Pitch and Timbre

Wu Xihong

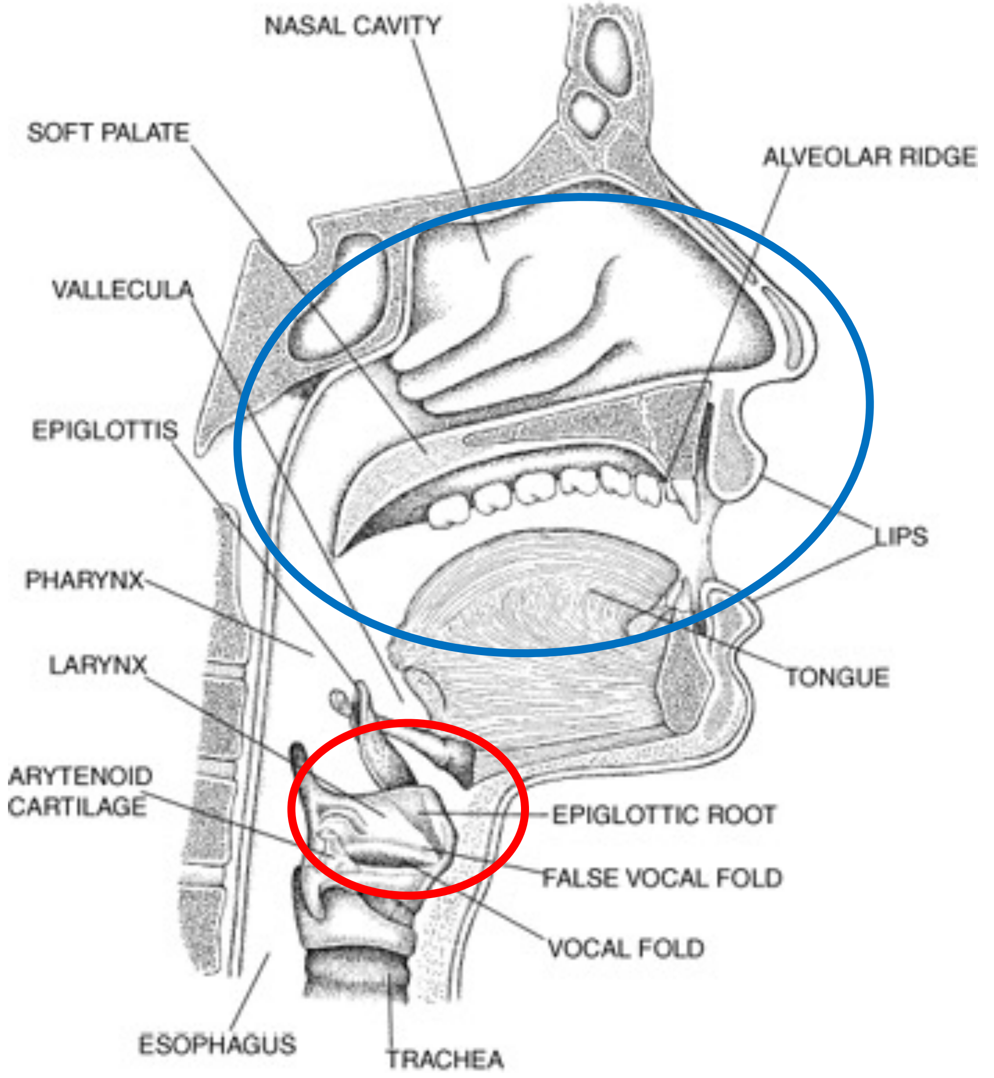
School of Artificial Intelligence

Physical Process of Sound Production

- Vibration of object → Wave → Sound
  - Common objects: Strings, rods, membranes, plates and shells
  - Different physical process: driving force, restoring force, object materials.
- Sound in Daily life:
  - Noise: thunder, footsteps, knocks at the door, drum, unvoicing speech, etc.
  - Music: piano, violin, voicing speech, etc.

"*The prime function of the auditory system is to tell the organism about its world by detecting, locating, and classifying sound sources*."

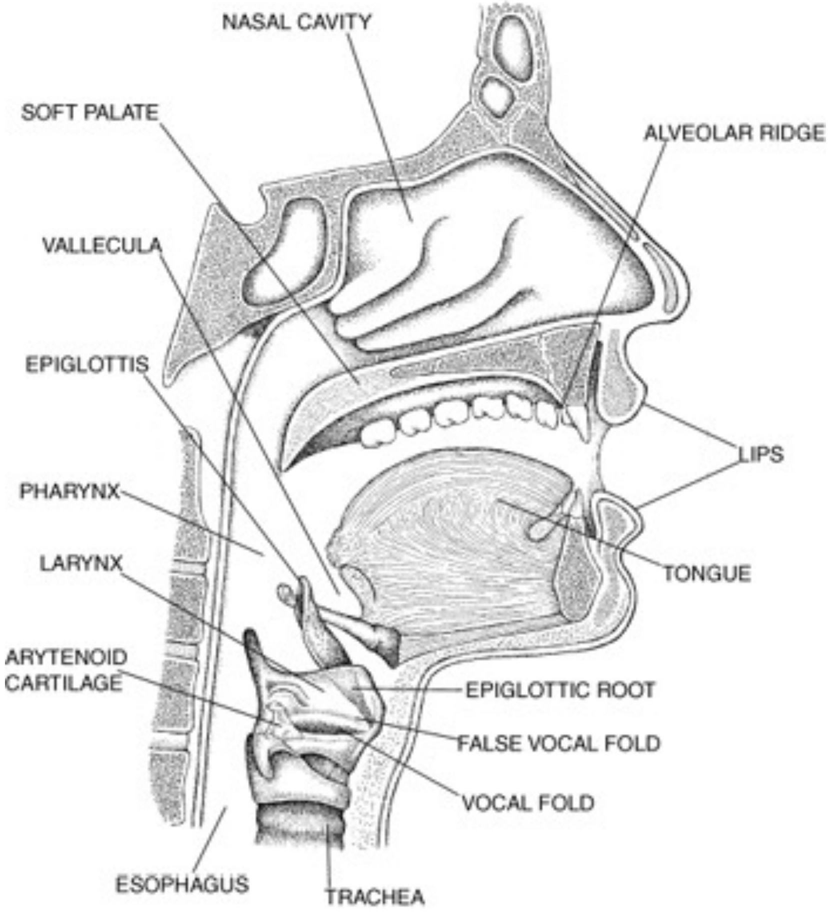
Vocal cords and Vocal tract



Vocal cords and Vocal tract

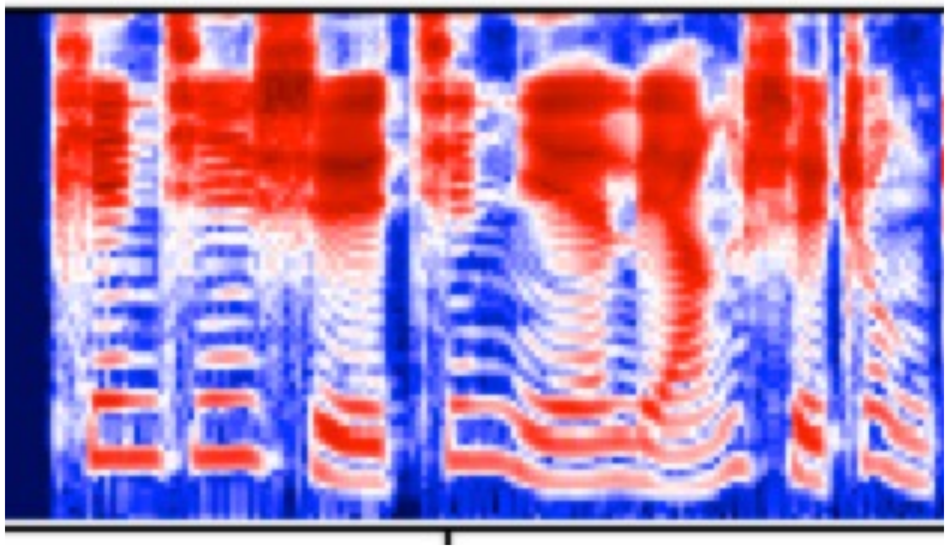
- Vocal cords:
  - Forced vibration: related with the length of vocal cords and strength.
  - Standing wave: overtone/harmonic.

Vocal cords and Vocal tract



- Vocal cords:
  - Forced vibration: related with the length of vocal cords and strength.
  - Standing wave: overtone, harmonic.
- Vocal tract:
  - Glottal wave resonates in the vocal tract
  - The gestures of vocal tract.
- Sound propagation:
  - Damping vibration
- Physical descriptions:
  - fundamental frequency, harmonic frequency and intensity.

Auditory Representations of Pitch



- *“In a complex sound environment, deciding what sound fragments should be grouped and identified as a single sources is perhaps the most difficult and important low-level processing task.”*

How are pitch relevant to auditory systems?

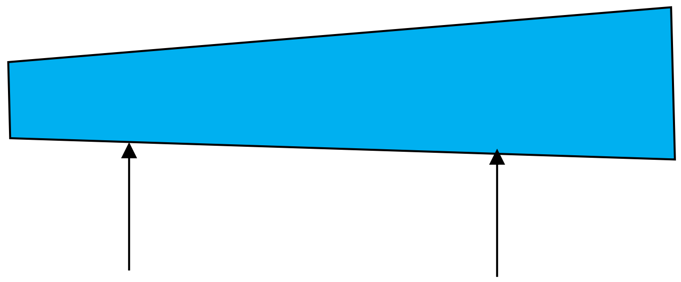
1. Pitch of pure tones

Pitch of pure tones

Place theory

Place of maximum in basilar membrane excitation (excitation pattern)

- <u>which</u> fibers excited



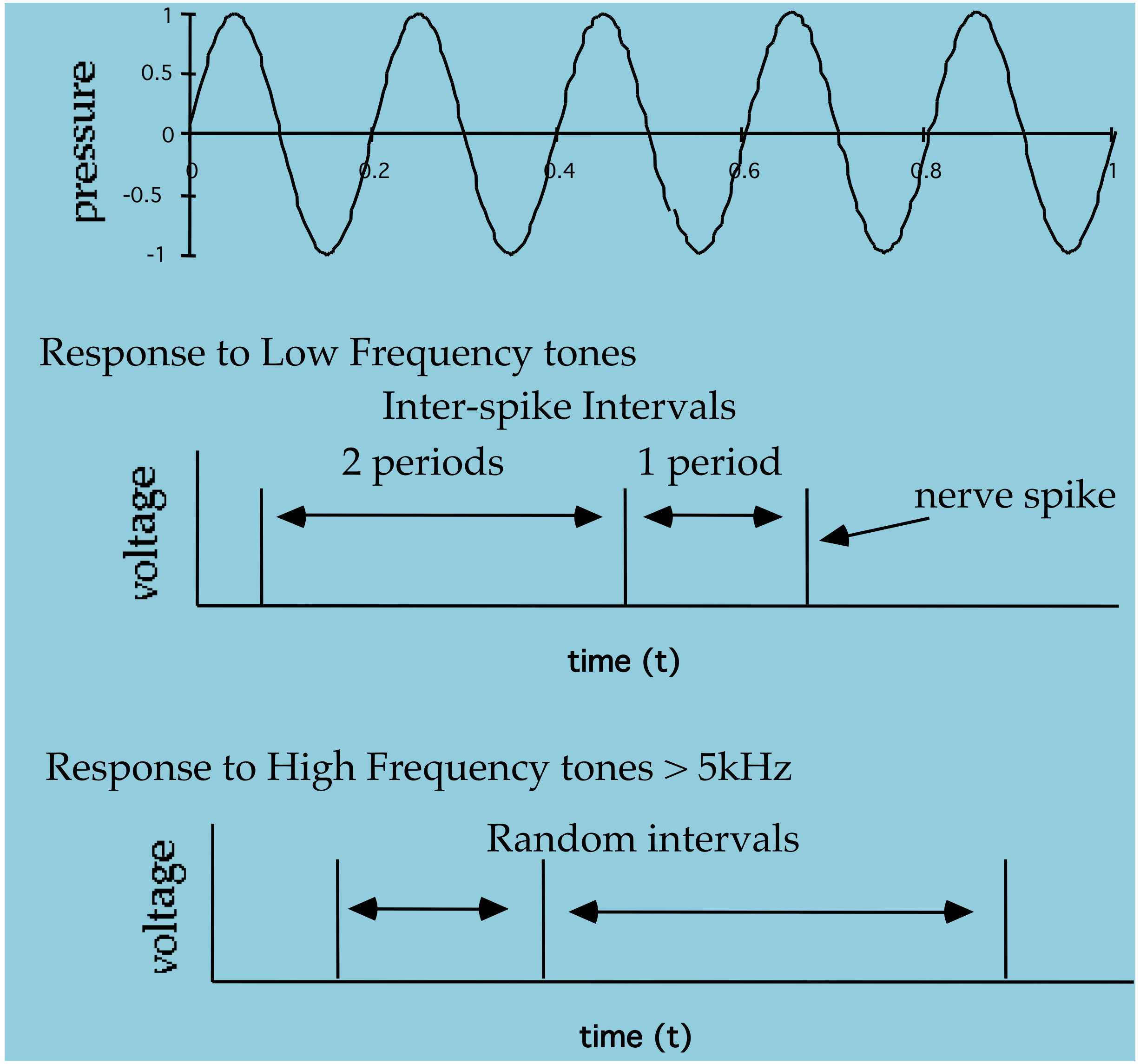
Timing theory

Temporal pattern of firing

- <u>how</u> are the fibers firing
- needs phase locking

Phase-locking

Auditory nerve connected to inner hair cell tends to fire at the <u>same phase</u> of the stimulating waveform.



Frequency difference thresholds increase above 4 kHz

BCJ Moore (1973) JASA.

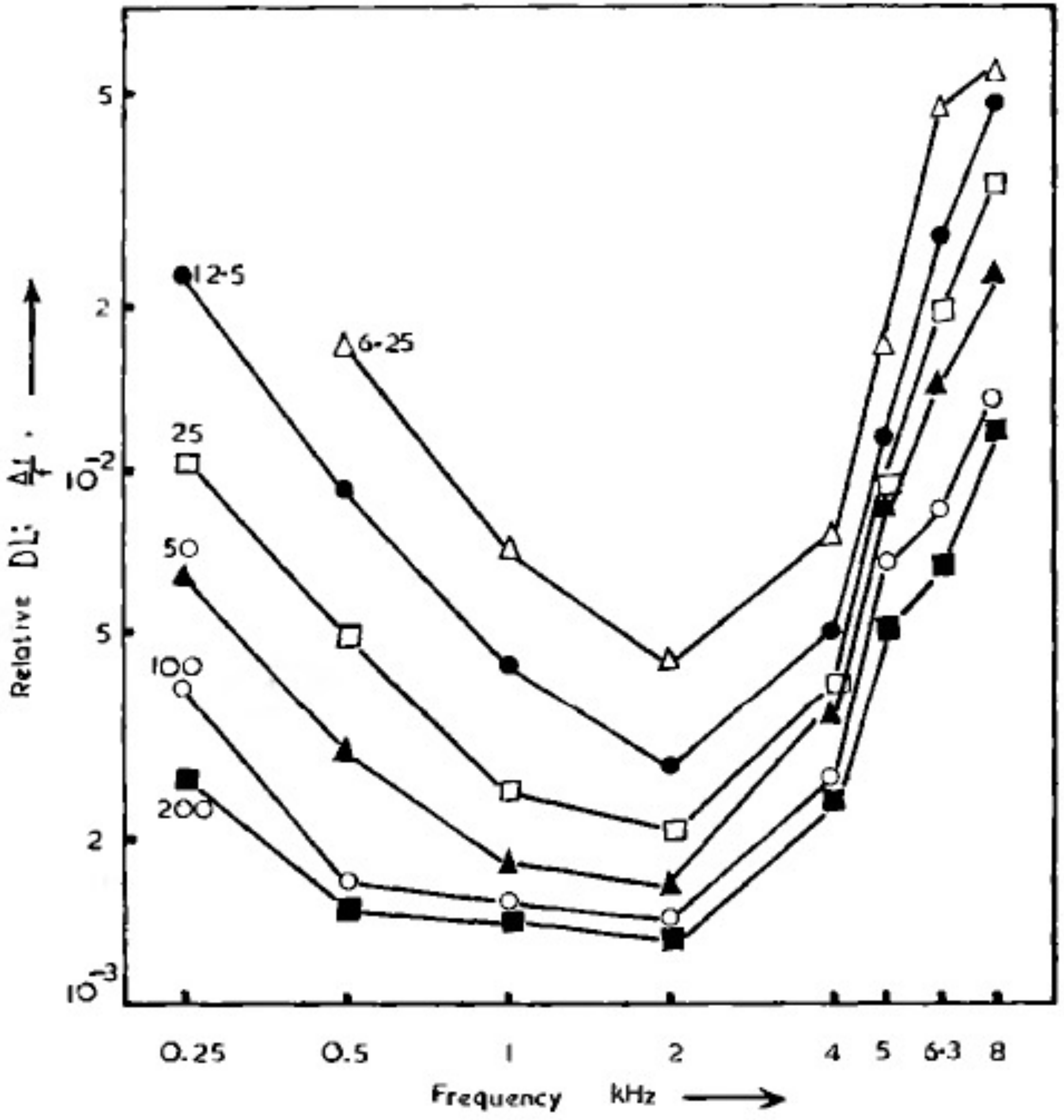


FIG. 2. Plot of the relative frequency DL ($\Delta f/f$) as a function of frequency. The parameter is duration in milliseconds. Results are for subject T.C. and were obtained using a PEST procedure.

