聲音有高度嗎？
音高之聽覺生理基礎

Do Sounds Have a Height?
Physiological Basis for the *Pitch* Percept

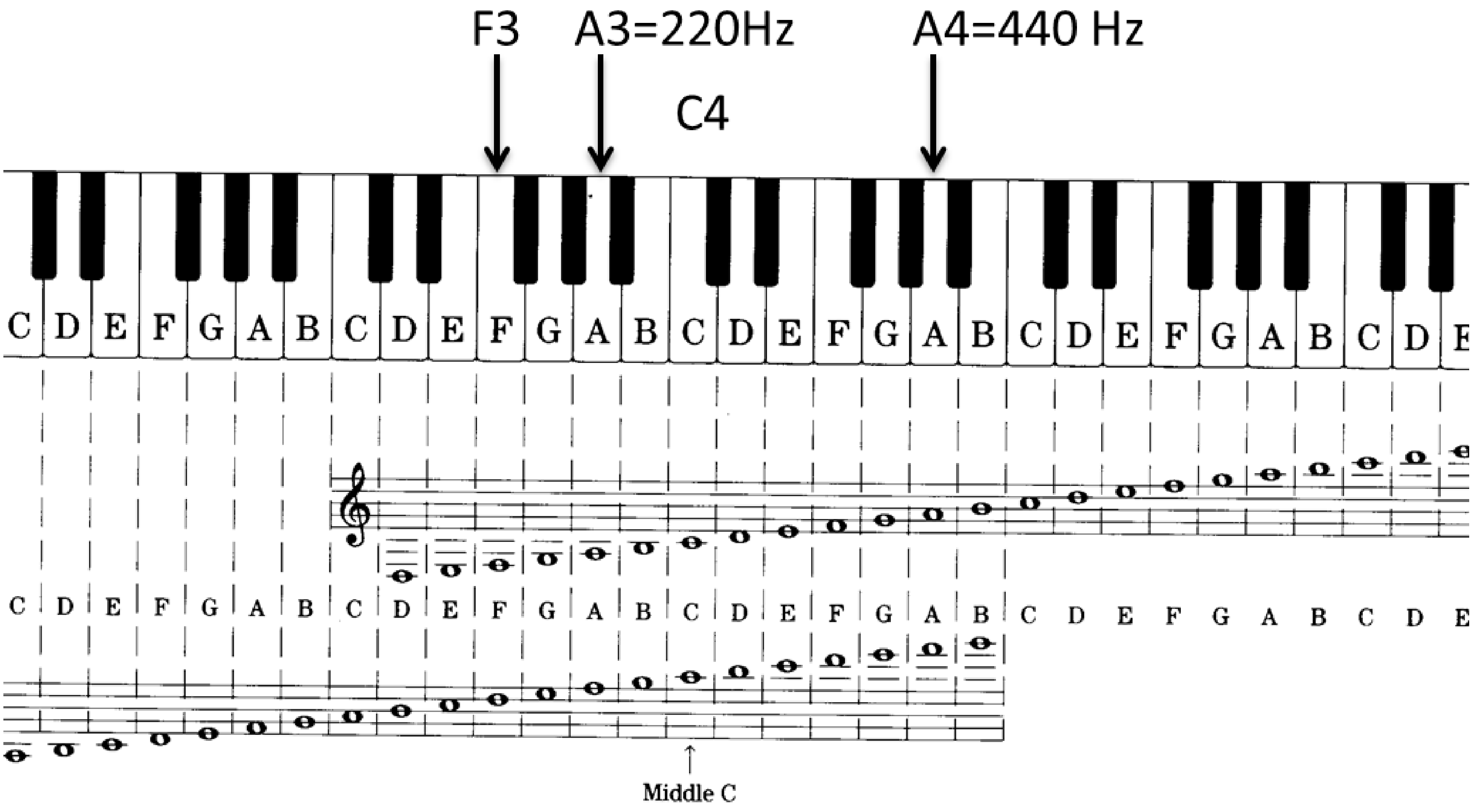
Yi-Wen Liu 劉奕汶
Dept. Electrical Engineering, NTHU

Updated Oct. 26, 2015

Do sounds have a height?
Not necessarily

- 樂音 vs. 噪音
- 語音 vs. 呢喃之音
- Let’s focus on sounds that do have pitch.
- Questions:
  - Definition of pitch?
  - How does the human auditory system encode the pitch?

Definition of musical pitch



Do-Re-Mi vs. C-D-E

- Note name: ABCDEFG. A4 = 440 Hz.
- Solfège: 教唱歌的唱法
  - 簡譜 1234567
- Musical Key: Every key can serve as the “Do”.
  - E.g. D-flat major.
- Major vs. minor scale
  - Do-Re-Mi-Fa-Sol-La-Ti-Do (全全半全全全半)
  - La-Ti-Do-Re-Mi-Fa(#)-Sol#-La (全半全全？？全)

Distance between adjacent *semitones*

- There are 12 semitones per octave
- So, in modern music, the semitones are “well-tempered”, meaning that:
- the frequency of C# is $\mathbf{2^{1/12}}$ times that of C, and so on.
    - $\mathbf{2^{1/12}}$ is approximately _____?
- In some literature, $\mathbf{2^{1/1200}}$ is called a *cent*.
    - How well can human tell a pitch is *off* ?

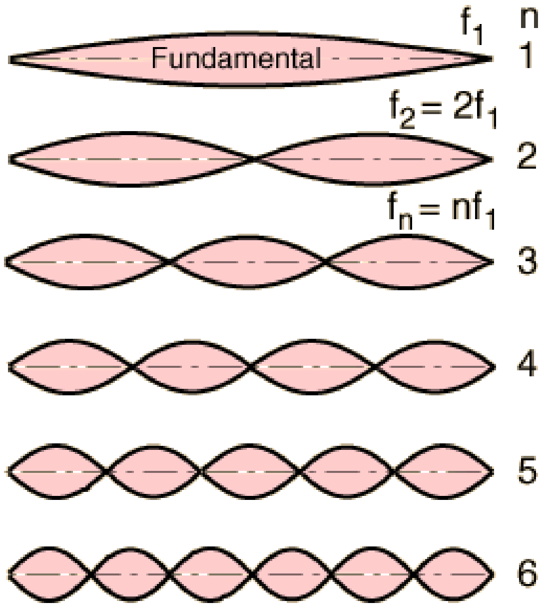
思考討論題
why 12 semitones per octave?

- Why not 10, 14, or other numbers?

Musical intervals

- major 5th = 7 semitones apart.
  - Frequency ratio = $2^{7/12}$, or approximately 3/2.
- Major 4th = 5 semitones apart.
  - Frequency ratio approx. 4/3.
- Major 3rd = 4 semitones, approx. 5/4.
- Minor 3rd = 3 semitones, approx. 6/5.

