrus, right lingual gyrus, cingulate gyrus, and thalamus**.

  Statistical results also identified a significant main effect of exhalation compared with rest period in the bilateral precentral gyrus, cerebellum, right lingual gyrus, thalamus, and left supramarginal gyrus.

  In addition, a significant main effect of phonation was found in the **bilateral superior temporal gyrus and right insula**, as well as **in the left midbrain periaqueductal gray** for high-pitched phonation only.

- **Conclusions**

  We demonstrated that a blocked design **fMRI is sensitive** enough to define a widespread network of activation associated with phonation involving **pitch variation**. The results of this study will be implemented in our future research on phonation and its disorders.

Kryshtopava M, Lierde K. V, Meerschman I, D'Haeseleer E, De Moor M, Vandemaele P, Vingerhoets G, Claeys S. Functional magnetic resonance imaging study < of brain activity asiciated with piycj adaptation during phonation healthy women withous voice disorders. 2016. J Voice. Vol.28(6), pp.688-93.

Voice outcomes of laryngopharyngeal reflux treatment: a systemic review of 1483 patients

- The aim of this study is to explore voice quality modifications in laryngopharyngeal reflux (LPR) disease and to understand better the pathophysiological mechanisms underlying the **development of communicative disability.**
- Biological Abstracts, BioMed Central, Cochrane database, PubMed and Scopus were assessed for subject headings using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) recommendations. Relevant studies published between January 1990 and December 2015 describing the evaluation of voice quality in LPR disease were retrieved. Issues of clinical relevance, such as LPR diagnosis method, treatment efficacy and outcomes, were evaluated for each study. We determined the grade of recommendation for each publication according to the Oxford Centre for Evidence-Based Medicine evidence levels. The search identified **145 publications,** of which 25 studies met the inclusion criteria for a total of 1483 LPR patients. Data were extracted by 2 independent physicians who identified 16 trials with a IIb evidence level, 7 trials with a IIa evidence level and 2 RCTs with a Ib evidence level where 4 patient-based instruments and 5 clinician-based instruments were used.
- The main voice assessment outcomes reported were **hoarseness** assessments by physicians or patients, followed by acoustic parameters; 15 and 14 articles, respectively, demonstrated significant improvements in subjective and objective voice assessments after treatment.
- The methodology used to measure acoustic parameters (i.e. **sustained vowel duration**, the sample portion choice for measurement, etc.) varied from one study to another. The majority of studies indicated that voice quality assessments (especially acoustic parameters) remain an interesting outcome to measure the effectiveness of treatment, but further studies using standardised and **transparent methodology** to measure acoustic parameters are necessary to confirm the place of each tool in the LPR disease evaluation.

Lechien JR, Finck C, Costa de Araujo P, Huet K, Delvaux V, Piccaluga Nm Harmegnies B, Saussez S. Voice outcomes of laryngopharyngeal reflux treatment: a systemic review of 1483 patients. 2016. Eur Arch Otorhinoaryngol. Vol.273, pp. 1-23.

Relationship between acoustic measurements and self-evaluation in patients with voice disorders

- **OBJECTIVE**

  The study aimed to determine whether **there is a relationship between acoustic measures and self-evaluation** in patients with voice disorders.

- **STUDY DESIGN**

  This is a descriptive, transversal, and observational study.

- **METHODS**

  Patients (257) who answered the Voice Handicap Index protocols (VHI) and the Voice Symptoms Scale (VoiSS) and recorded the vowel /ε/ were included. Standard deviation (SD) measures of the fundamental frequency (F_0), jitter, shimmer, and the glottal to noise excitation ratio (GNE) vowel /ε/ were taken.

- **RESULTS**

  There was a weak positive correlation between all scores of VoiSS and the SD of the F_0 and jitter. The overall scores, physical limitation, and VoiSS showed weak positive correlations with shimmer. The overall scores, limitation, and emotional VoiSS showed weak negative correlations with the GNE. The VHI did not correlate with any of the acoustic measurements. There was no difference in the mean of the acoustic measures of the SD of F_0, jitter, and GNE because of a voice problem detected from the cutoff points of VoiSS. There was no difference in any of the acoustic measurements when patients with and without voice problems were compared from VHI cutoffs.