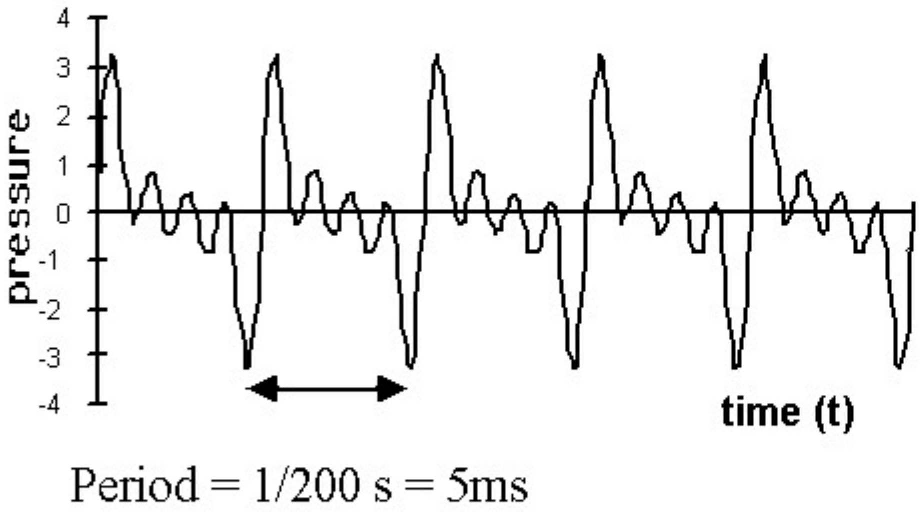
Pure tones: place vs timing

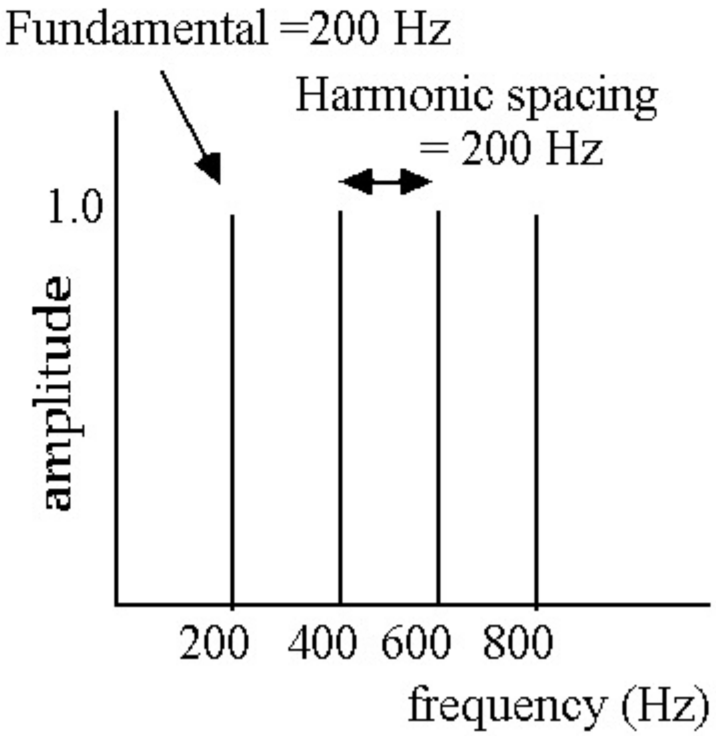
Low frequency tones	High frequency tones
Place & timing	Place only

1. Phase locking only for tones below 4 kHz
2. Frequency difference threshold increases rapidly above 4 kHz.
3. Musical pitch becomes absent above 4-5 kHz (top of piano)

2. Pitch of complex tones

Pitch of complex tones: fundamental & harmonics

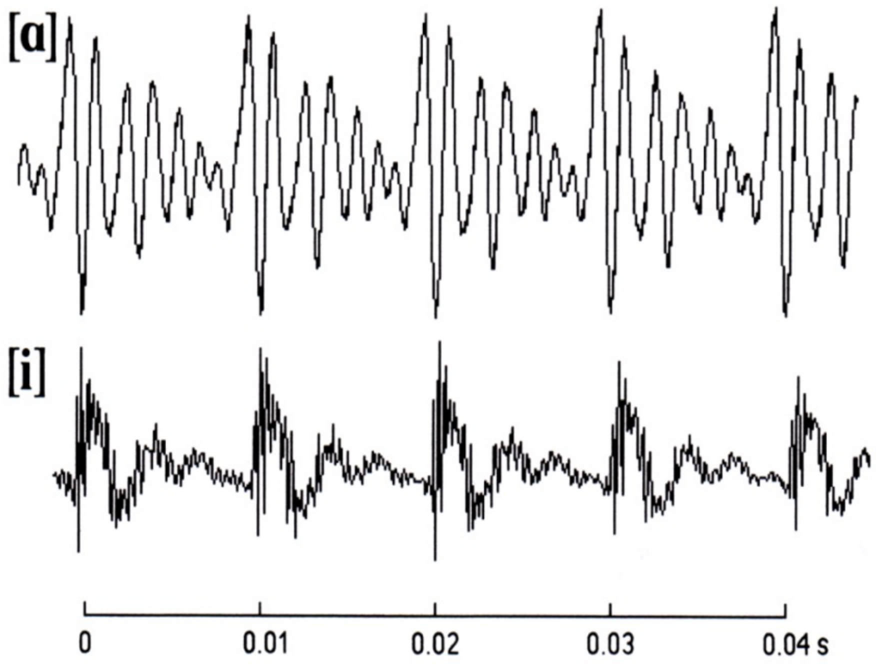




Pitch

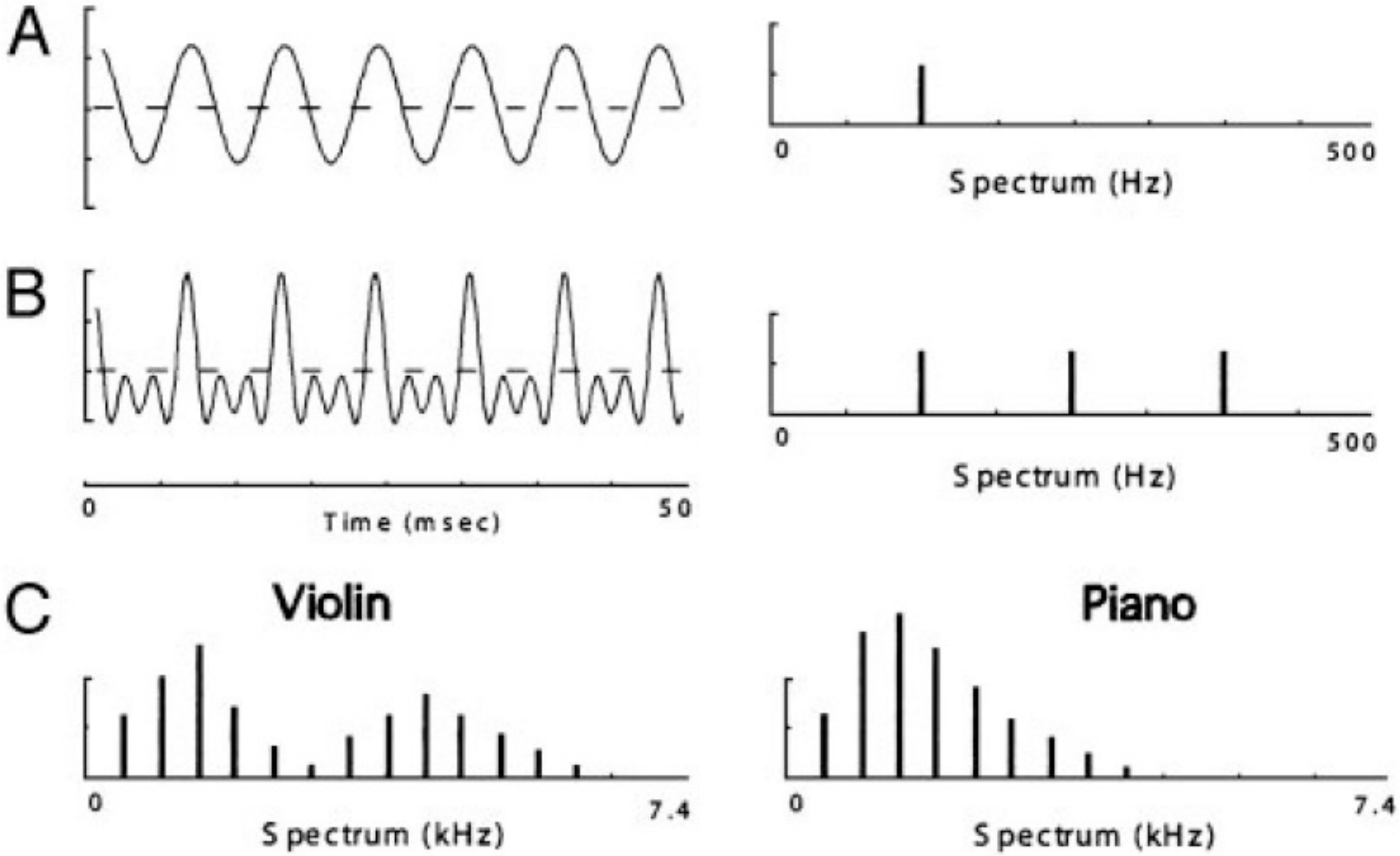
Pitch is perceptual frequency: the frequency of a sine wave with the same perceptual pitch.

- Often equal to the rate of the repetition
- Pitch is only defined for relatively fast repetitions (>20-50 Hz)
- Slower repetitions are perceived as a distinct sequence of sounds
  - in speech: sequence of phonemes
  - in music: sequence of notes
  - within these sounds, pitch can be perceived



Two vowels with the same pitch

More examples

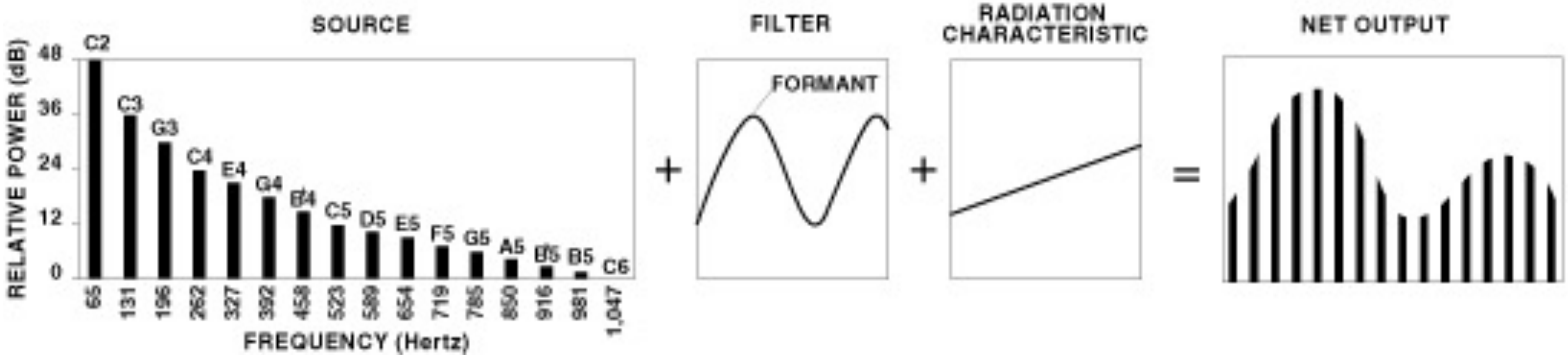


Source and Filter

Vocal cords

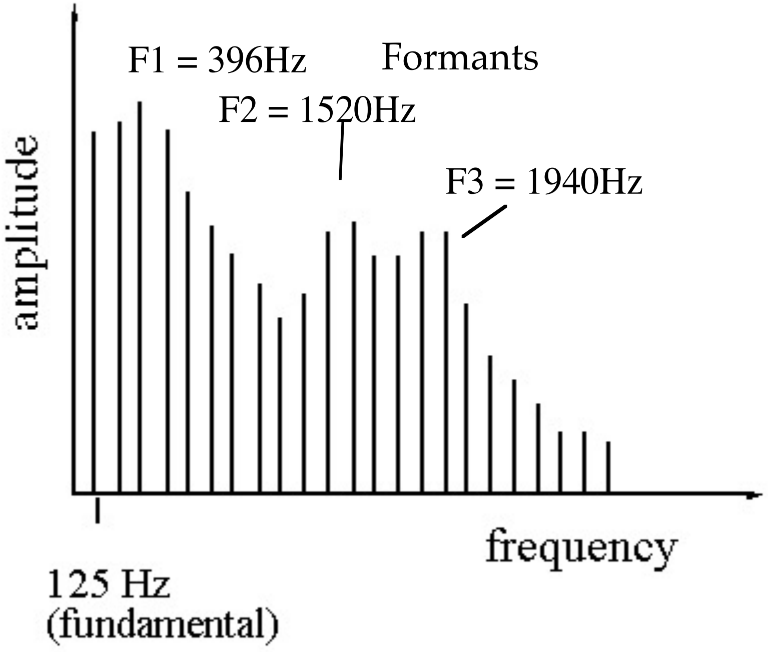
Vocal tract

Output sound

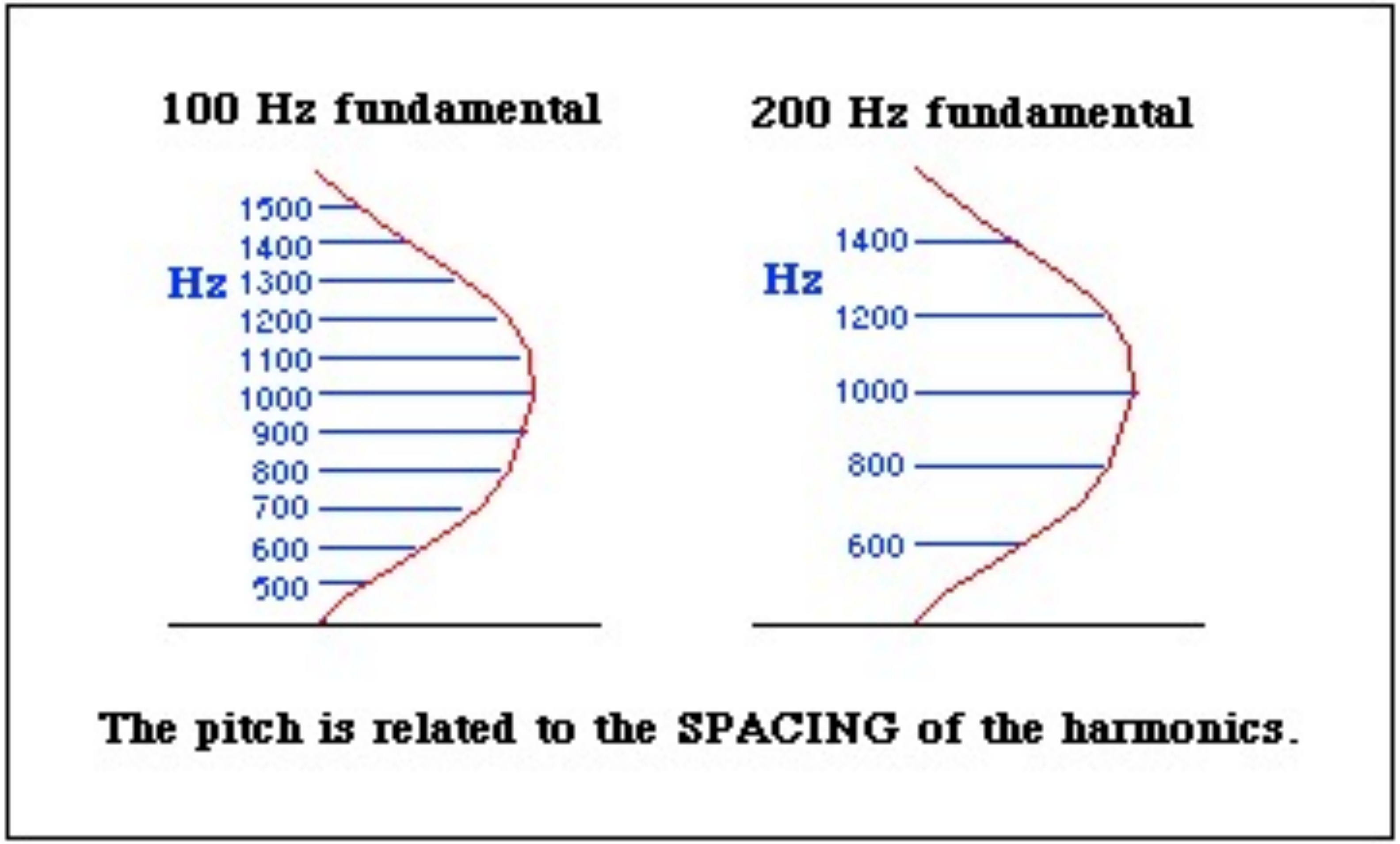


Pitch and Formants

1. Harmonics (giving pitch) produced by vocal cord vibration
2. Formant frequencies: resonances of the vocal tract

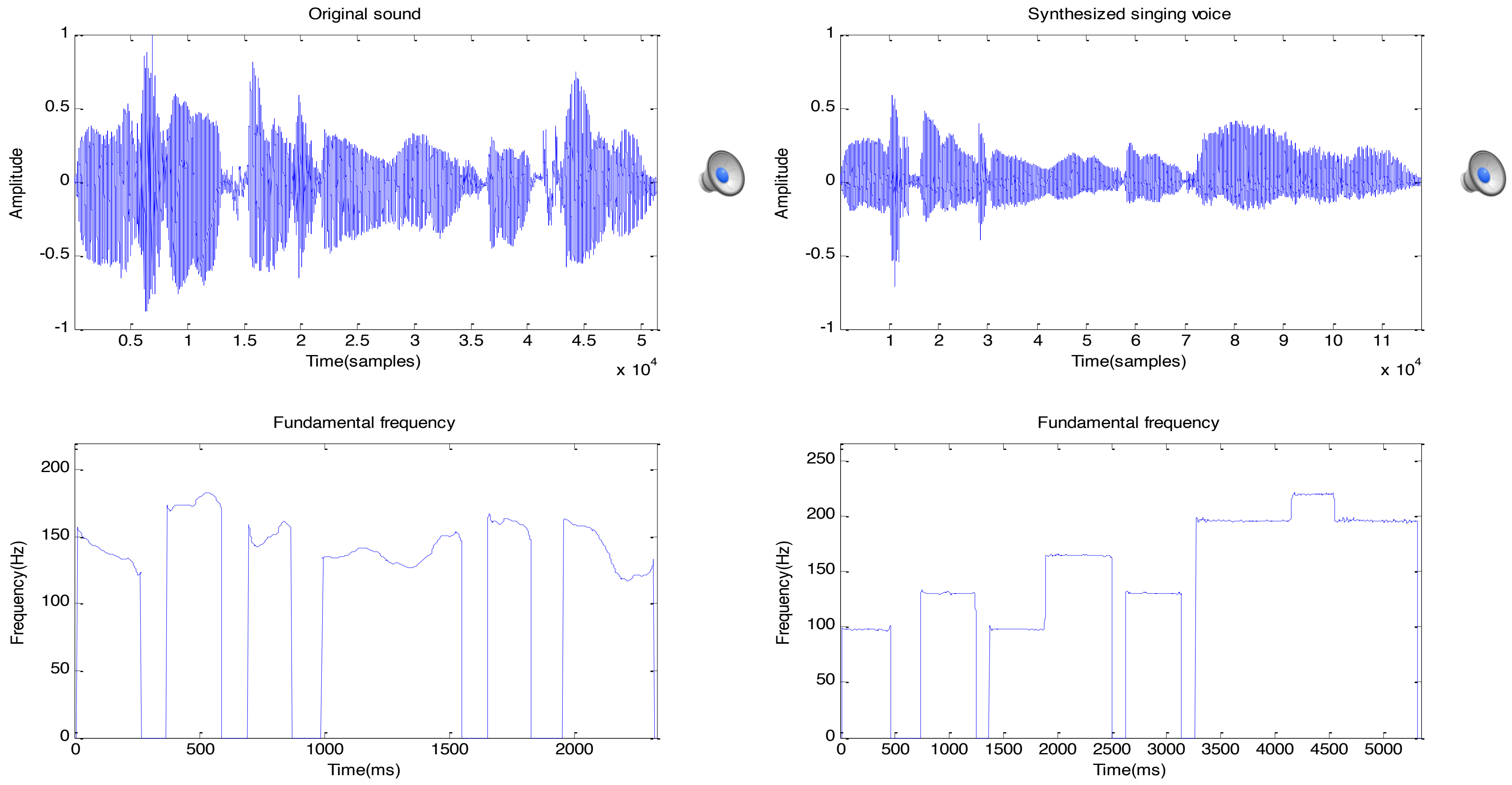


Same timbre - different pitch



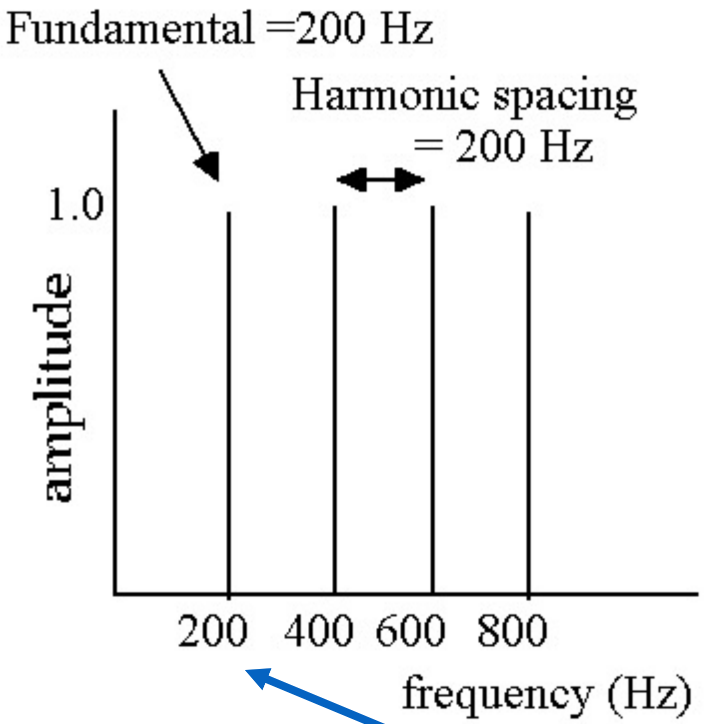
Music score

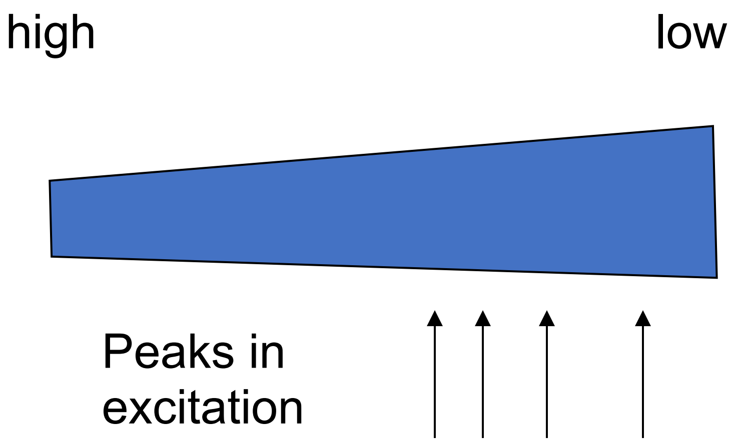

0 5. 5 1 5 3 1 5. 6 5 0
五 星 红 旗 迎 风 飘 扬


Original sound
Amplitude
Time(samples)
x 10^4
Synthesized singing voice
Amplitude
Time(samples)
x 10^4
Fundamental frequency
Frequency(Hz)
Time(ms)
Fundamental frequency
Frequency(Hz)
Time(ms)

