Hearing Things

How Your Brain Works

Prof. Jan Schnupp
wschnupp@cityu.edu.hk

HowYourBrainWorks.net//hearing

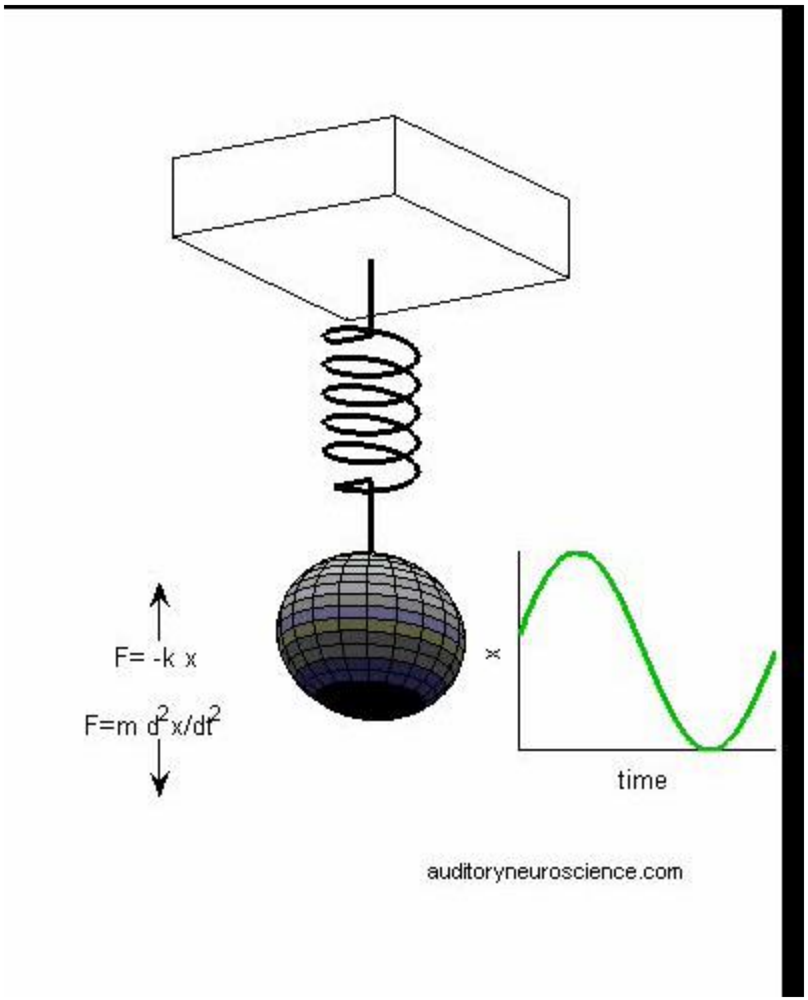
香港城市大學
City University of Hong Kong

1) Things Make Sounds, and Different Things Make Different Sounds

Sound Signals

- Many physical objects emit sounds when they are “excited” (e.g. hit or rubbed).
- Sounds are just pressure waves rippling through the air, but they carry a lot of information about the objects that emitted them.
  (Example: what are these two objects? Which one is heavier, object A or object B ?)
- The sound (or signal) emitted by an object (or system) when hit is known as the *impulse response.*
- Impulse responses of everyday objects can be quite complex, but the *sine wave* is a fundamental ingredient of these (or any) complex sounds (or signals).

Vibrations of a Spring-Mass System



Undamped

1. $F = -k \cdot y$ (Hooke's Law)
2. $F = m \cdot a$ (Newton's 2^{nd})
3. $a = dv/dt = d^2y/dt^2$

$\Rightarrow -k \cdot y = m \cdot d^2y/dt^2$

$\Rightarrow y(t) = y_o \cdot \cos(t \cdot \sqrt{k/m})$

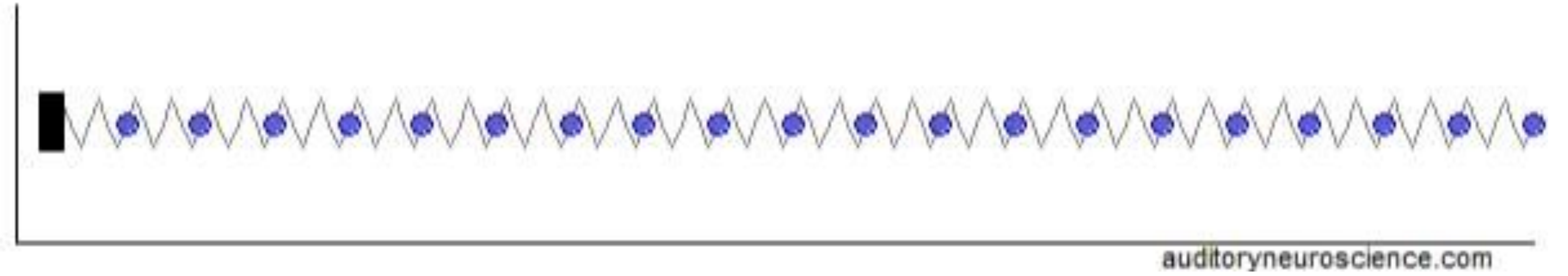
Damped

4. $-k \cdot y - r\, dy/dt = d^2y/dt^2$

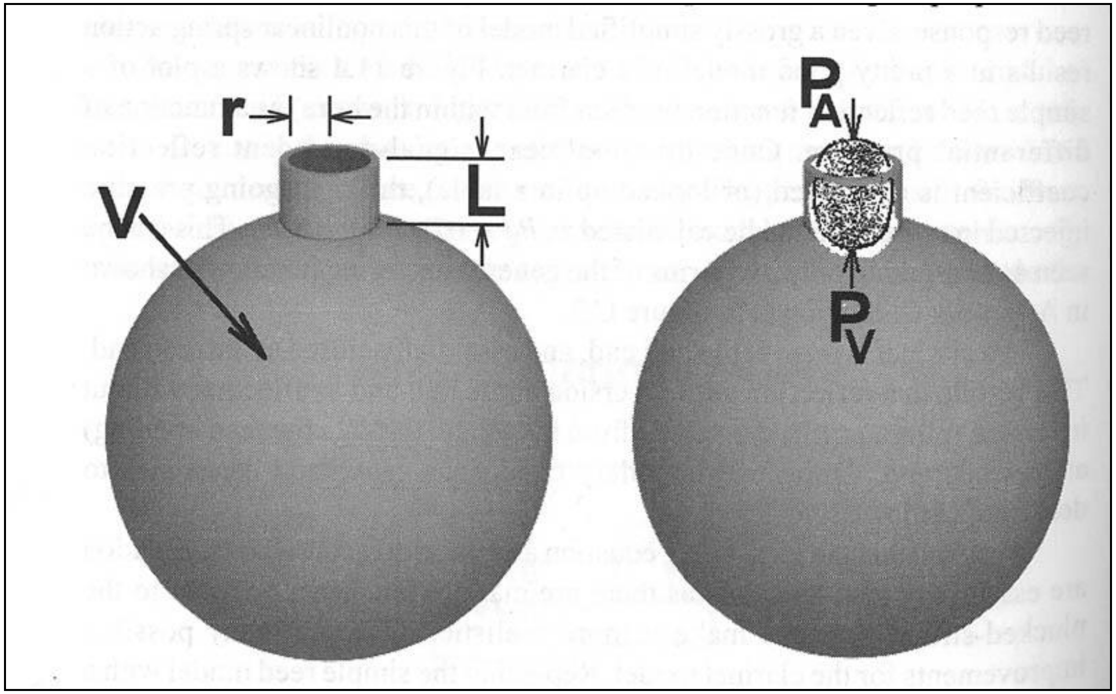
Don't worry about the formulae! Just remember that mass-spring systems like to vibrate at a rate proportional to the square-root of their "stiffness" and inversely proportional to their weight.

https://auditoryneuroscience.com/shm

Sound wave propagation



Resonant Cavities



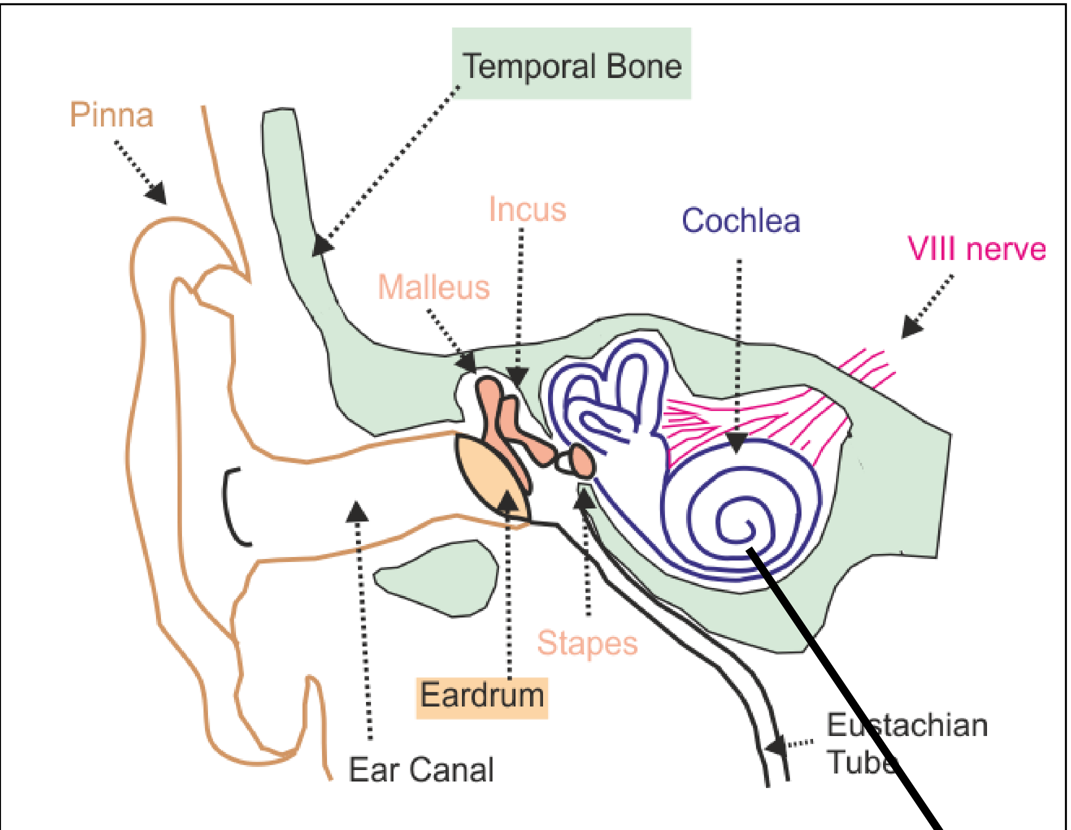
- In resonant cavities, “lumps of air” at the entrance/exit of the cavity oscillate under the elastic forces exercised by the air inside the cavity.
- The preferred resonance frequency is inversely proportional to the square root of the volume. (Large resonators => deeper sounds).

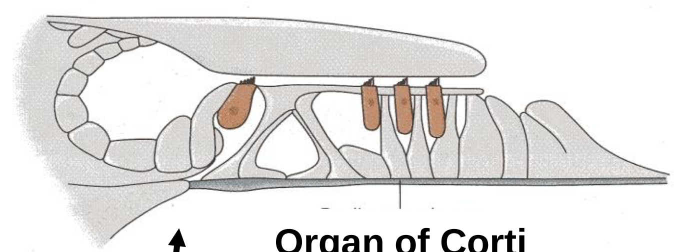
2) The Ear as a Frequency Analyser

The Ear as a Frequency Analyser

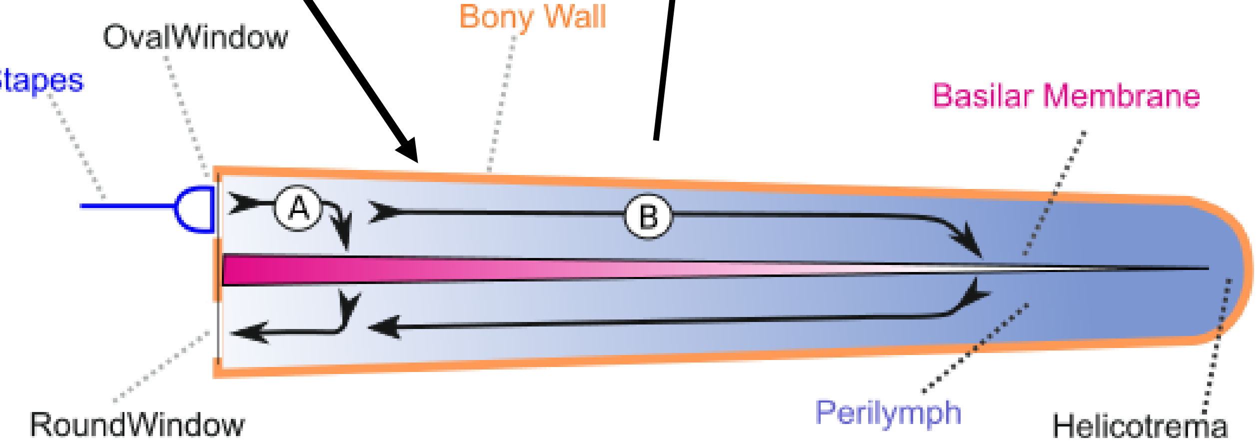
- As we have just seen, physical properties of objects, such as size, mass, stiffness, are reflected in the frequency spectra they emit when they make sounds.
- An important job of the ear is to perform a time-frequency analysis of incoming sounds.
- This results in:
- A place code for frequency (tonotopy)
- A rate code for intensity
- A time code for temporal structure (including “fine structure”)

The Ear





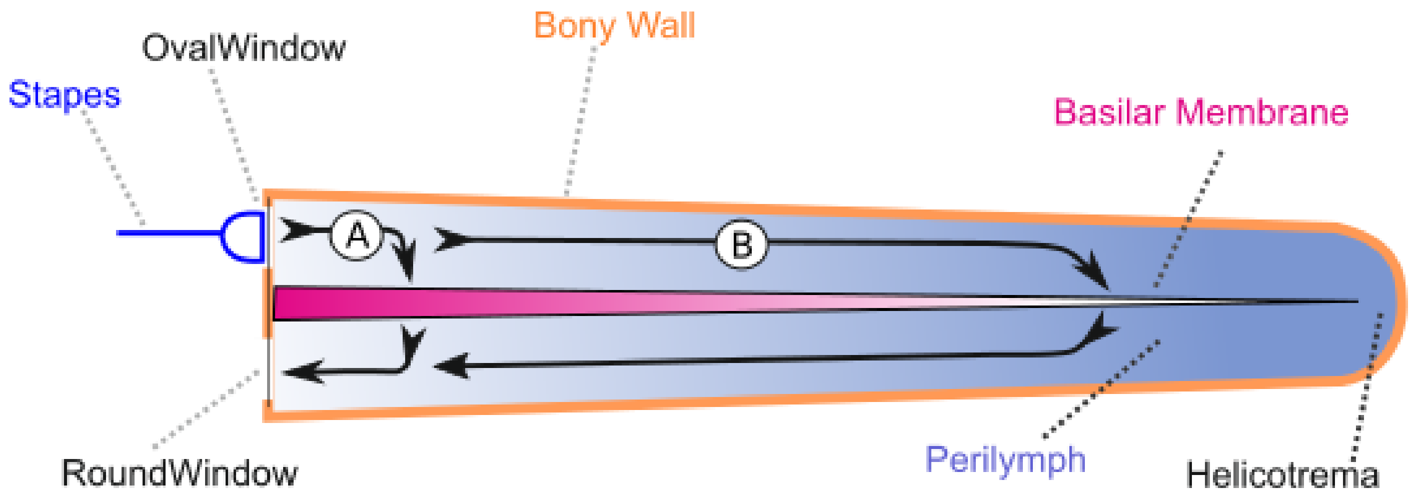
Organ of Corti



Cochlea “unrolled”

auditoryneuoscence.com

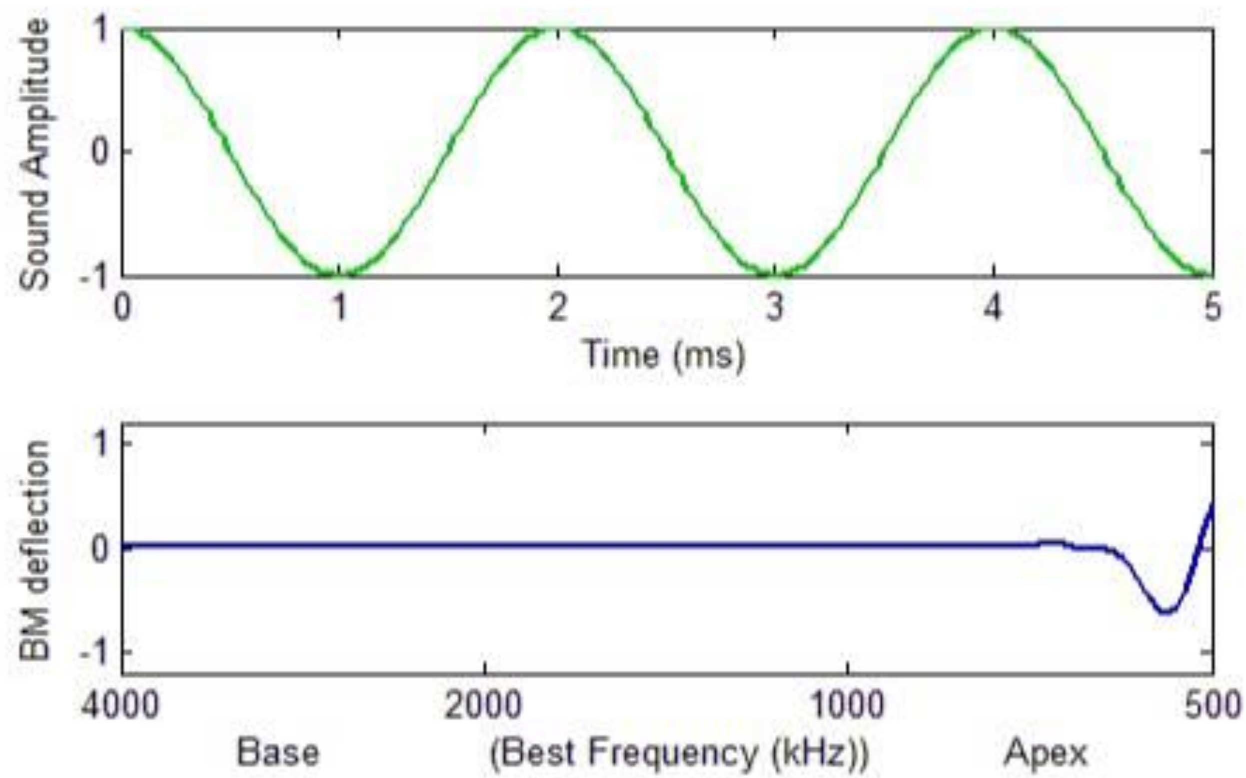
Basilar membrane mechanics: a trade-off between stiffness and inertia



auditoryneuoscence.com

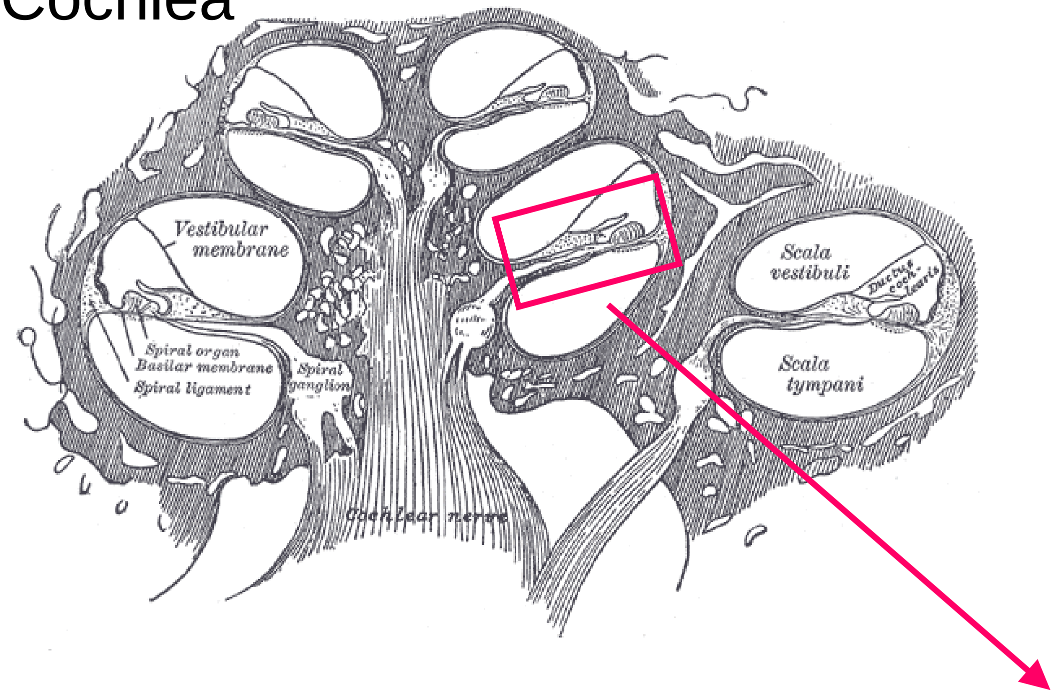
- The basilar membrane is stiff at the base and floppy at the apex.
- Vibrations travelling from the stapes to the round window must either take a short route through stiff membrane or a longer route through more inert fluid.
- -> a mass-spring filter bank.