Physics of the (struck) string instruments in a nutshell



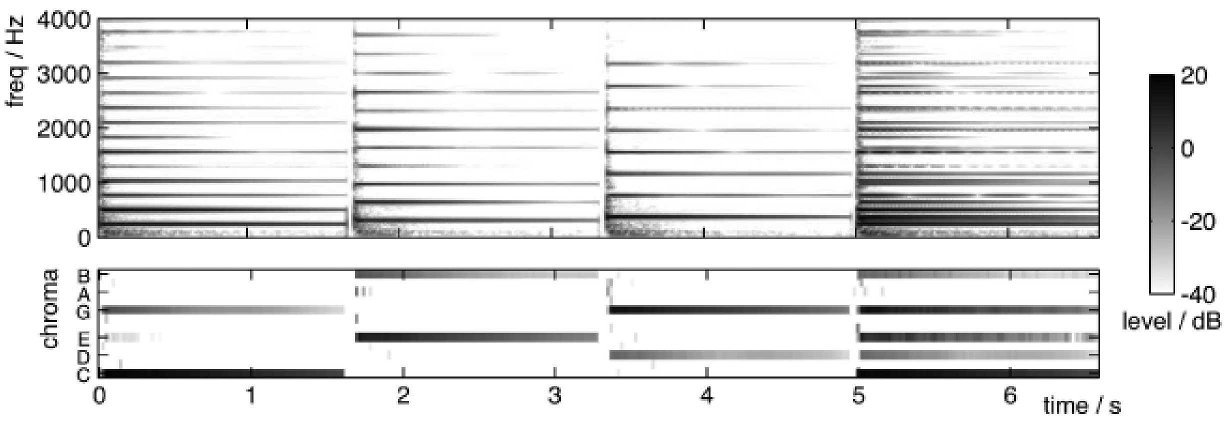


Fig. 2. Middle C, followed by the E and G above, then all three notes together—a C Major triad—played on a piano. Top pane shows the spectrogram; bottom pane shows the chroma representation.

延伸討論

Why certain **chords (和絃)** sound more "harmonic" than other?

Consonance vs. dissonance

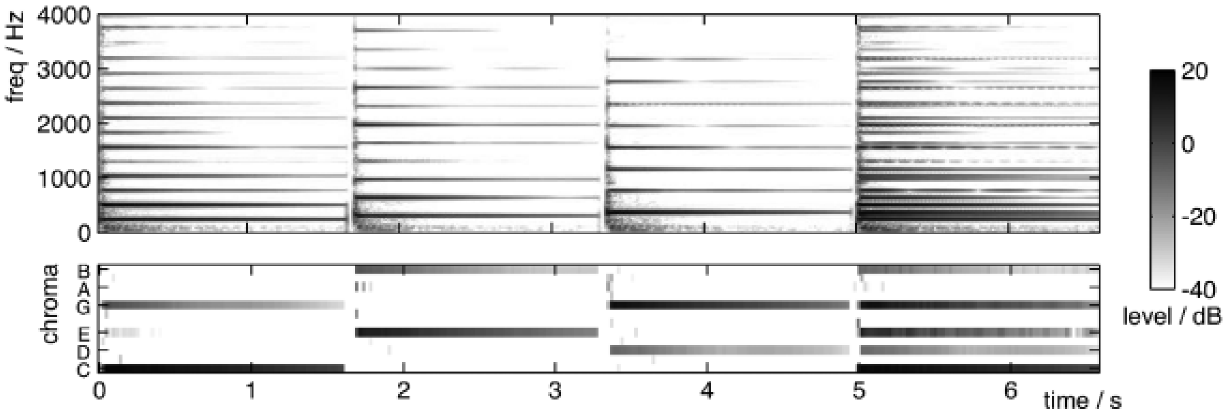


Fig. 2. Middle C, followed by the E and G above, then all three notes together—a C Major triad—played on a piano. Top pane shows the spectrogram; bottom pane shows the chroma representation.

延伸討論2: *Timbre*

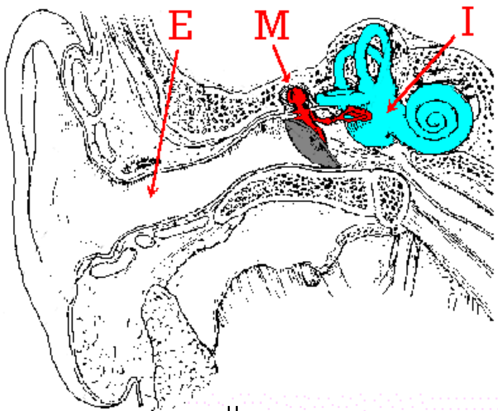
- Why do different instruments sound different?

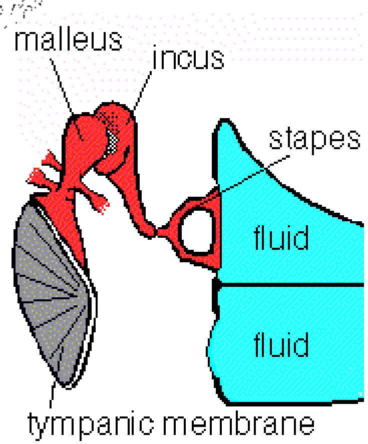
- Why do different people’s voices sound different?

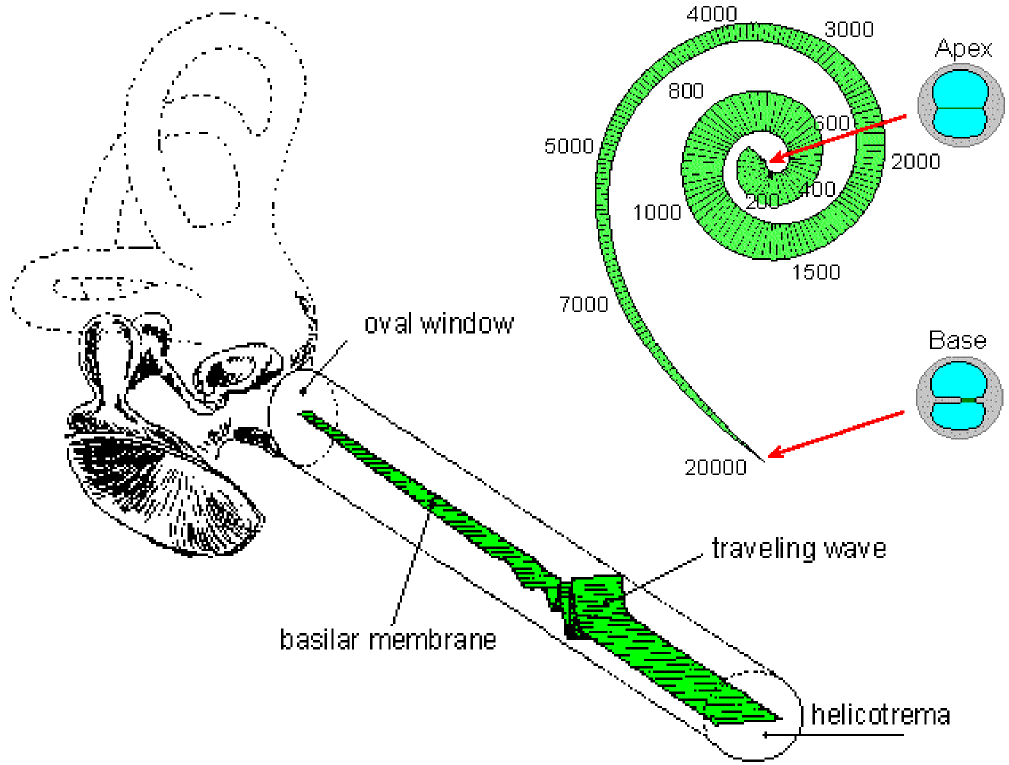
Frequency-to-place mapping in the auditory system

- **Cochlea**, the spectral analyzer
- Auditory nerve
- Auditory brainstem
- Midbrain – thalamus – (primary) **auditory cortex**