- **CONCLUSIONS**

  There is a correlation between the scores of VoiSS and acoustic measurements. Patients with self-reported voice problems in VoiSS present greater deviations in acoustic measures, mainly in jitter. **There is no correlation between the VHI scores and the acoustic measures and no difference in the averages of these measures between patients with and without voice problems detected from the VHI cutoffs.**

Lopes, L. W, Silva J. D, Simoes L. B, Evangelista D. S, Silva P. O. C, Almeida A. A, Lima-Silva M. F. Relationship between acoustic measurements and self-evaluation in patients with voice disorders. 2016. J Voice.

Vocal performance of group fitness instructors before and after instruction: changes in acoustic measures and self-ratings

- **OBJECTIVES**

  (1) To quantify acute changes in acoustic parameters of the voices of group fitness instructors (GFIs) before and after exercise instruction. (2) To determine whether these changes are discernible perceptually by the instructor.

- **STUDY DESIGN**

  This is a **pilot prospective cohort study.**

- **METHODS**

  Participants were six female GFIs, based in Brisbane, Australia. Participants performed a series of vocal tasks before and after instruction of a 60-minute exercise class. Data were obtained pertaining to fundamental frequency (pitch), intensity (volume), jitter, shimmer, harmonic-to-noise ratio (HNR), maximum duration of sustained phonation (MDSP), and pitch range. Additionally, self-ratings of voice quality were obtained before and after instruction. Data were analyzed using the Wilcoxon signed rank test.

- **RESULTS**

  Significant increases ($P \leq 0.05$) were found in fundamental frequency and intensity after instruction. No significant changes in jitter, shimmer, HNR, or MDSP were found before and after instruction. For the group, no significant change in self-ratings of voice quality occurred before and after instruction.

- **CONCLUSIONS**

  Statistically significant changes in **pitch and volume** were found on acoustic analysis. However, these subtle changes remained **within the limits of what is considered normal** and representative of the participant's age and gender. Further research into the effects of exercise instruction on the voice is needed.

Dallaston K, Rumbach AF. Vocal performance of group fitness instructors before and after instruction: changes in acoustic measure and self-ratings. 2016. J Voice. Vol.30(1),pp.127-8.

Subcortical modulation in auditory processing and auditory hallucinations

- Hearing perception in individuals with auditory hallucinations has not been well studied. Auditory hallucinations have previously been shown to involve primary auditory cortex activation. This activation suggests that auditory hallucinations activate the terminal of the auditory pathway as if auditory signals are submitted from the cochlea, and that a hallucinatory event is therefore perceived as hearing. **The primary auditory cortex is stimulated by some unknown source that is outside of the auditory pathway.**
- The current study aimed to assess the outcomes of stimulating the primary auditory cortex through the auditory pathway in individuals who have experienced auditory hallucinations. **Sixteen patients with schizophrenia** underwent functional magnetic resonance imaging **(fMRI)** sessions, as well as hallucination assessments. During the fMRI session, auditory stimuli were presented in one-second intervals at times when scanner noise was absent. Participants listened to auditory stimuli of sine waves (SW) (4-5.5kHz), English words (EW), and acoustically reversed English words (arEW) in a block design fashion. The arEW were employed to deliver the sound of a human voice with minimal linguistic components. Patients' auditory hallucination **severity was assessed by the auditory hallucination item** of the Brief Psychiatric Rating Scale (BPRS). During perception of arEW when compared with perception of SW, **bilateral activation of the globus pallidus correlated with severity of auditory hallucinations.**
- EW when compared with arEW did not correlate with auditory hallucination severity. Our findings suggest that the **sensitivity of the globus pallidus to the human voice is associated with the severity of auditory hallucination.**

Ikuta T, DeRosse P, Argyelan M, Karlsgodt KH, Kingsley PB, Szeszko PR, Malhotra AK. Subcortical modulation in auditory processing and auditory hallucinations. 2015. Behav Brain Res. Vol. 295, pp.78-81.