Basilar Membrane Tuning Animation

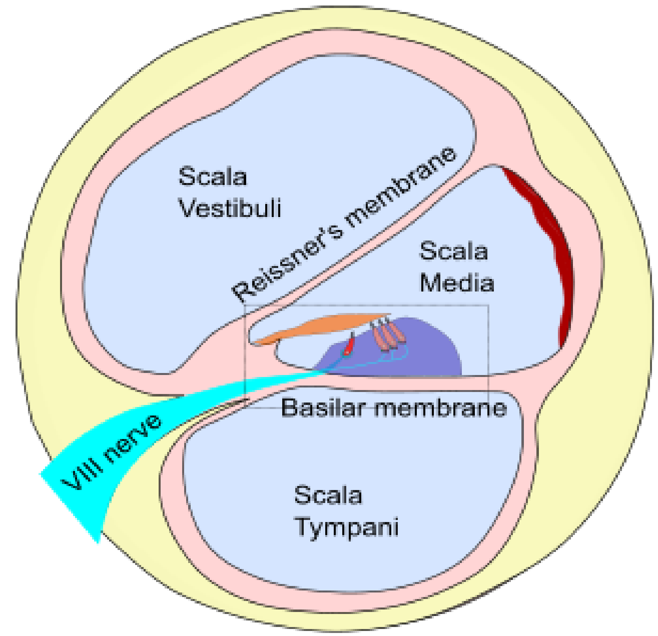


See auditoryneuroscience.com | The Ear

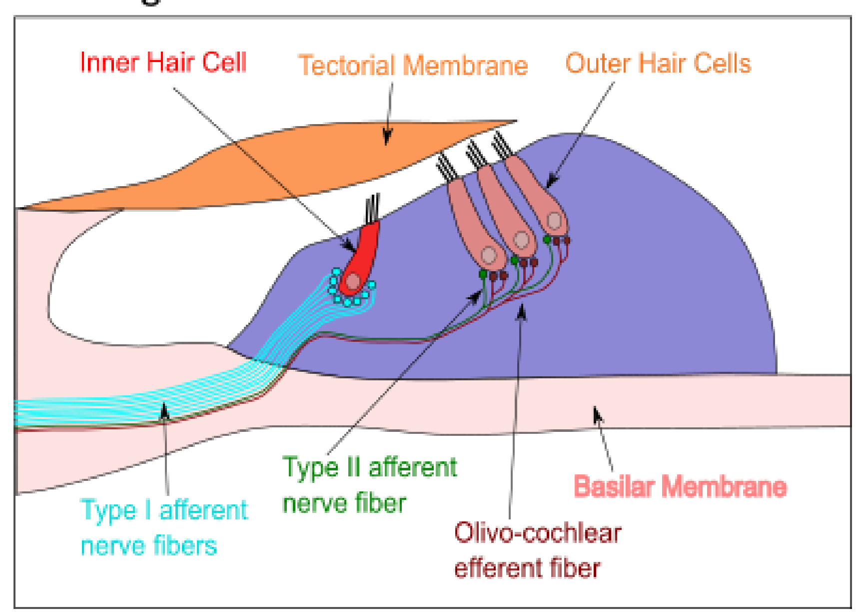
Cochlea



A Cochlear Transsection

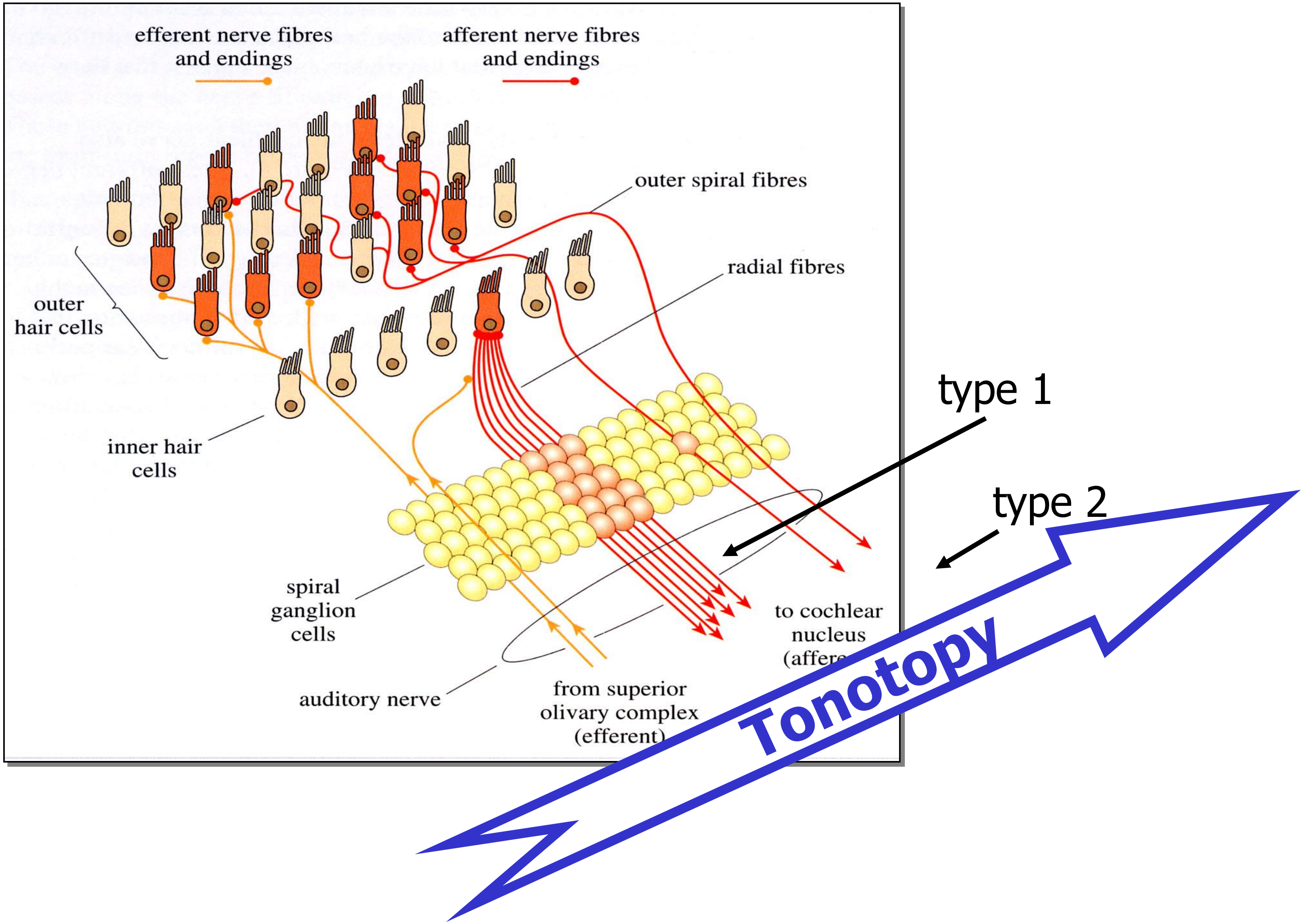


B Organ of Corti

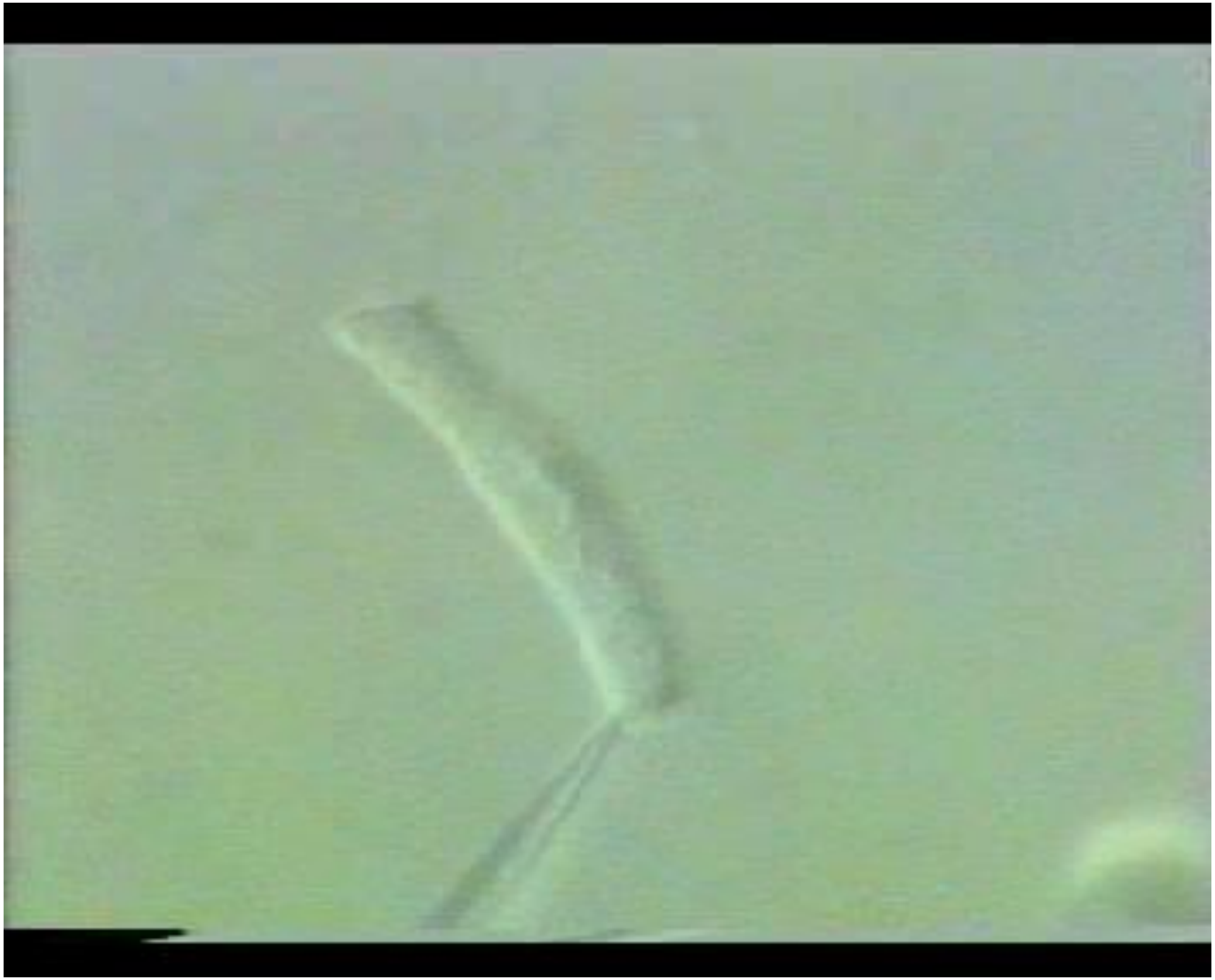


Sound Transduction

Afferent Innervation of the Hair Cells



The Outer Hair Cell's Special Trick



https://auditoryneuroscience.com/ear/dancing-outer-hair-cell

Hair Cells

- Hair cells in the Organ of Corti transduce the mechanical vibration of the basilar membrane into electrical signals.
- Inner hair cells transmit these signals to auditory nerve fibres.
- Outer hair cells are mechanical feedback devices which amplify the signal on a tuneable manner.

3) Real World Sounds are Rich in Frequencies

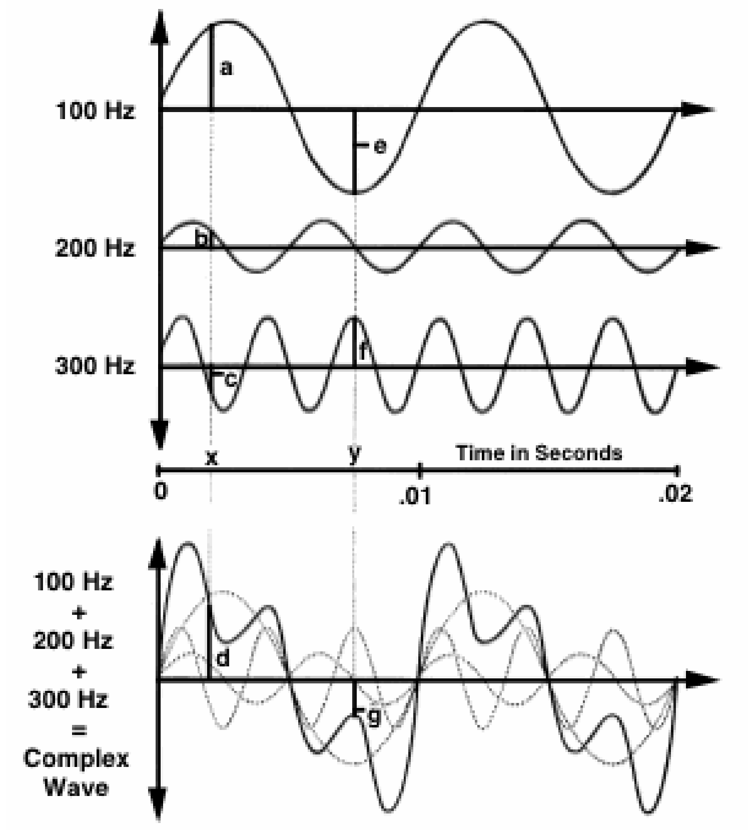
The Trouble with Sound Frequency 1

- **The word “frequency” can mean different things in different contexts**. The frequency with which I walk my dog or brush my teeth is simply the number of events per unit time. But sound waves don’t have obvious events that should count and non-events that don’t count.
- Consider the sound wave here. Imagine it is 0.1 s long. Are there 4 “events” and therefore its “frequency” is 40 Hz? Or do I need to count each little peak (of which there are 38) and its frequency is 380 Hz?

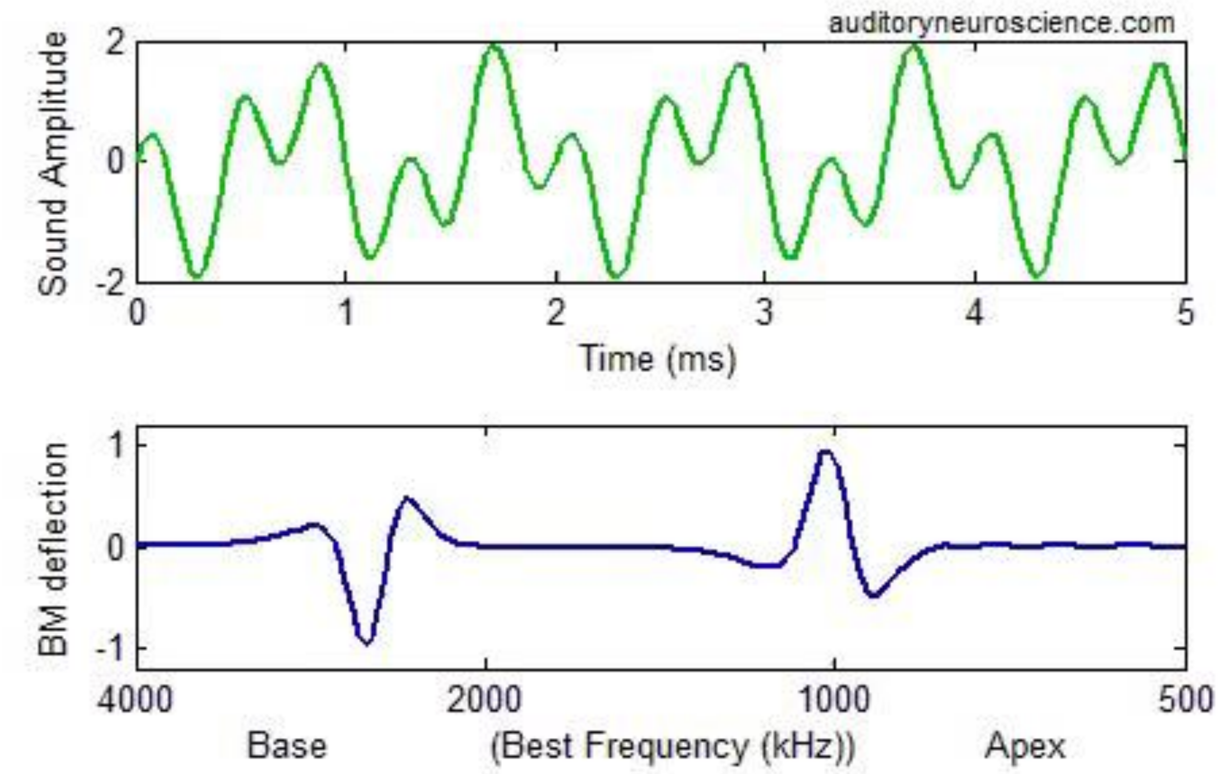
- To avoid such ambiguity, **when hearing researchers talk about sound frequencies, they usually don’t refer to “event rates” but to “Fourier (sine wave) components”.**

The Trouble with Sound Frequency 2

- Mathematicians often deal with the issue of frequencies of a complex wave shape by using “Fourier’s theorem”, which says that any complex wave can be produced by summing together several (sometimes very many!) sine waves.
- In terms of their “Fourier spectra”, such waves then don’t have just one frequency, but several and potentially very many: one of each Fourier component sine wave that has non-zero amplitude.



“Fourier Analysis” by the Basilar Membrane?



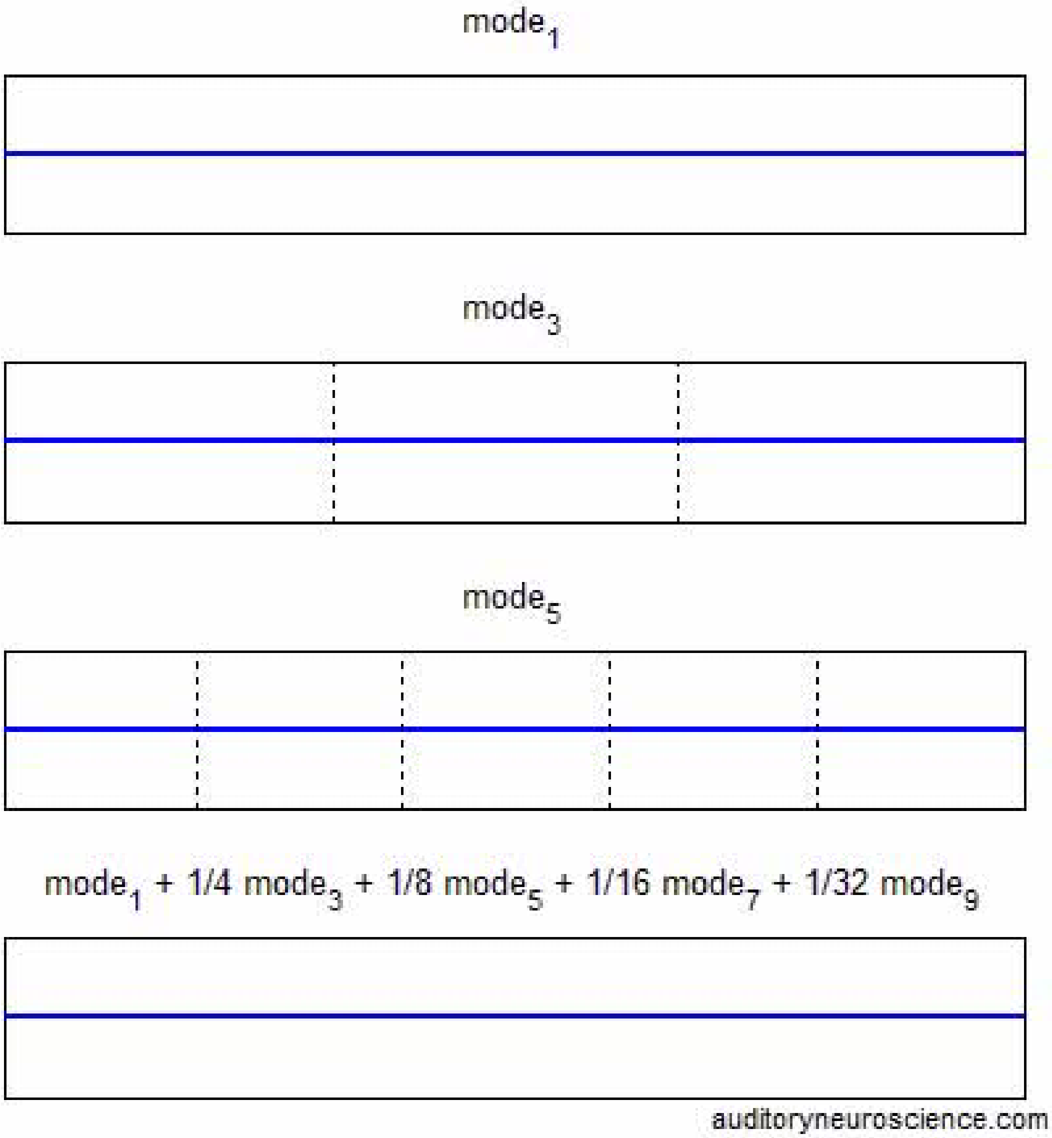
BM response to a “complex tone” made from a 1000 Hz and a 3000 Hz sine wave.

To some extent, frequency filtering by the basilar membrane can “decompose” complex sounds into its constituent frequency components. However, if a complex sound had many frequency components, the spatial resolution of this “**place code for frequency**” becomes quickly overwhelmed.
http://auditoryneuroscience.com/ear/bm1-travelling-wave

More about Real World Sounds

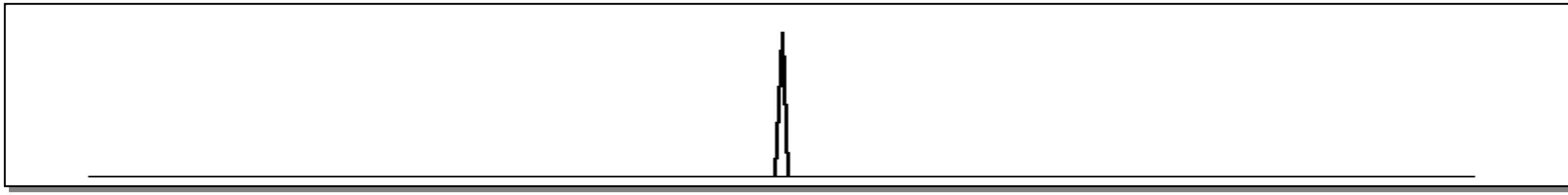
- Physical Objects tend to have many “Modes of Vibration”.
- This leads to sounds with many frequency components “**overtones**” which are often “**harmonics**”.
- “Harmonic” means the frequency components are all integer multiples of a lowest, common “fundamental” frequency.
- Periodic pulse trains are also harmonic
- Irregular, non-periodic sounds don’t have harmonic structure and are noise-like.
- Human speech comprises both periodic (“voiced”) and non-periodic, noise-like (“unvoiced”) speech sounds.

Many Real-World Objects Vibrate at Multiple Frequencies

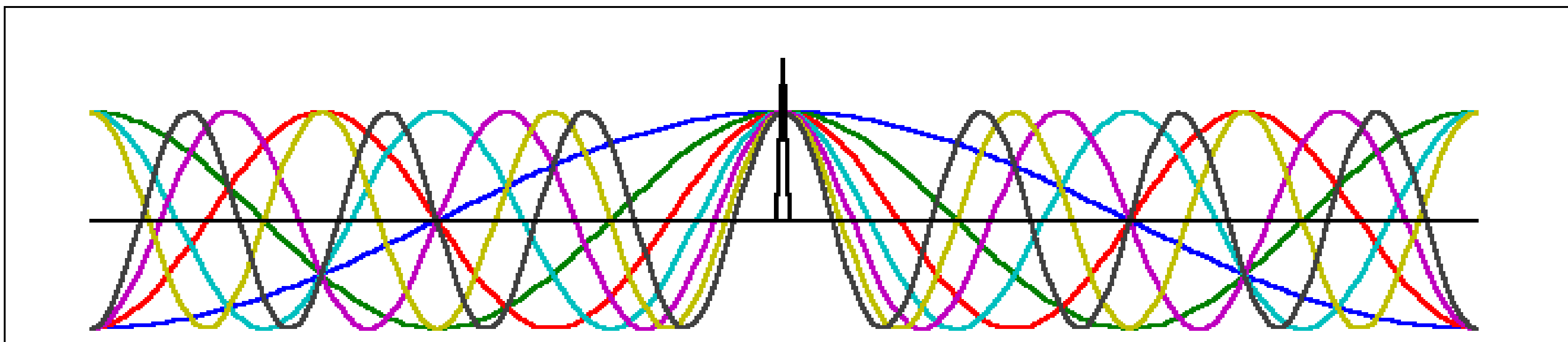


http://auditoryneuroscience.com/acoustics/modes_of_vibration

The Impulse (or “Click”)

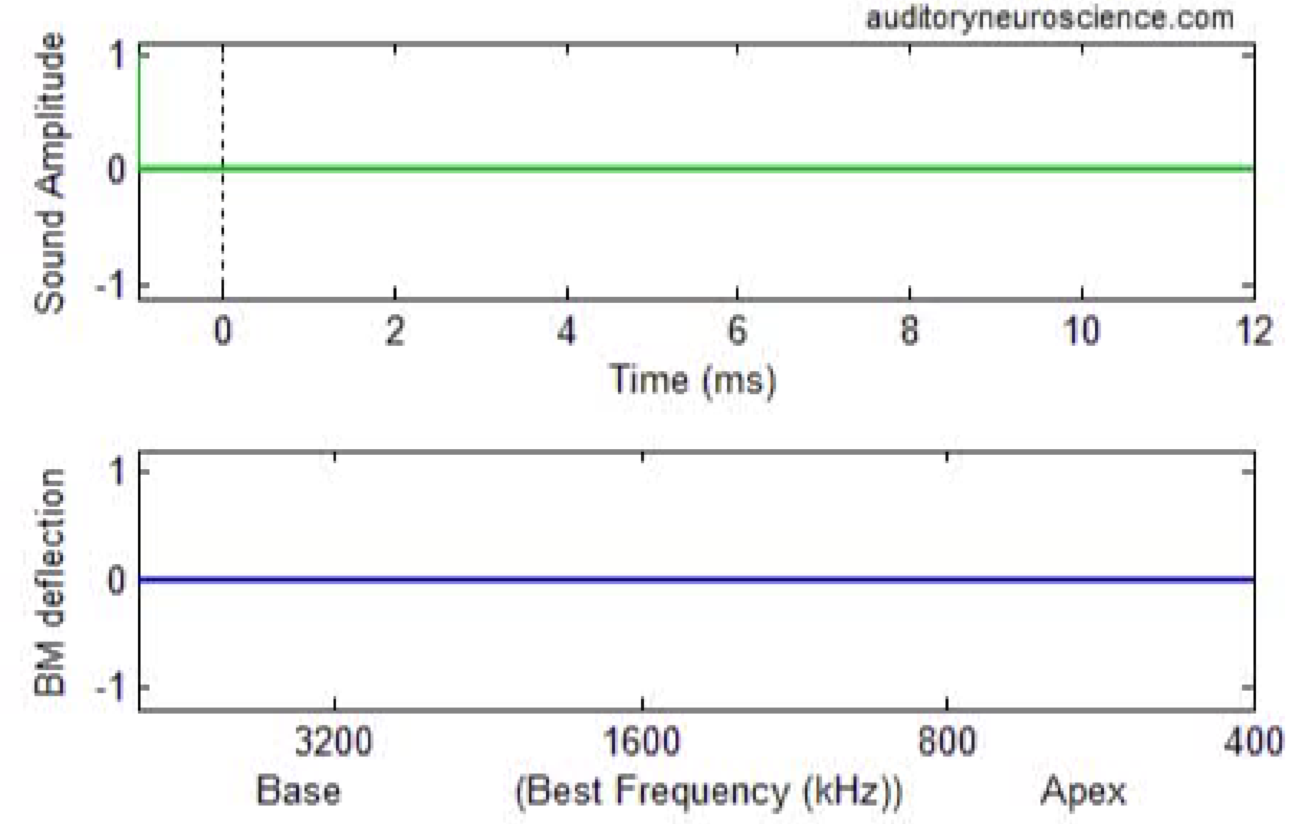


- The “ideal click”, or impulse, is an infinitesimally short signal.

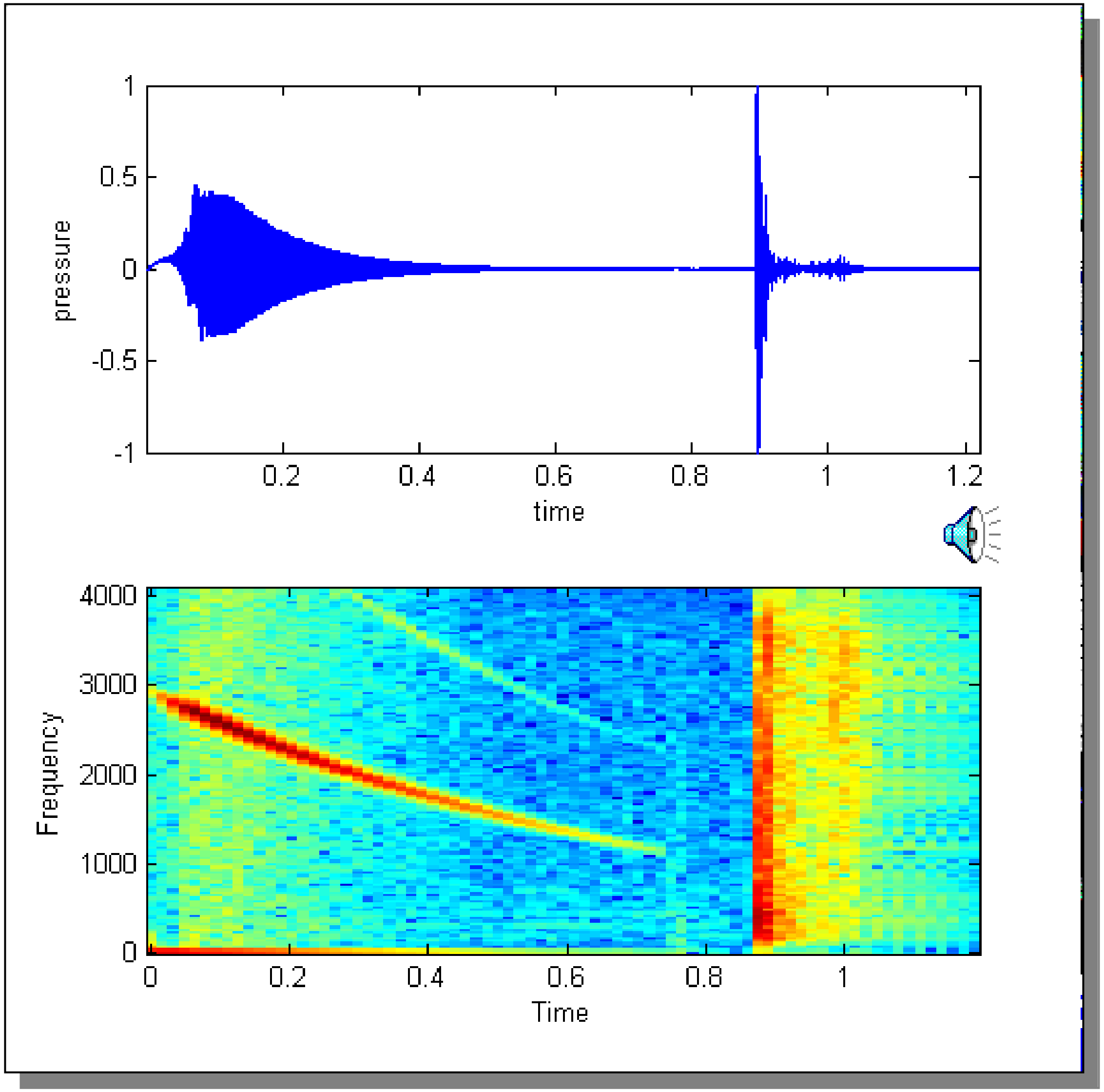


- The Fourier Transform encourages us to think of this click as an infinite series of sine waves, which have started at the beginning of time, continue until the end of time, and all just happen to pile up at the one moment when the click occurs.