Tonotopic organization in the Cochlea

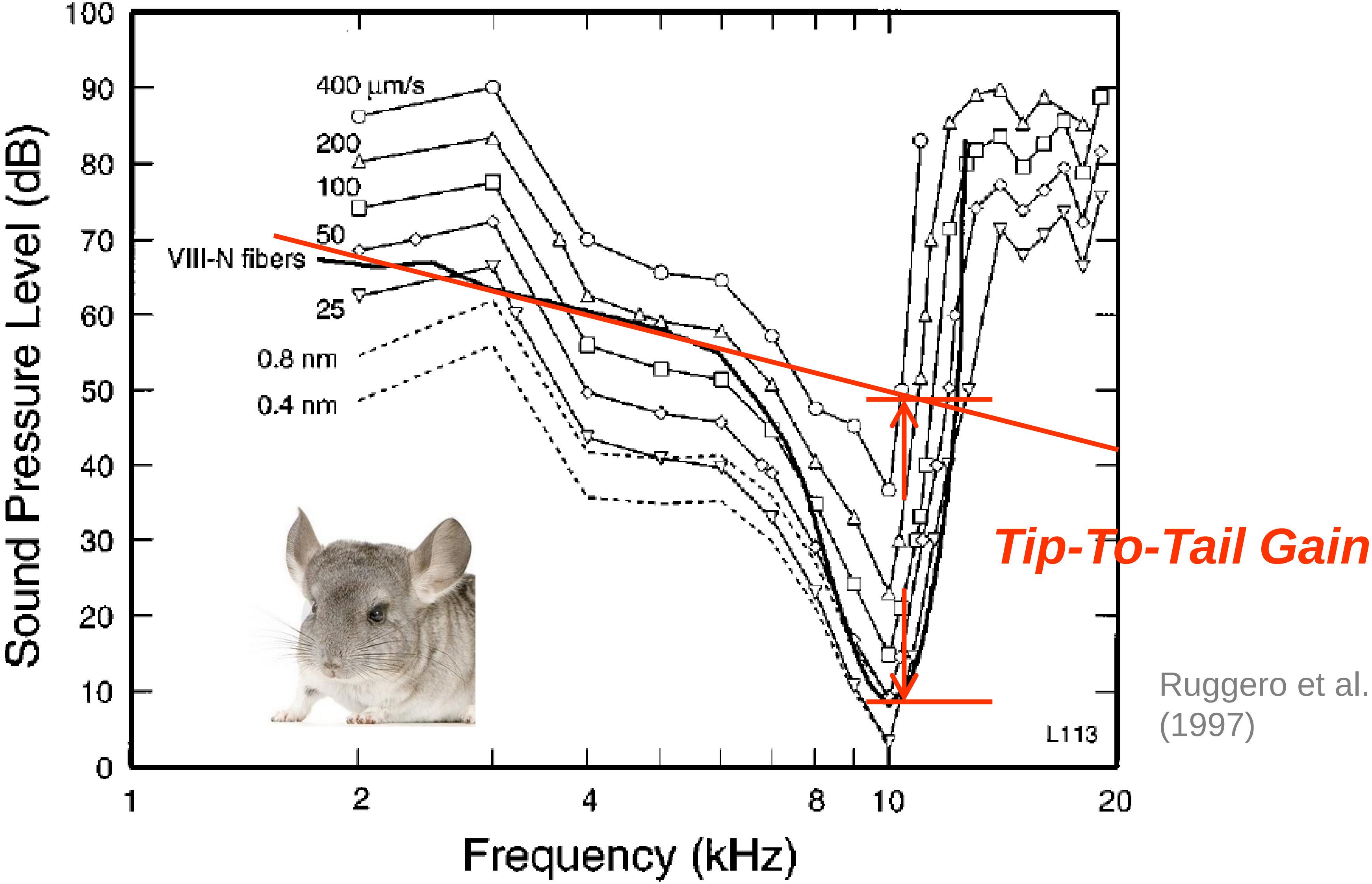






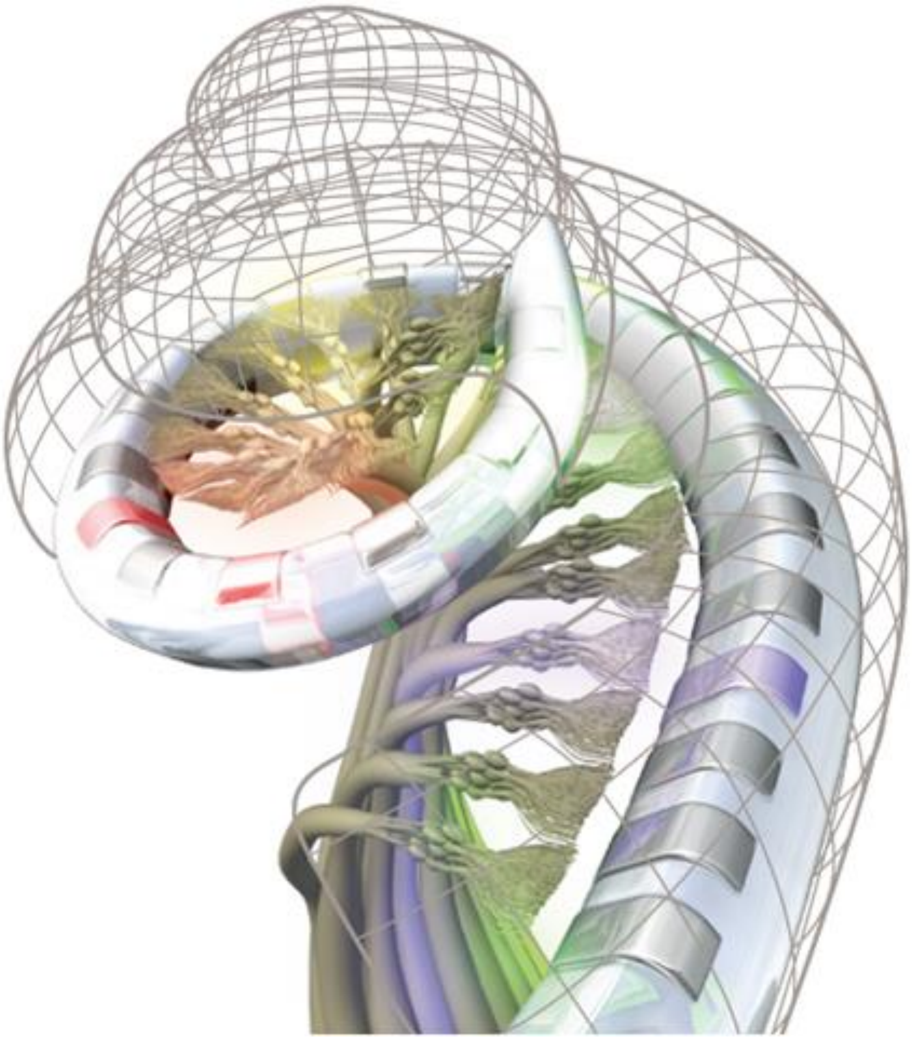
http://www.vimm.it/cochlea/cochleapages/theory/