Helmholtz’s place theory

Helmholtz suggested that the ear reintroduces energy at the fundamental by a process of distortion that produces energy **at frequencies corresponding to the difference between two components physical present.**

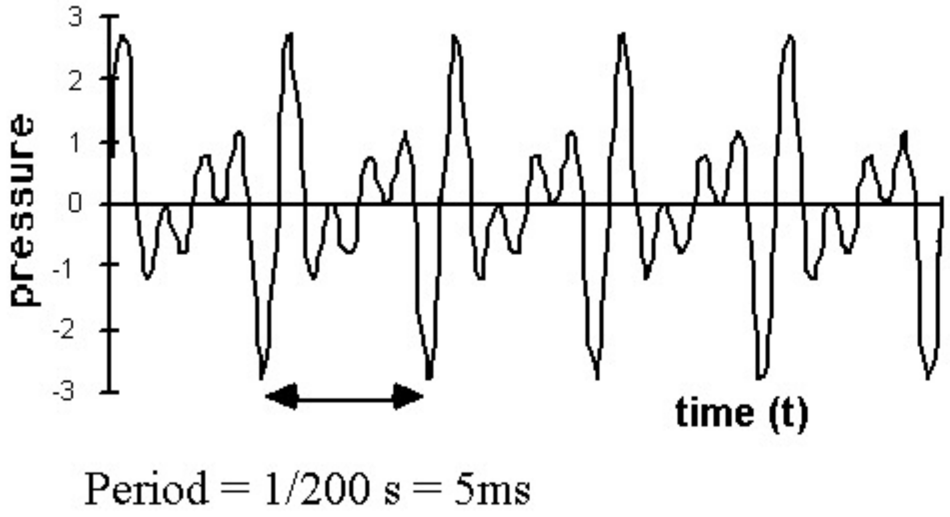


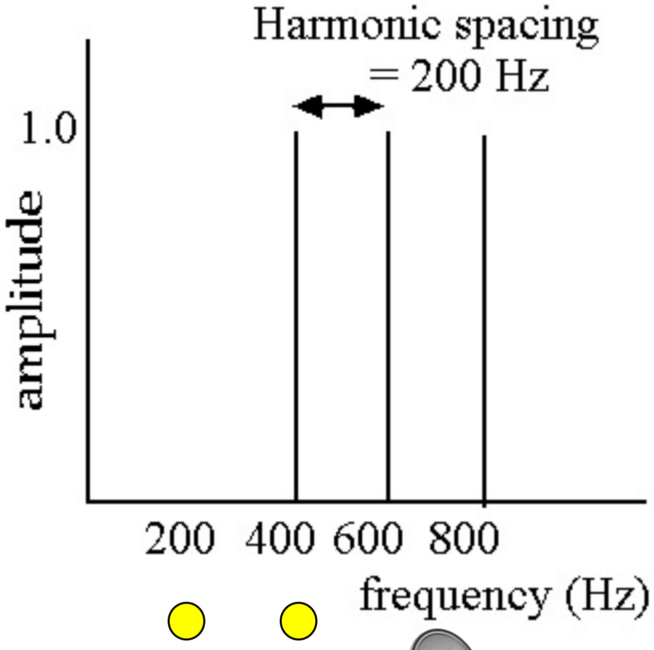


Pitch = fundamental frequency coded by place of excitation

Arguments against Helmholtz

1. Fundamental not necessary for pitch -- Missing fundamental

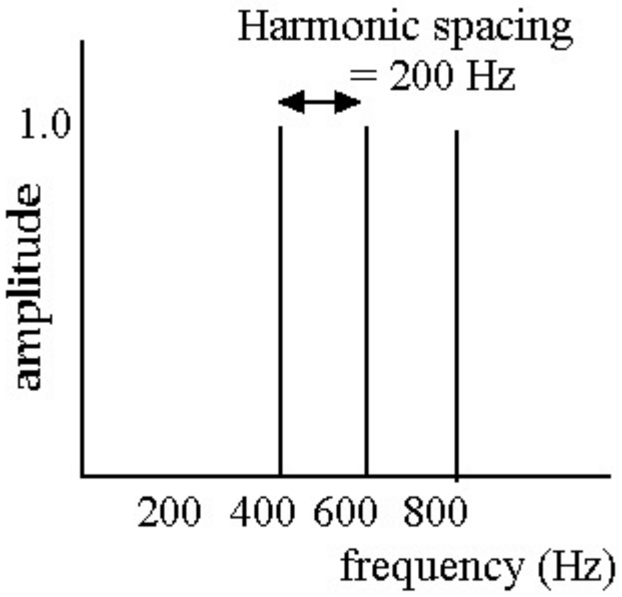




No fundamental but you still hear the pitch at 200 Hz

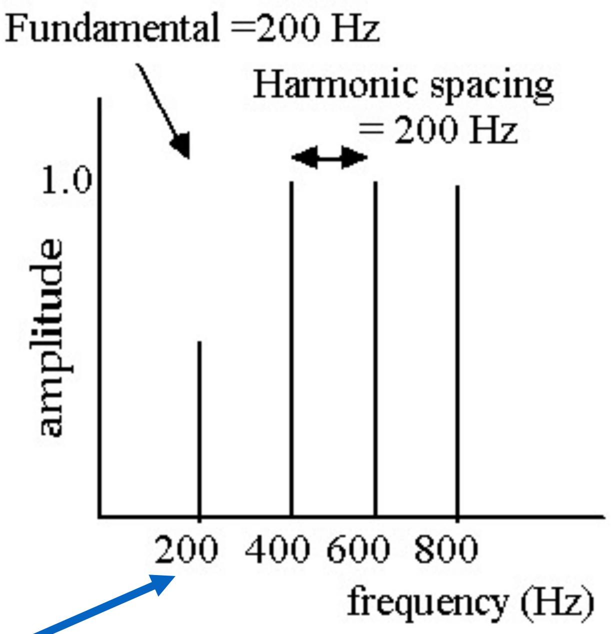
Distortion: Helmholtz fights back

Sound stimulus



Middle-ear distortion

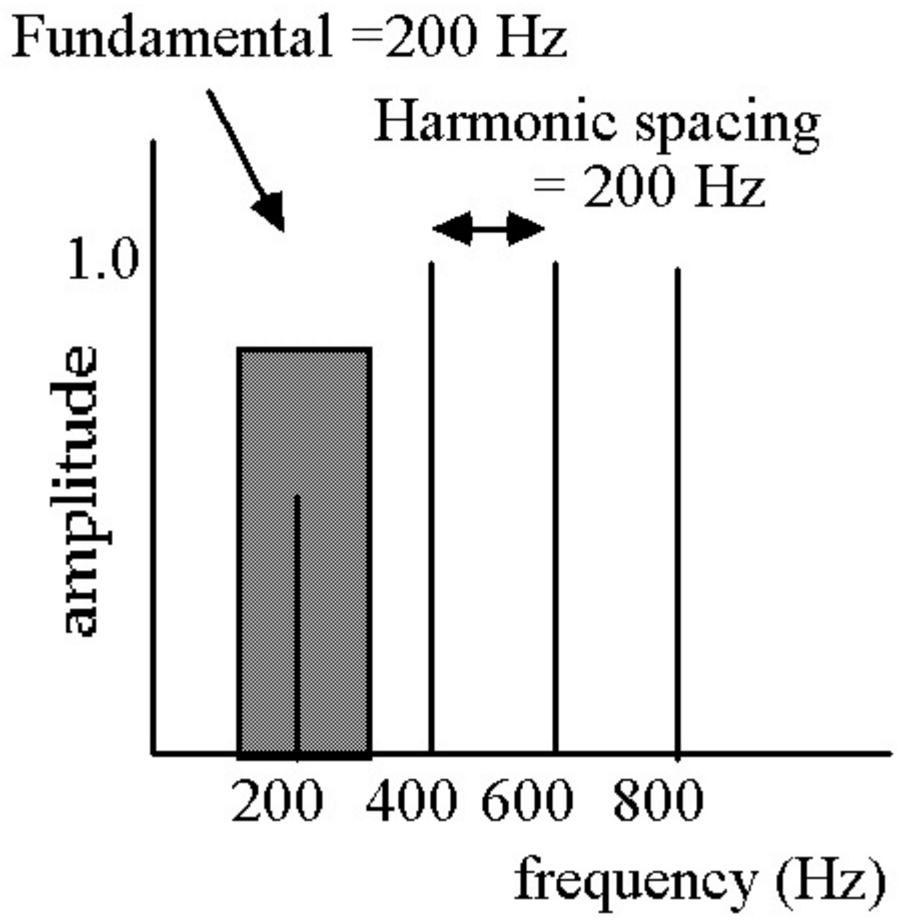
Sound going into cochlea



Produces f2 - f1

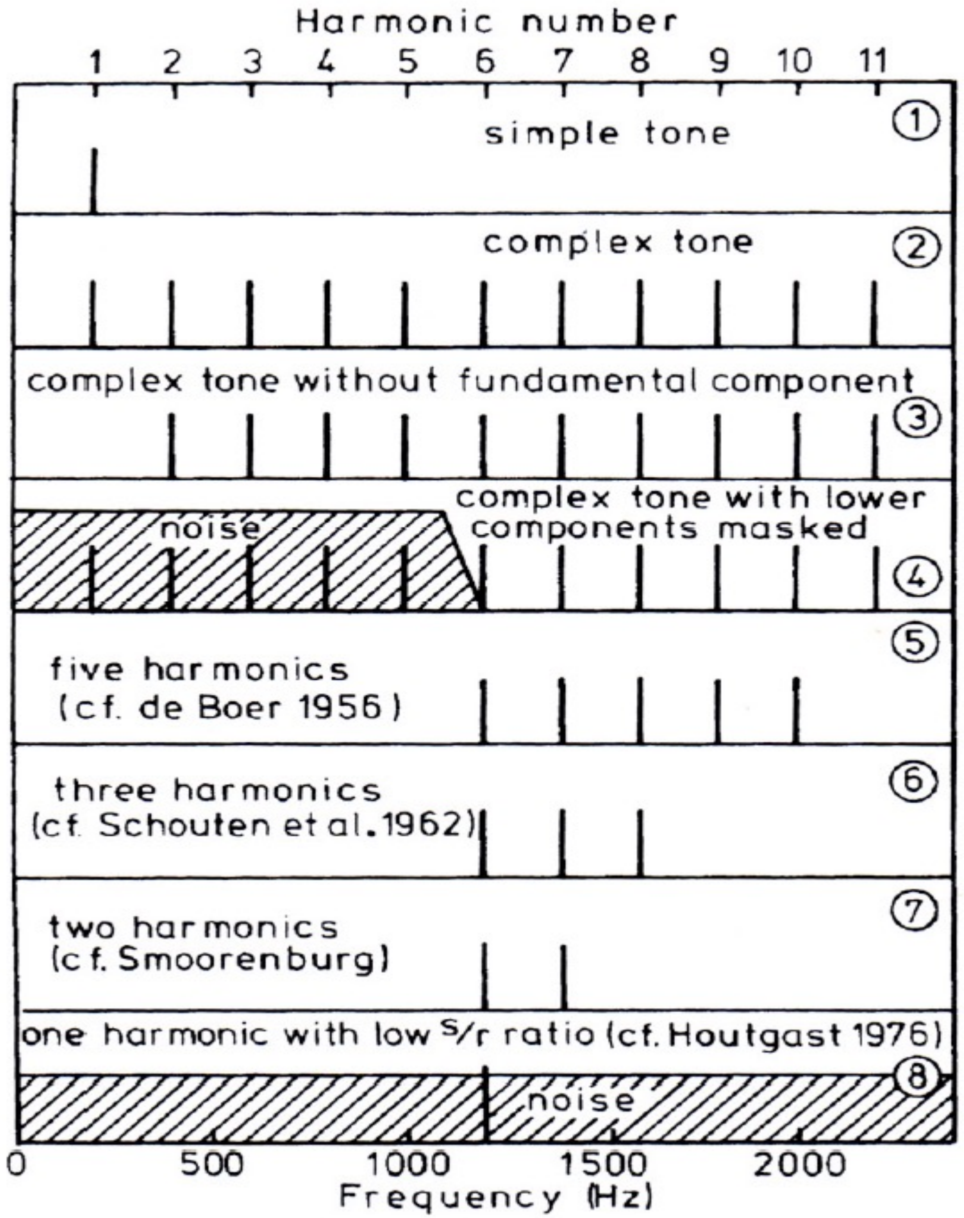
600 - 400

Against Helmholtz: Masking the fundamental



Masked complex still has a pitch of 200 Hz

Eight signals with the same low pitch



Phenomenon of Missing Pitch

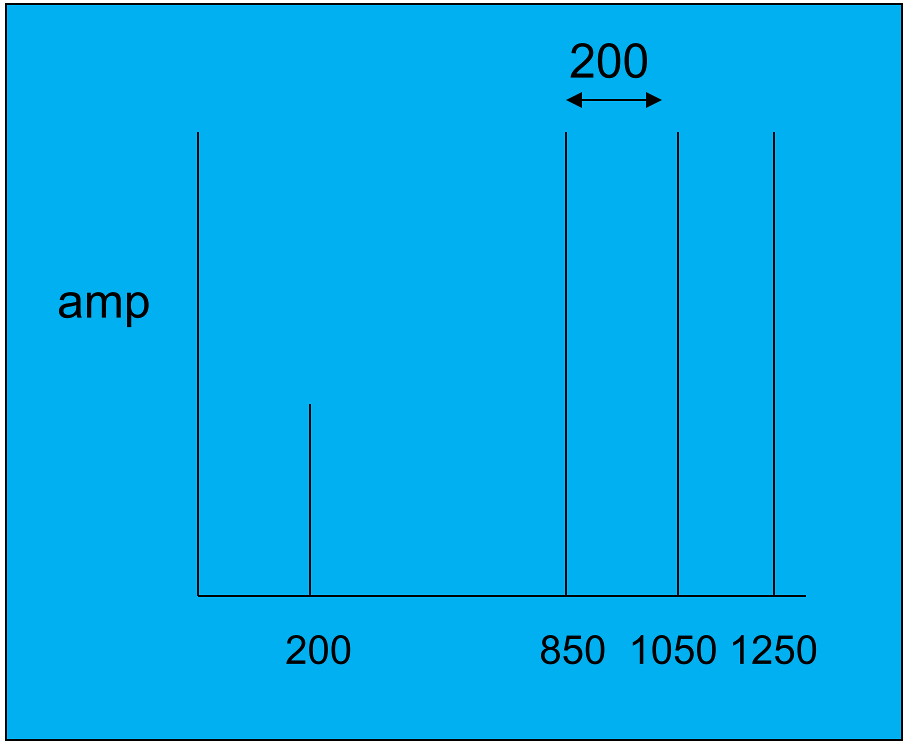
- clicks 200 times per second
  - has low pitch very close to 200 Hz pure tone
  - contains harmonics with frequencies of 200, 400, 600, 800, etc Hz
  - can filter to remove 200 Hz, pitch unchanged, timbre slightly different
  - can even eliminate all but 1800, 2000, 2200, low pitch remains, timbre very different
  - low pitch is present even when masked by low freq noise
- low pitch is called “residue”
  - also called “periodicity pitch”, “virtual pitch”, and “low pitch”
  - perception of residue pitches are what we normally hear when we listen to complex tones
- **Theories of residue pitch**
  - pattern recognition models
  - time interval models

Against Helmholtz: Enharmonic sounds

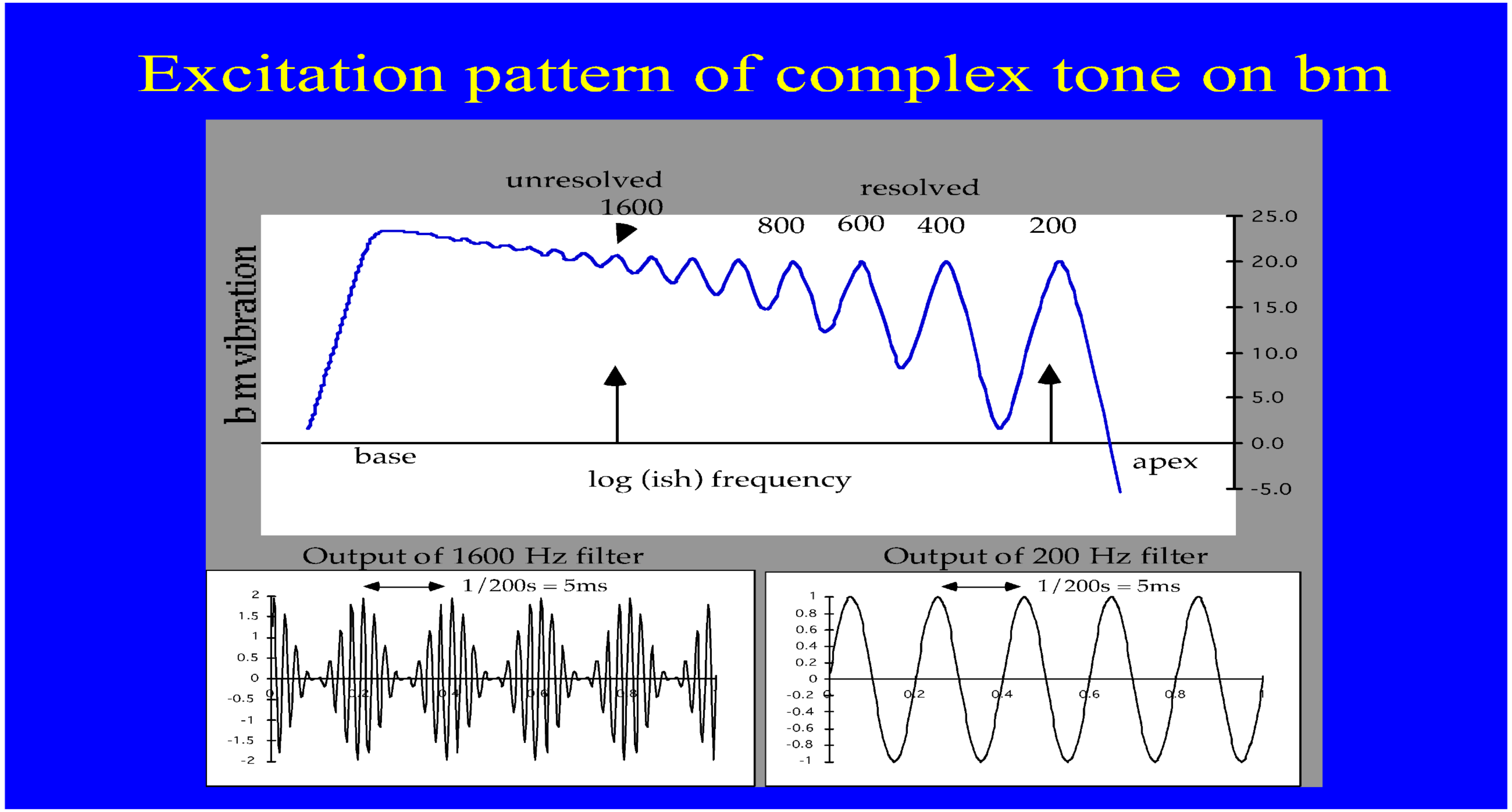
Middle-ear distortion gives difference tone (1050 - 850 = 200)