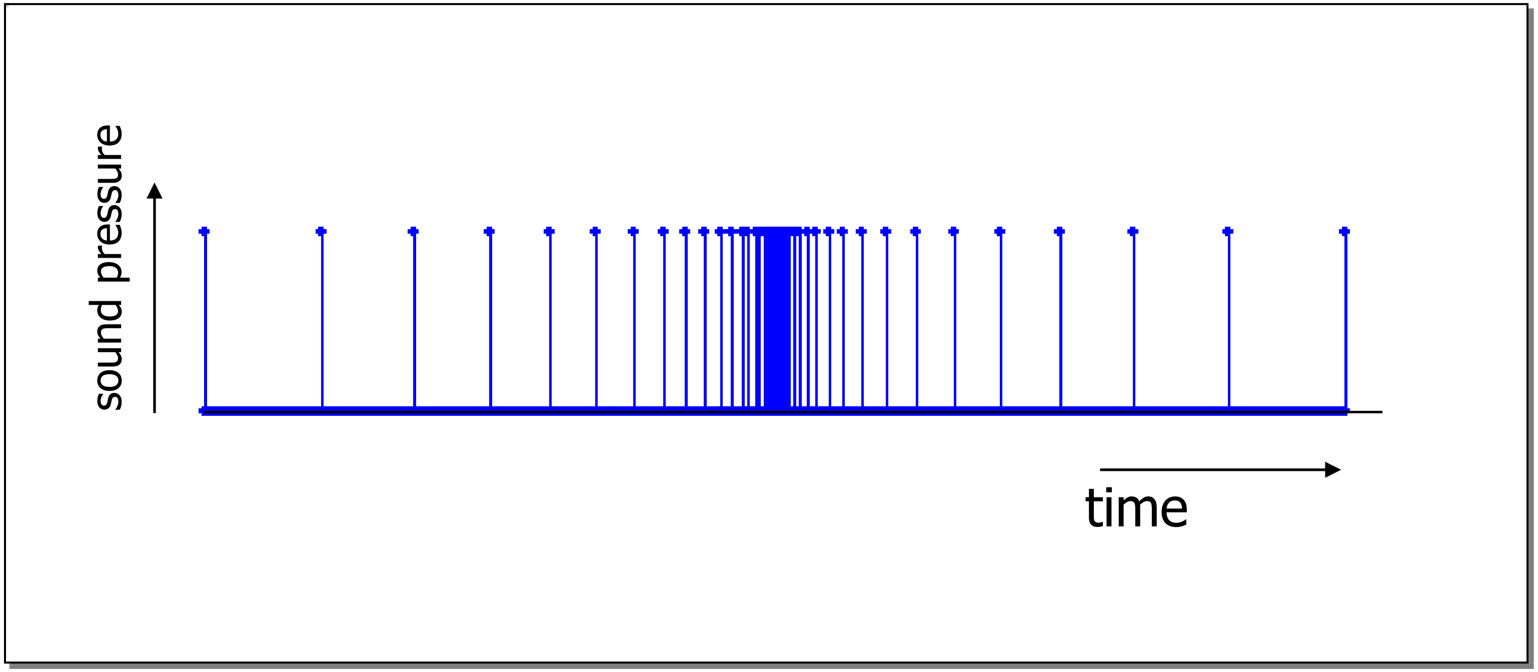
Basilar Membrane Response to Clicks



The Spectrogram



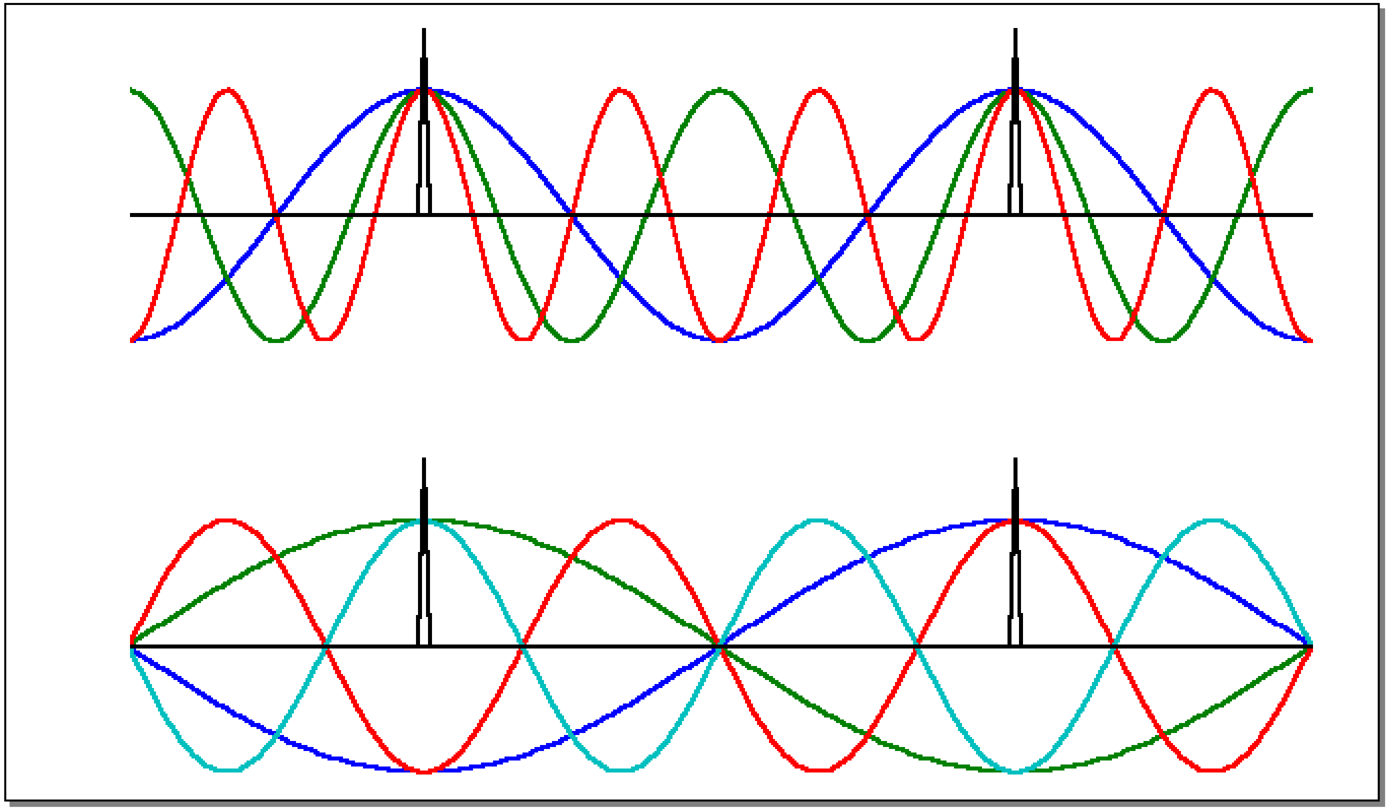
Click Trains & the "30 Hz Transition"



- At frequencies up to ca 30-50 Hz, each click in a click train is perceived as an isolated event.
- At frequencies above ca 30-50 Hz, individual clicks fuse, and one perceives a continuous hum with a strong pitch.

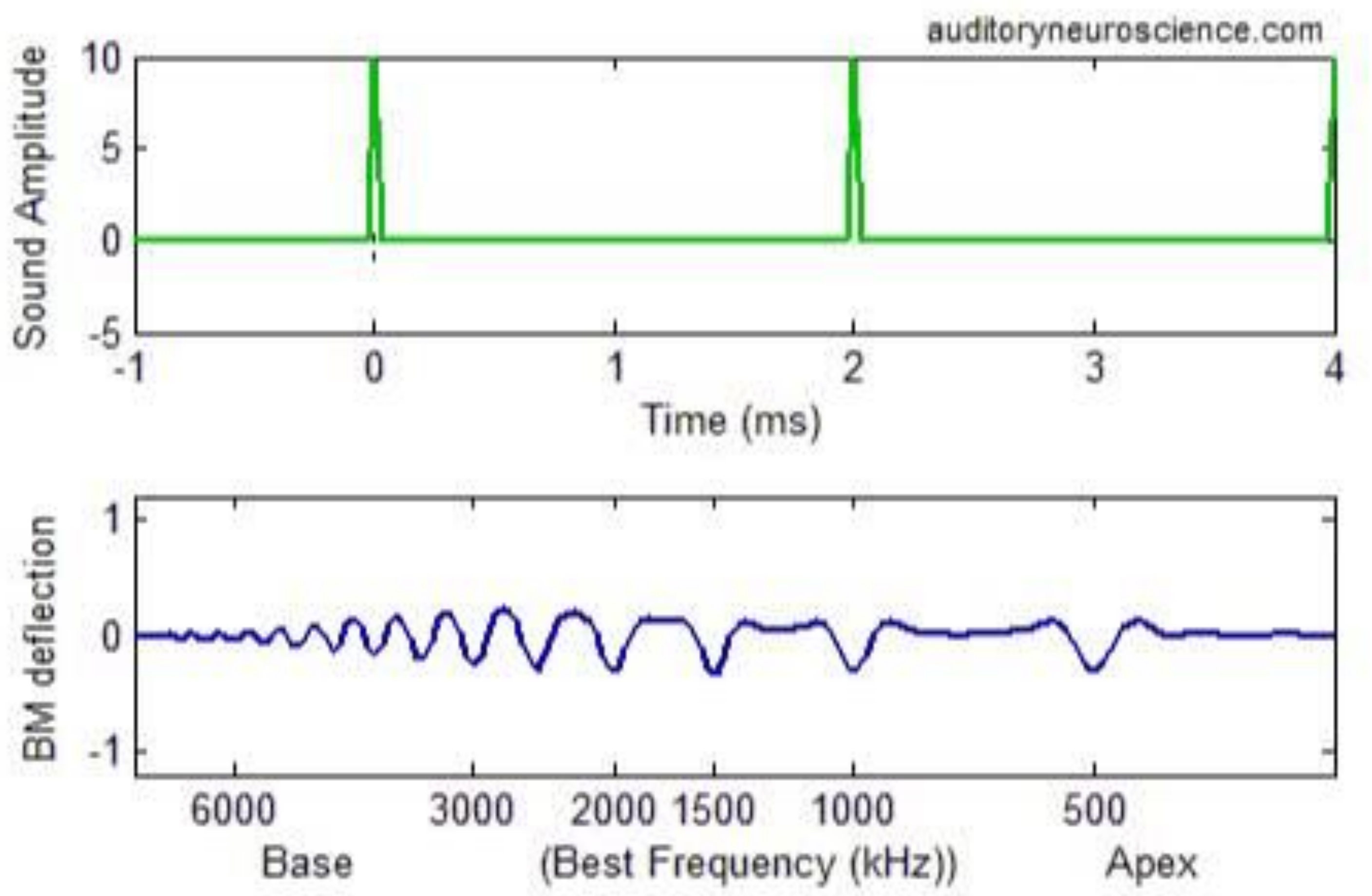
http://auditoryneuroscience.com/pitch/click_train

Harmonic Structure of Click Trains



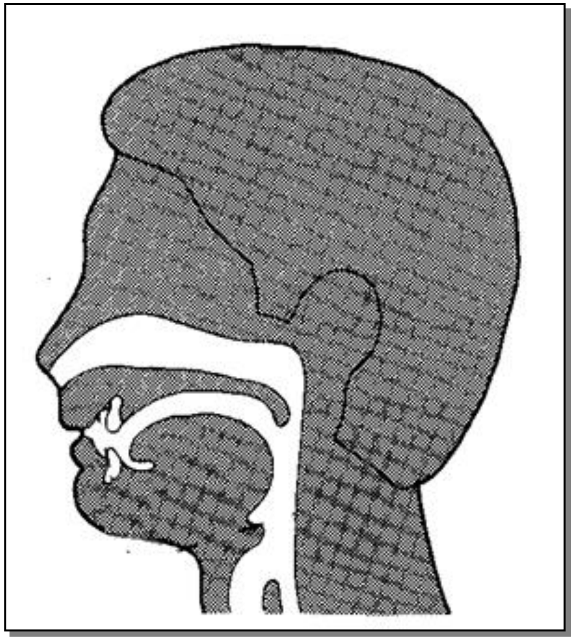
- If we represent each click in a train by its Fourier Transform, then it becomes clear that certain sine components will add (top) while others will cancel (bottom). This results in a strong harmonic structure.

Basilar Membrane Response to Click Trains



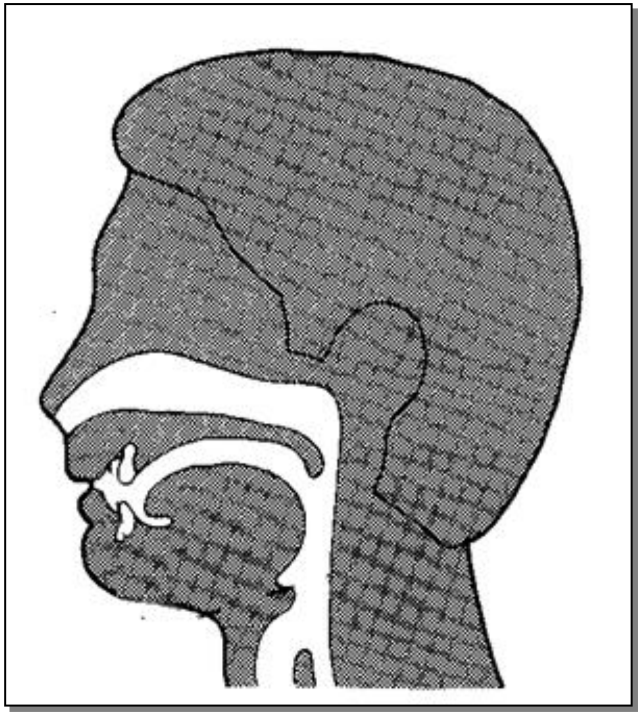
https://auditoryneuroscience.com/ear/bm3_click_trains

Vocal Folds in Action



https://auditoryneuroscience.com/vocfld

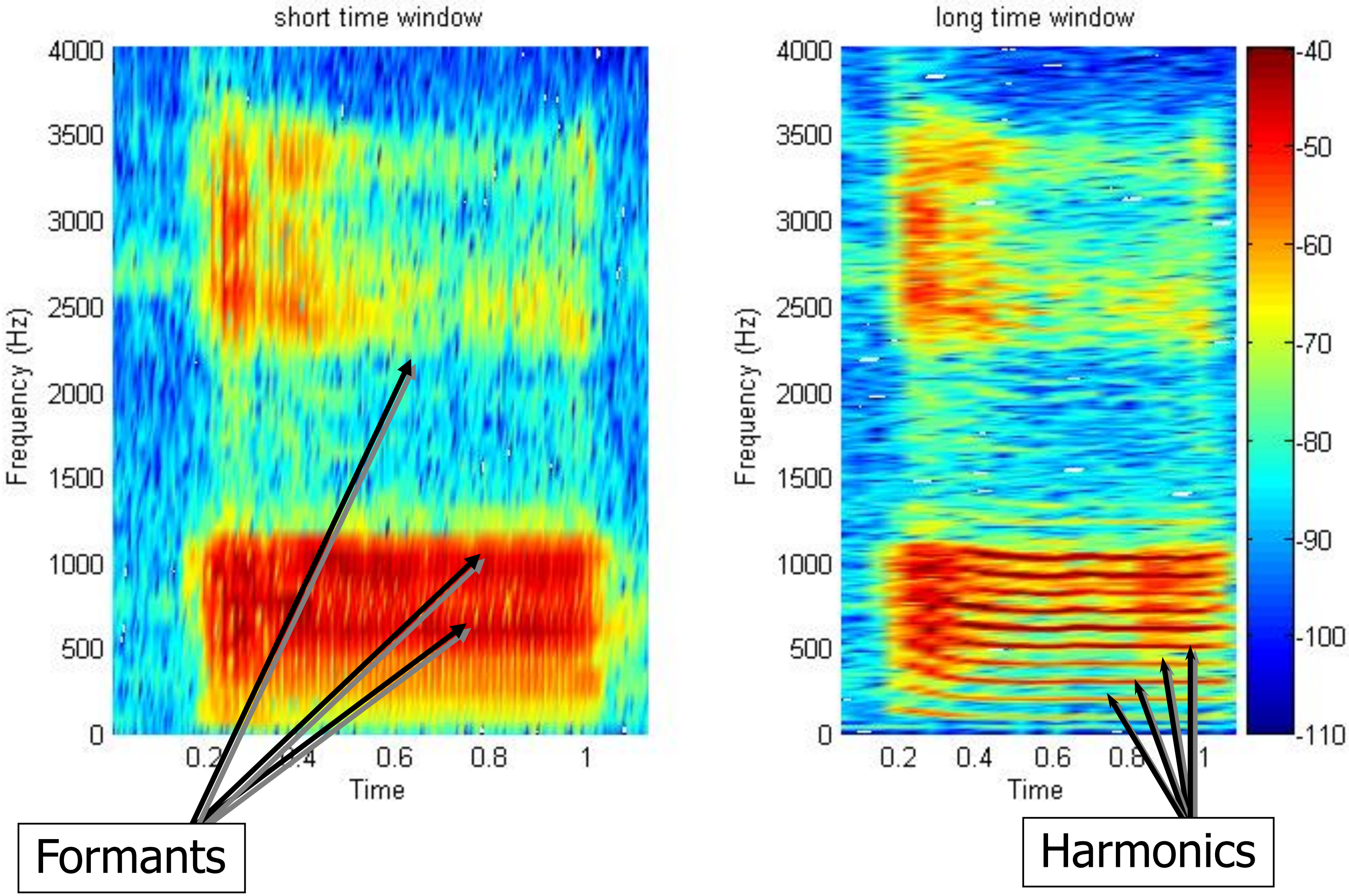
Articulation



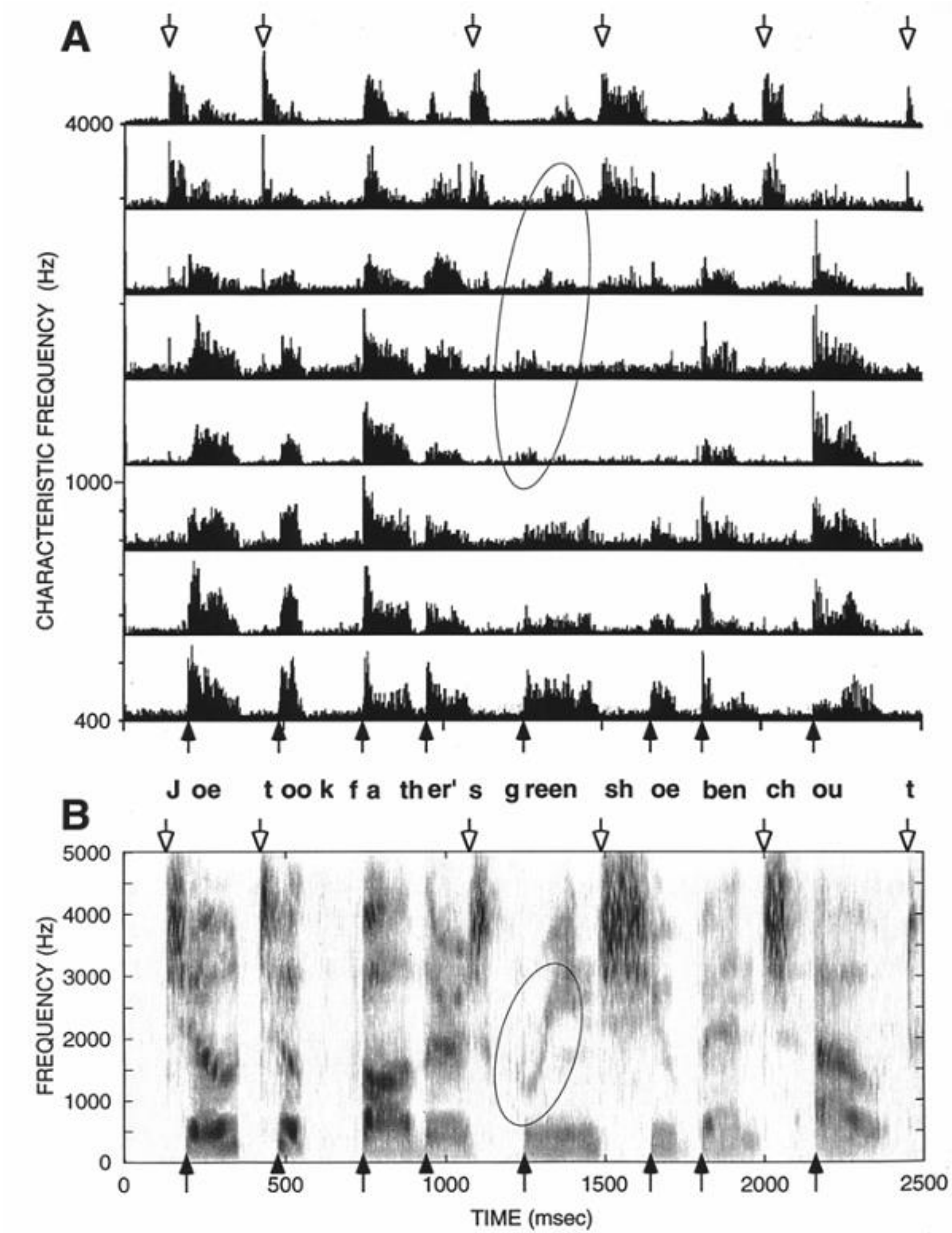
- Articulators (lips, tongue, jaw, soft palate) move to change resonance properties of the vocal tract.

https://auditoryneuroscience.com/artic

Harmonics & Formants of a Vowel



4) Place Codes and Temporal Codes for Properties of Sounds



The "Neurogram"

As a crude approximation, one might say that it is the job of the ear to produce a spectrogram of the incoming sounds, and that the brain interprets the spectrogram to identify sounds.

This figure shows histograms of *auditory nerve fibre discharges* in response to a speech stimulus. Discharge rates *depend on the amount of sound energy near the neuron's characteristic frequency.*

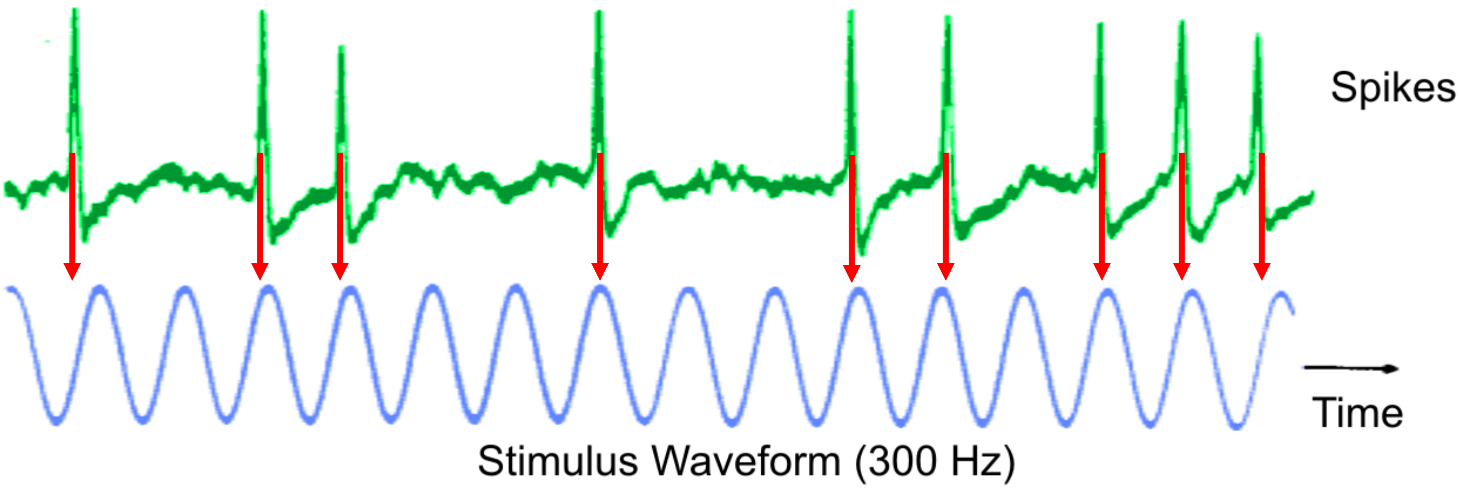
The speech formants are well resolved.

Place coding and temporal coding in the auditory nerve

- Basilar membrane tuning sets up a place coding for frequency: Tonotopy.
- But there is a lot of both rate and spike timing coding going on: Phase locking.