Selectivity of cochlear frequency responses

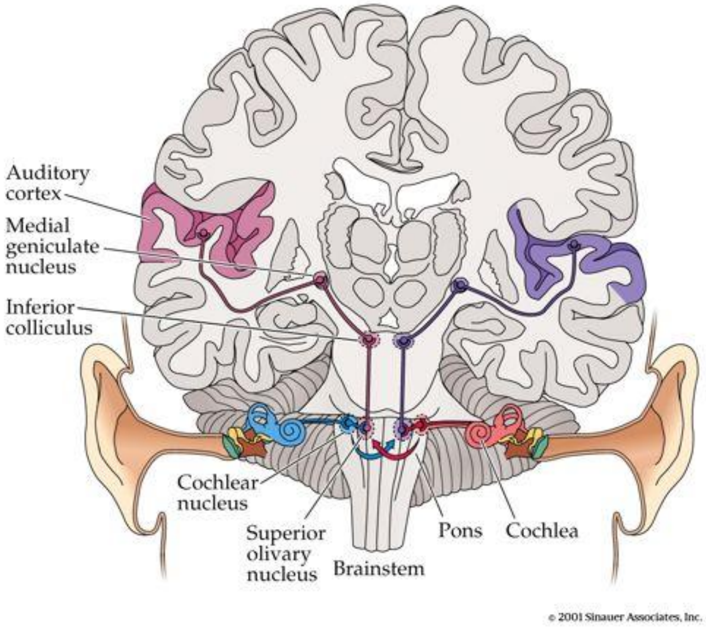


Ruggero et al. (1997)

Tonotopic organization in auditory nerves, and beyond



Ascending auditory pathways

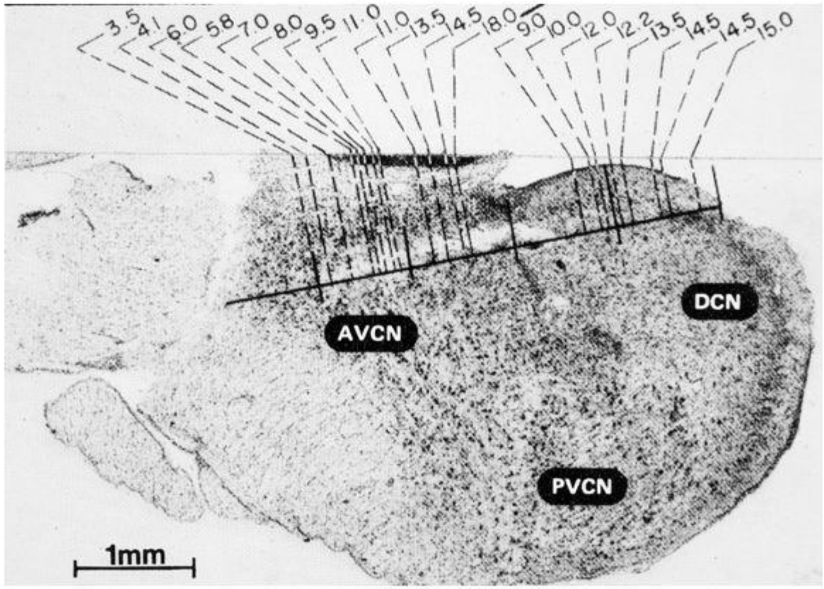


http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/localization/

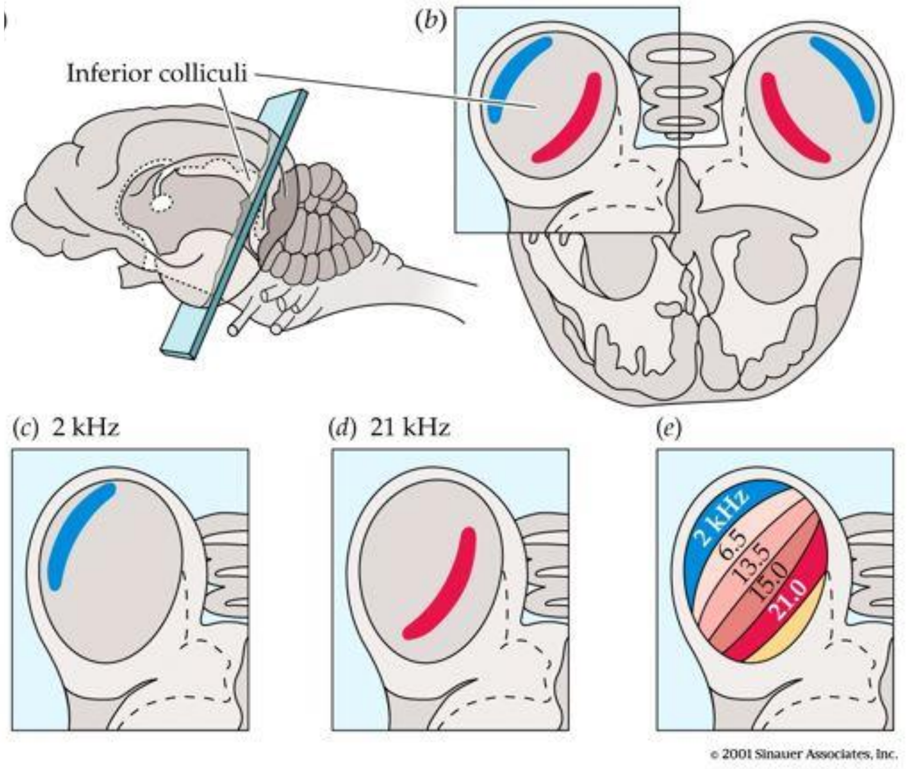
http://pronews.cochlearamericas.com/2013/02/cochlear-nucleus-electrodes-maximize-performance/

Tonotopic organization in the central auditory system

Cochlear nucleus



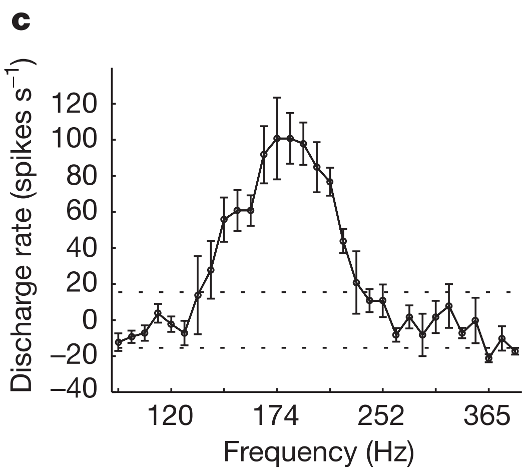
Inferior colliculus



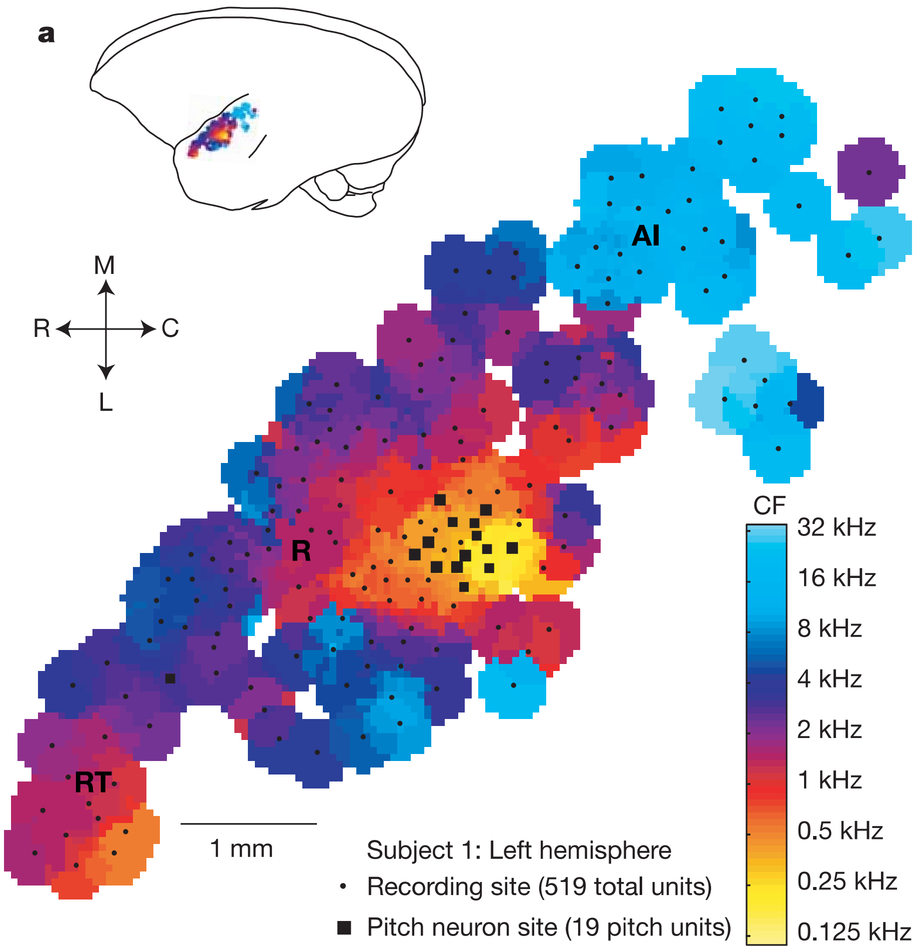
http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/localization/