BUT

Pitch heard is actually about 210



Schouten’s theory

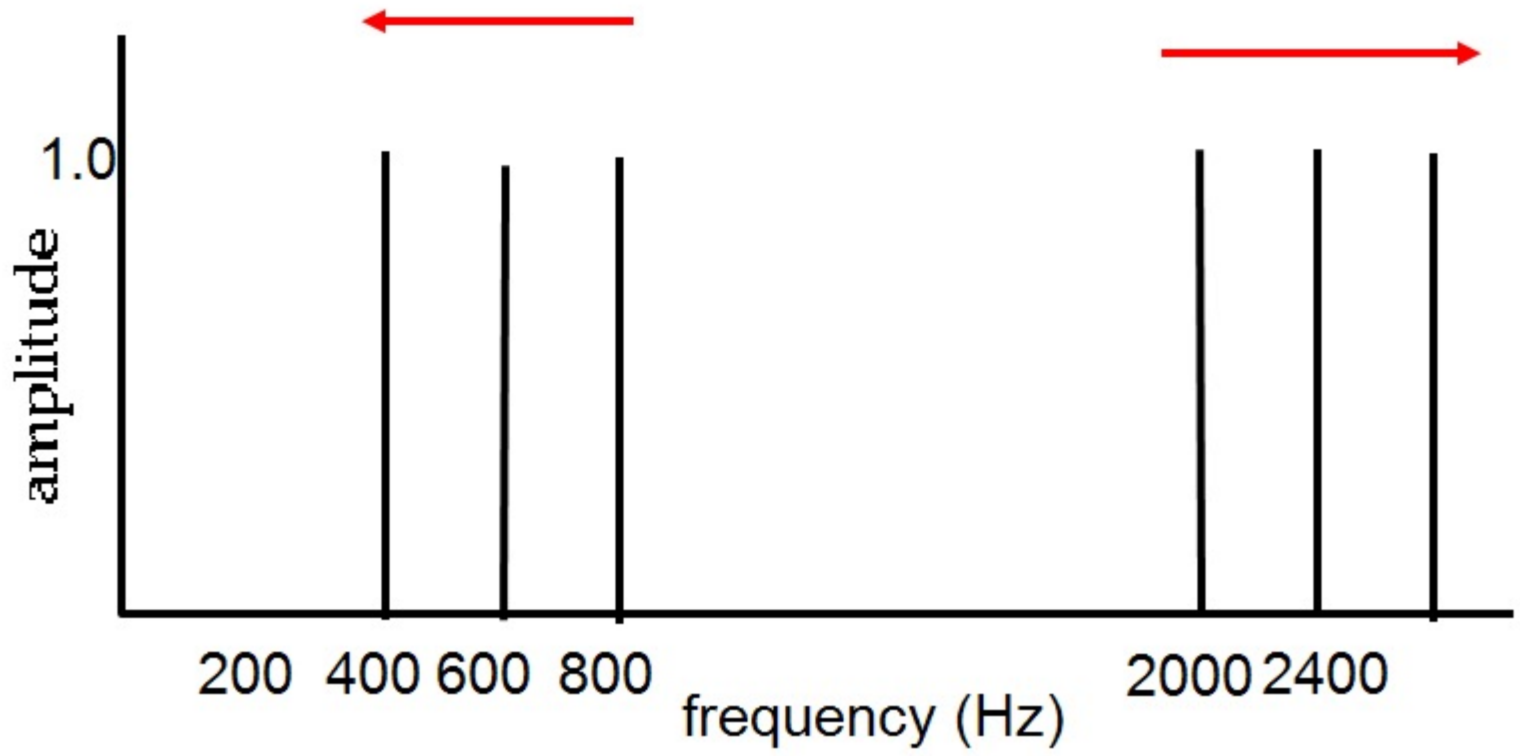


• Notice that the excitation patterns of the higher numbered harmonics are closer together than those of the low-numbered harmonics.

• **Schouten proposed that the brain times the intervals between beats of the unresolved harmonics of a complex sound, in order to find the pitch**.

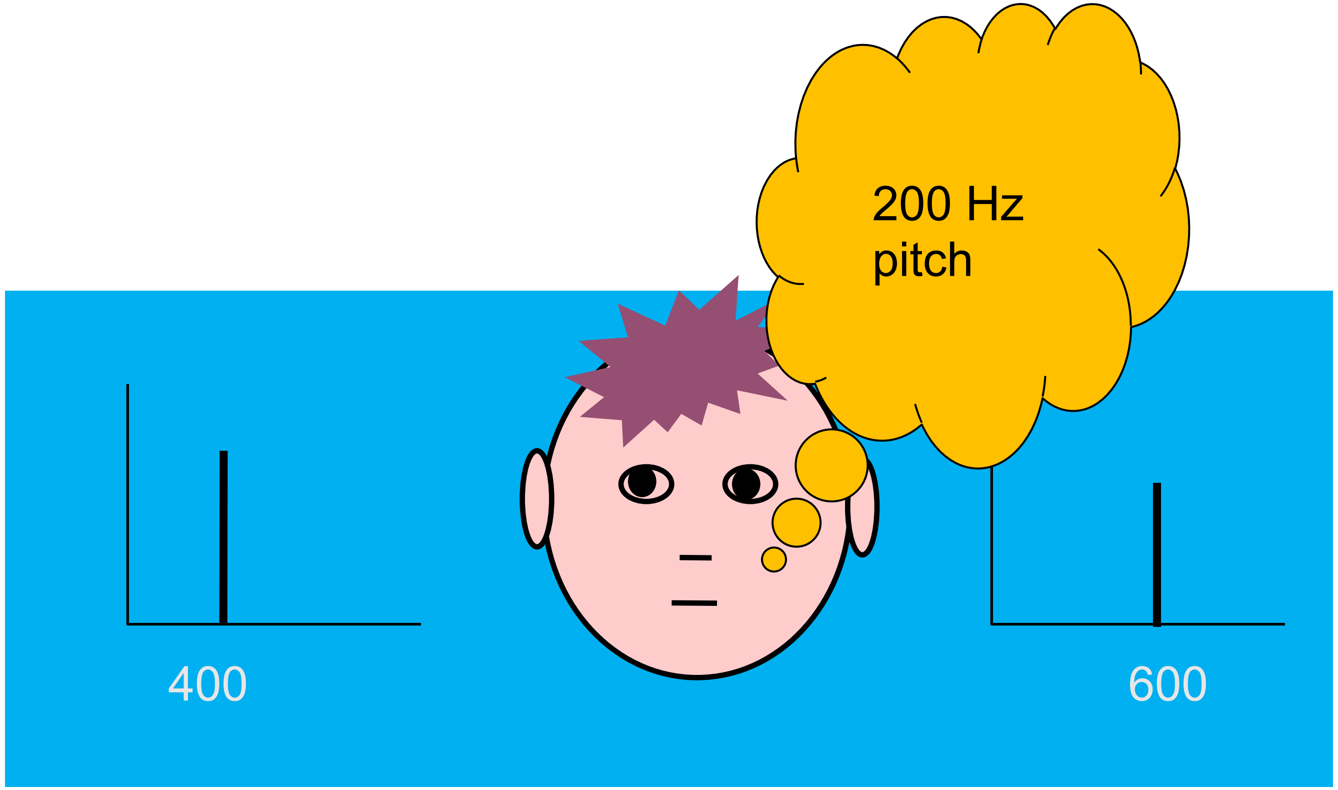
Problems with Schouten's theory (1)

- <u>Resolved</u> harmonics **dominant** in pitch perception - not unresolved
  - Musical pitch is weak for complex sounds consisting only of unresolved harmonics
  - Pitch difference harder to hear for unresolved than resolved complexes



Against Schouten (2): Dichotic harmonics

- Pitch of complex tone still heard with one harmonic to each ear
(Houtsma & Goldstein, 1972, JASA)



No chance of distortion tones or physical beats

Goldstein's theory

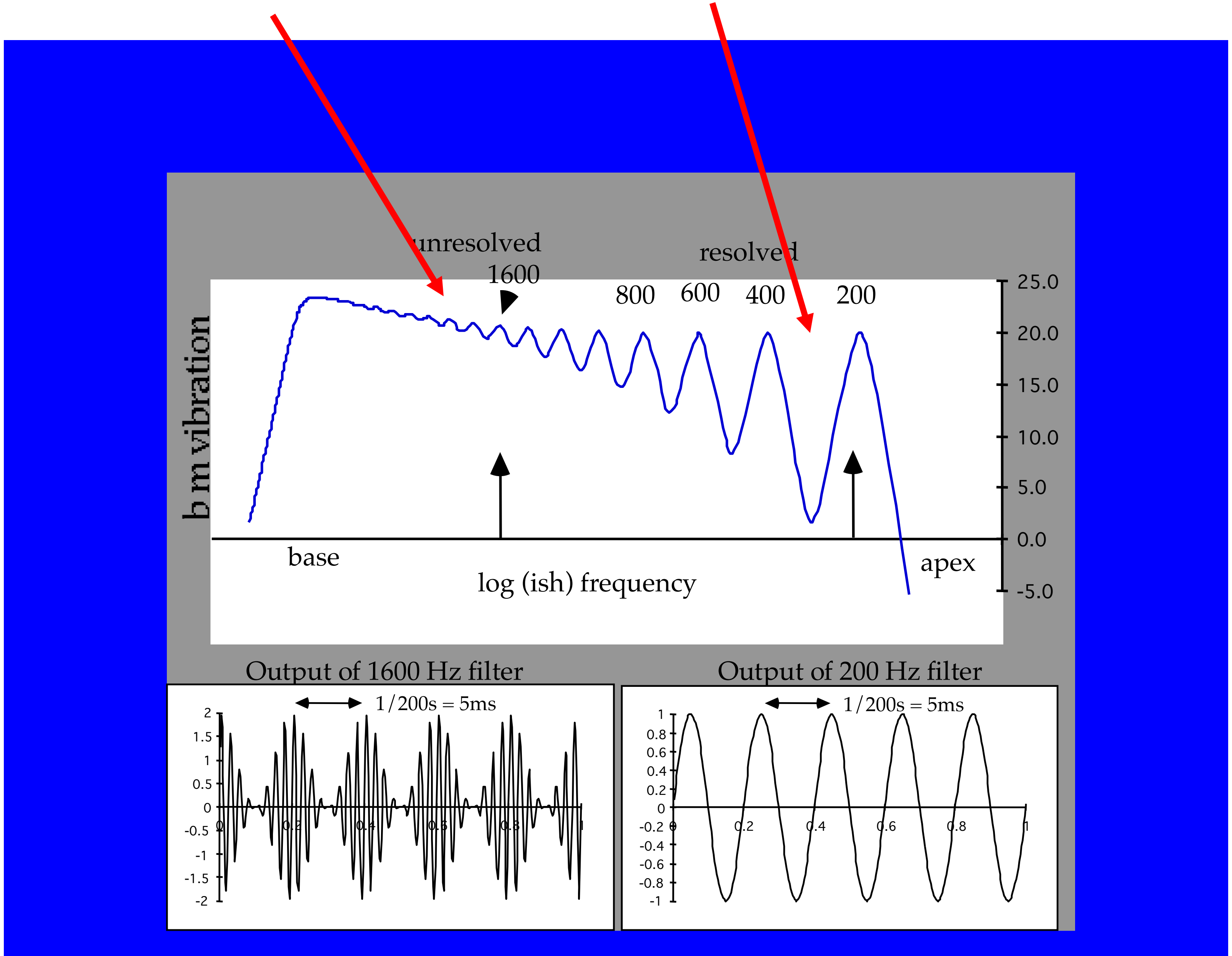
- Pitch based on resolved harmonics
- Brain estimates frequencies of resolved harmonics (eg 402 597 806)
  - could be by a place mechanism, but more likely through phase-locked timing information near appropriate place.
- Then finds the ***best-fitting consecutive harmonic series*** to those numbers (eg 401 602 804) -> pitch of 200.5

Two pitch mechanisms ?

- Goldstein has difficulty with the fact that unresolved harmonics have a pitch <u>at all</u>.
- So: Goldstein's mechanism could be good as the <u>main</u> pitch mechanism…

  …with Schouten's being a <u>separate</u> (weaker) mechanism for unresolved harmonics

Schouten's + Goldstein's theories

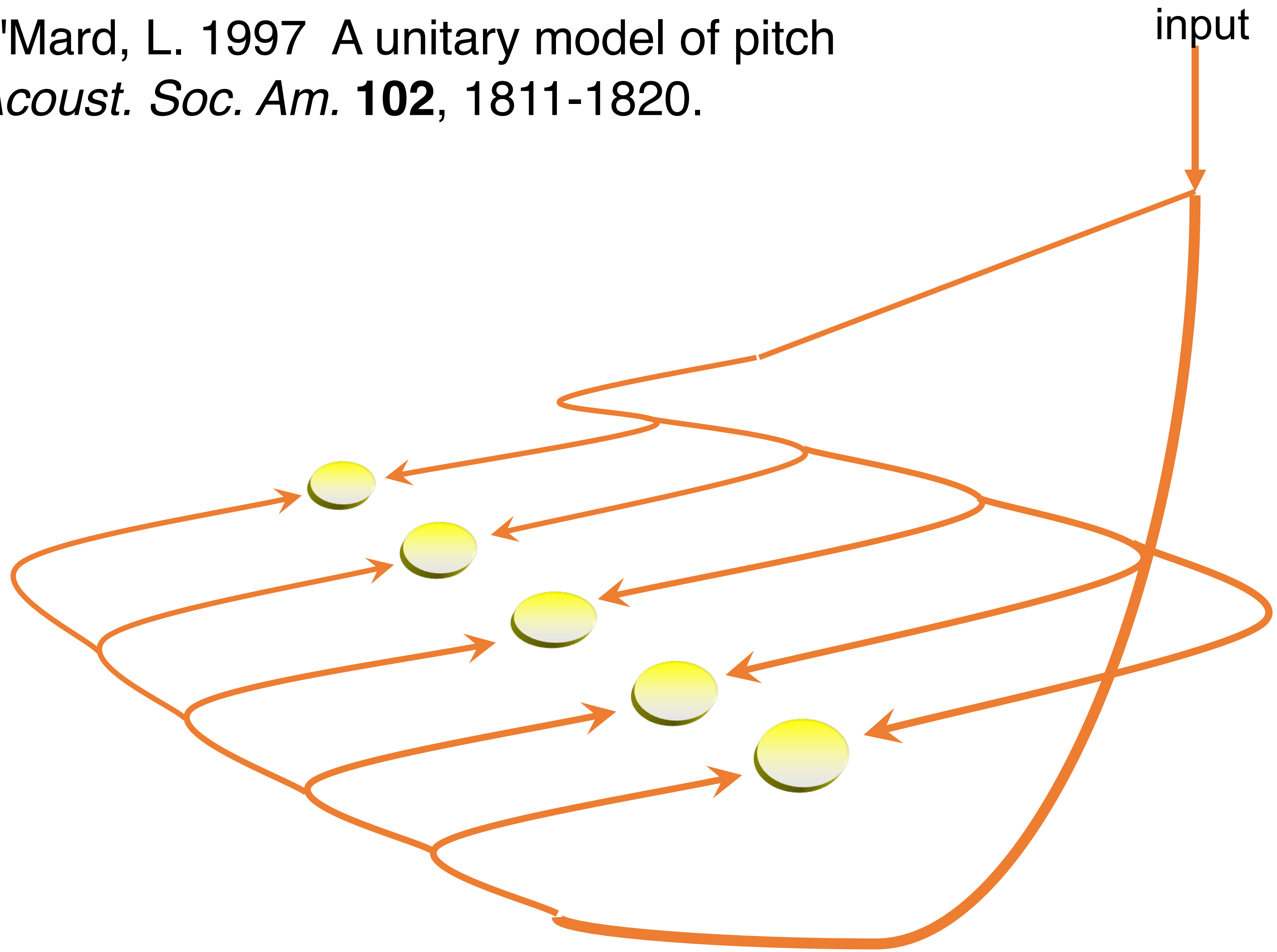


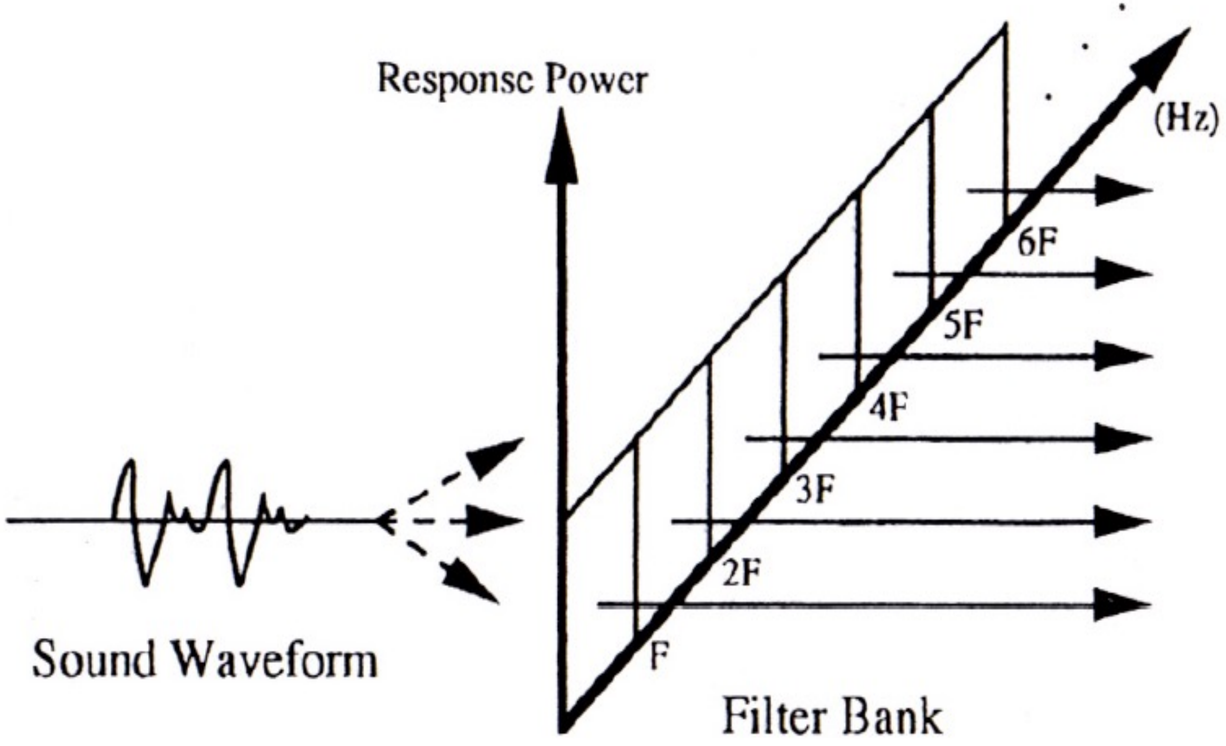
3. Computational models of Pitch Perception

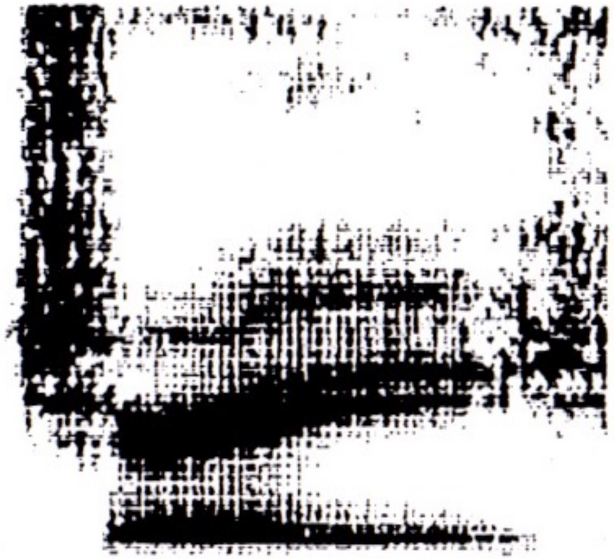
Autocorrelator

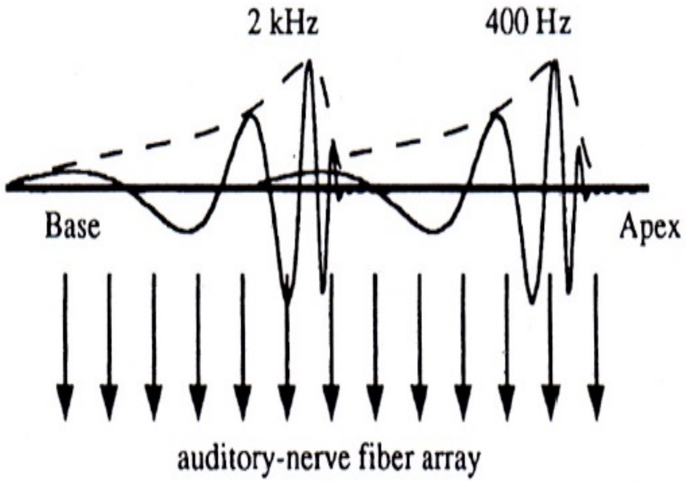
JCR Licklider

Meddis, R. & O'Mard, L. 1997 A unitary model of pitch perception. *J. Acoust. Soc. Am.* **102**, 1811-1820.