Phase Locking

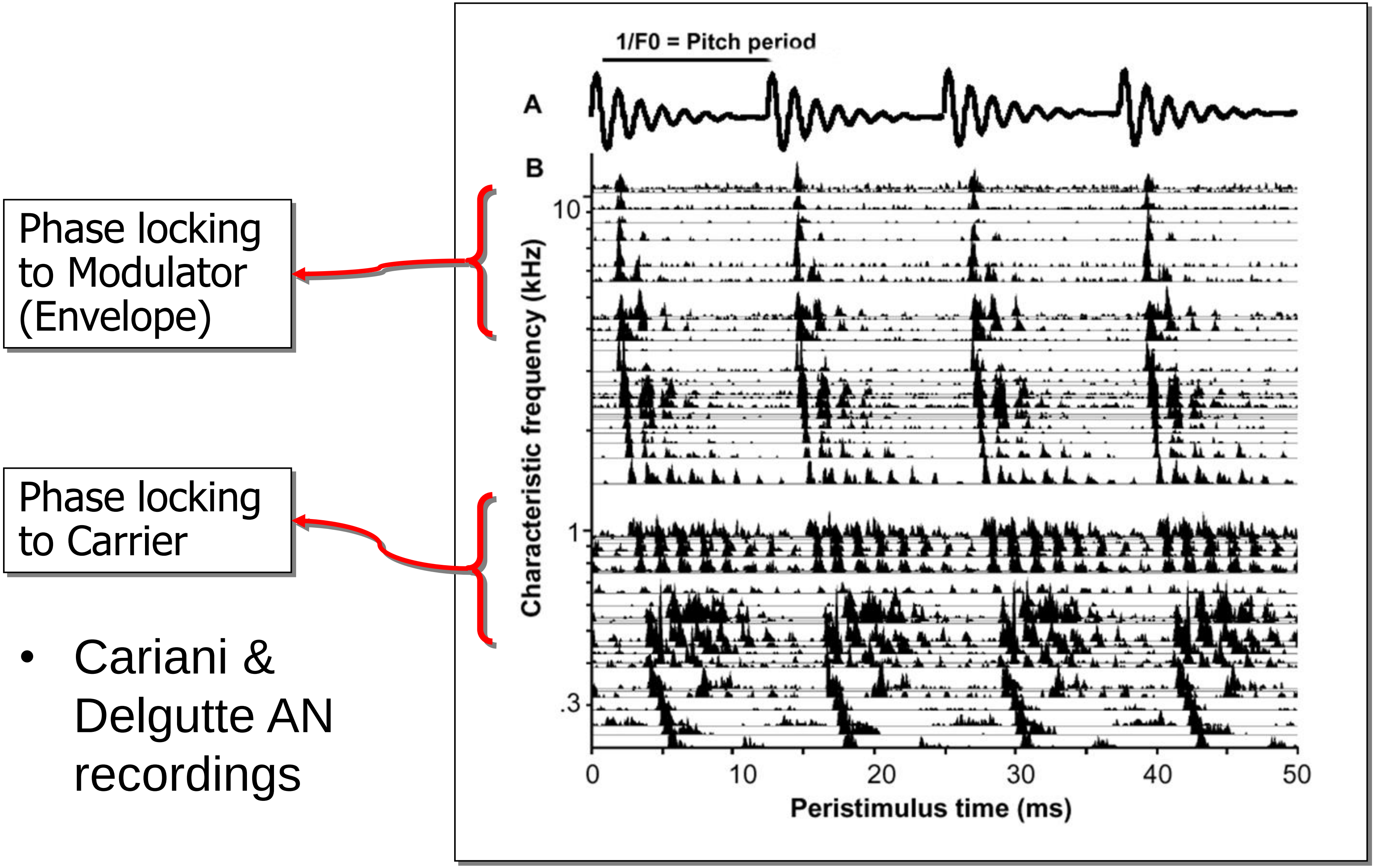
The discharges of cochlear nerve fibres to low-frequency sounds are not random; they occur at particular times (*phase locking*). This encodes information about the temporal structure of sounds that is used for hearing of pitch and spatial hearing.



http://auditoryneuroscience.com/ear/phase-locking

Evans (1975)

AN Phase Locking to Artificial "Single Formant" Vowel Sounds



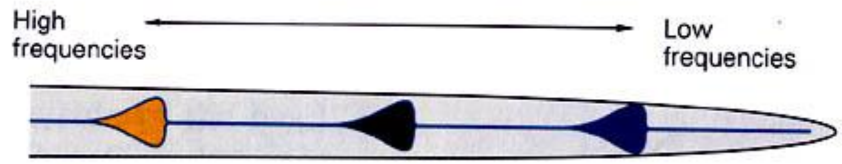
- Cariani & Delgutte AN recordings

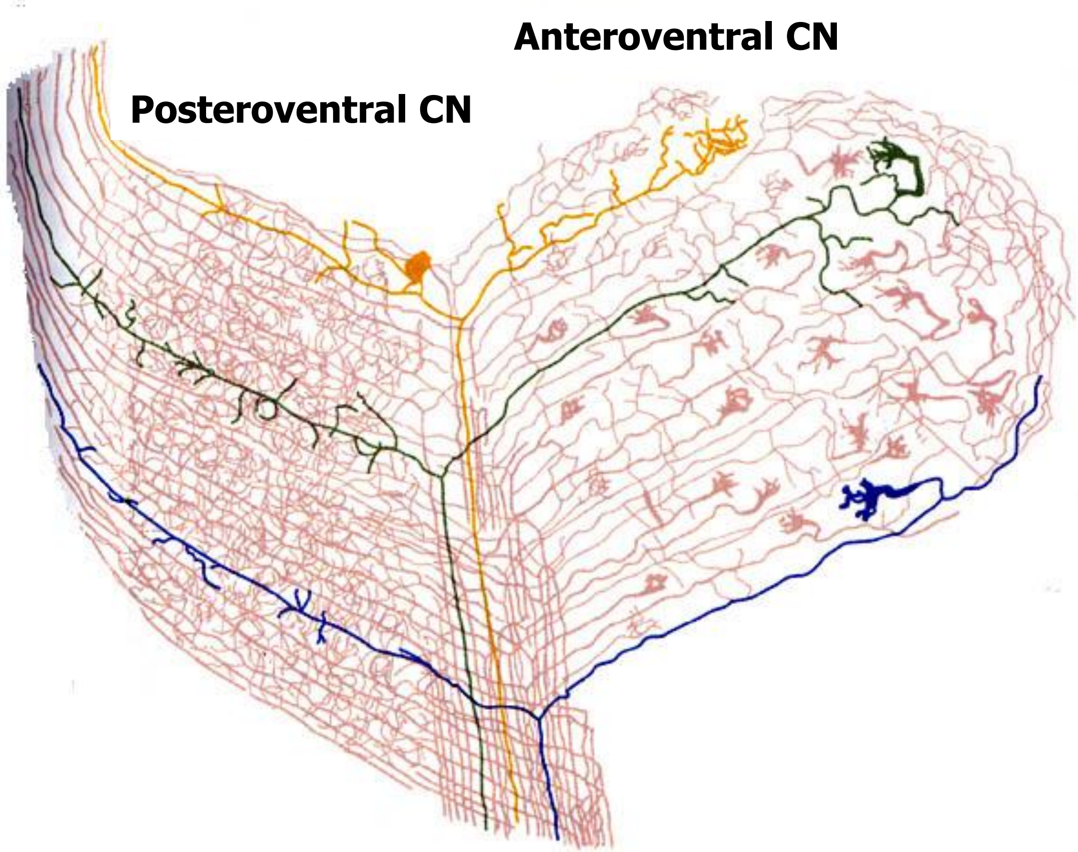
The Auditory Pathway

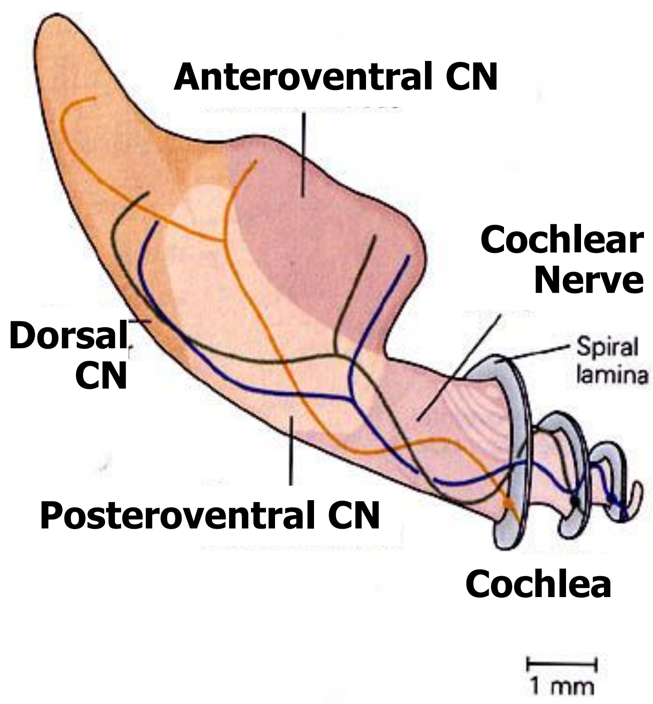
- Once the auditory nerve fibres reach the brain, things get quite complicated very quickly.
- Unlike in the visual system, where much information goes from eye straight to thalamus and cortex, in the auditory system there are dozens of nuclei in the brainstem (cochlear and superior olivary nuclei) and midbrain (lateral lemniscus and inferior colliculus) before we reach thalamus.

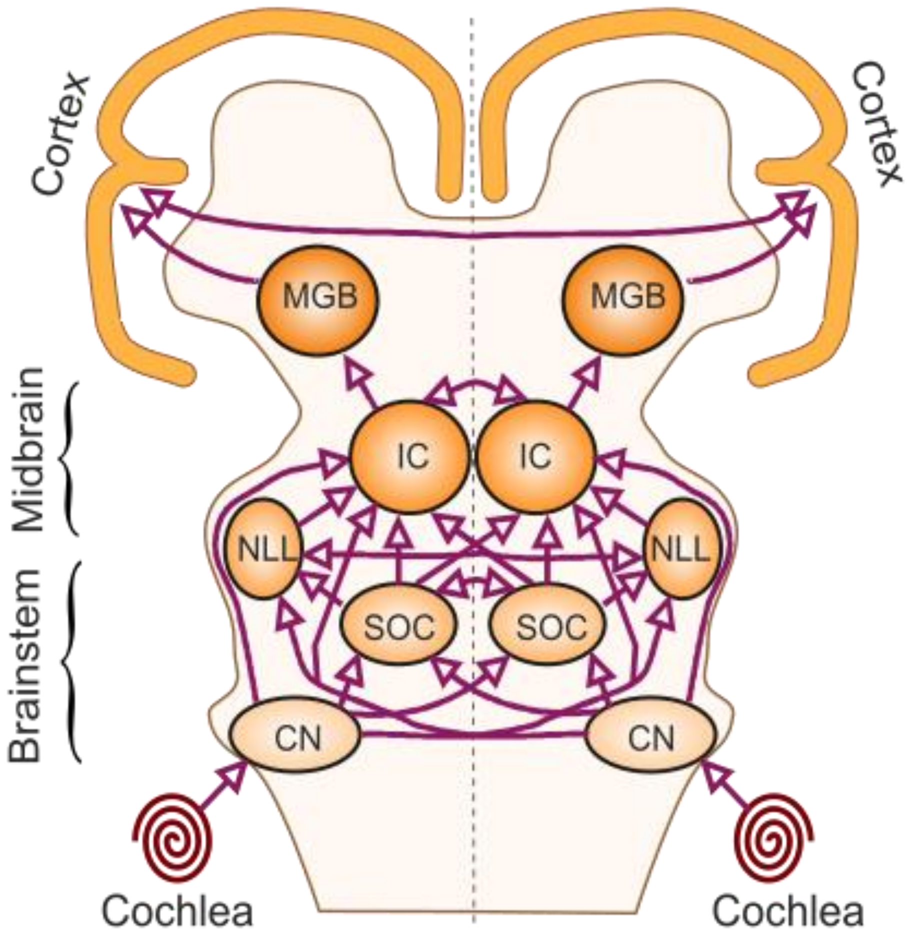
Tonotopy in the Cochlear Nucleus

The base of the BM projects to medial CN, the apex to lateral CN









Simplified schematic of the ascending auditory pathway

Break

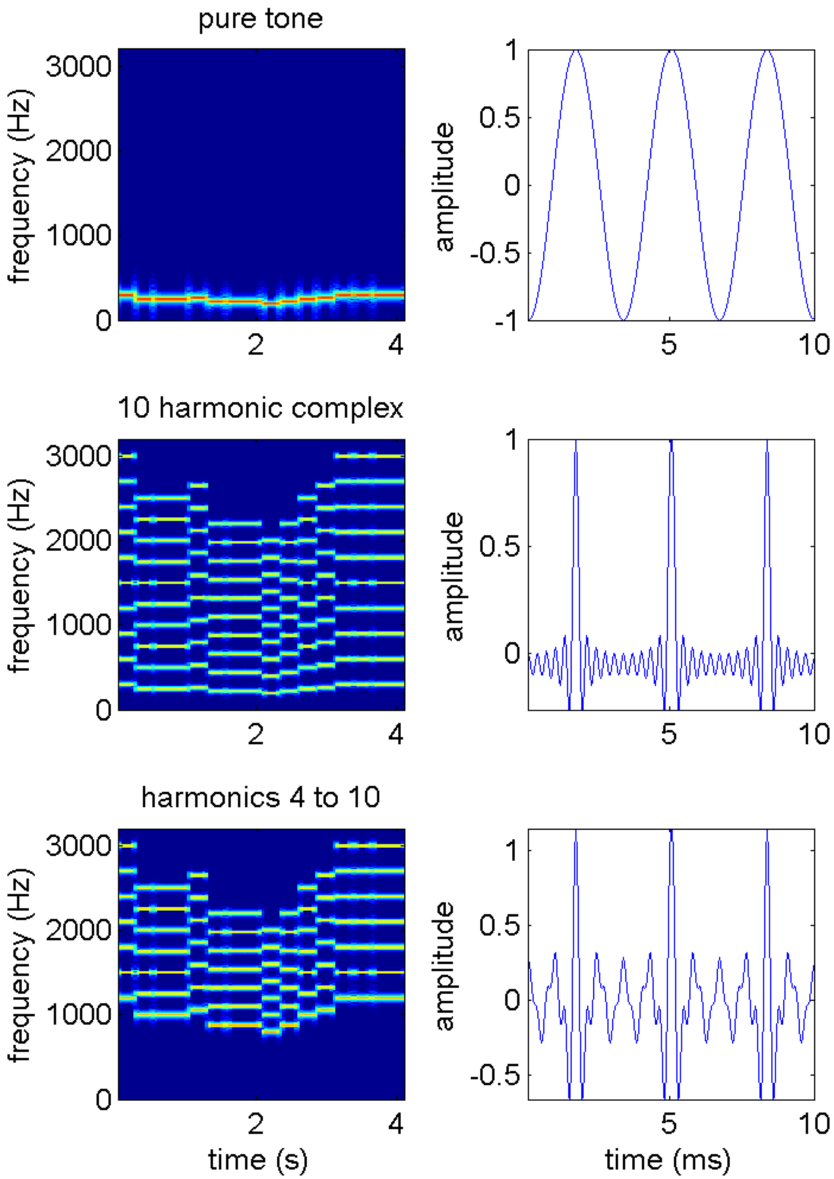
A “Functional Approach”

- To delve a bit deeper into central auditory processing, let’s look at:
- Periodicity pitch
- Spatial hearing
- Speech

- There are many aspects of hearing we won’t have time to cover (e.g. musical rhythm and emotion, scene analysis,…)

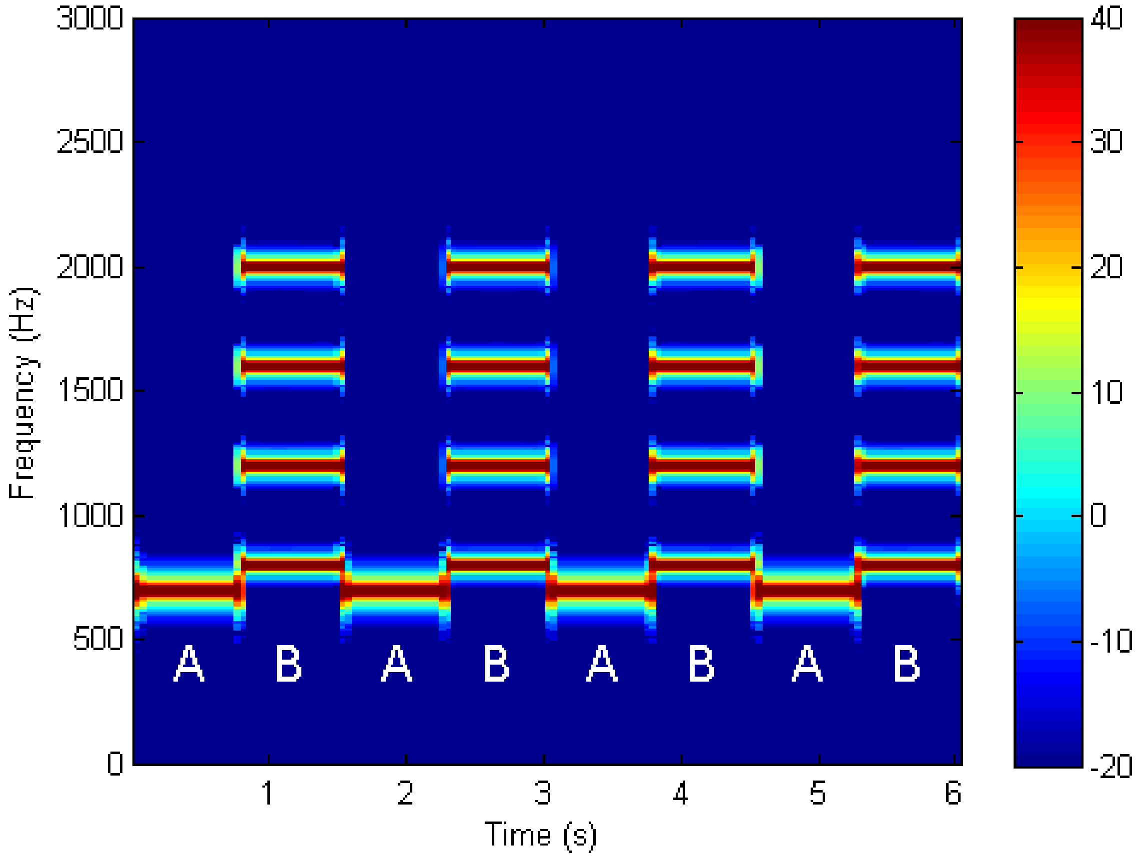
5) Periodicity Pitch

Missing Fundamental Sounds



- https://auditoryneuroscience.com/topics/missing-fundamental

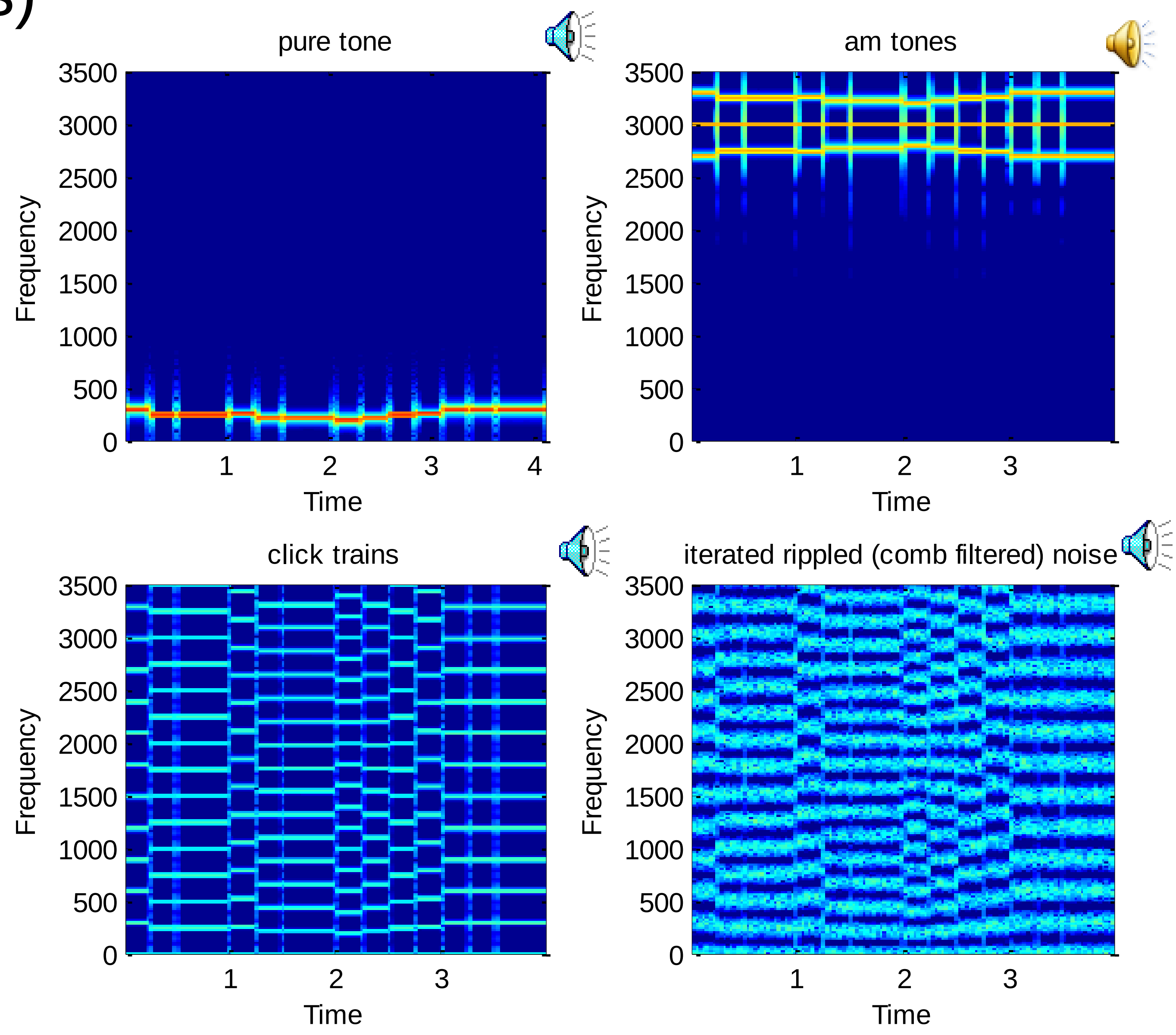
Counter-intuitive Missing Fundamental



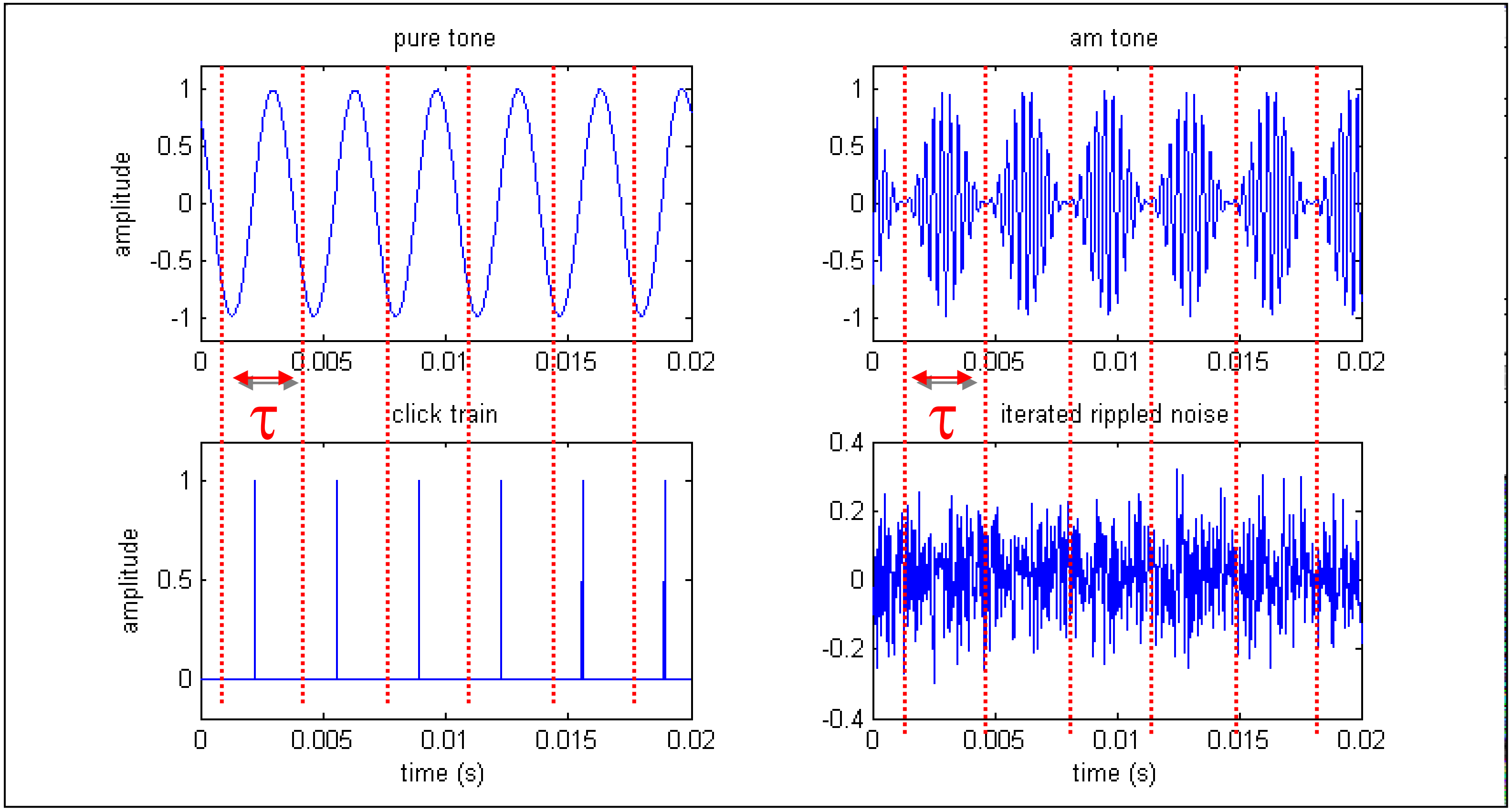
Do you hear this as up-down-up-down
or as down-up-down-up ?

- http://auditoryneuroscience.com/topics/why-missing-fundamental-stimuli-are-counterintuitive

The Pitch of "Complex" Sounds (Examples)



The Periodicity of a Signal is a Major Determinant of its Pitch



- Iterated rippled noise can be made more or less periodic by increasing or decreasing the number of iterations. The less periodic the signal, the weaker the pitch.