Tonotopic organization in the auditory cortex

- Single-unit extracellular recordings.
- Awake marmosets.



http://commons.wikimedia.org/wiki/File:White-eared_Marmoset_3.jpg



Bendor and Wang. (2005). Nature 436: 1161-65.

音高之聽覺生理基礎。
MYSTERY EXPLAINED?

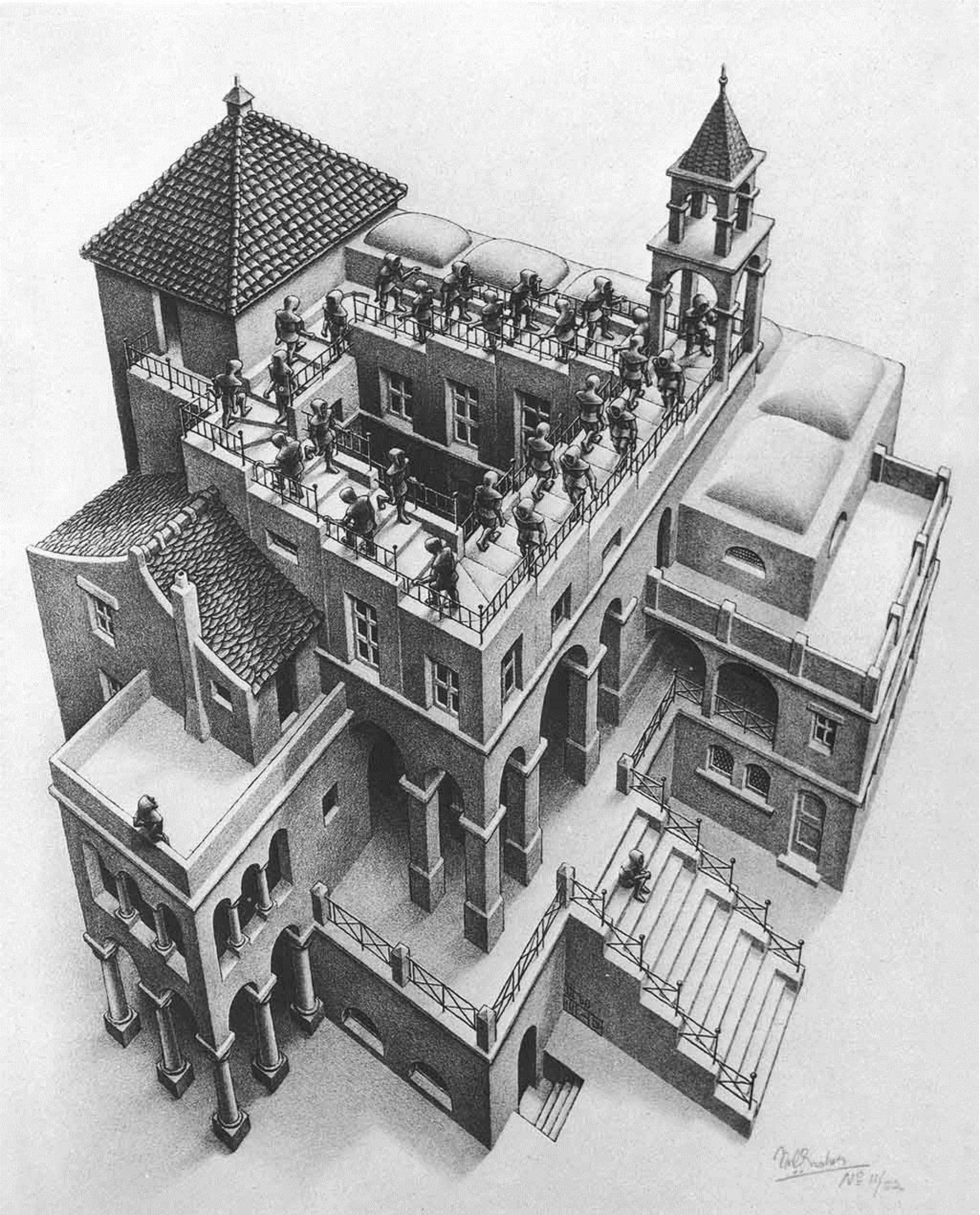
A few hard things to explain

- Octave similarity
  - 學習論
  - 物理論

- Violation of pitch ranking
  - 音高不見得具有絕對的高低順序

Violation of pitch ranking: Shepard's Tone

http://vimeo.com/34749558



Comments on Shepard's tone

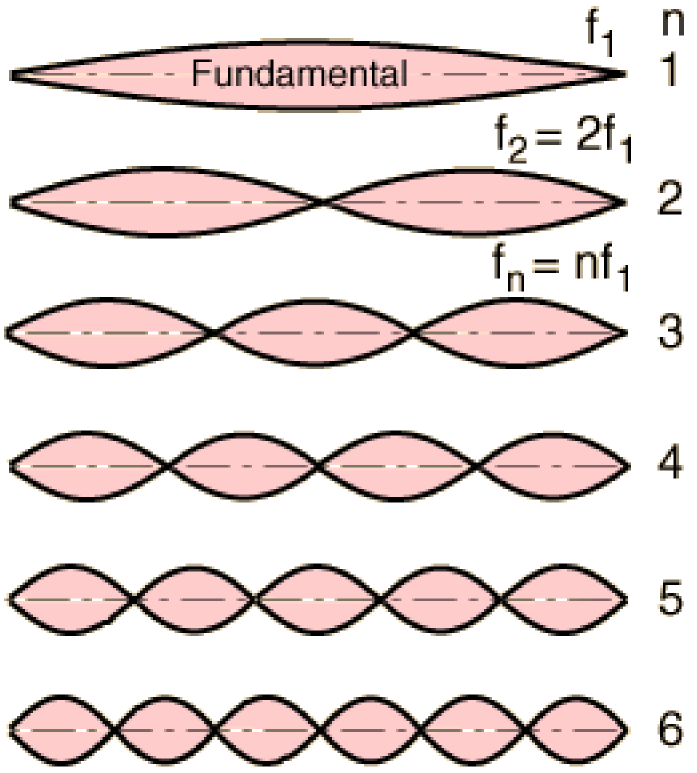
- Sounds can be digitally manipulated so their pitch relation becomes *circular*.
  - Algebraic structure of a modulo-12 system.
  - Don't try it at home.
- Pitch ranks can be context-dependent.
  - Distance between C and F# is the farthest apart.