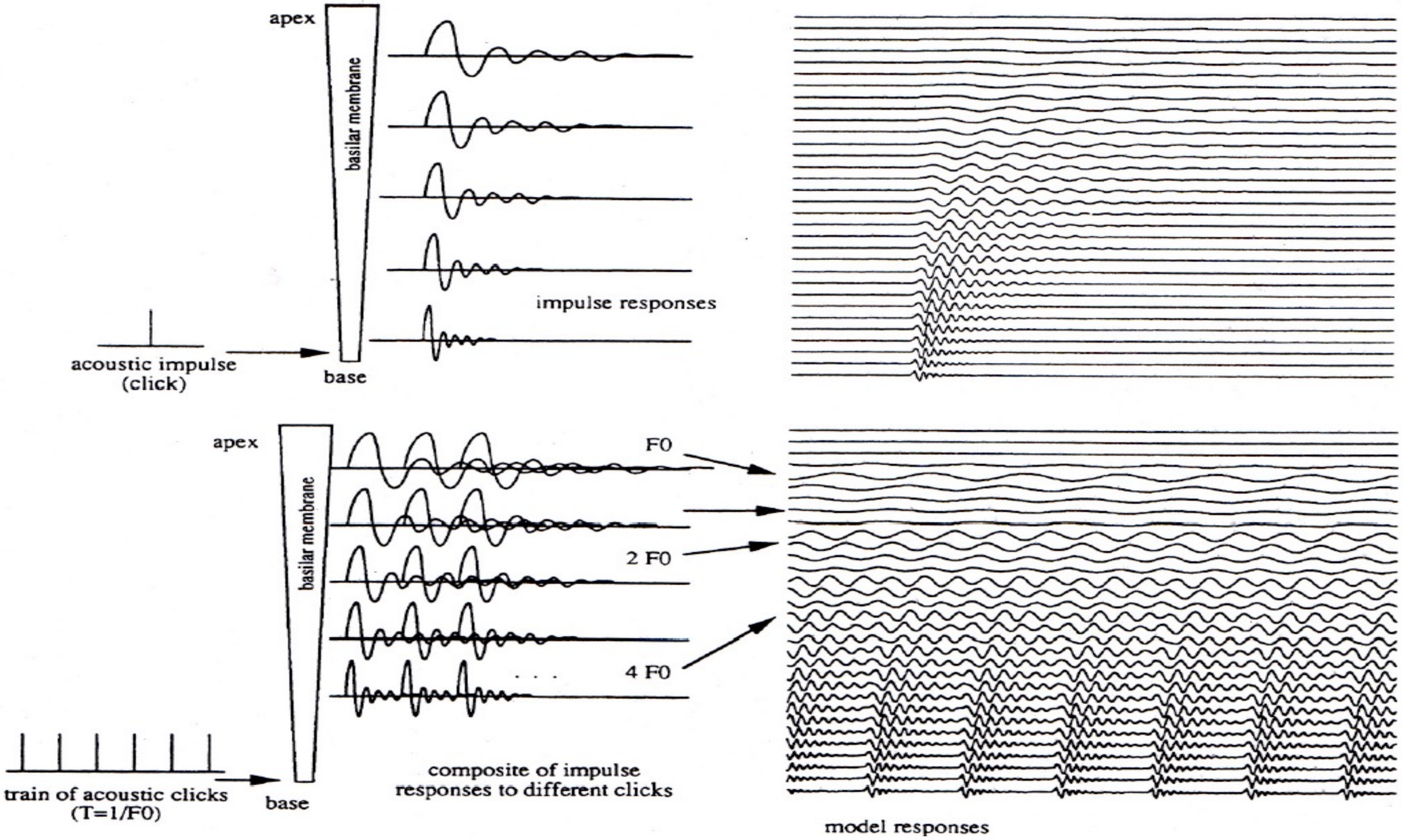




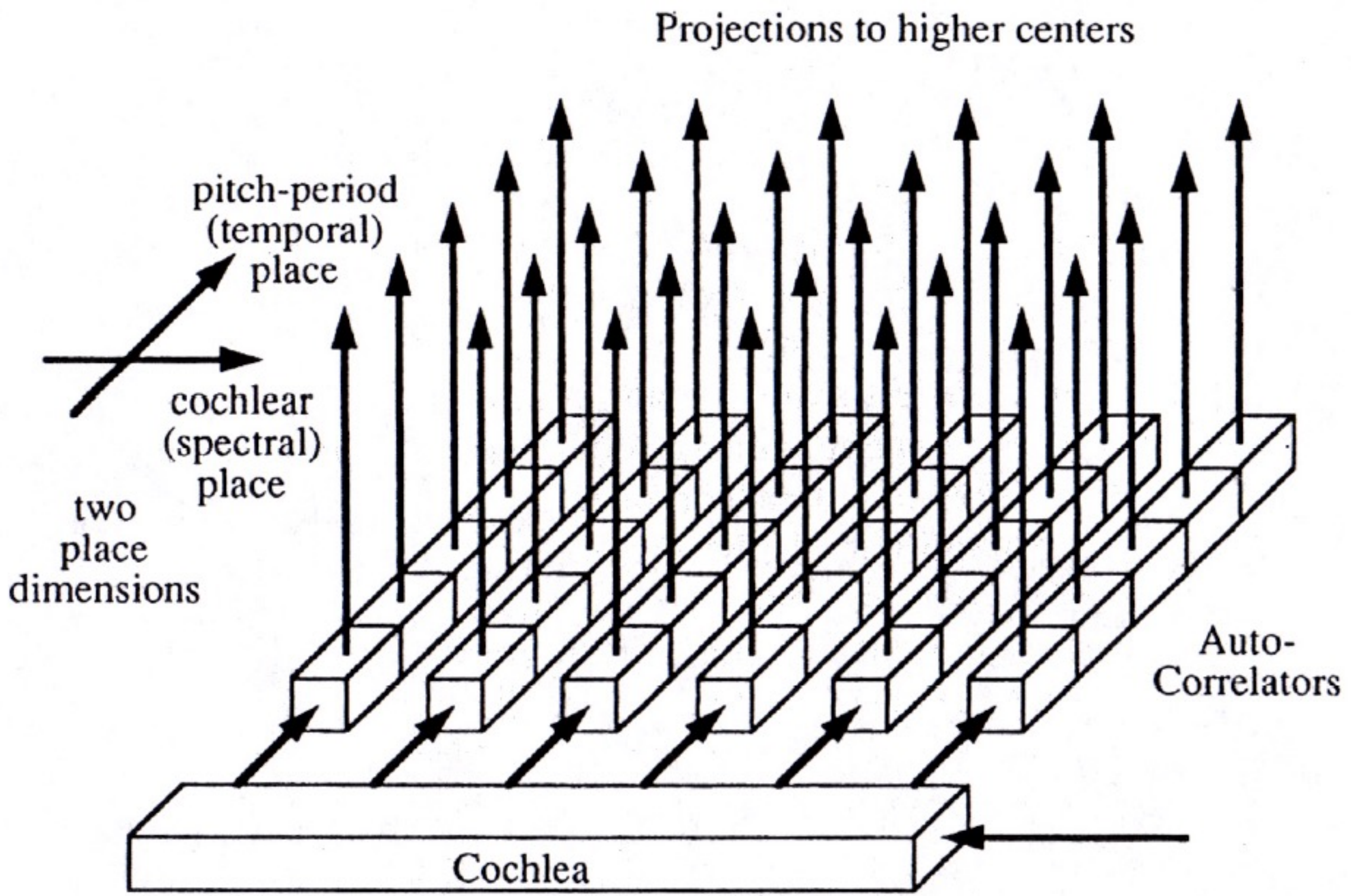




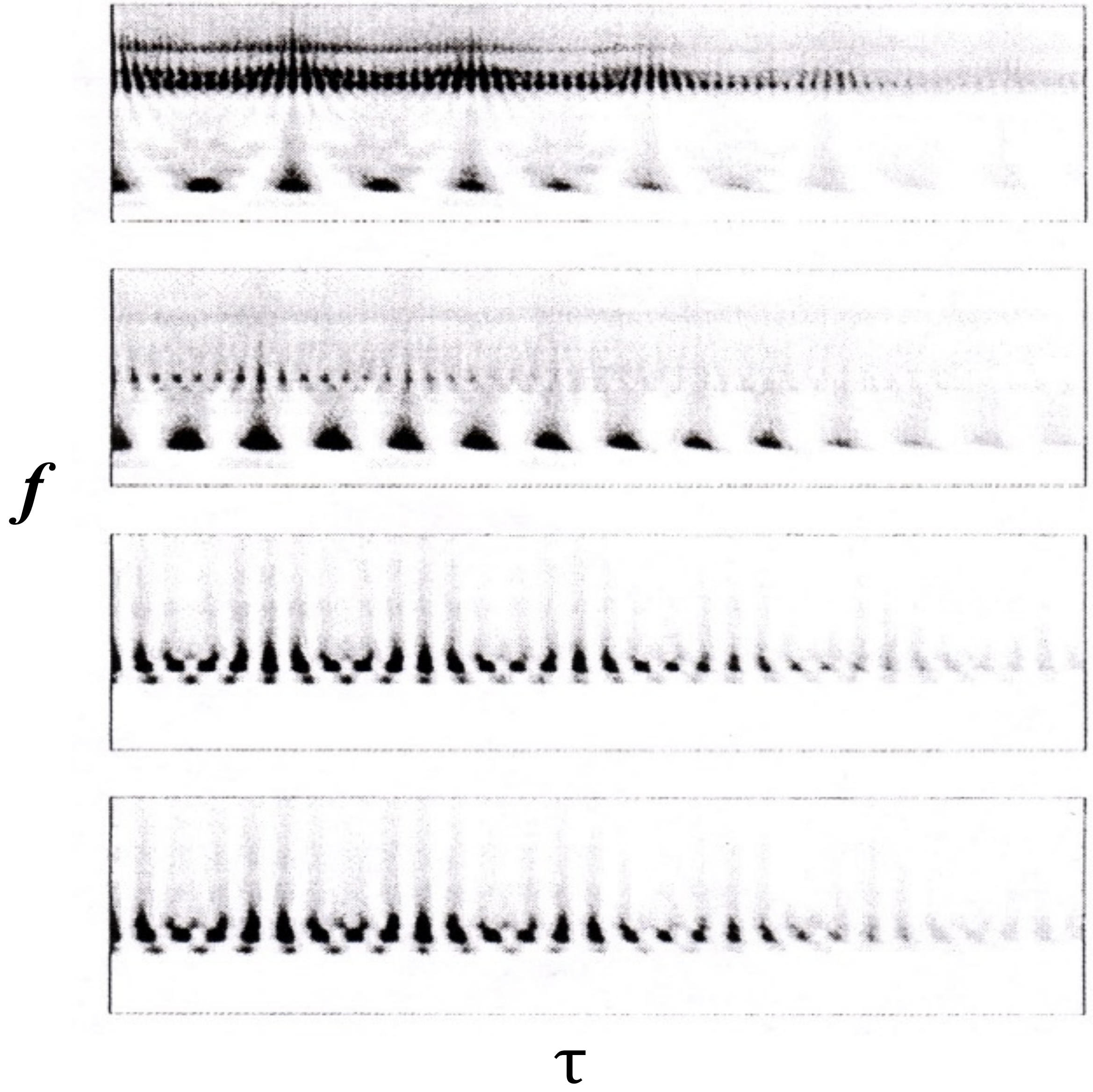
- Implementation details
  - cochlear filters aren't uniformly spaced, they're logarithmic for f > 500 Hz
  - cochlear filters aren't of equal bandwidth
  - spectrograms assume fixed window size
- Spectrogram is not a literal representation
  - what is role played by shapes and phases of cochlear filters?
  - the fine temporal structure of their outputs?


apex
basilar membrane
impulse responses
acoustic impulse
(click)
base
apex
basilar membrane
F0
2 F0
4 F0
. . .
train of acoustic clicks
(T=1/F0)
base
composite of impulse
responses to different clicks
model responses

Calculating periodicity: the correlogram

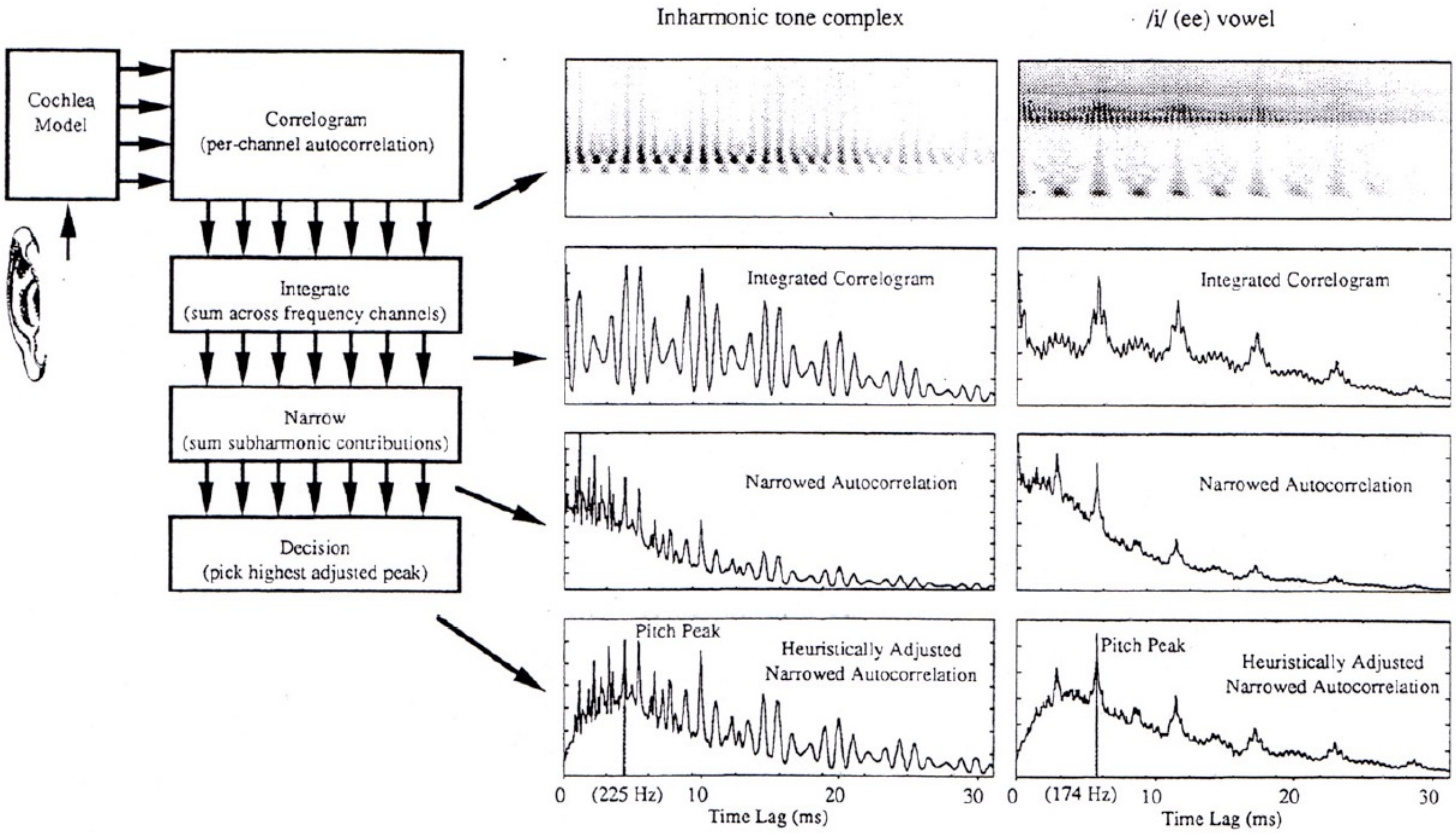


Correlograms



- /i/ vowel [ee]
- /u/ vowel [oo]
- three tone harmonic complex
- three tone inharmonic complex

Slaney-Lyon pitch model

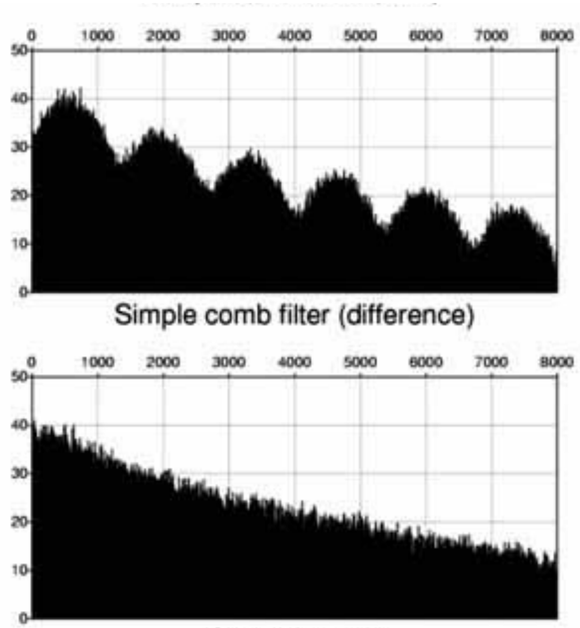


Some other sounds that give pitch

- **SAM Noise**: envelope timing - no spectral
  - Sinusoidally amplitude modulated noise

100 200 300 400 500

- Binaural interactions
- Rippled noise: envelope timing - spectral
  - $\sin(\omega t) + \sin(\omega(t-T)) = 2\sin(\omega t - \omega T/2)\cos(\omega T/2)$
    Comb-filter -> sinusoidal spectrum
    (high pass to remove resolved spectral structure)
  - Huygens @ the steps from a fountain



Original sound

Huygen’s repetition pitch

Christian Huygens in 1693 noted that the noise produced by a fountain at the chateau of Chantilly de la Cour was reflected by a stone staircase in such a way that it produced a musical tone. He correctly deduced that this was due to the successively longer time intervals taken for the reflections from each step to reach the listener's ear.