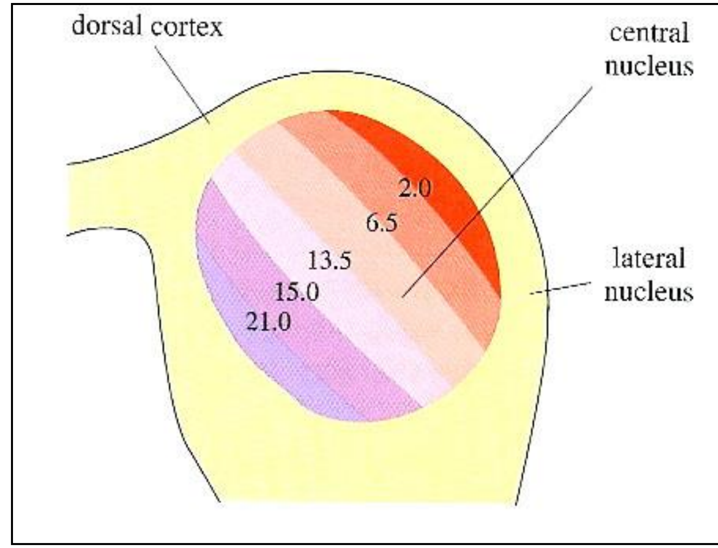
Periodicity Maps in the Midbrain?

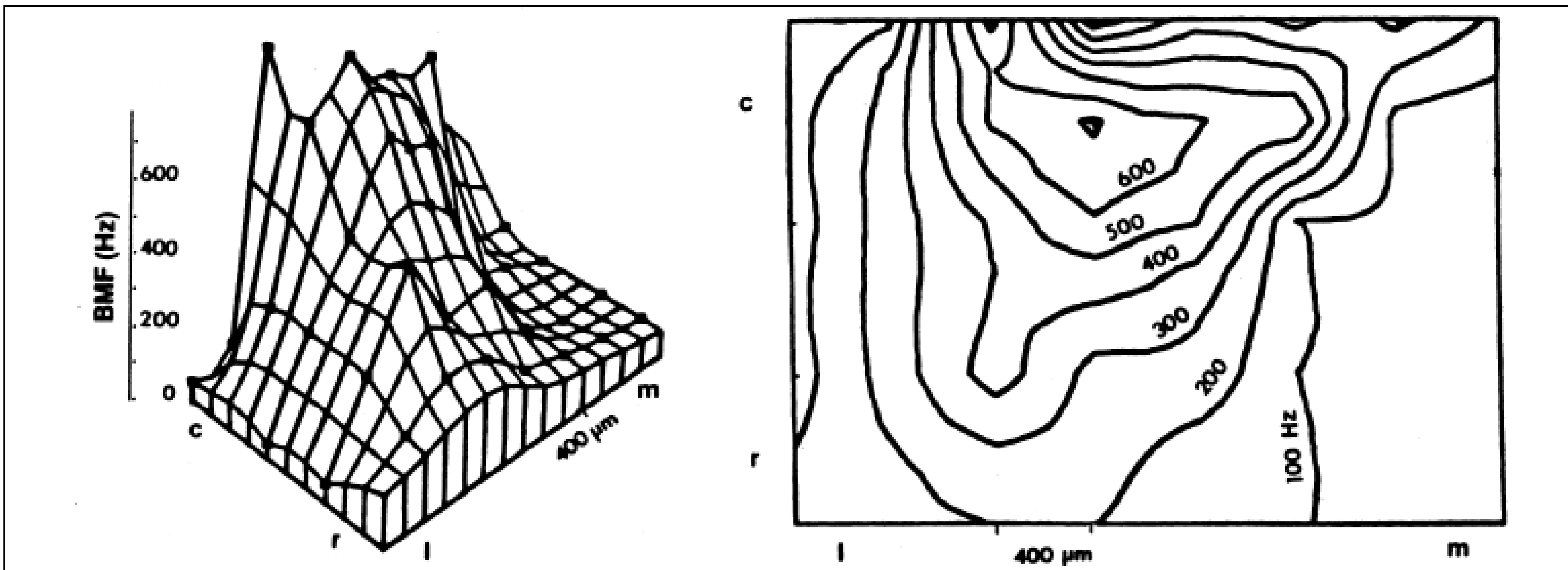
Neurons in the midbrain or above show much less phase locking to AM than neurons in the brainstem.

Transition from a *timing* to a *rate* code.

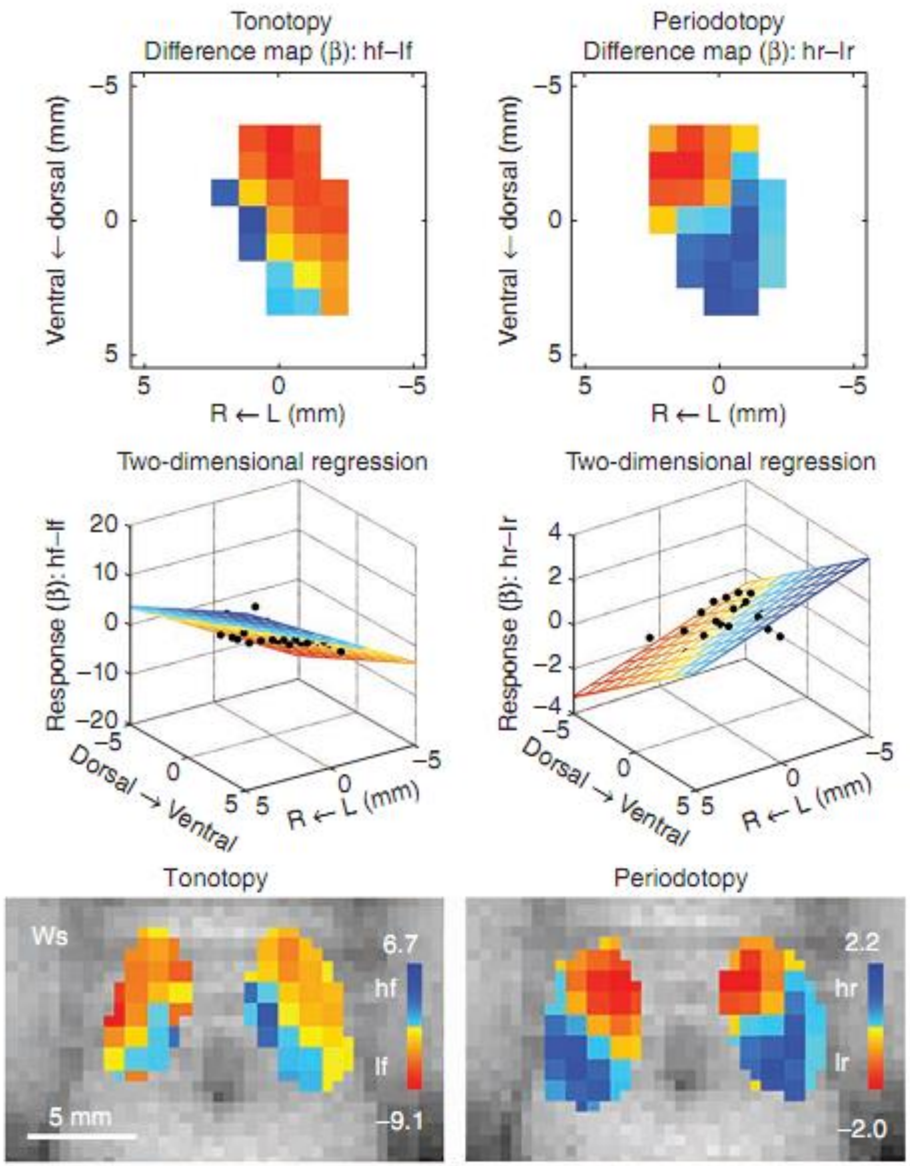
Some neurons have *bandpass MTFs* and exhibit "best modulation frequencies" (BMFs).

Topographic maps of BMF may exist within isofrequency laminae of the ICc, ("periodotopy").



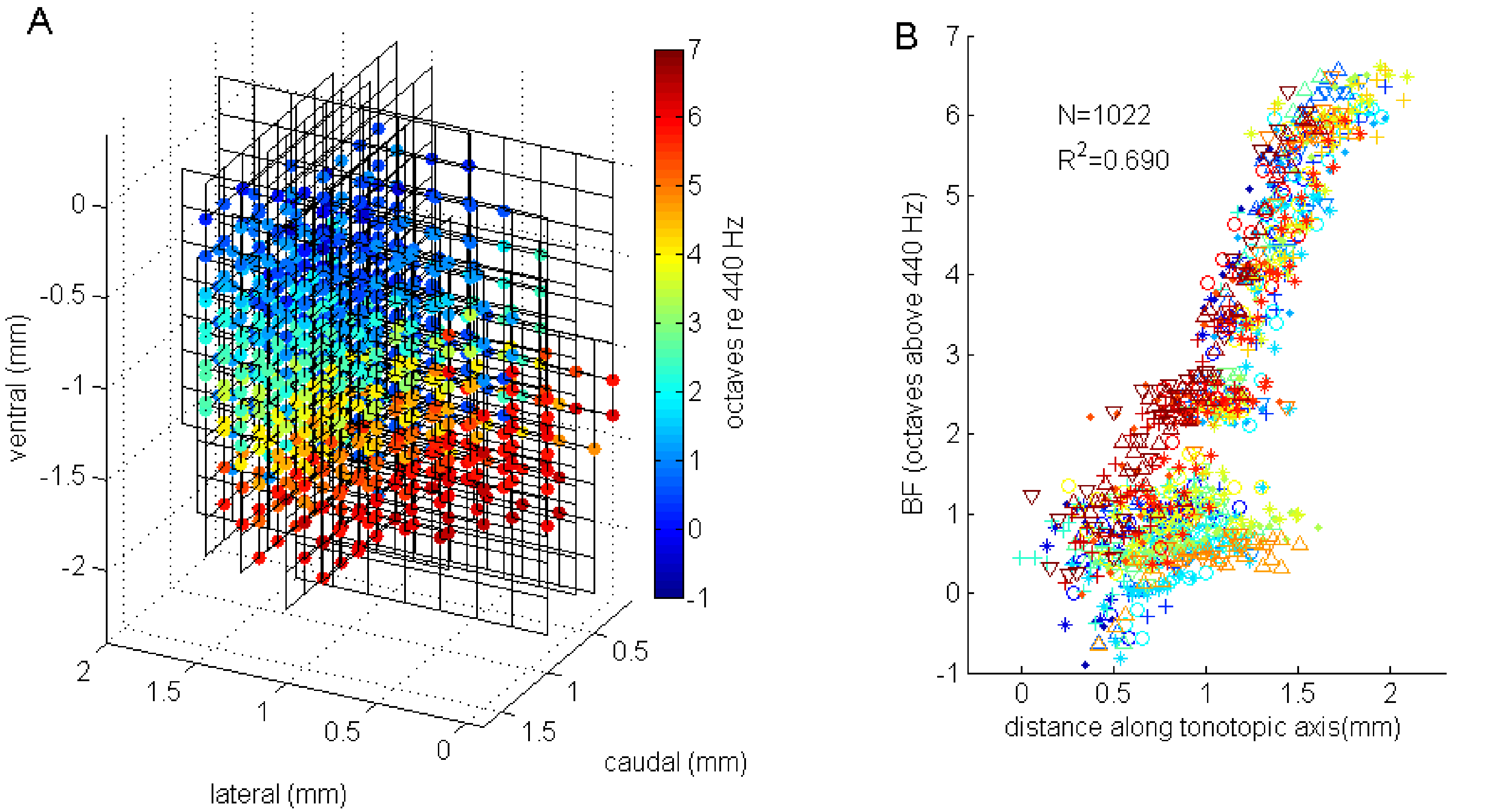


Periodotopic maps via fMRI

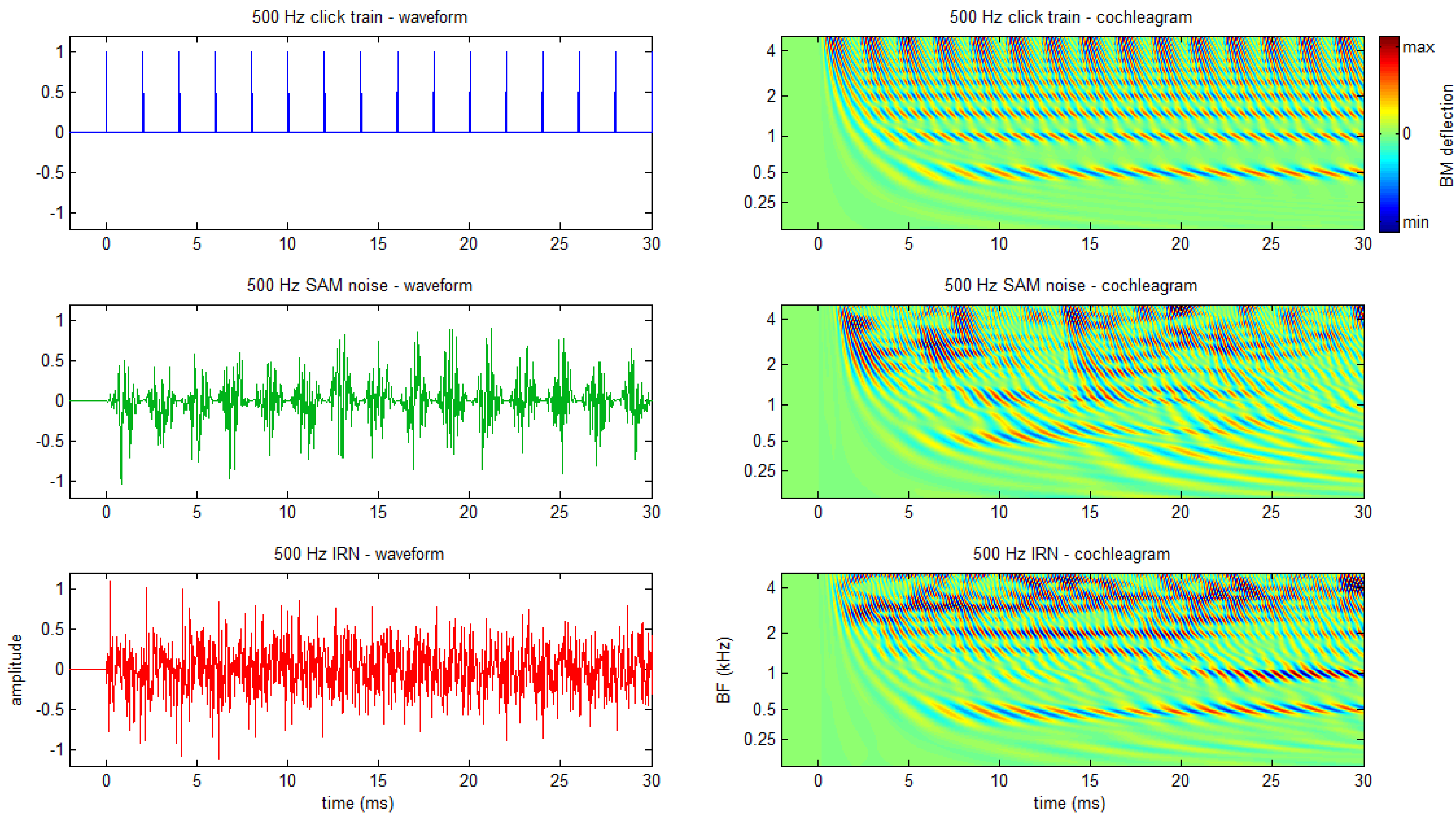


- Baumann et al Nat Neurosci 2011 described periodotopic maps in monkey IC obtained with fMRI.
- They used stimuli from 0.5 Hz (infra-pitch) to 512 Hz (mid-range pitch).
- Their sample size is quite small (3 animals – false positive?)
- The observed orientation of their periodotopic map (medio-dorsal to latero-ventral for high to low) appears to *differ* from that described by Schreiner & Langner (1988) in the cat (predimonantly caudal to rostral)

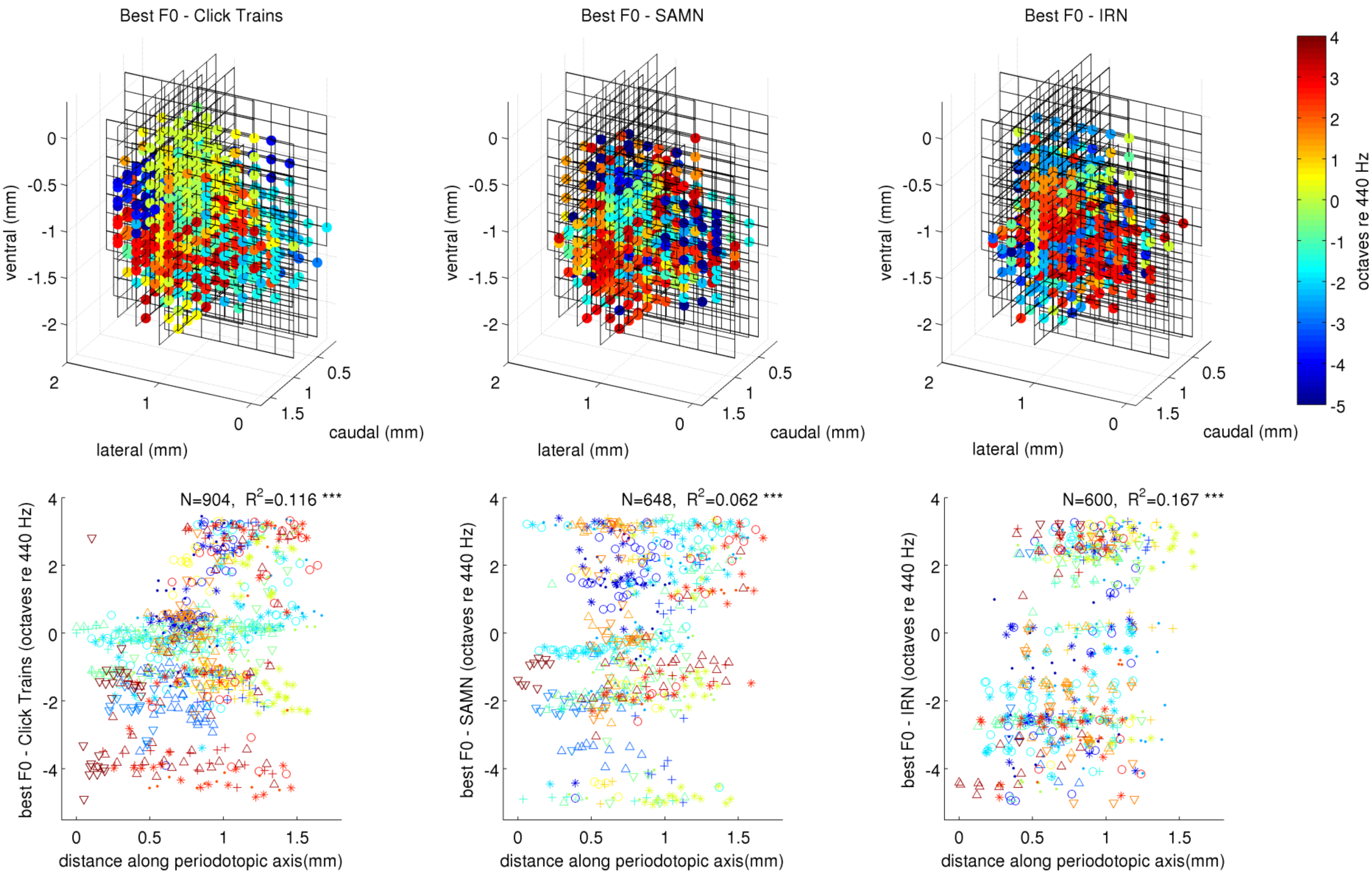
Tonotopy in Inferior Colliculus



3 Flavours of (Quasi-)Periodic Sounds



Periodotopy in IC?



Schnupp et al (2015) Frontiers in Neural Circuits

Auditory cortex of Ferret (A), Cat (B) and Macaque Monkey (C)

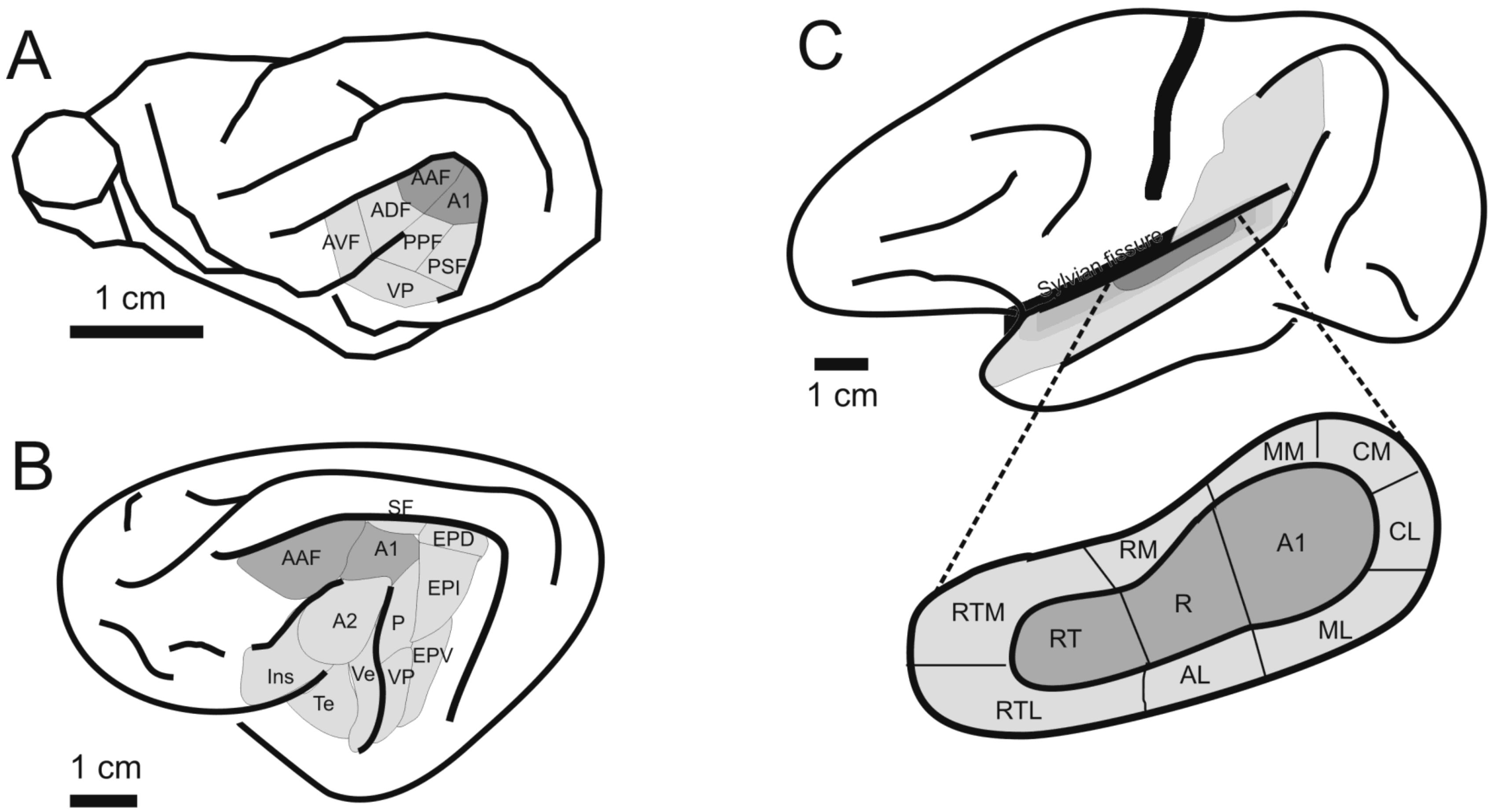
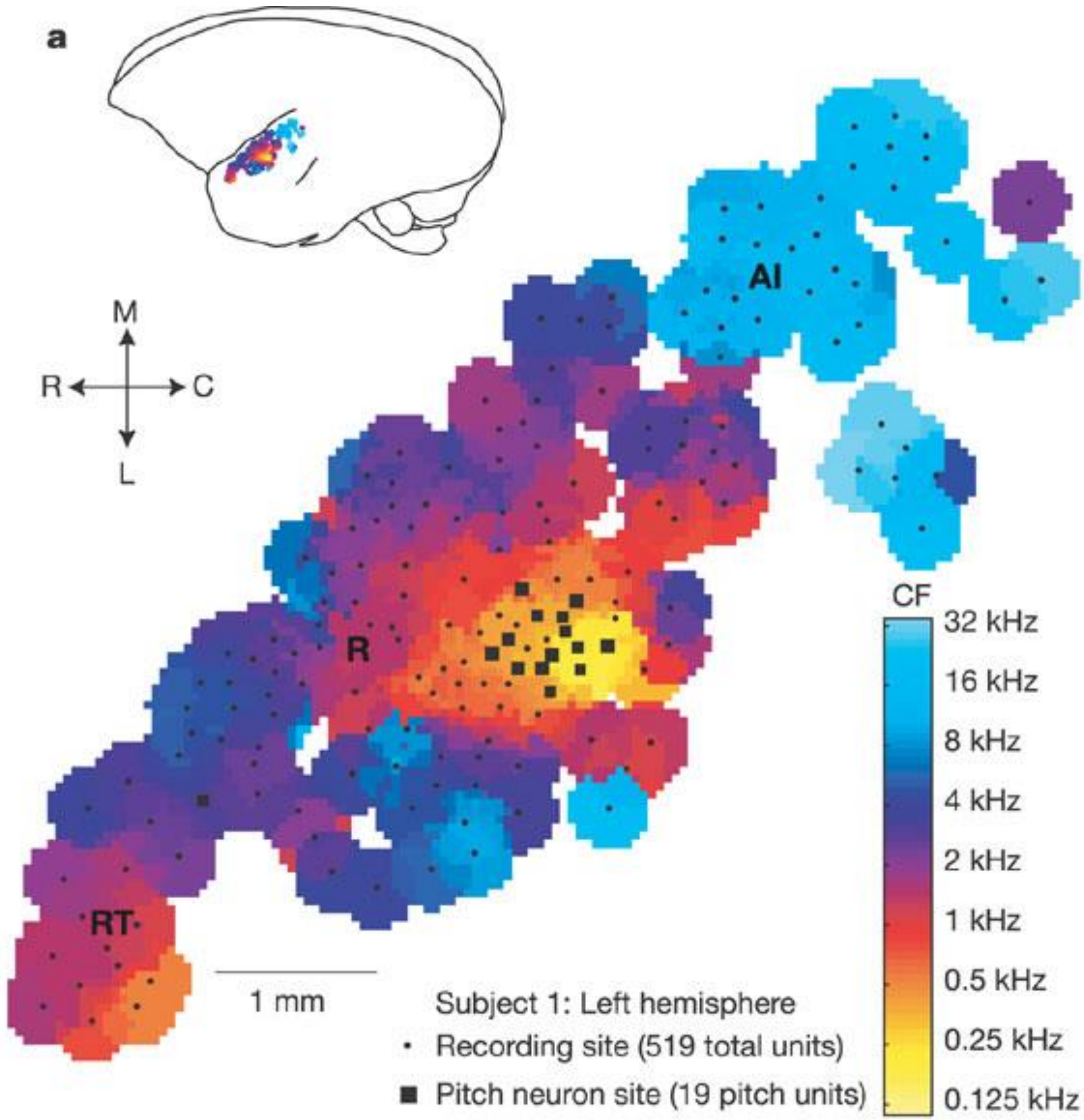