Objective measurement of high level auditory cortical function in children

- **OBJECTIVE**

  This study examined whether the **N2 latency of the cortical auditory evoked potential (CAEP)** could be used as an objective **indicator of temporal processing ability in normally hearing children.**

- **METHODS**

  The N2 latency was evoked using three temporal processing paradigms: (1) differences in **voice-onset-times** (VOTs); (2) **speech-in-noise** using the CV/da/embedded in broadband noise (BBN) with varying signal-to-noise ratios (SNRs); and (3) 16Hz amplitude-modulated (AM) BBN presented (i) alone and (ii) following an unmodulated BBN, using four modulation depths. **Thirty-four** school-aged children with normal hearing, speech, language and reading were stratified into two groups: **5-7 years** (n=13) and **8-12 years** (n=21).

- **RESULTS**

  The N2 latency shifted significantly and systematically with differences in VOT and SNR, and was significantly different in the two AM-BBN conditions.

- **CONCLUSIONS**

  For children without an N1 peak in the cortical waveform, the N2 peak can be used as a sensitive measure of temporal processing for these stimuli.

- **SIGNIFICANCE**

  N2 latency of the CAEP can be used as **an objective measure of temporal processing ability** in a paediatric population with temporal processing disorder who are difficult to assess via behavioural response.

Almeqbel A, McMahon C. Objective measurement of high-level auditory cortical function in children. 2015. Int J pediatr Otorhinolaryngol. Vol. 79(7), pp. 1055-62.

Voice emotion recognition by cochlear-implanted children and their normally-hearing peers

- Despite their remarkable success in bringing spoken language to hearing impaired listeners, the signal transmitted through cochlear implants (CIs) remains impoverished in **spectro-temporal fine structure**. As a consequence, **pitch-dominant information such as voice emotion**, is diminished. For young children, the ability to correctly identify the mood/intent of the speaker (which may not always be visible in their facial expression) is an important aspect of **social and linguistic development**.
- Previous work in the field has shown that children with **cochlear implants (cCI)** have significant deficits in voice emotion recognition relative to their normally hearing peers (cNH).
- Here, we report on voice emotion recognition by a cohort of **36 school-aged cCI**. Additionally, we provide for the first time, a comparison of their performance to that of cNH and NH adults (aNH) listening to CI simulations of the same stimuli. We also provide comparisons to the performance of adult listeners with CIs (aCI), most of whom learned language primarily through normal acoustic hearing. Results indicate that, despite strong variability, on average, **cCI perform similarly to their adult counterparts**; that both groups' mean performance is similar to aNHs' performance with 8-channel noise-vocoded speech; that cNH achieve excellent scores in **voice emotion recognition** with full-spectrum speech, but on average, **show significantly poorer scores** than aNH with 8-channel noise-vocoded speech. A strong developmental effect was observed in the cNH with noise-vocoded speech in this task.
- These results point to the considerable benefit obtained by cochlear-implanted children from their devices, but also underscore the need for further research and development in this important and neglected area.

Chatterjee M, Zion DJ, Deroche ML, Burianek BA, Limb CJ, Goran AP, Kulkarni AM, Christensen JA. Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. 2015. Hear res. Vol.322, pp. 151-62.

Acoustic properties of vocal singing in prelingually-deafened children with cochlear implants or hearing aids

- **OBJECTIVE**

  The purpose of the present study was to investigate vocal **singing performance** of hearing-impaired children with cochlear implants (CI) and hearing aids (HA) as well as to evaluate the relationship between **demographic** factors of those hearing-impaired children and their singing ability.

- **METHODS**

  Thirty-seven prelingually-deafened children with CIs and **31 prelingually-deafened children with HAs, and 37 normal-hearing (NH)** children participated in the study. The fundamental frequencies **(F0)** of each note in the recorded songs were extracted and the **duration of each sung note was measured.** Five metrics were used to evaluate the **pitch-related and rhythm-based** aspects of singing accuracy.