Problem we haven’t addressed

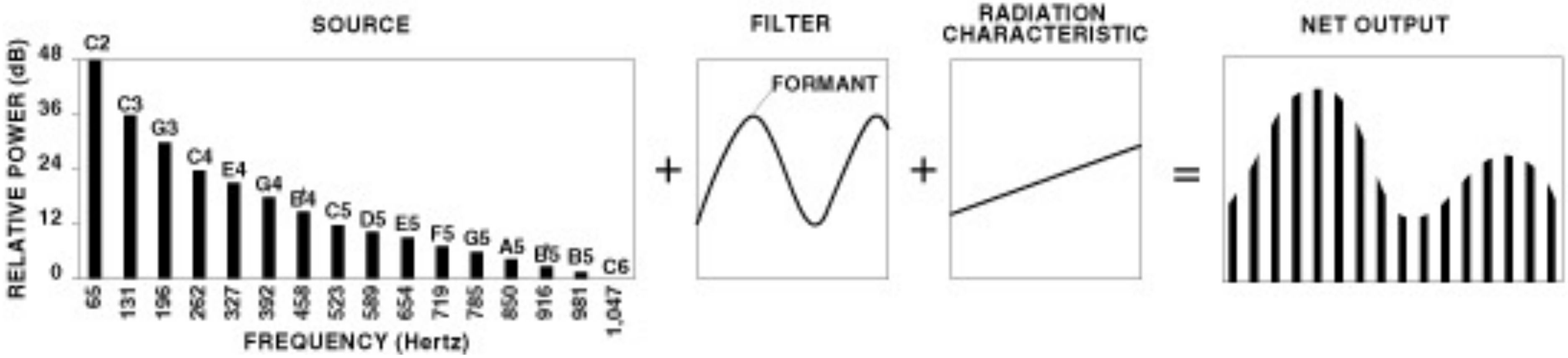
- What happens when we have two simultaneous pitches - as with two voices or two instruments - or just two notes on a piano?
- How do you know which harmonic is from which pitch?

Auditory scene analysis!

4. Timbre

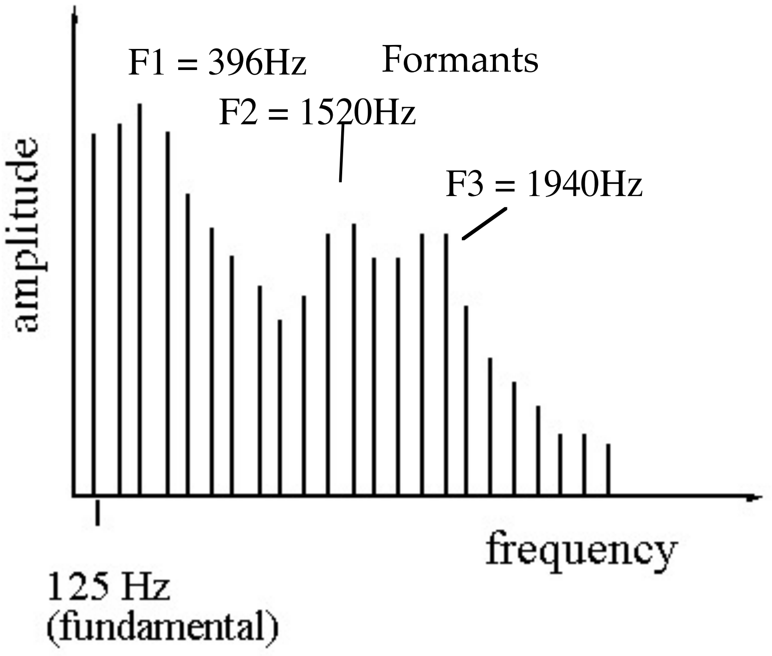
Source and Filter

Vocal cords | Vocal tract | Output sound



Pitch and Formants

1. Harmonics (giving pitch) produced by vocal cord vibration
2. Formant frequencies: resonances of the vocal tract
3. Formant frequencies change as you change the shape of your vocal tract



Vowel production

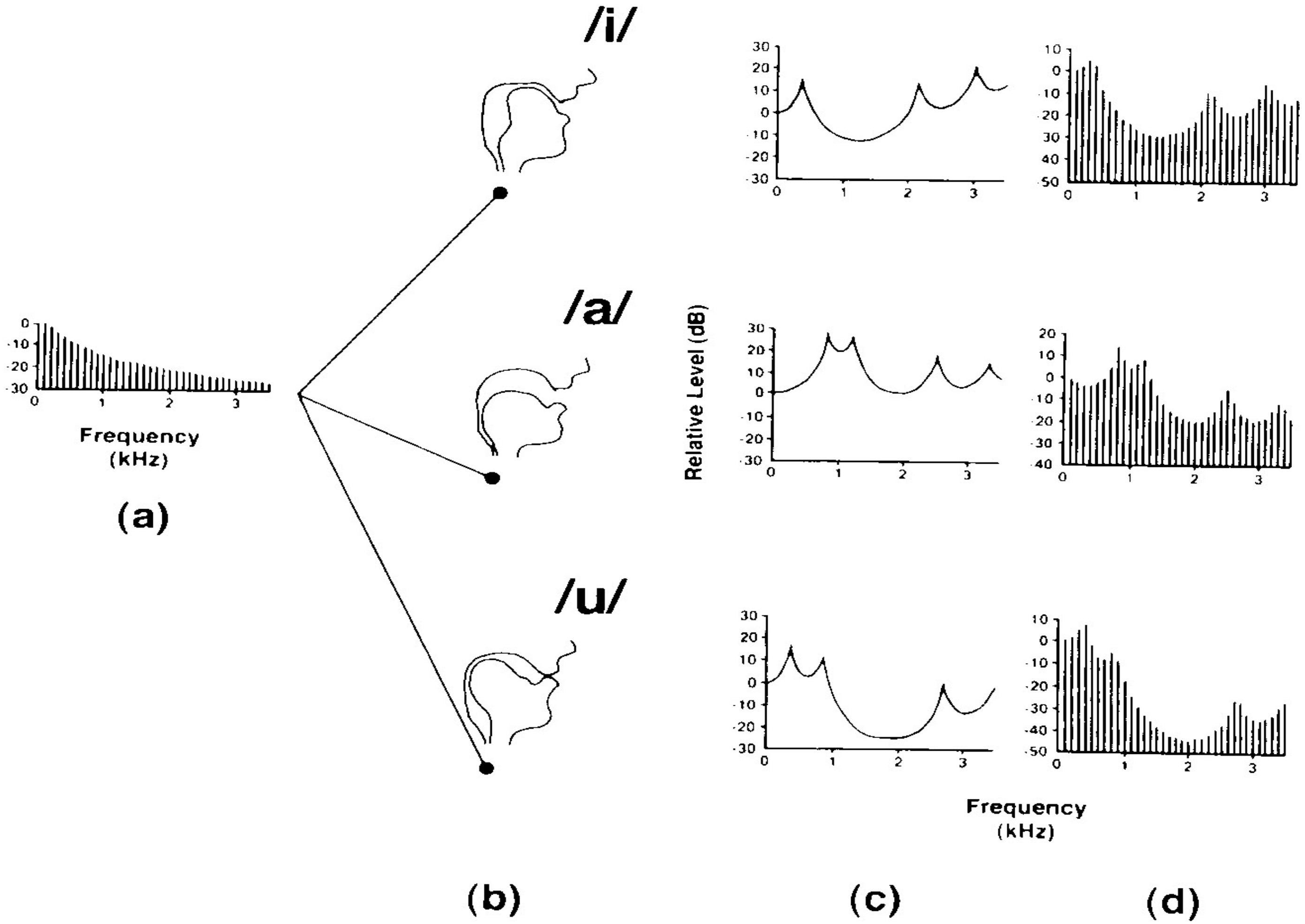
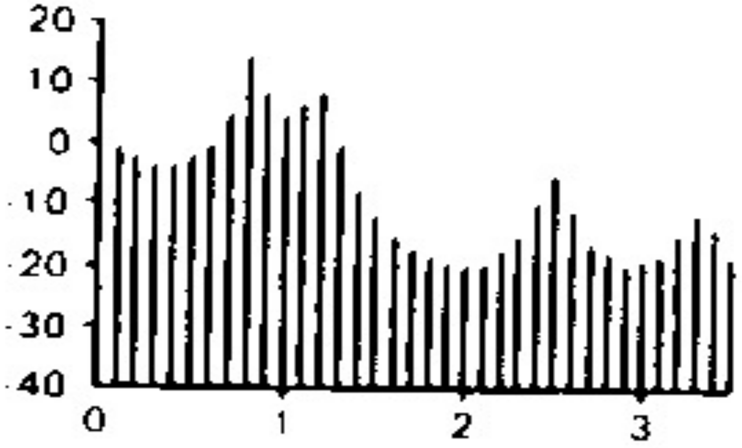


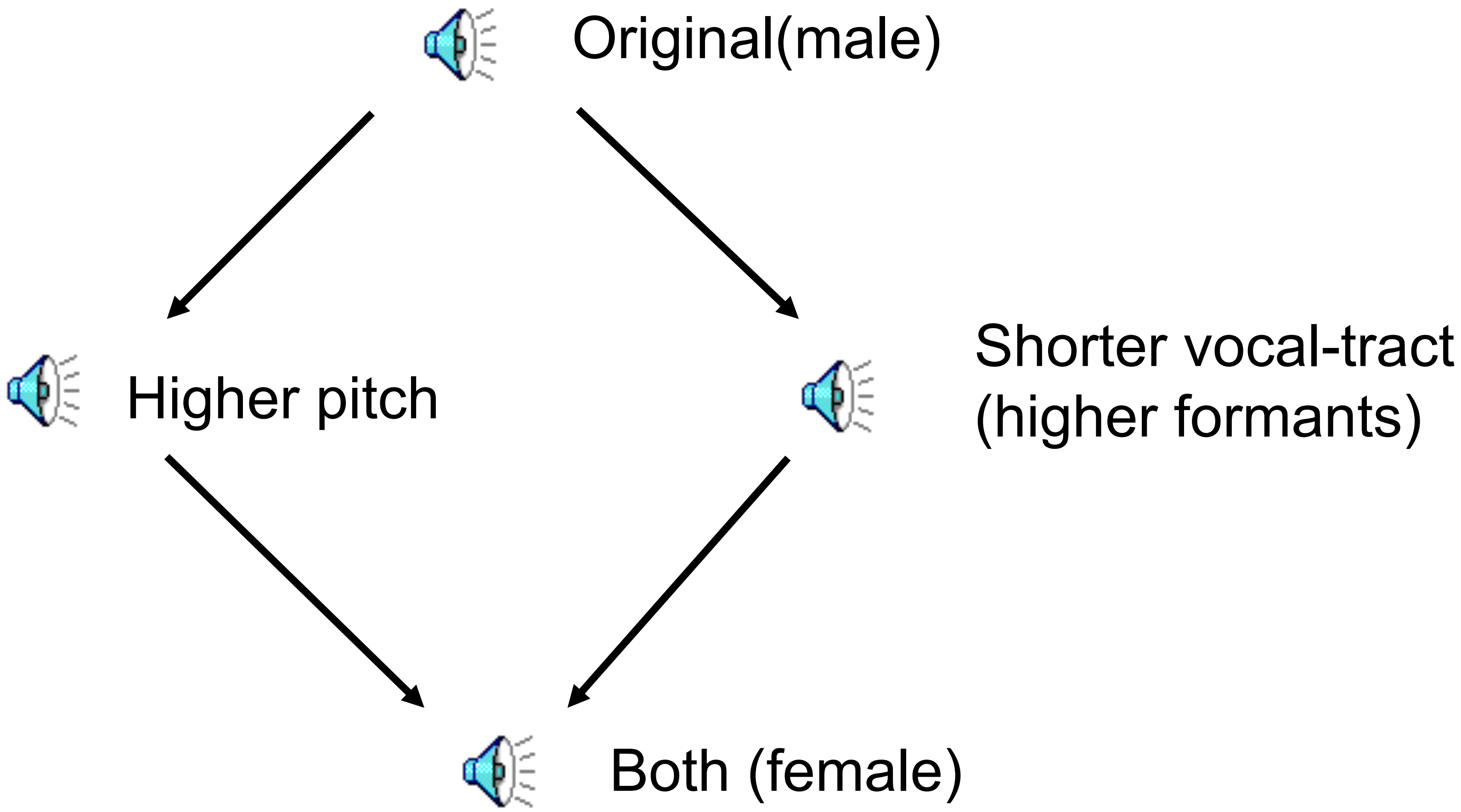
FIG. 8.1 Illustration of how three different vowel sounds are produced. Part (a) shows the spectrum of the sound produced by vibration of the vocal folds. It consists of a series of harmonics whose levels decline with increasing frequency. Part (b) shows schematic cross-sections of the vocal tract in the positions appropriate for the three vowels. Part (c) shows the filter functions or transfer functions associated with those positions of the vocal tract. Part (d) shows the spectra of the vowels resulting from passing the glottal source in panel (a) through the filter functions in panel (c). Adapted from Bailey (1983) by permission of the author.

Timbre - Vowel sounds

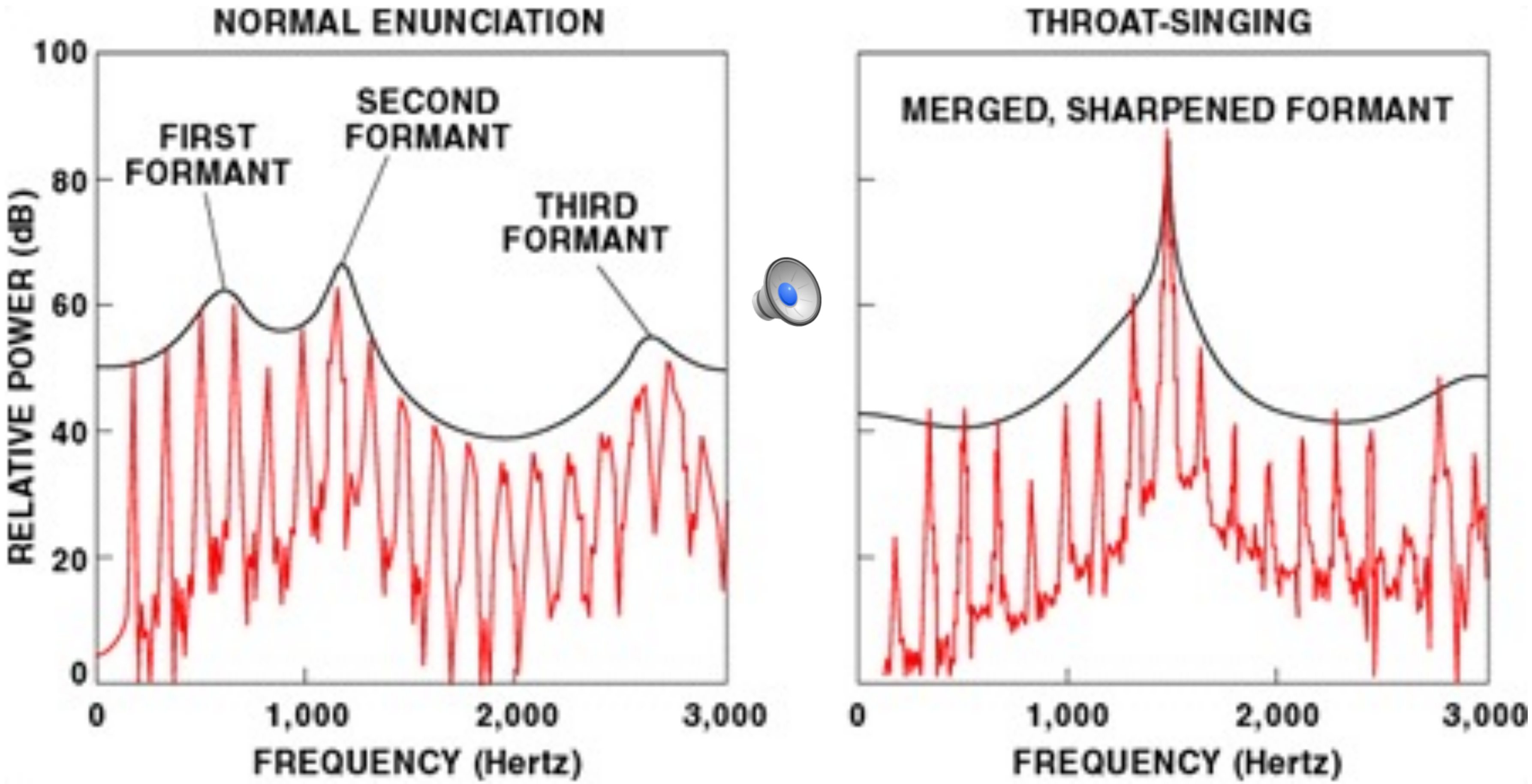
- Vowel Sounds in speech differ in the relative amplitudes of their harmonics.
- A particular vowel has harmonics that have a greater amplitude near the *formant* frequencies.
  - As you change the pitch of a vowel, you change the fundamental frequency and the spacing of the harmonics, but the formant frequencies stay the same.
  - If you change the vowel without changing the pitch of the voice, the fundamental and the harmonic spacing stay the same but the formant frequencies change.