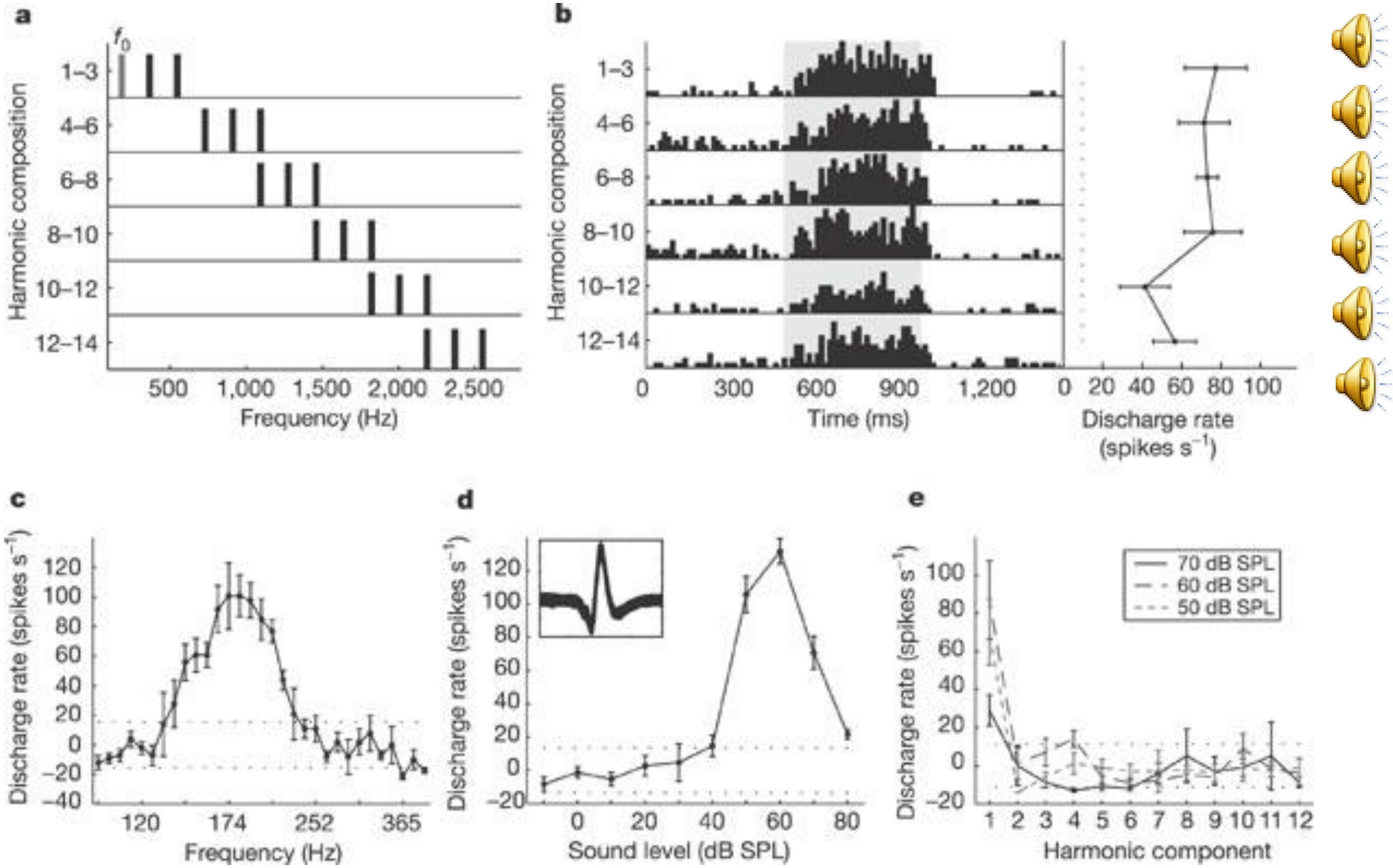
Figure 2-18 of Schnupp, Nelken, King “Auditory Neuroscience”

A pitch area in primates?

- In marmoset, Pitch sensitive neurons are most commonly found on the boundary between fields A1 and R.
- Fig 2 of Bendor & Wang, *Nature 2005*



A pitch sensitive neuron in marmoset A1?



- Apparently pitch sensitive neurons in marmoset A1.
- Fig 1 of Bendor & Wang, *Nature 2005*

Periodicity Pitch - Summary

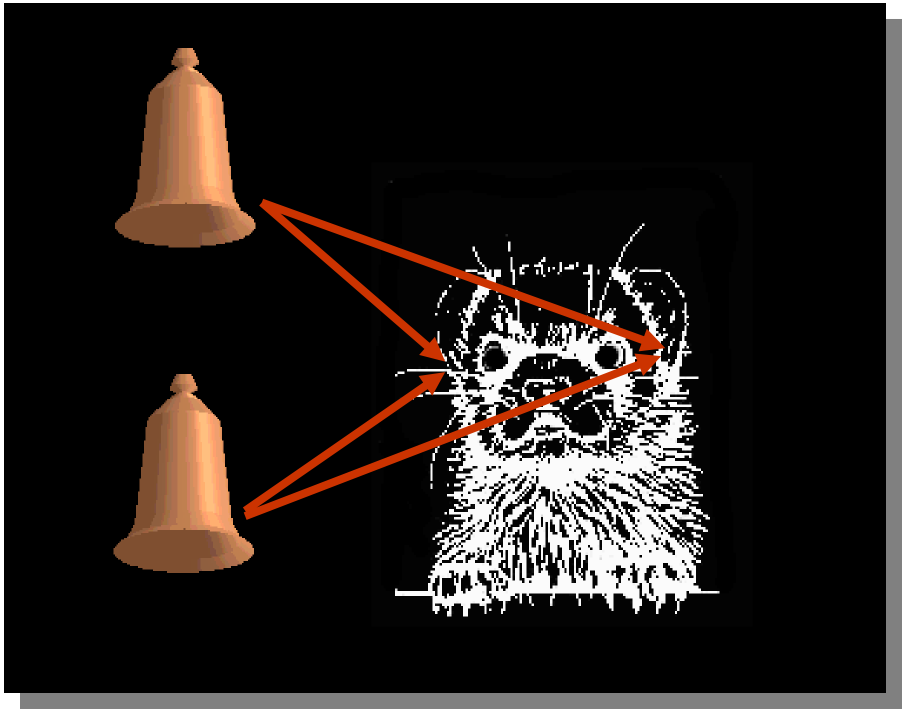
- Pitchy sounds are periodic. That also means they have many “harmonic” frequency components.
- Missing fundamental stimuli show that there is no straight-forward relationship between tonotopy and perceived pitch. Temporal encoding of periodicity is clearly important.
- Earlier studies have postulated the existence of “periodotopic maps”. More recent work sheds doubt on that.
- Be sceptical of the notion of “pitch neurons” in the brain.

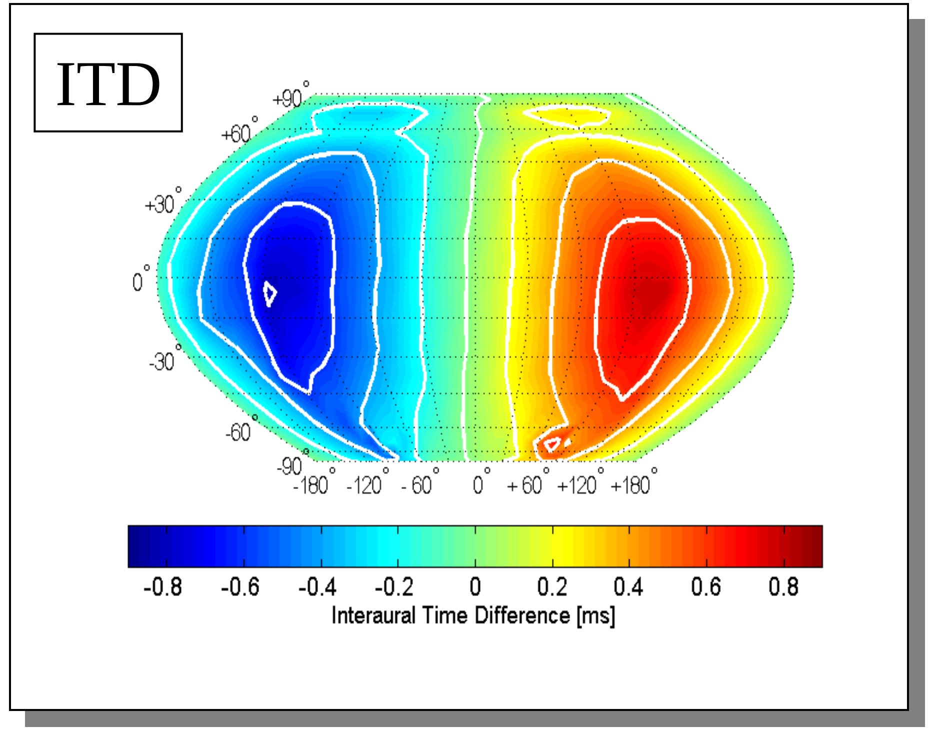
6) Spatial Hearing

Earning One's Supper

http://auditoryneuroscience.com/foxInSnow

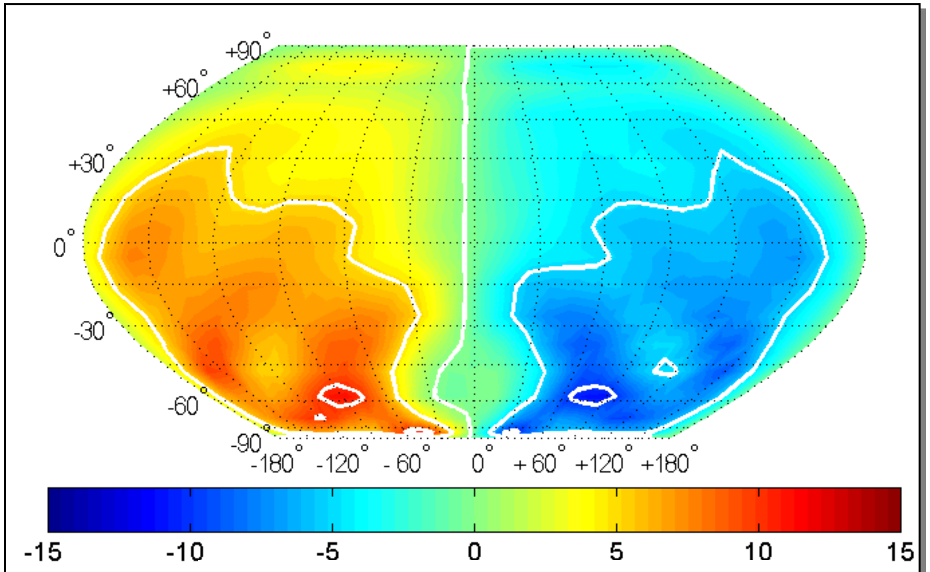
Interaural Time Difference (ITD) Cues



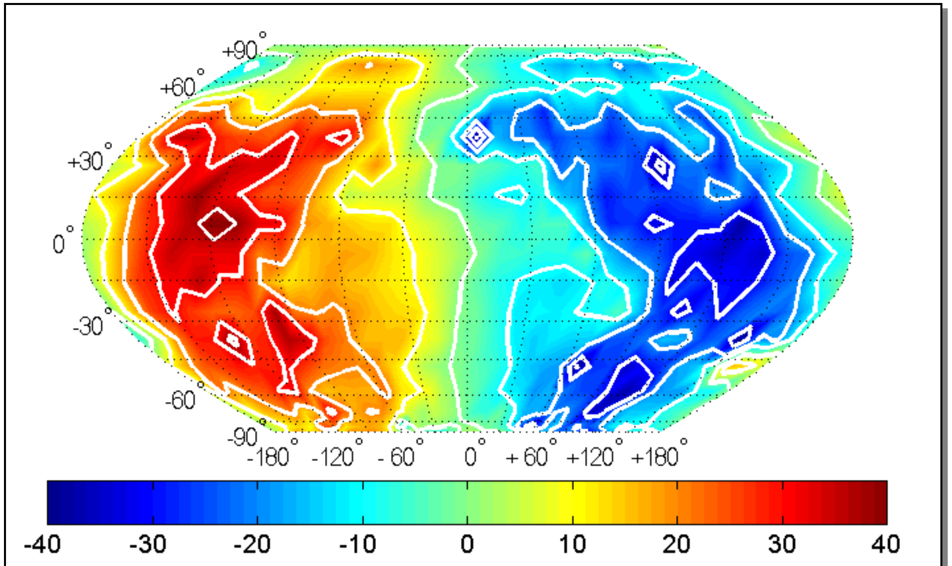


ITDs are powerful cues to sound source direction, but they are ambiguous (“cones of confusion”)

Interaural Level Cues (ILDs)



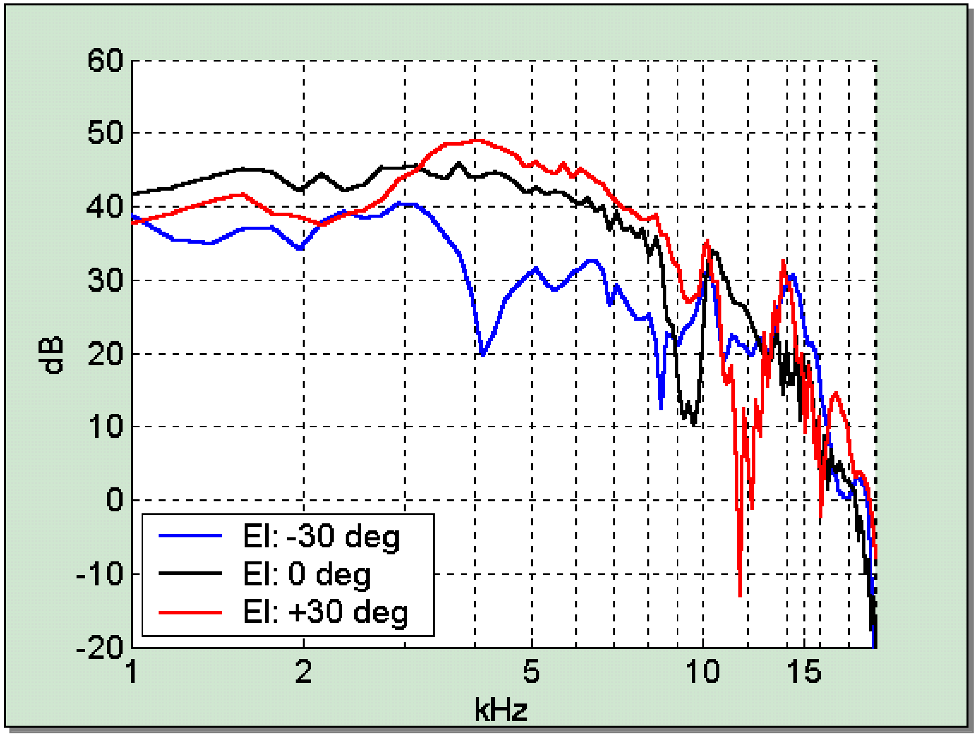
ILD at 700 Hz

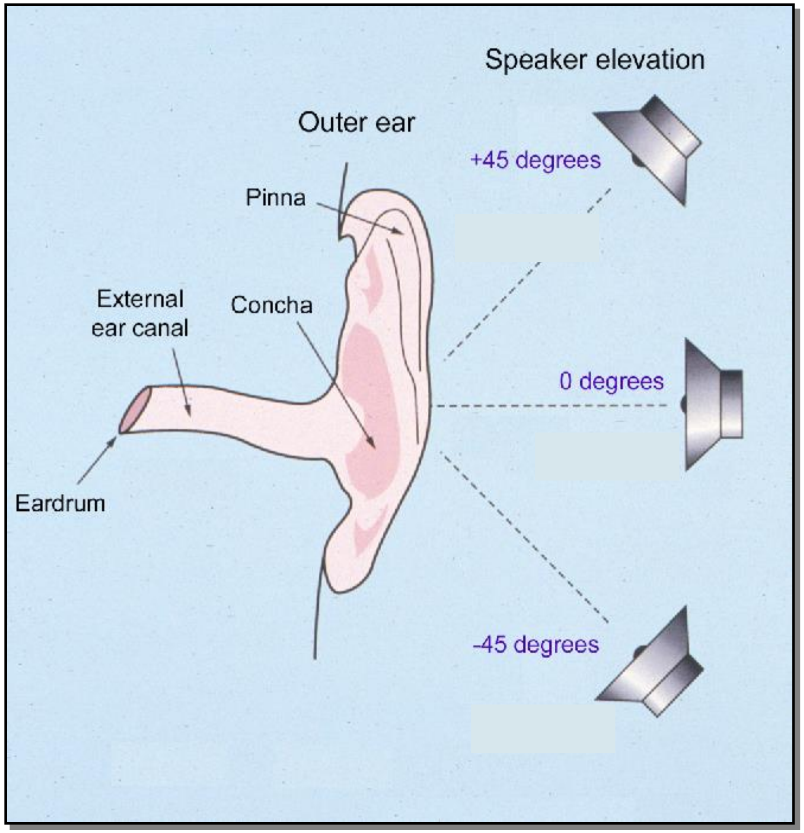


ILD at 11000 Hz

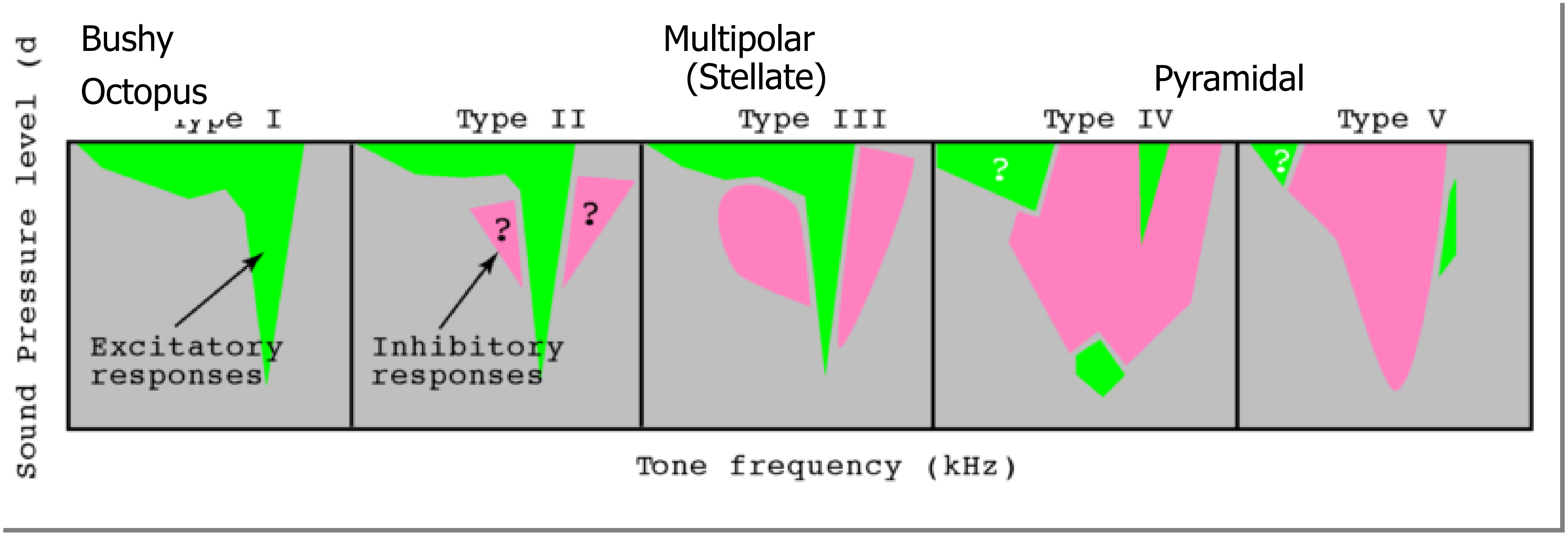
Unlike ITDs, ILDs are highly frequency dependent. At higher sound frequencies ILDs tend to become larger, more complex, and hence potentially more informative.

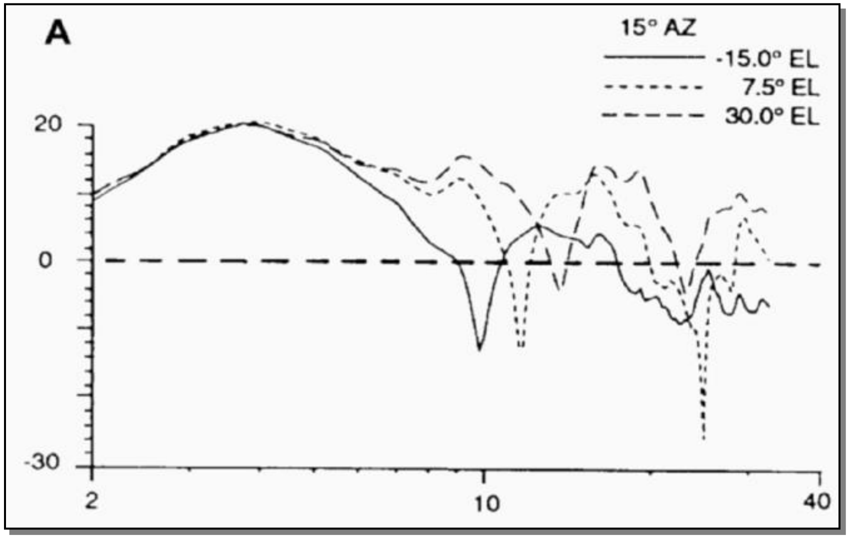
Spectral (Monaural) Cues



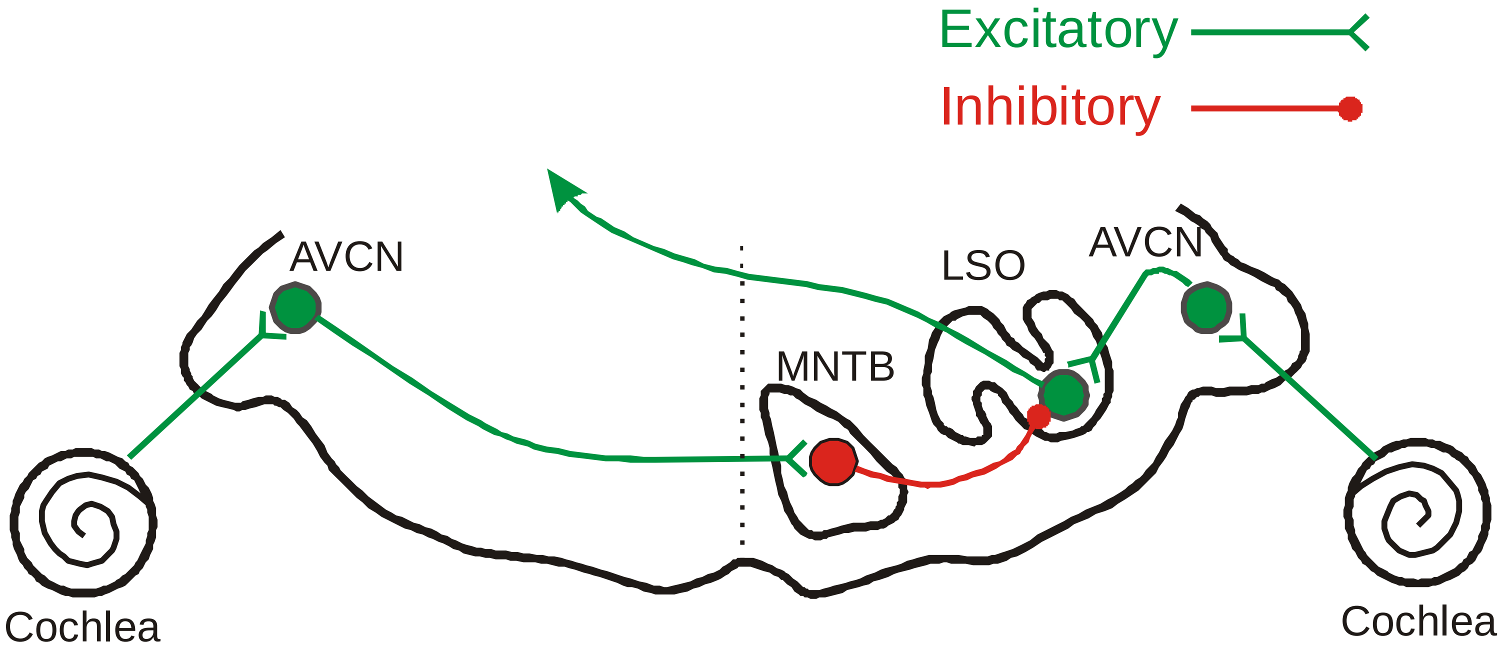


Spectral Cues and the Dorsal Cochlear Nucleus

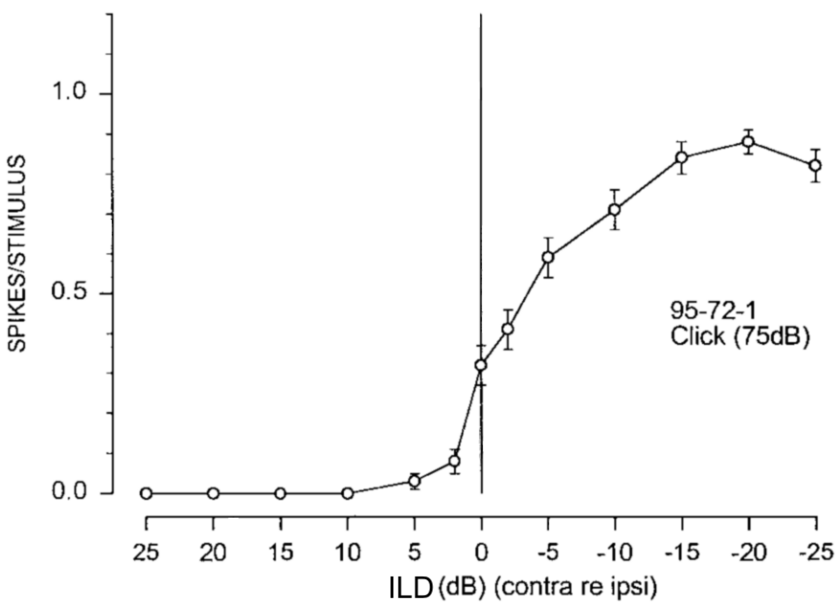




"Type IV" neurons in the dorsal cochlear nucleus often have inhibitory frequency response areas with excitatory sidebands. This makes them sensitive to "spectral notches" like those seen in spectral localisation cues.