A modified definition of the pitch

- Pitch is a percept that can be compared against that of a pure tone.
  - It often is the **fundamental frequency.**
  - Intentionally vague definition, so that A > B, B > C does not necessarily imply A > C.
  - Question: What then is the physiological basis for pitch?
    - Place coding vs. Time coding
    - Time-place conversion

Place coding vs. Time coding: the issue of *harmonic resolvability*

- Musical sounds are often periodic. Think of the vibration of a string.
  - Signal consists of components at f_0, $2f_0$, $3f_0$, etc.
- Cochlear filter bandwidth increases from low to high frequency.
- Therefore, higher harmonics can fall into the same filter, thus becoming *unresolved*.

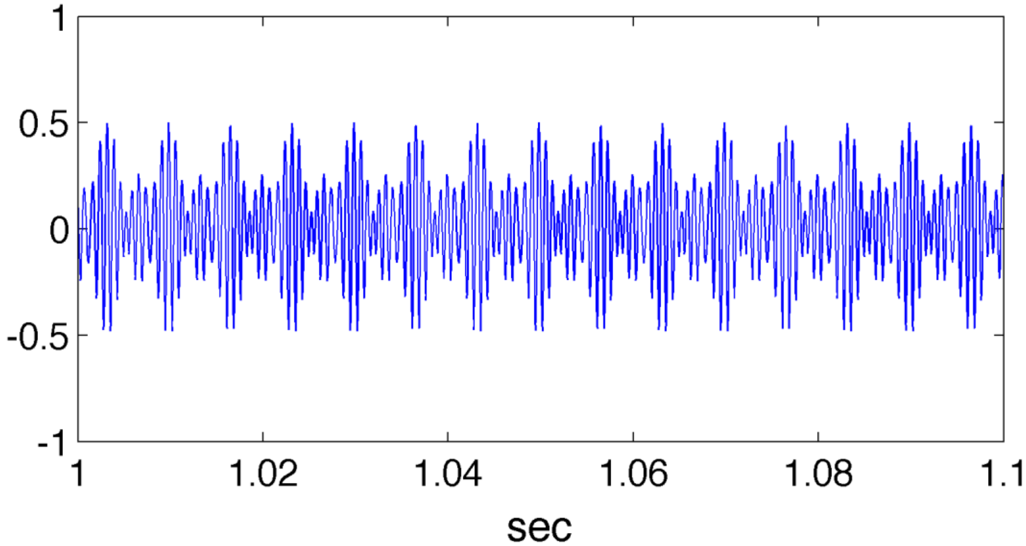


http://hyperphysics.phy-astr.gsu.edu/hbase/waves/string.html

Being unresolvable actually enables *time-coding*

- When multiple harmonics pass through one cochlear filter, they can encode the fundamental frequency via the timing information in neural firing patterns.

Example:
f_0 = 150 Hz;
sum of harmonics #8 to #10 (i.e., 1200, 1350, and 1500 Hz).



- Can explain consonance and dissonance
  -- In particular, octave similarity

Psychological evidence of time coding: The case of *missing fundamental*

Pure tone at 150 Hz

Tone complex with 10 harmonics

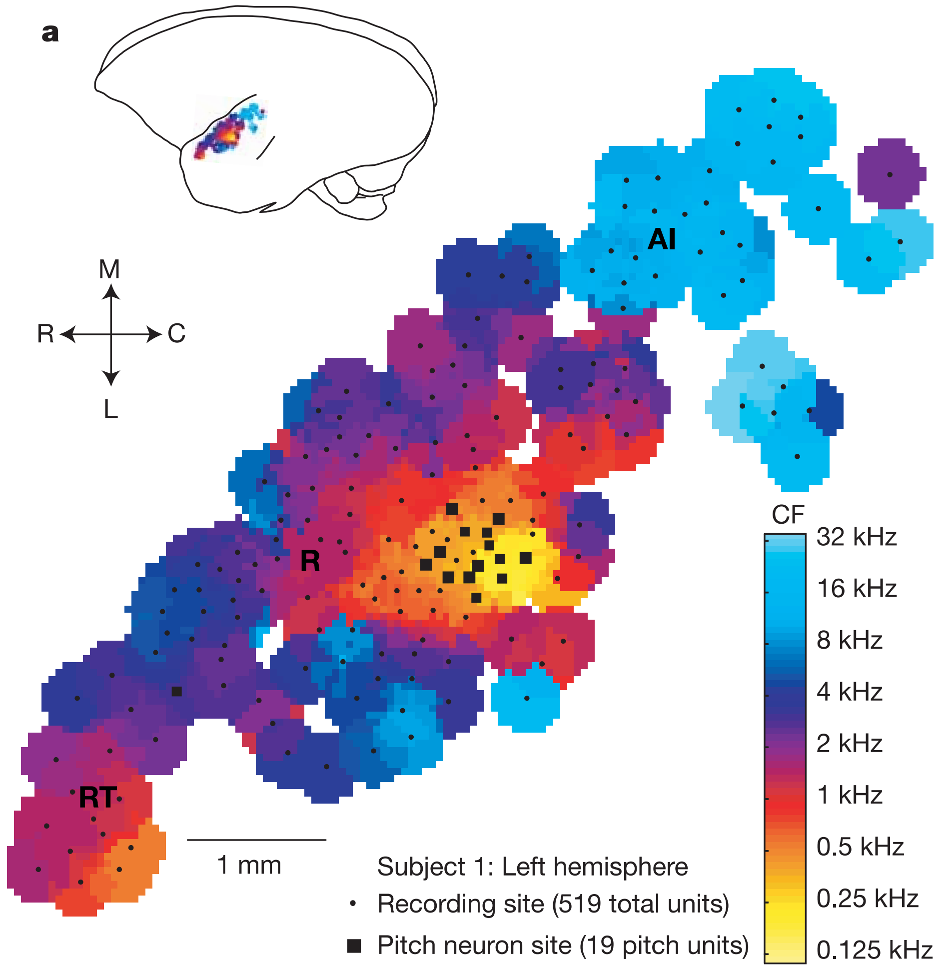
Harmonic number =
10, 9, 8, 7, 6, 5, 4, 3.

- Caution: Pitch percept could also be caused by “distortion product”

How about in the cerebral cortex?