Vocal tract change

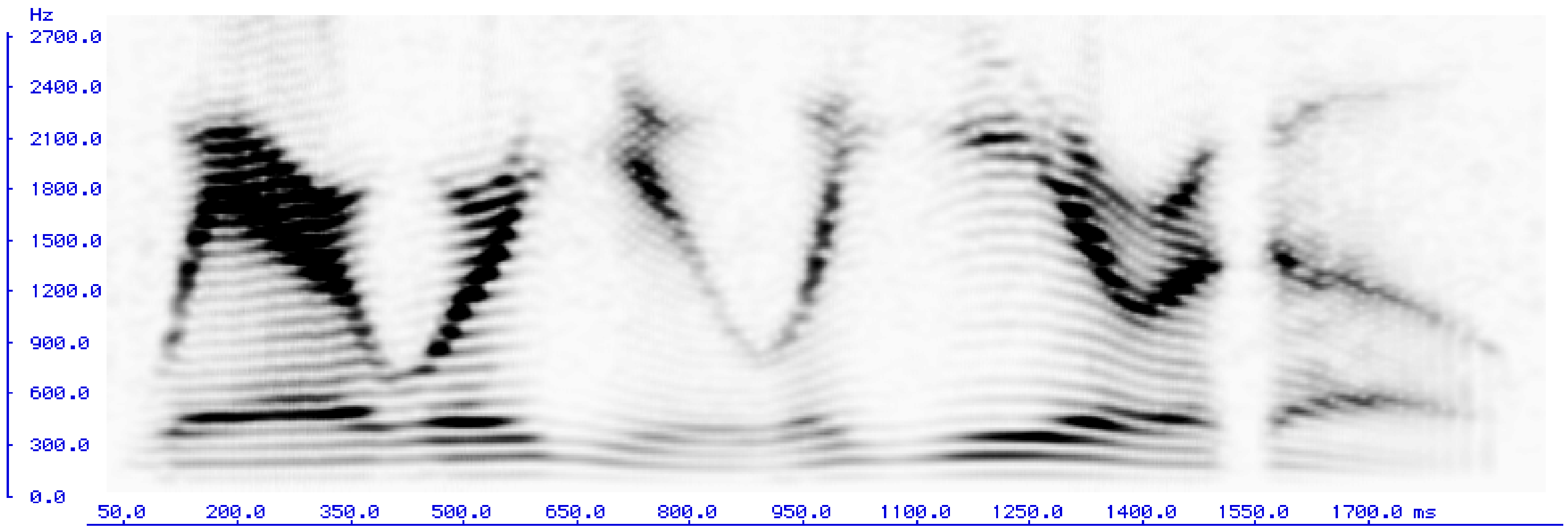


Tuvan Throat Music

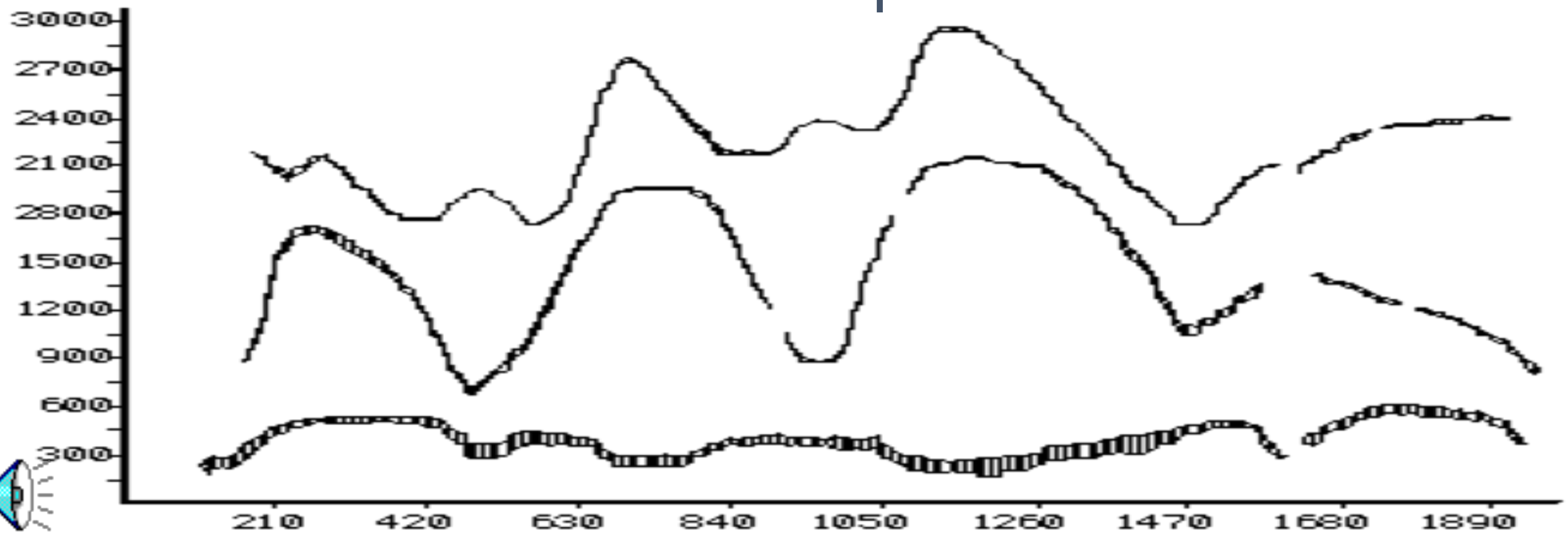


Tuvan Throat Music

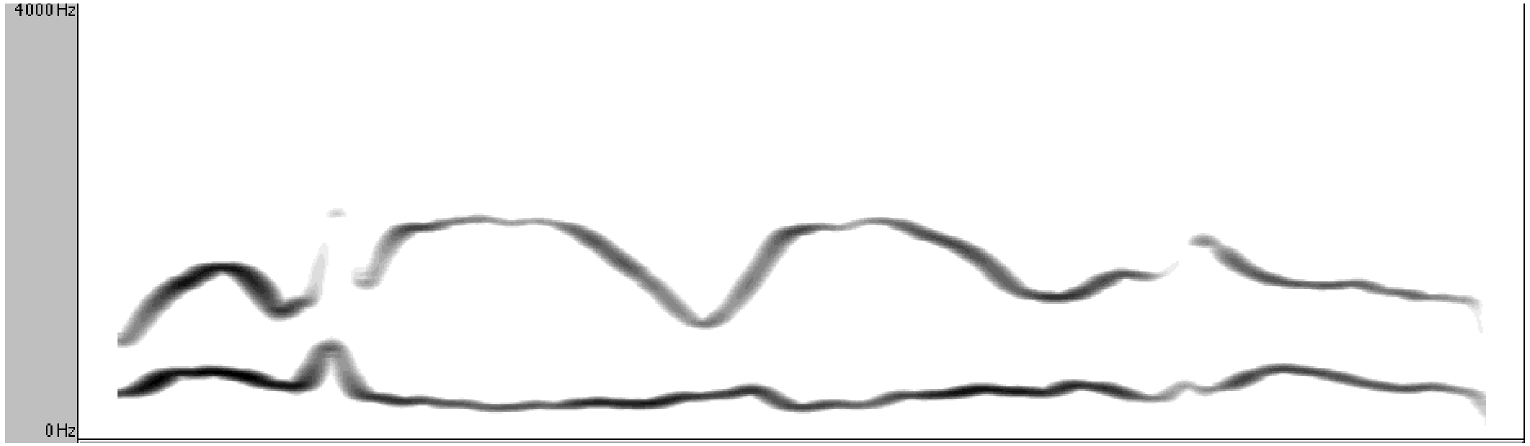
narrow-band spectrogram

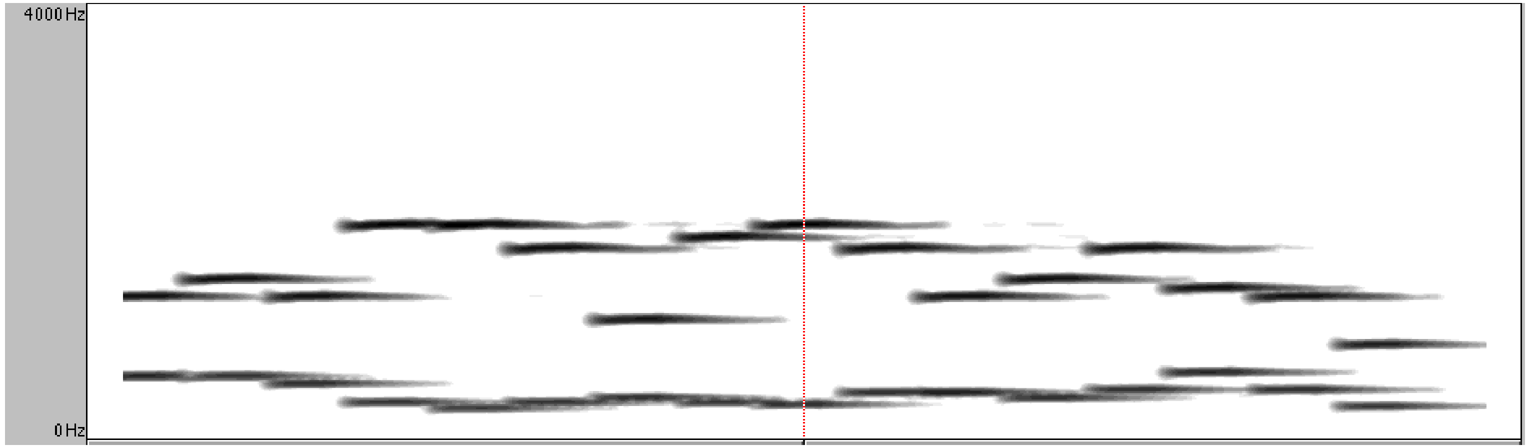


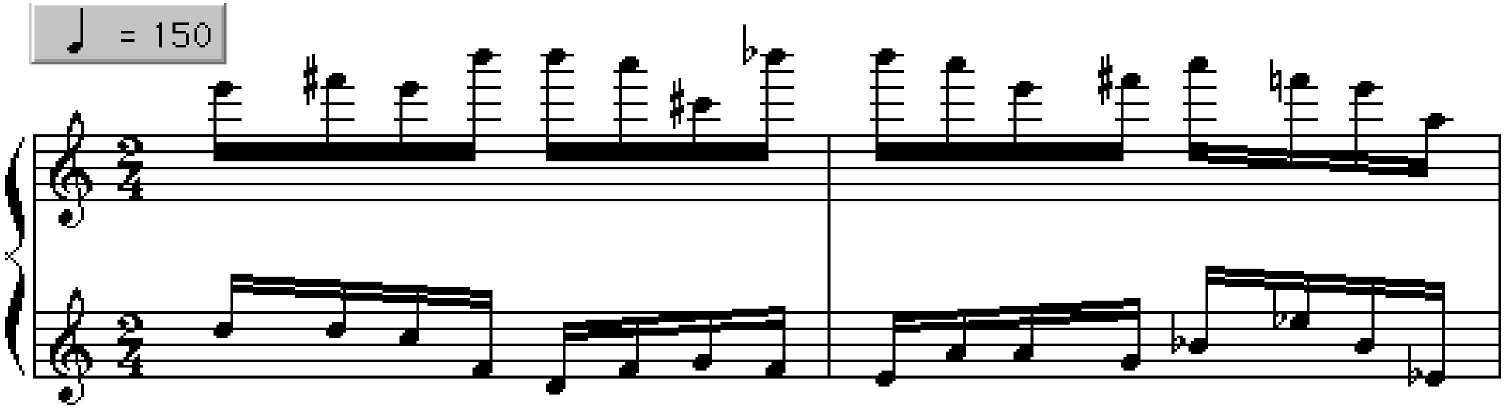
sine-wave speech



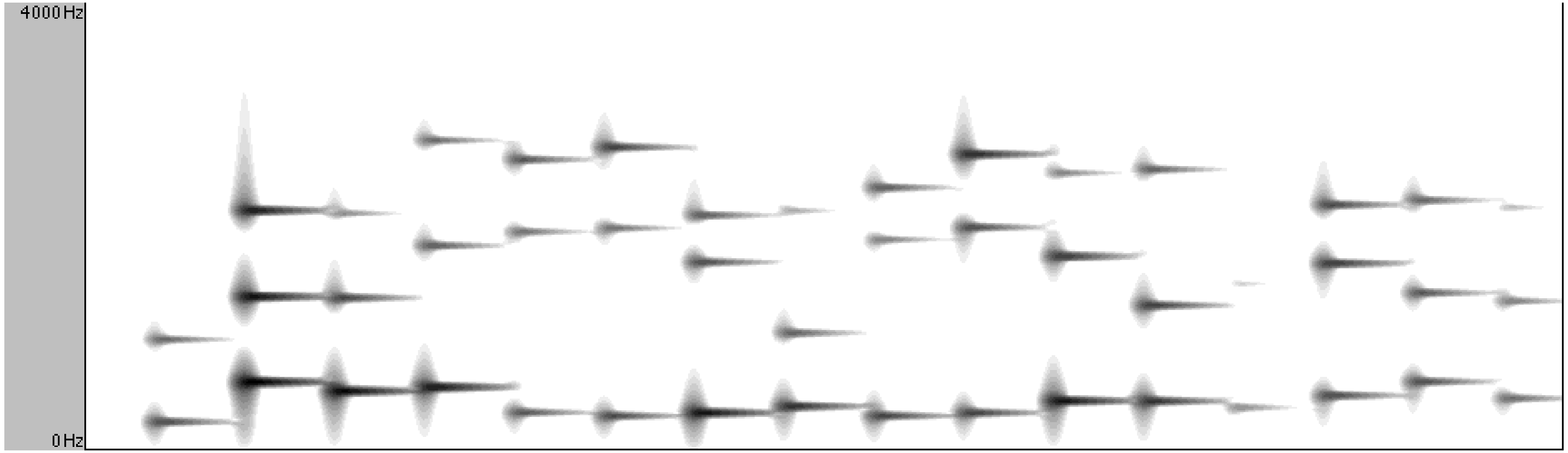
Speech
music

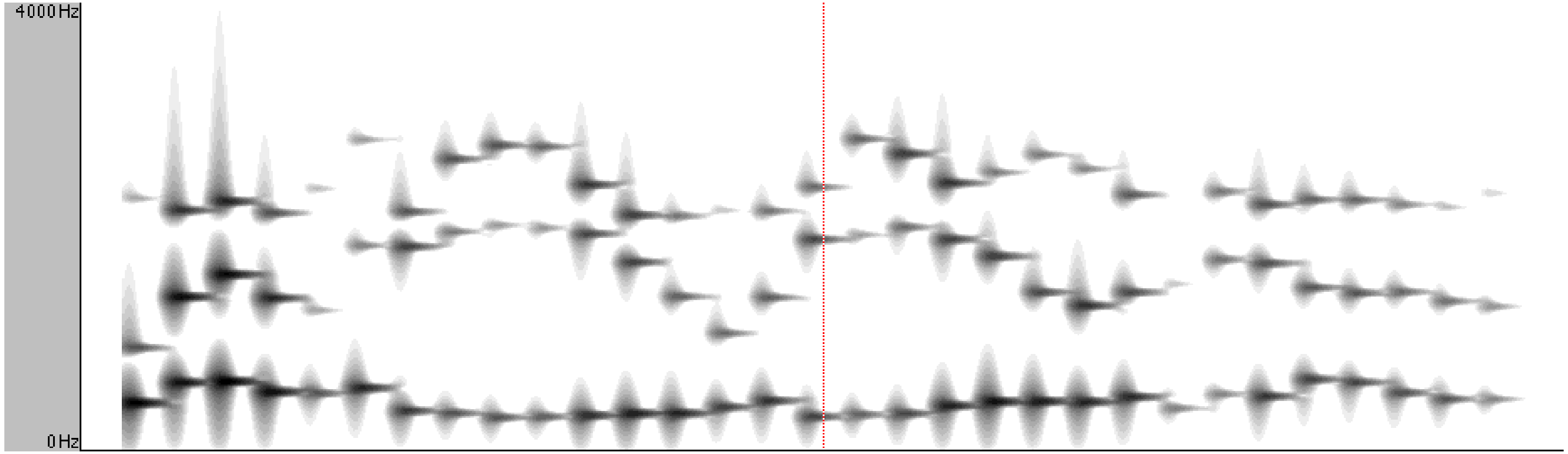




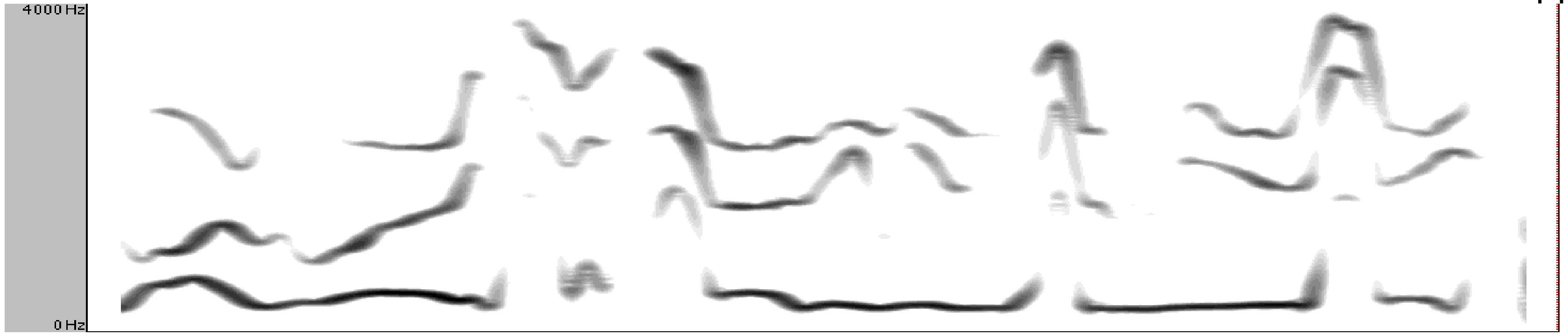


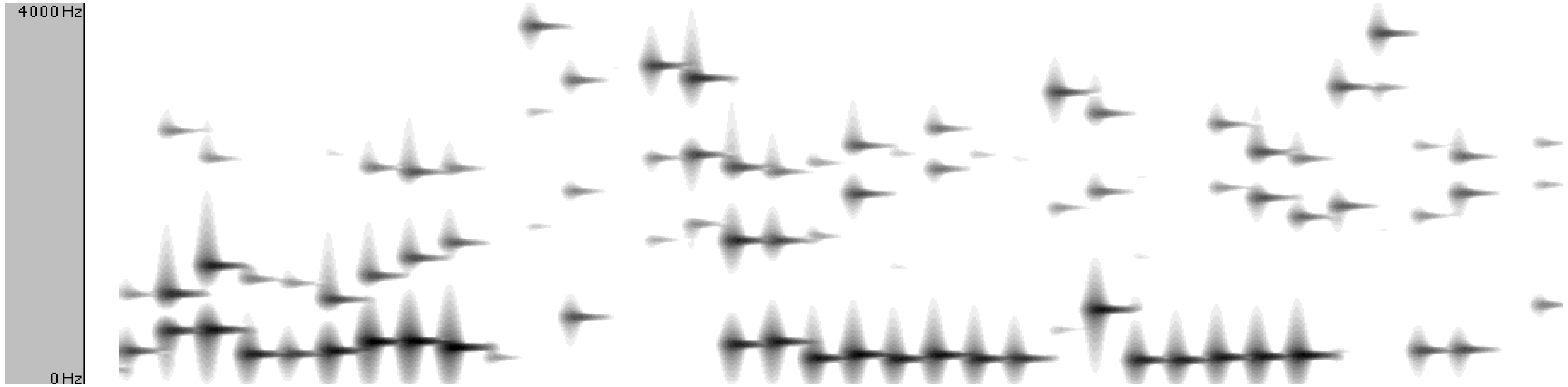
Speech
music

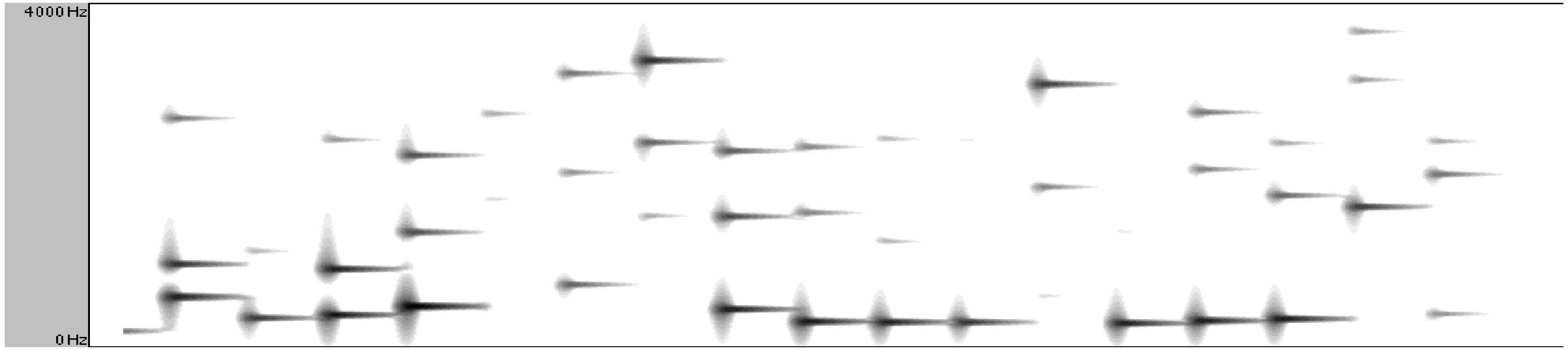




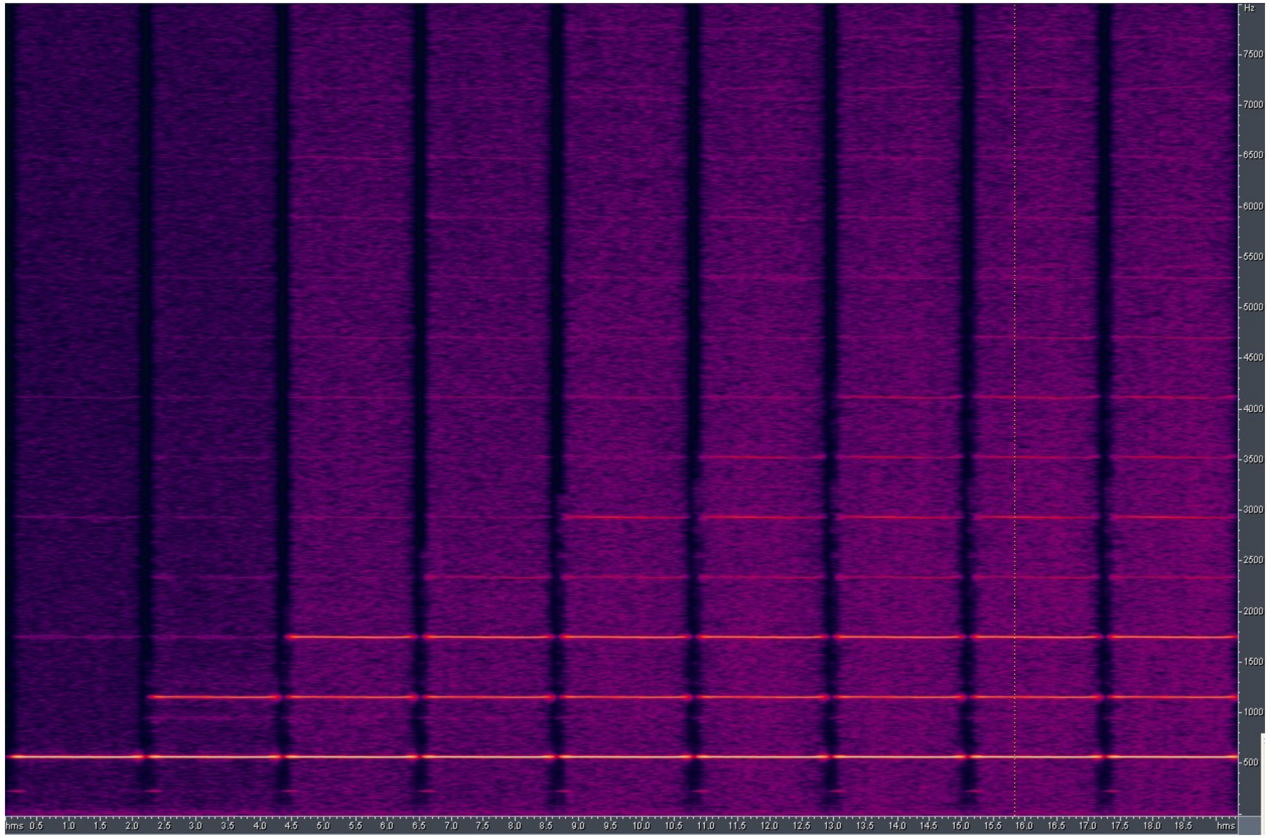
Speech
music



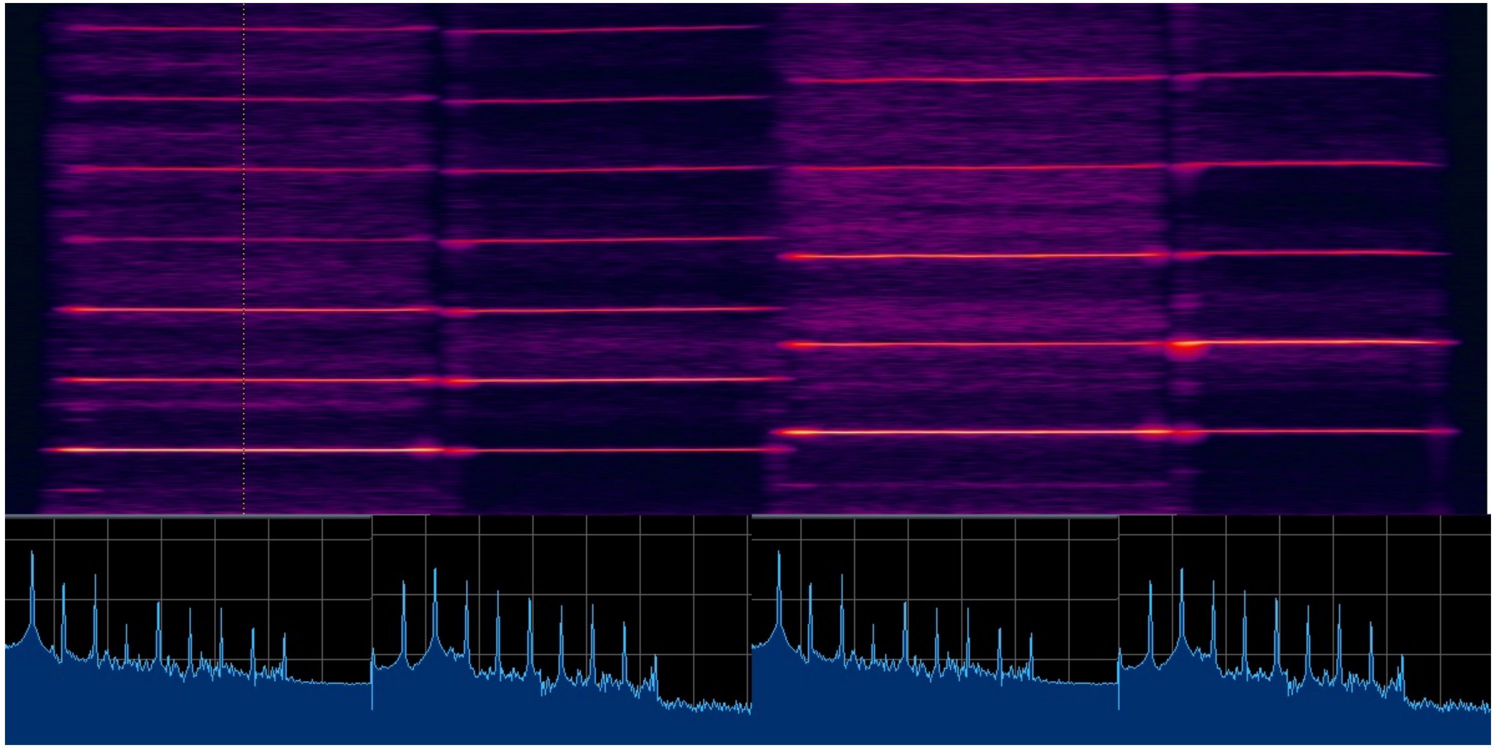




Adding harmonics to make an instrument's timbre



Different notes on clarinet and oboe



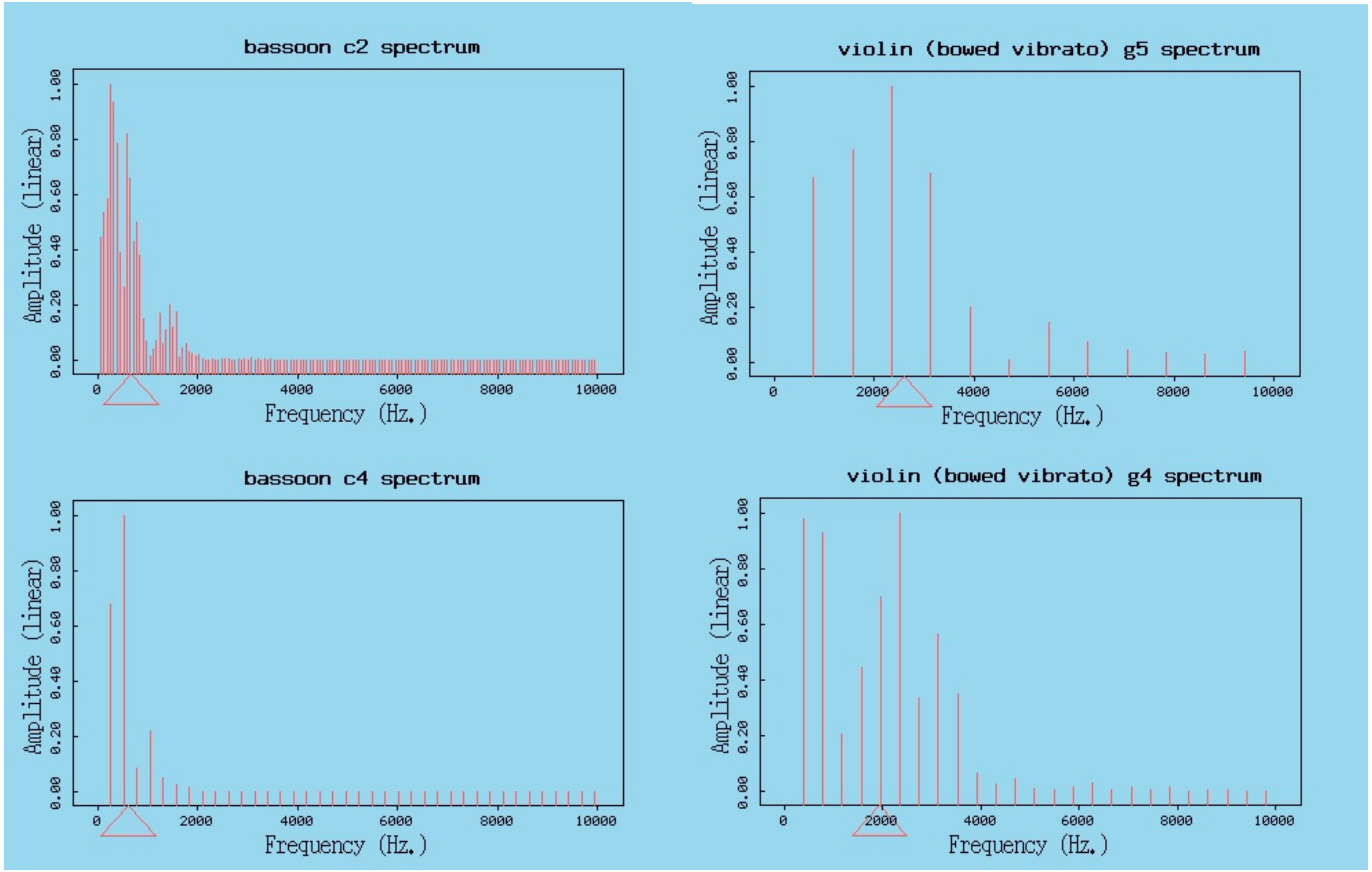
What determines an instrument’s timbre

1. “formant” frequencies

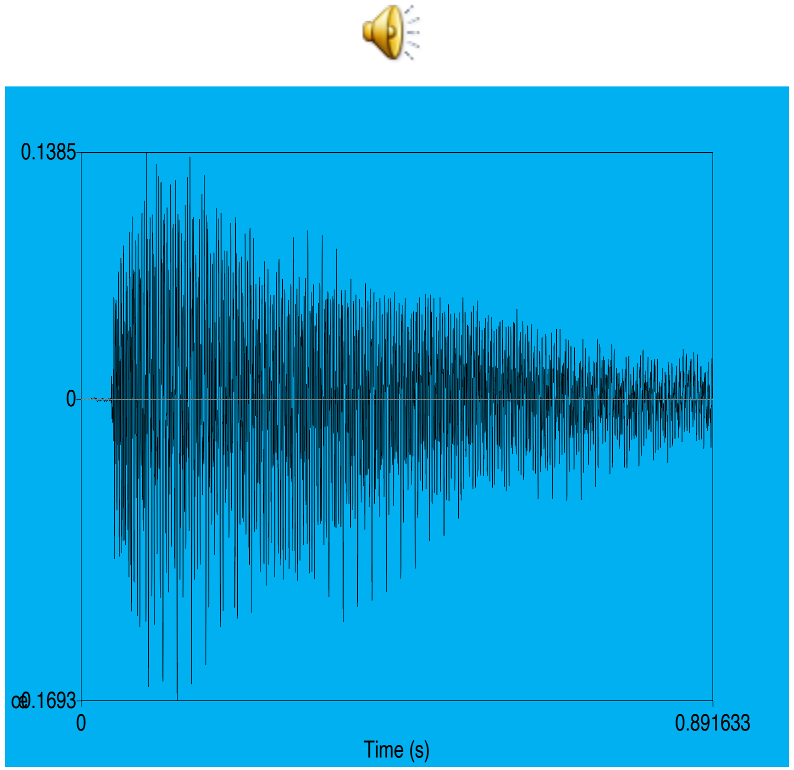
2. Amplitude envelope

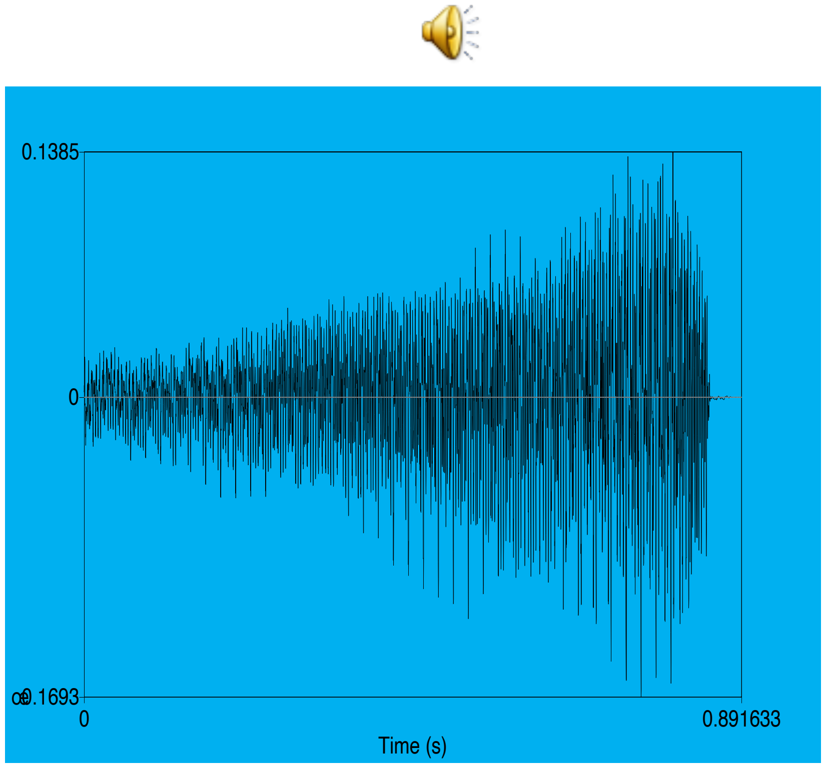
3. Onset / offset transients

Instrument timbre is more like speech formants