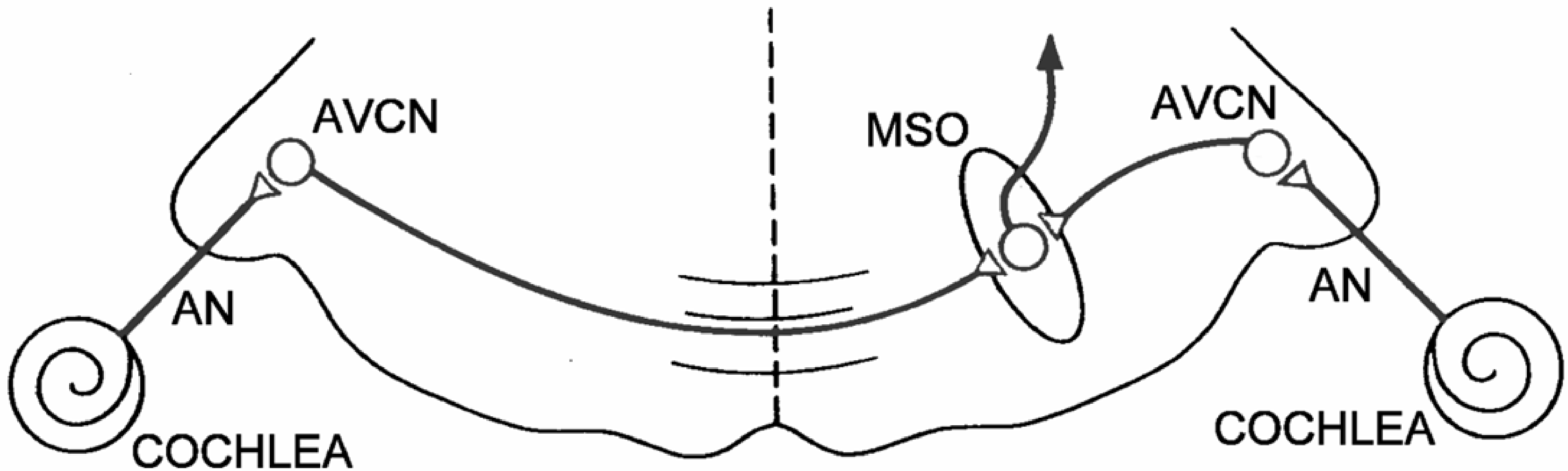


Excitatory
Inhibitory
AVCN
LSO
AVCN
MNTB
Cochlea
Cochlea

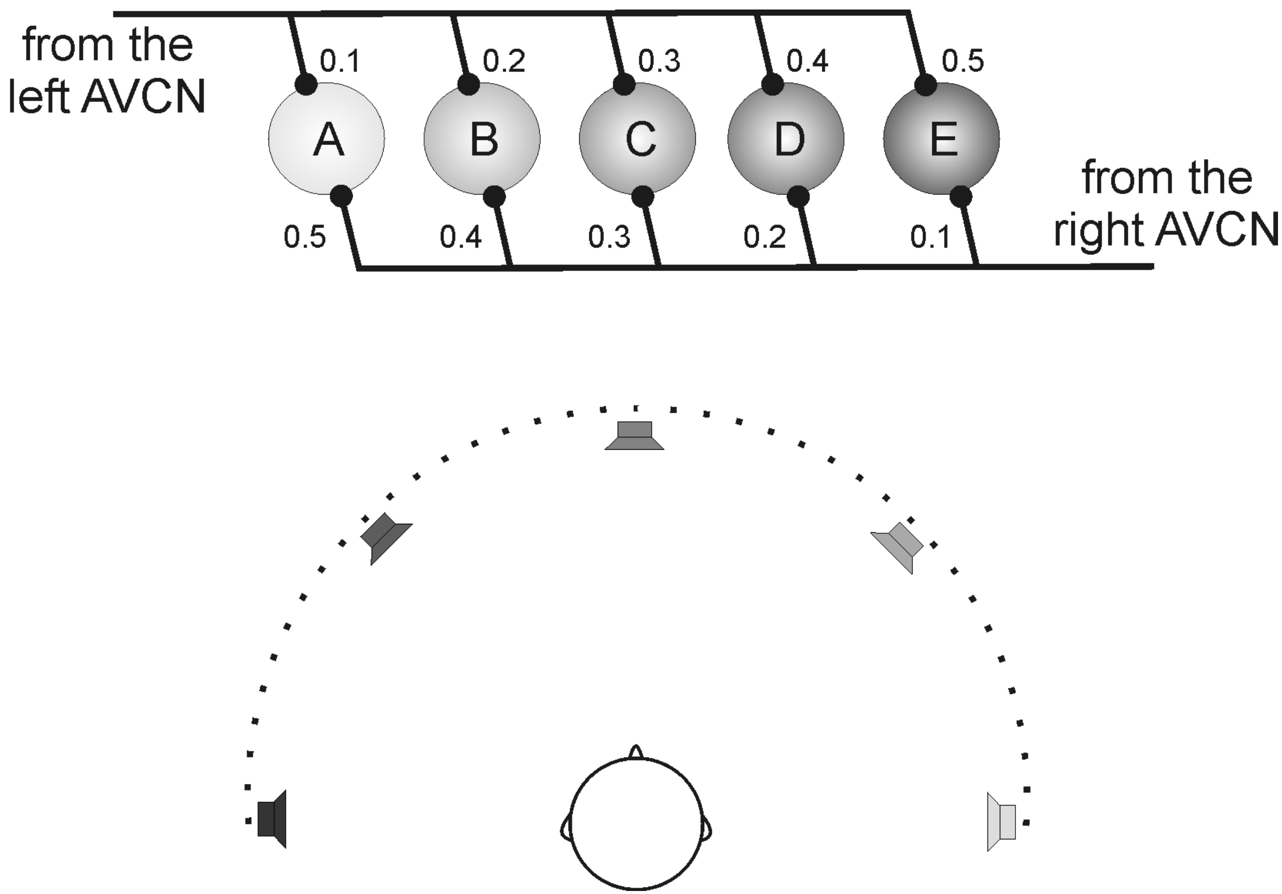


SPIKES/STIMULUS
1.0
0.5
0.0
25
20
15
10
5
0
-5
-10
-15
-20
-25
ILD (dB) (contra re ipsi)
95-72-1
Click (75dB)

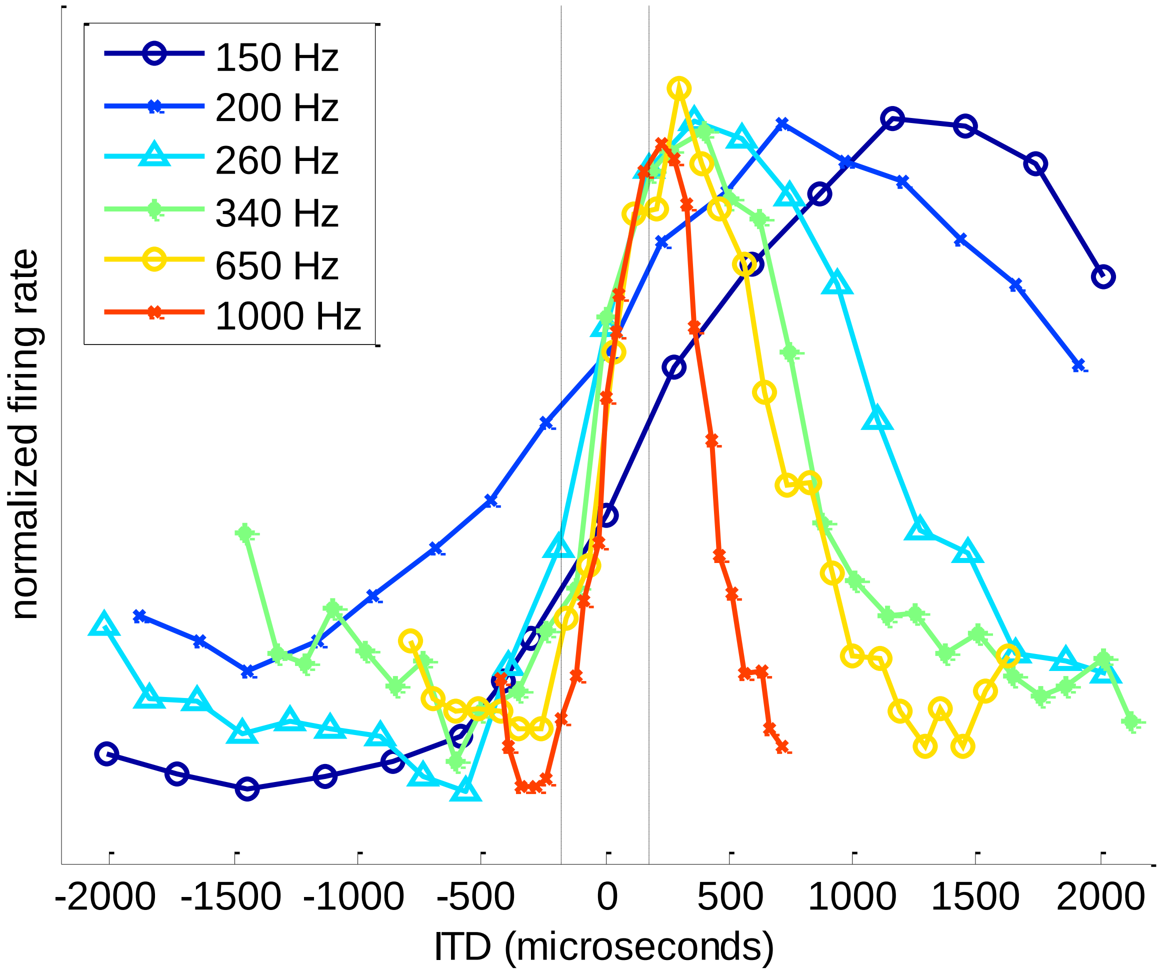
ITD PATHWAY



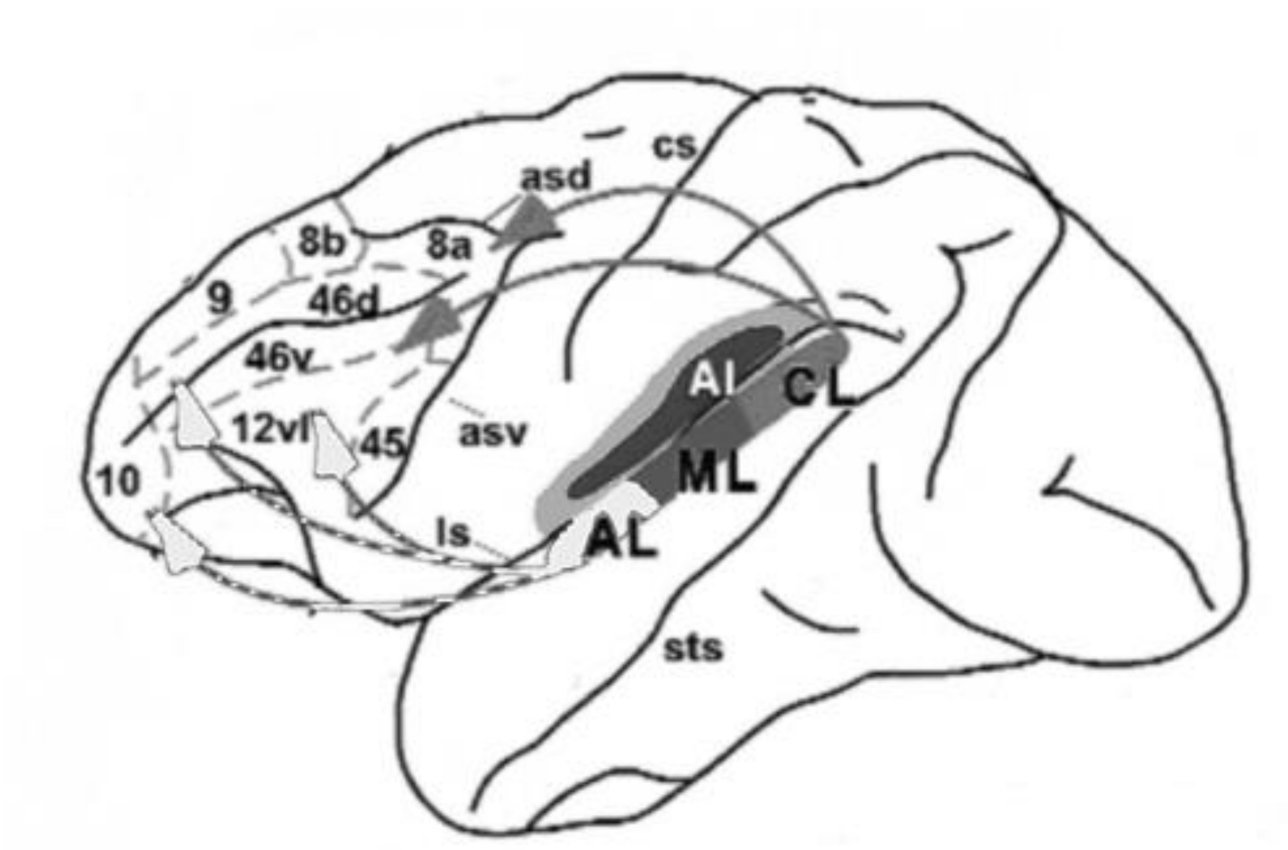
The Jeffress Model



Guinea Pig ITD Tuning Curves

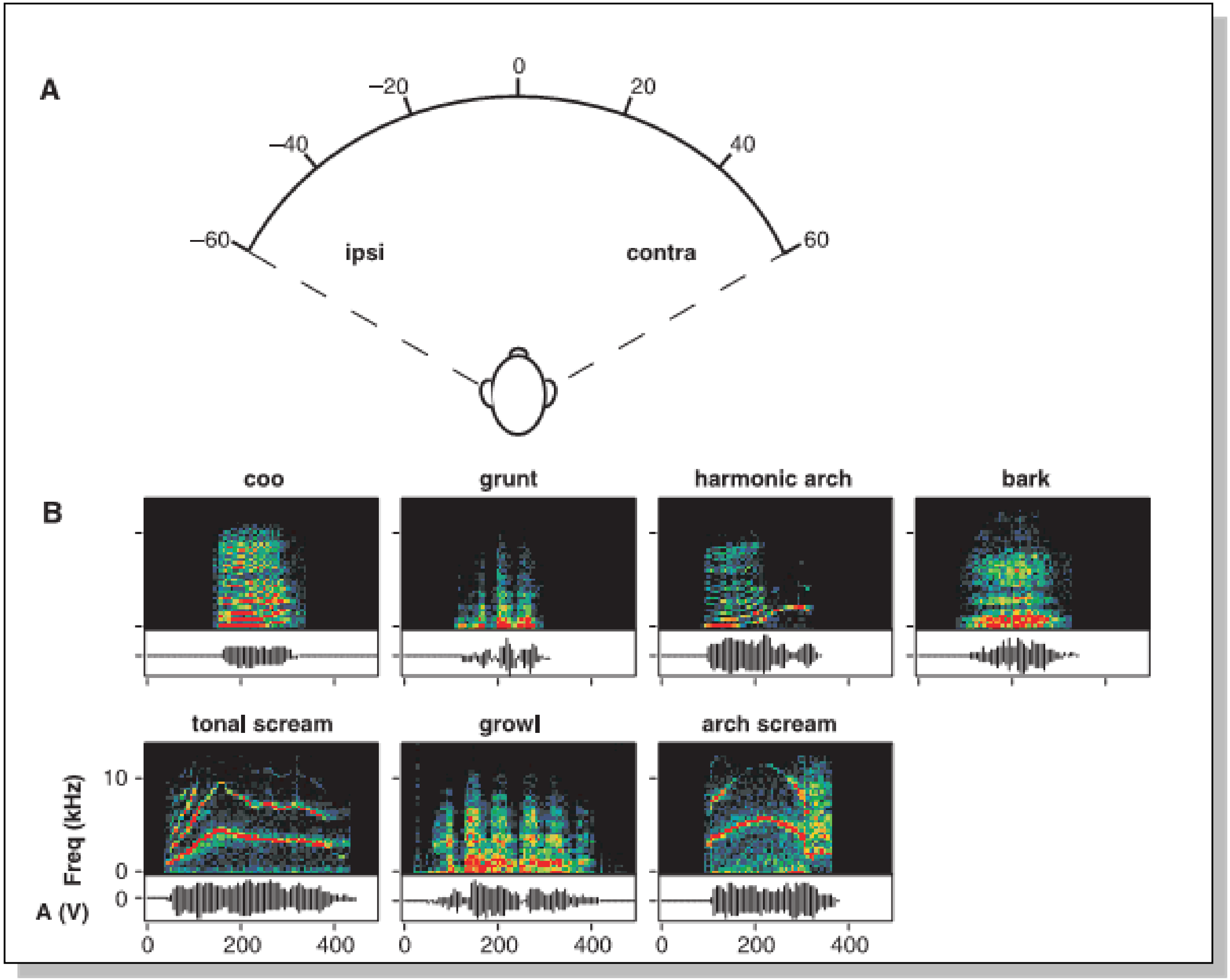


Where and What Streams in Auditory Cortex?

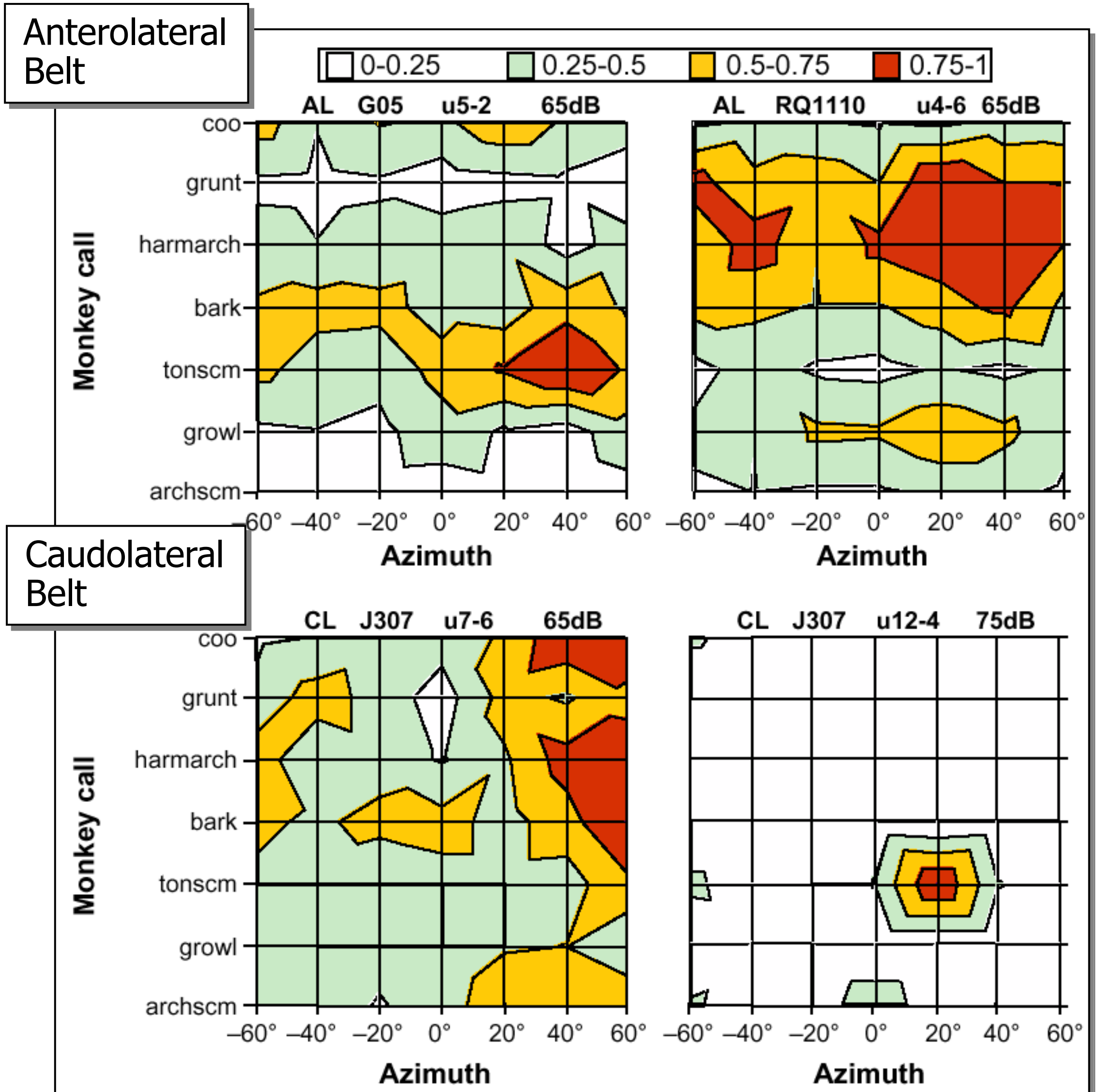


After Romanski et al

The Rauschecker & Tian 2001
"What vs. Where experiment"

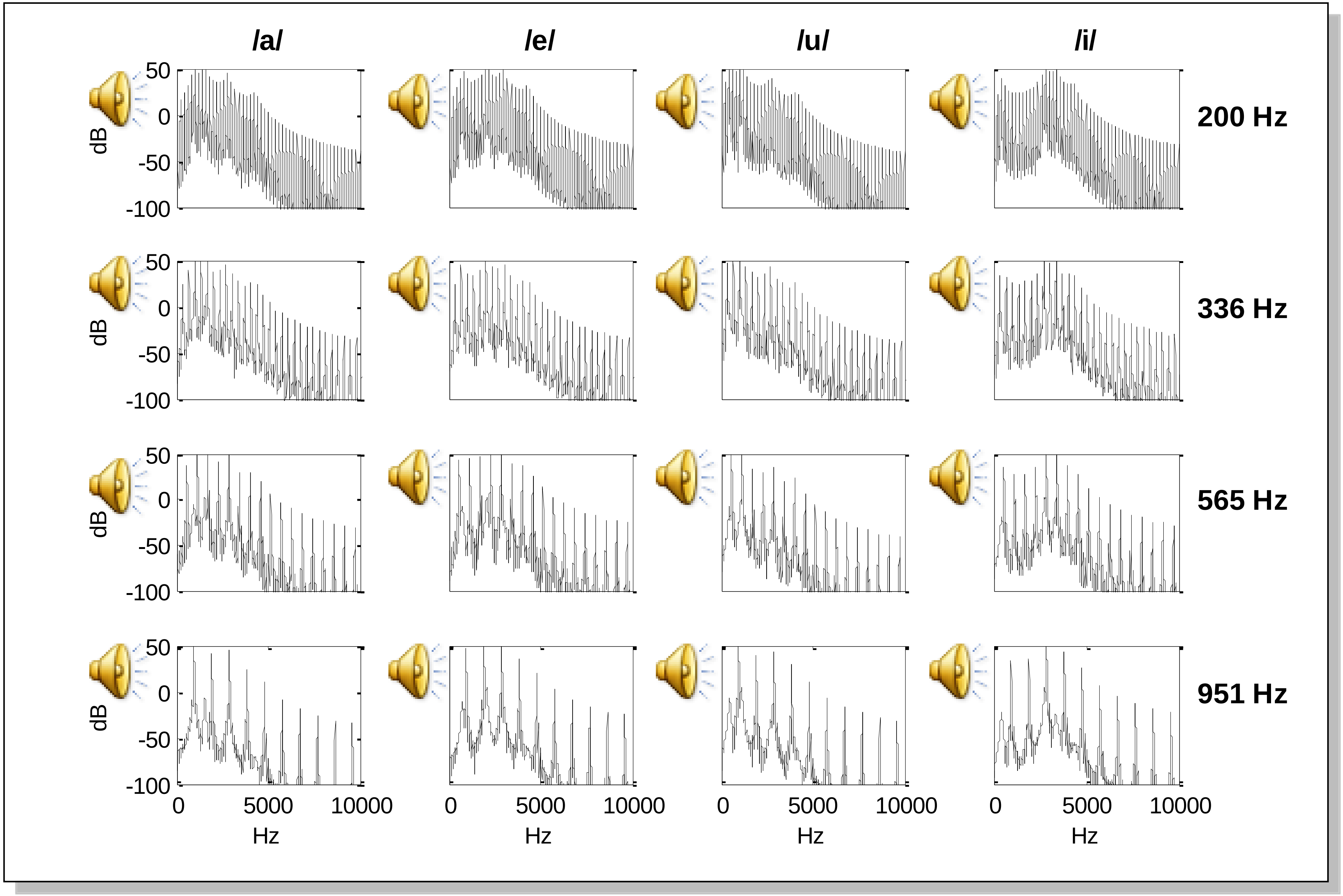


Are there “What” and “Where” Streams in Auditory Cortex?