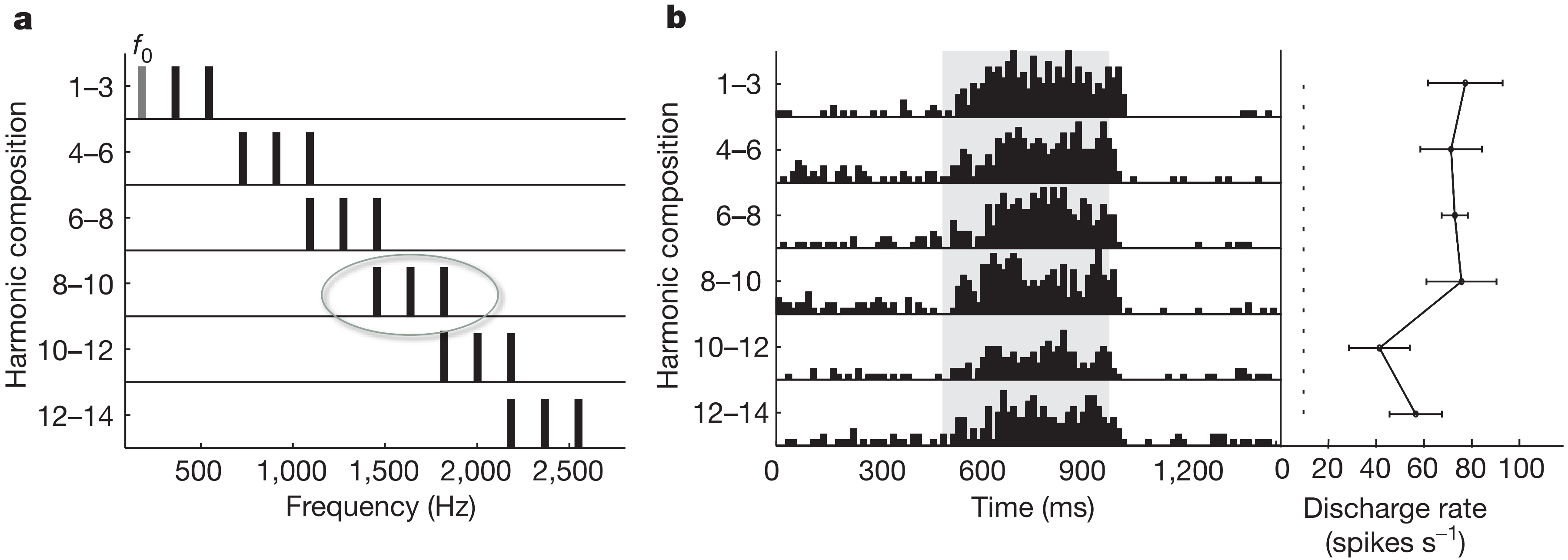
- Is pitch encoded by specialized neurons, or collectively by network oscillation?
  - Grandma’s cell for every pitch?

Pitch neurons in the auditory cortex!

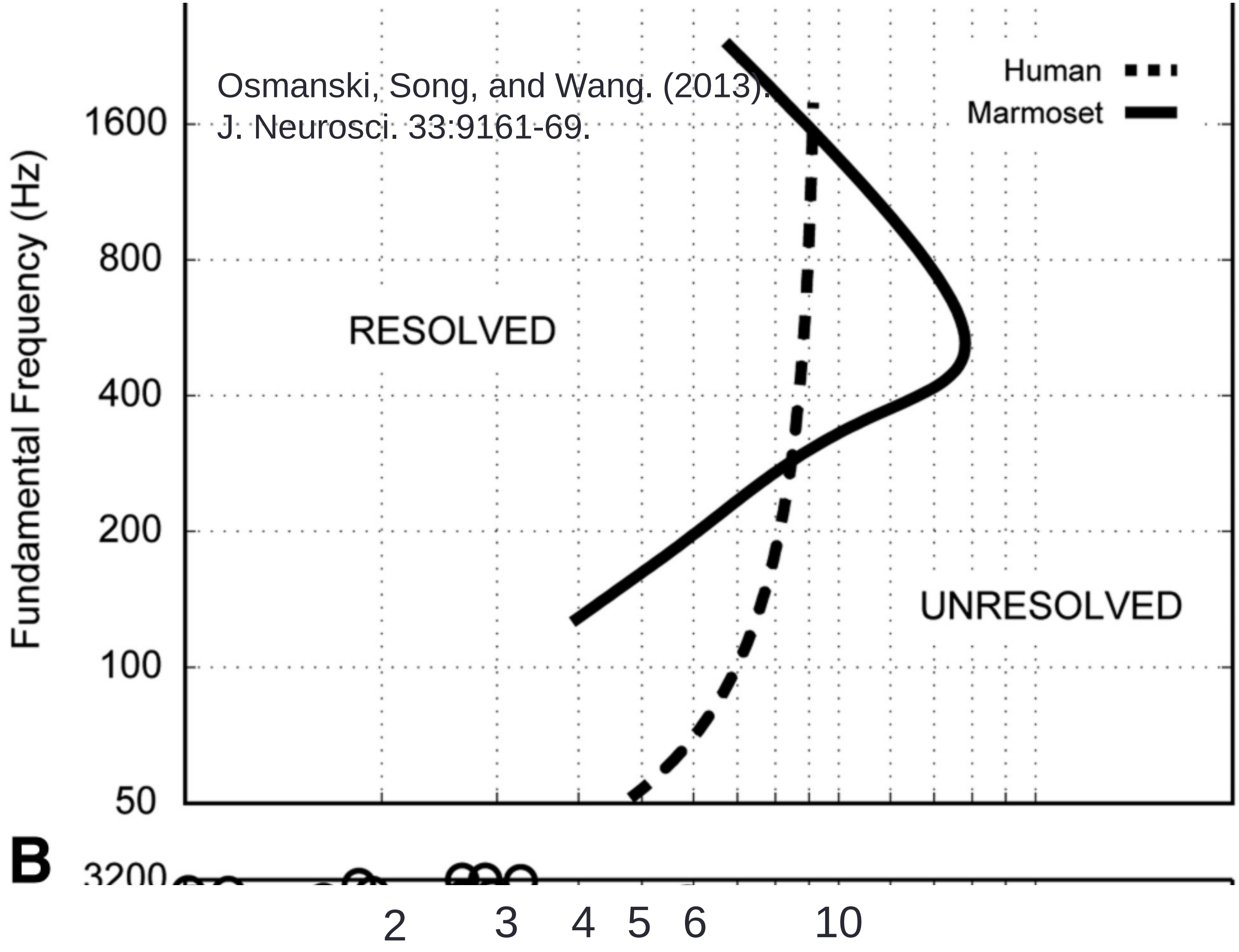


Bendor and Wang. (2005). Nature 436: 1161-65.

Pitch neurons: Stimulus and responses



Harmonic resolvability is inversely proportional to cochlear filter bandwidth



Comments on pitch neurons

- Now there are neurons that would specifically fire when the stimulus has a certain pitch.
  - **Regardless of** the harmonic composition (or ***timbre***).
  - Pitch information must have been processed at earlier stages along the auditory pathway.
  - But how?
    - (Of interests to engineers, too.)

Where and how do pitch neurons acquire the pitch information?