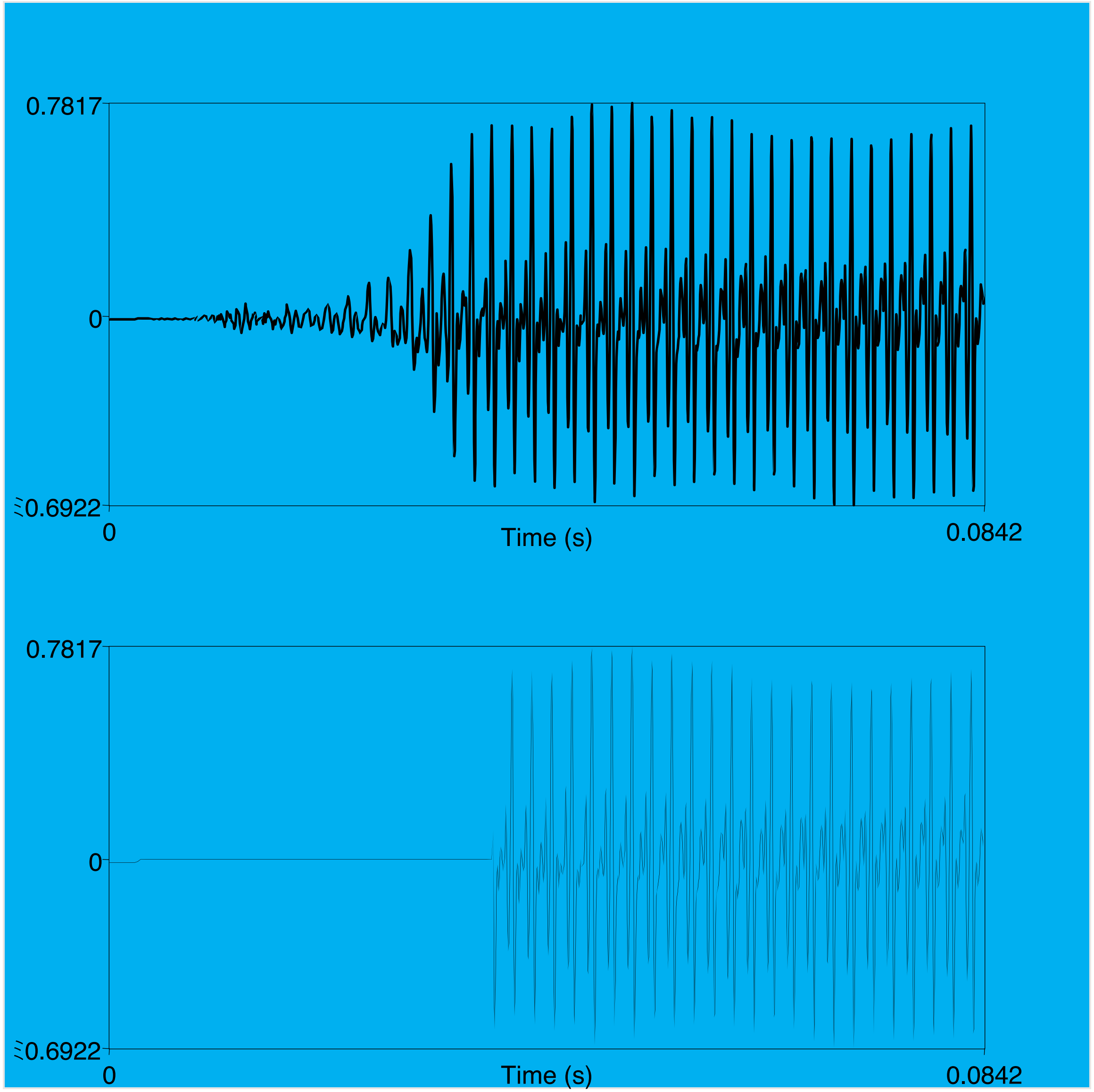


Forwards & backwards temporal envelopes





Onset transients



Synthesis of Speech and Song

Speech

Song

Terms using in different fields

- Physical attributes:
  - fundamental frequency(基频), harmonics(谐波), intensity(强度)
- Auditory attributes:
  - pitch(基音), timbre(音色), loudness(响度)
- Speech Science:
  - prosody: pitch(音高), intensity(音强), duration(音长), stop(停顿)
- Music:
  - Pitch(音高), timbre(音色), intensity(音强)

Pitch and Timbre in Speech and Music

Speech

- Pitch
  - Perception of auditory object
    - Integrating harmonic component into an object under noisy environment
    - Speech transmission far away
  - Pitch variation
    - Mandarin tones
    - Paralinguistic information (Interrogative, imperative, exclamation) ；
- Timbre
  - Shape of vocal tract
  - Speech transmitted content

Music

- Pitch
  - Perception of auditory object
    - Integrating harmonic component into an object under noisy environment
  - Pitch variation
    - Melody
- Timbre
  - Music instrument
  - Rhythmic

Q&A