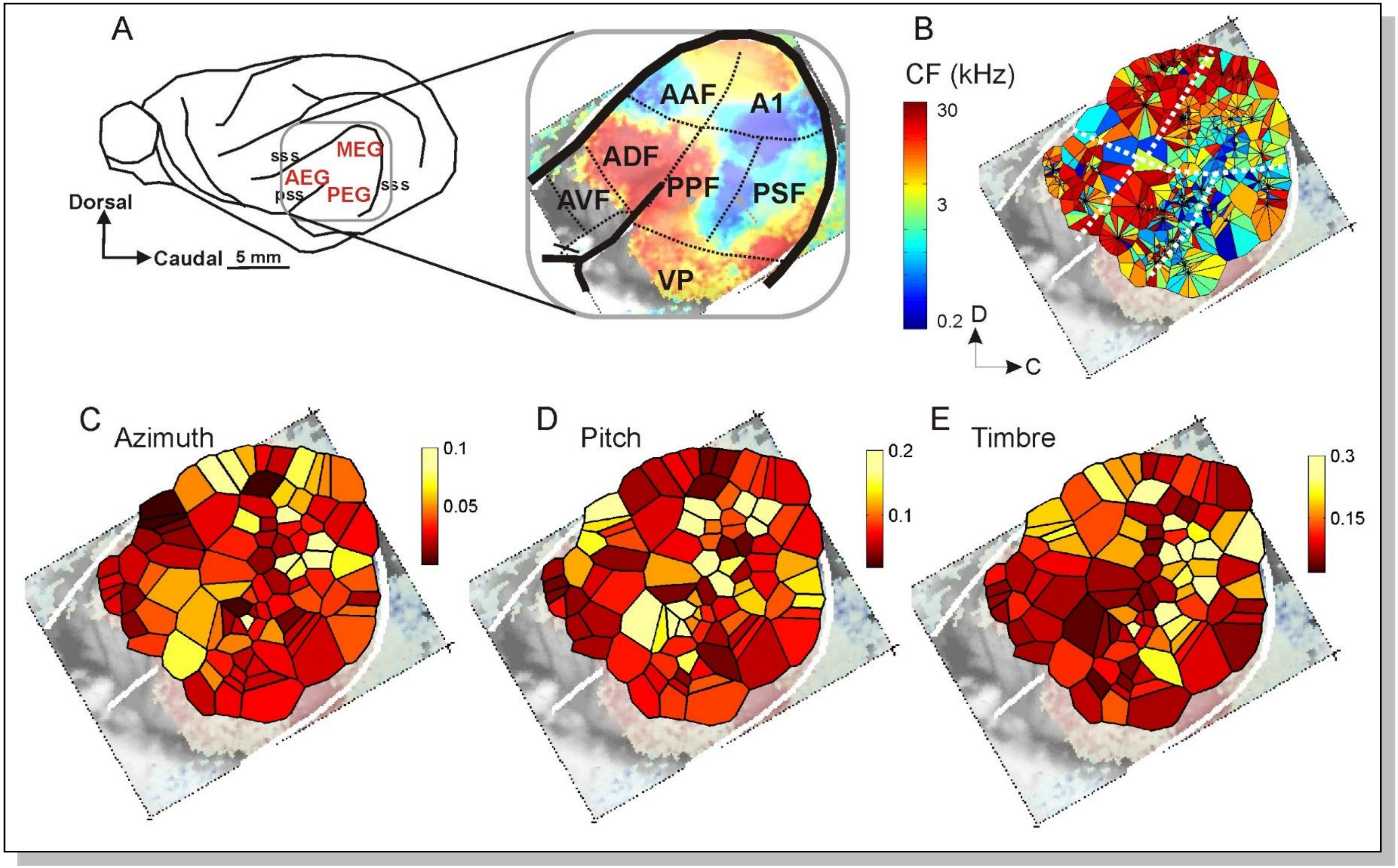


- Some reports suggest that anterior cortical belt areas may more selective for sound identity and less for sound source location, while caudal belt areas are more location specific.
- It has been hypothesized that these may be the starting positions for a ventral “what” stream heading for inferotemporal cortex and a dorsal “where” stream which heads for postero-parietal cortex.
- Figure 2 from Tian & Rauschecker, Science 2001

Artificial Vowel Sounds

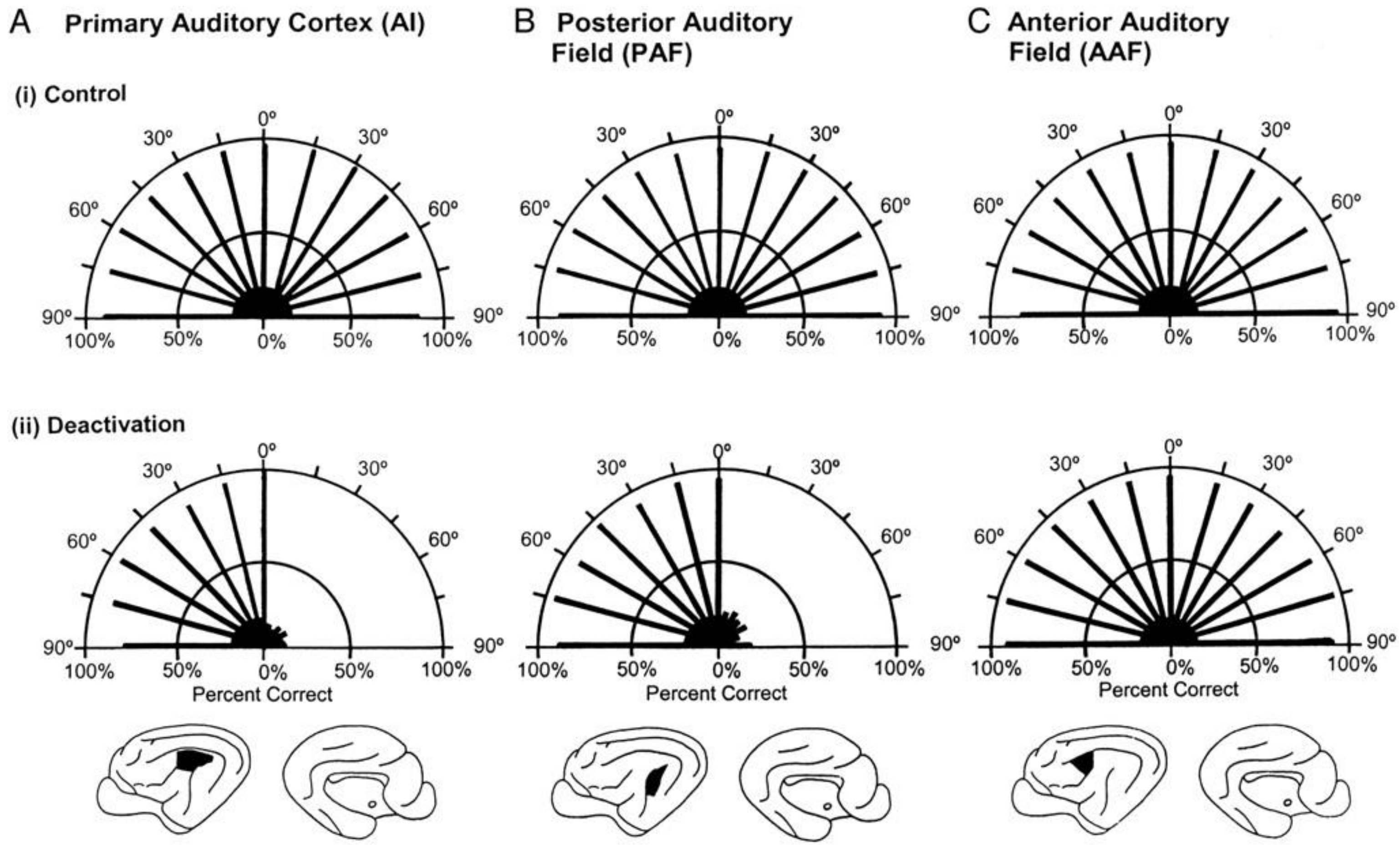


Azimuth, Pitch and Timbre Sensitivity in Ferret Auditory Cortex



JK Bizley, KMM Walker, BW Silverman, AJ King and JWH **Schnupp,** (2009)
Interdependent encoding of pitch, timbre and spatial location in auditory cortex. J Neurosci 29(7):2064-75.

Cortical Deactivation



- Deactivating some cortical areas (A1, PAF) by cooling impairs sound localization, but impairing others (AAF) does not.
- Lomber & Malhorta *J. Neurophys* (2003)

Spatial Hearing Summary

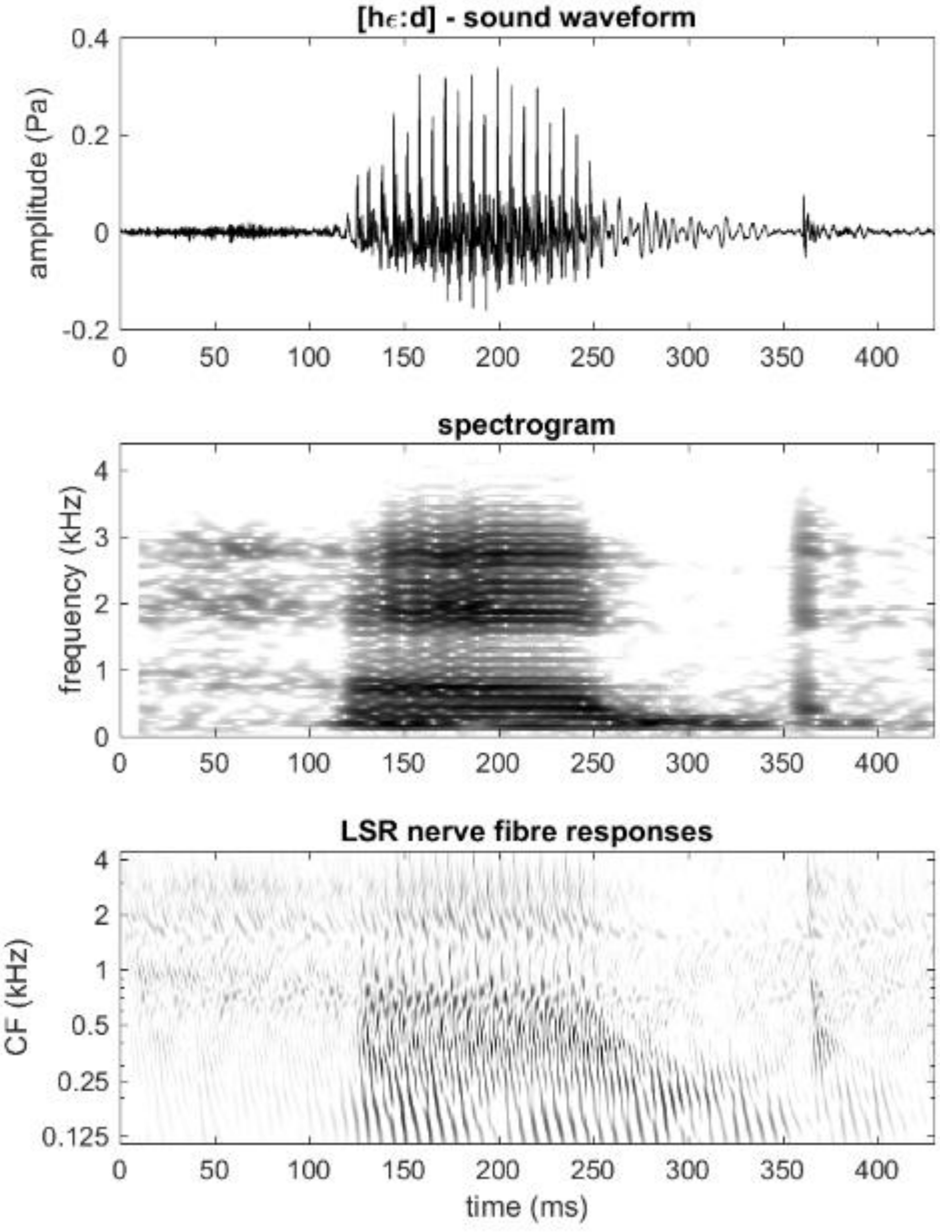
- Our brain processes binaural (ITD and ILD) as well as spectral cues to determine sound source direction.
- The brain's ITD sensitivity is remarkably precise, down to tens of microseconds.
- The lateral and medial superior olive are brainstem nuclei where binaural information about ILDs and ITDs respectively is first computed.
- Higher up in the brain, information from different cues and pathways in combined. Except for the superior colliculus, there is no "map of auditory space" in the mammalian brain.
- The degree to which different cortical areas specialize in spatial hearing is controversial.

7) Speech

Hearing Speech

- As we have seen earlier, speech sounds are created when vibrating vocal folds or temporary or partial obstructions of the vocal tract excite “formant” resonances in cavities in our vocal tracts.
- Change in vocal fold vibration frequency -> pitch change, important for melody, prosody, voice tone in tonal languages. (See earlier parts of this lecture.) Pitch not important in semantics in Indo-European languages. http://auditoryneuroscience.com/content/pitchInSpeech
- Change in resonances (formants) through “articulation” is responsible for most of the semantic category distinctions in speech. (Dynamics also play a role, e.g. “voice onset time”)
- http://auditoryneuroscience.com/topics/two-formant-artificial-vowels
- Brain transforms acoustic to phonemic to semantic information as information travels from primary auditory cortex to superior temporal gyrus and on to many other brain regions.

Encoding of the word "head" by auditory nerve fibers

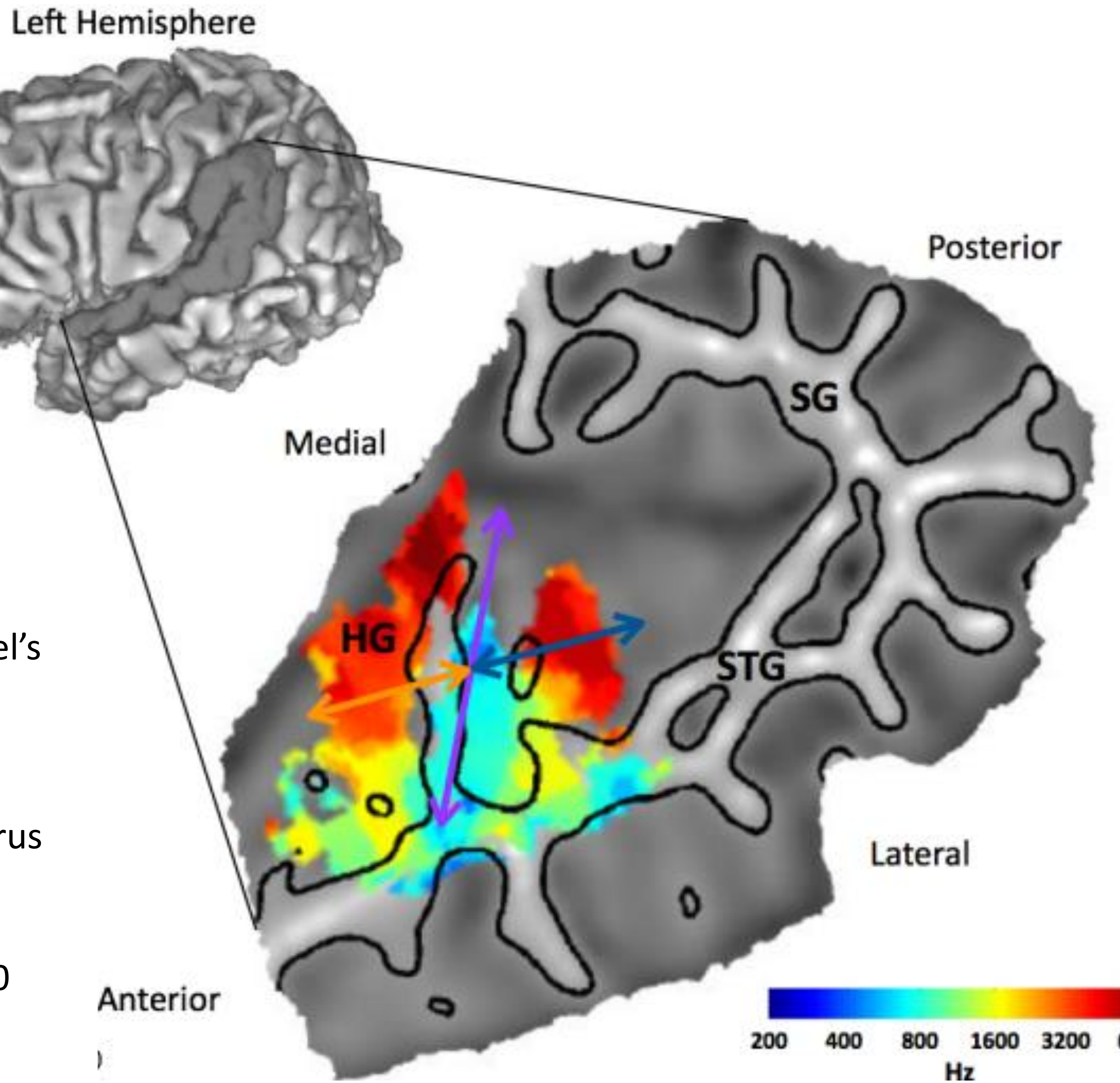


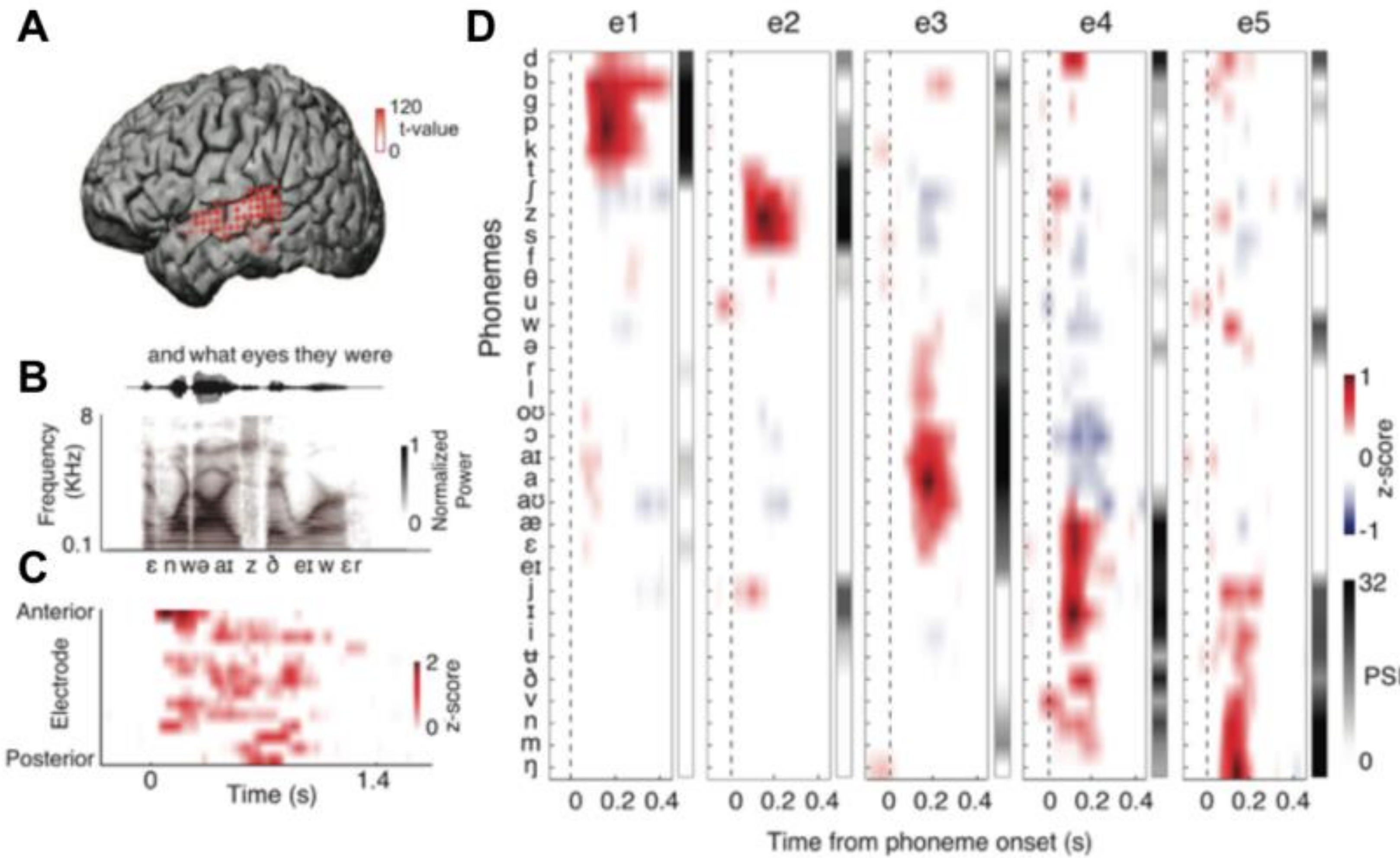
Auditory Cortex of the Human Brain.

Primary cortex is on Heschel's Gyrus.

HG: Heschel's Gyrus
STG: Superior Temporal Gyrus
SG: Supramarginal Gyrus

From Humphreys et al 2010 Neuroimage.





Human STG is sensitive to phonetic feature such as “manner of articulation”.

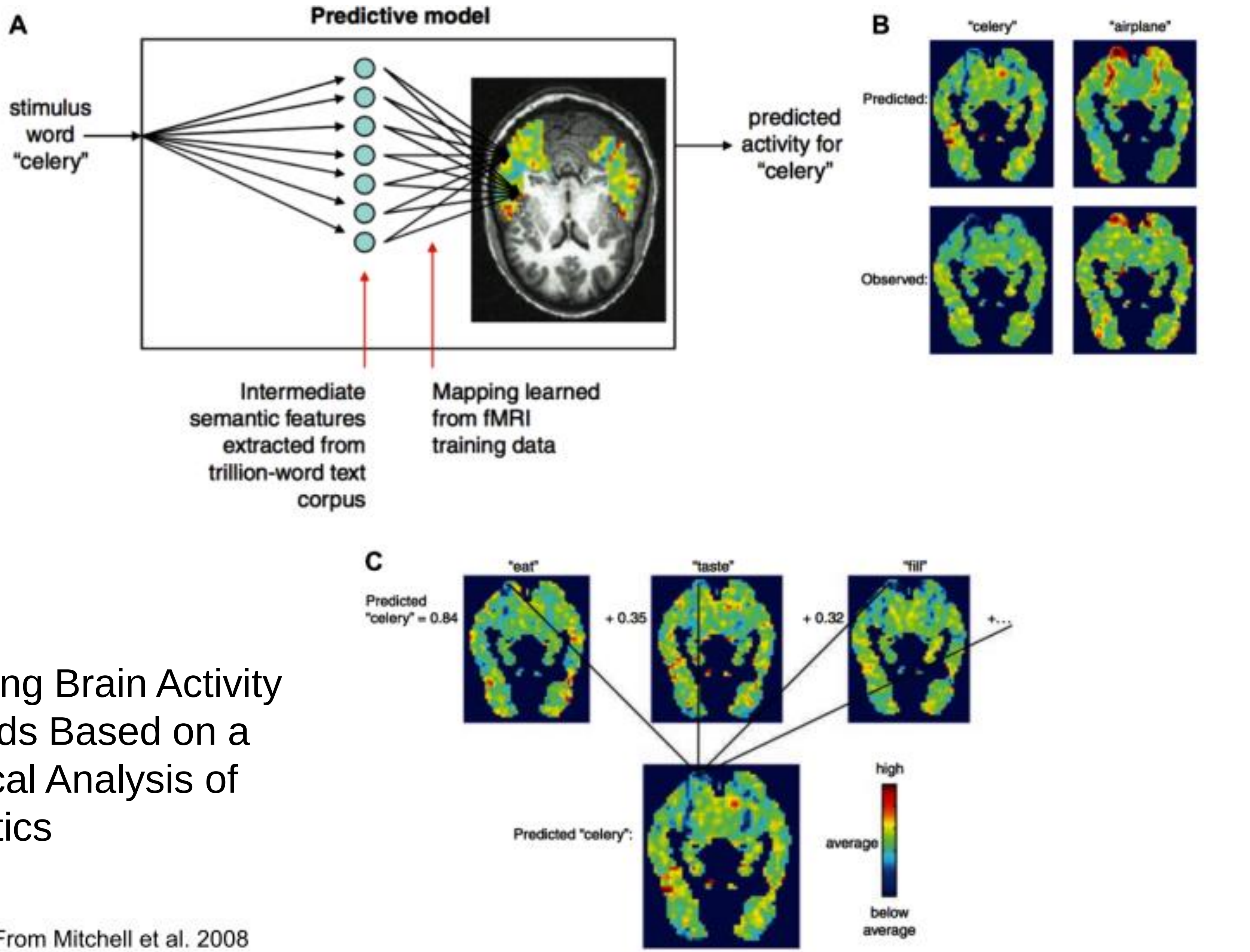
ECoG electrode arrays placed straight on the brain of epilepsy patients.

Electrode e1-> plosives. e2-> fricatives. e3 -> vowels. e5 -> nasals.

From Mesgarani et al (2014) Science.

Neural representations of the meanings of speech sounds

- “Meaning” is a big and somewhat poorly defined concept, which can make rigorous scientific study difficult, but it is possible to define semantic categories by statistical analyses of very large sets (corpora) of sentences, looking for words that occur in very similar contexts with other words (e.g. “strawberry” and “celery” may both co-occur with “kitchen”, “eat”, “grocery”, while “boat” and “plane” may both co-occur with “ride”, ”travel”, “course”, “navigate” etc.
- Mitchell et al (2008) argued that words which have similar meanings according to a statistical semantic analysis should produce similar brain activations in those parts of the brain involved in interpreting meaning. They built a “predictive model” which was able to predict evoked brain activity over large parts of the brain, including bilateral occipital and parietal lobes, fusiform and middle frontal gyri, and sensory cortex, as well as left inferior frontal gyrus, medial frontal gyrus and anterior cingulate. Meaning is “rich”, so it is perhaps not surprising that much of the brain is involved in encoding it.



Predicting Brain Activity for Words Based on a Statistical Analysis of Semantics

From Mitchell et al. 2008 Science, Figures 1 and 2