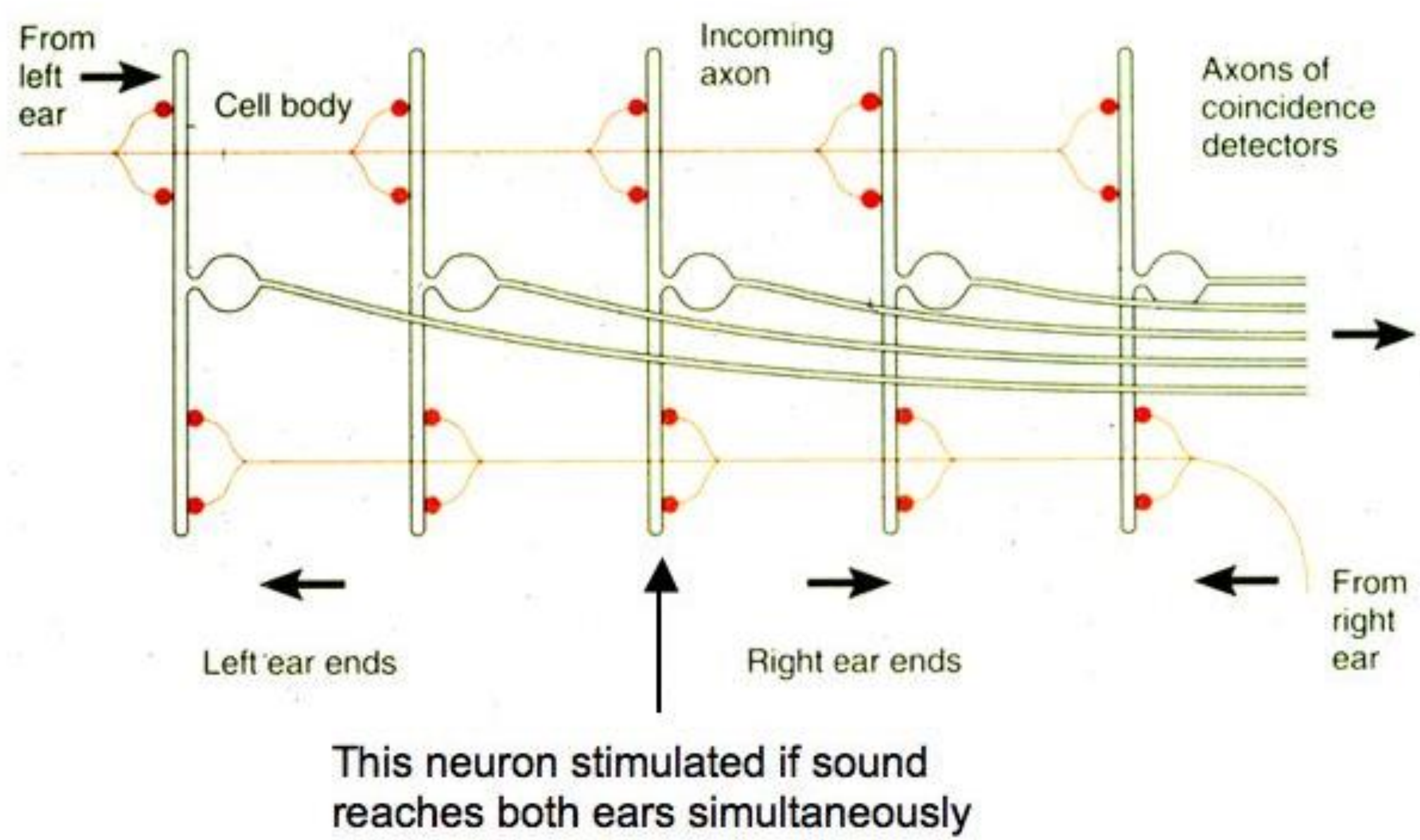
Time-to-place conversion

- Assume that time-coding would cause certain cochlear filter to fire at the rate of f_0.

- It was suggested that the periodic temporal firing pattern can be converted to maximal output at a certain place.
  - Might be achievable through *time-delay coincidence detector*

    - Licklider, JCR (1959). Three auditory theories, In S. Koch (Ed.), *Psychology: A study of a science*. Study I, Vol. I (pp. 41-144).

Time-to-place conversion by a coincidence detector



http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/localization/

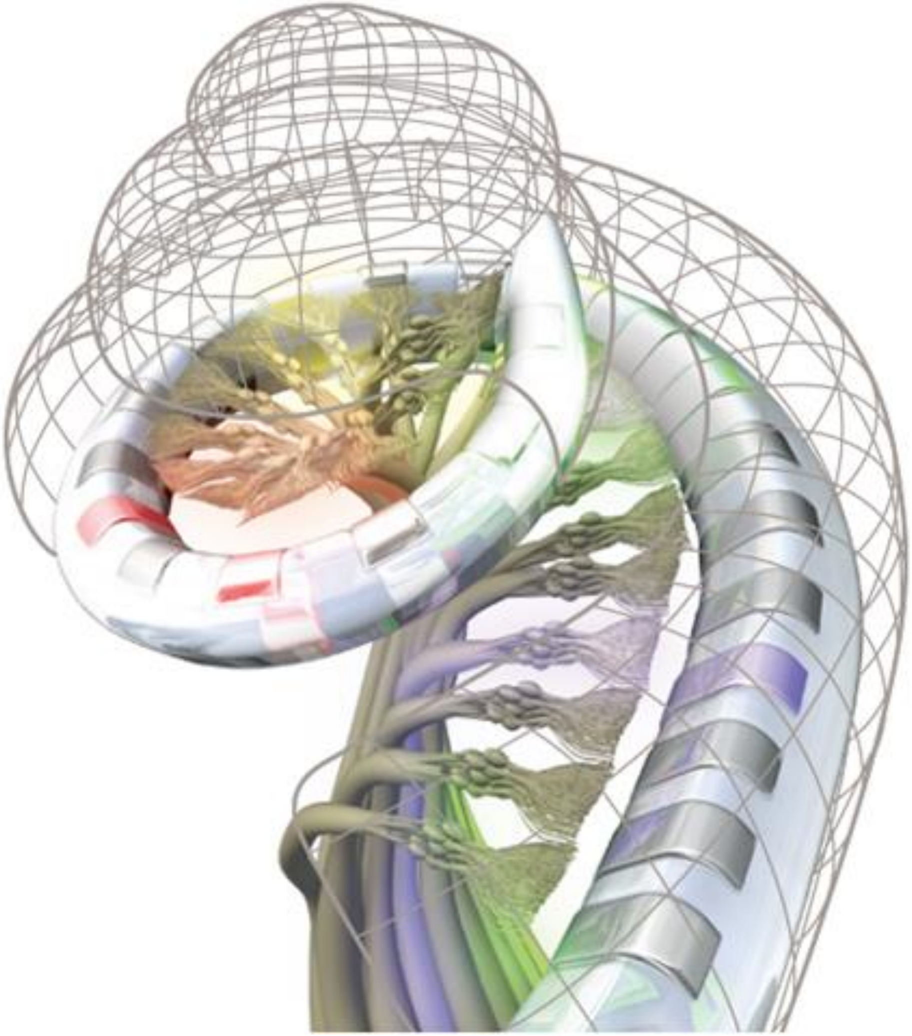
Summary: One pitch, two mechanisms

- Sounds with pitch are comprised of harmonics
- If f_0 is high, all audible harmonics are resolved and pitch is place coded.
- Otherwise, higher harmonics could be un-resolved, enabling the pitch to be time-coded.
  - Actually, at $f_0 < 500$ Hz, pitch might solely rely on time coding.
- Existence of pitch neurons in the auditory cortex suggests time-to-place conversion happens somewhere.

Open questions

- How does auditory system process **multiple pitch**?
  - Computational modeling and engineering applications
  - Measurement techniques?
    - fMRI?
    - MEG?
    - Electrode array recording?
- Relation to other functions in speech and music processing
  - Hemispheric difference

Final comment:
Pitch, the holy grail in auditory prosthesis



References

- Müller et al. (2011). "Signal processing for music analysis," IEEE J. Selected Topics in Signal Process., 5(6): 1088-1110.
- Poeppel et al. (2012). *The Human Auditory Cortex*, New York: Springer.

- Bendor D and Wang X (2005). "The neural representation of pitch in primate auditory cortex," Nature, 436:1161-65.
- Osmanski MS, Song X and Wang X. (2013). "The Role of harmonic resolvability in pitch perception in a vocal nonhuman primate, the common marmoset (*Callithrix jacchus*)," J. Neurosci. 33:9161-69.

Online materials

- Huron D. (2012). Shepard's Tone Phenomenon, video demo available at www.vimeo.com
- Prof. David Heeger's website at New York University http://www.cns.nyu.edu/~david/