- **RESULTS**

  Children with CIs and HAs showed significantly **poorer performance in either the pitch-based assessments or the rhythm-based measure** than the NH children. No significant differences were seen between the CI and HA groups in all of these measures except for the mean deviation of the **pitch intervals**. For both hearing-impaired groups, **length of device use was significantly correlated with singing accuracy.**

- **CONCLUSIONS**

  There is a marked deficit in vocal singing ability either in pitch or rhythm accuracy in a majority of prelingually-deafened children who have received CIs or fitted with HAs. Although an increased length of device use might facilitate singing performance to some extent, the chance for the hearing-impaired children fitted with either HAs or CIs to reach high proficiency in singing is **quite slim**.

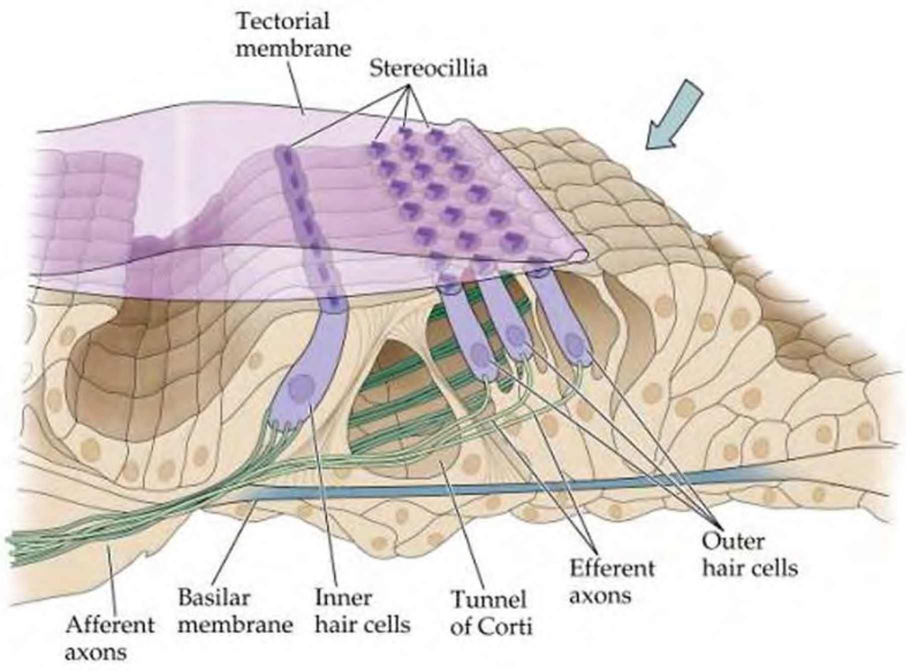
Mao Y, Zhang M, Nutter H, Zhang Y, Zhou Q, Liu Q, Wu W, Xie D, Xu L. Acoustica properties of vocal singin in prelingually-deafened children with cochlear implants or hearing aids. 2013. Int J pediatr otorhinolaryngol. Vol.77(11), pp. 1833-40.

Auditory perceptual analysis of voice in abused children and adolescents

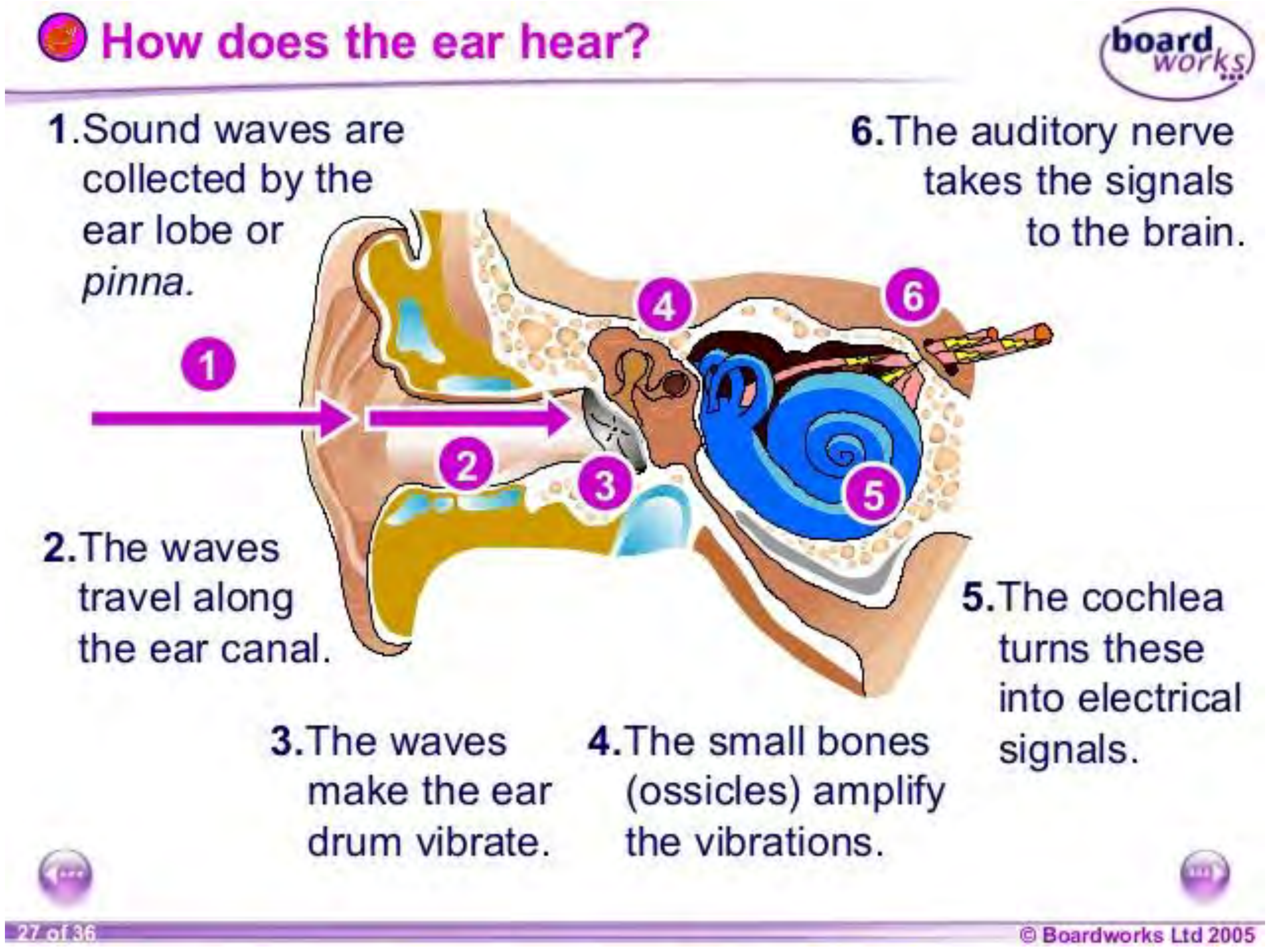
- This study aimed to analyze the prevalence of vocal changes in abused children and adolescents. Through auditory-perceptual analysis of voice and the study of the association between vocal changes, communication disorders, psychiatric disorders, and global functioning.
- The prevalence of vocal change was greater than that observed in general population, with significant associations with communication disorders and global functioning. The results demonstrate that the situations these children experience can intensify **the triggering of abusive vocal behaviors** and consequently, of vocal changes.

Stivanin L, Santos FP, Oliveira CC, Santos Bd, Ribeiro ST, Scivoletto S. Auditory perceptual analysis of voice in abused children and adolescents. 2015. Braz J Otorhinolaryngology. Vol. 81(1), pp. 71-8.

Ekstra billeder



http://www.prosoundweb.com/article/perception_is_reality_psychoacoustics_from_an_audio_engineers_perspective/P2/



http://www.slideshare.net/cartlidge/8-l-sound-hearing-boardworks