

MIT OpenCourseWare
<http://ocw.mit.edu>

24.910 Topics in Linguistic Theory: Laboratory Phonology
Spring 2007

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

24.910

Laboratory Phonology

Basic Audition

- No class next week (Tuesday is a Monday)
- Readings for 2/27: Johnson chs 5 & 6
- Assignments (due 2/27):
 - Basic acoustics.
 - VOT and laryngeal contrasts in Mandarin and English.

Audition

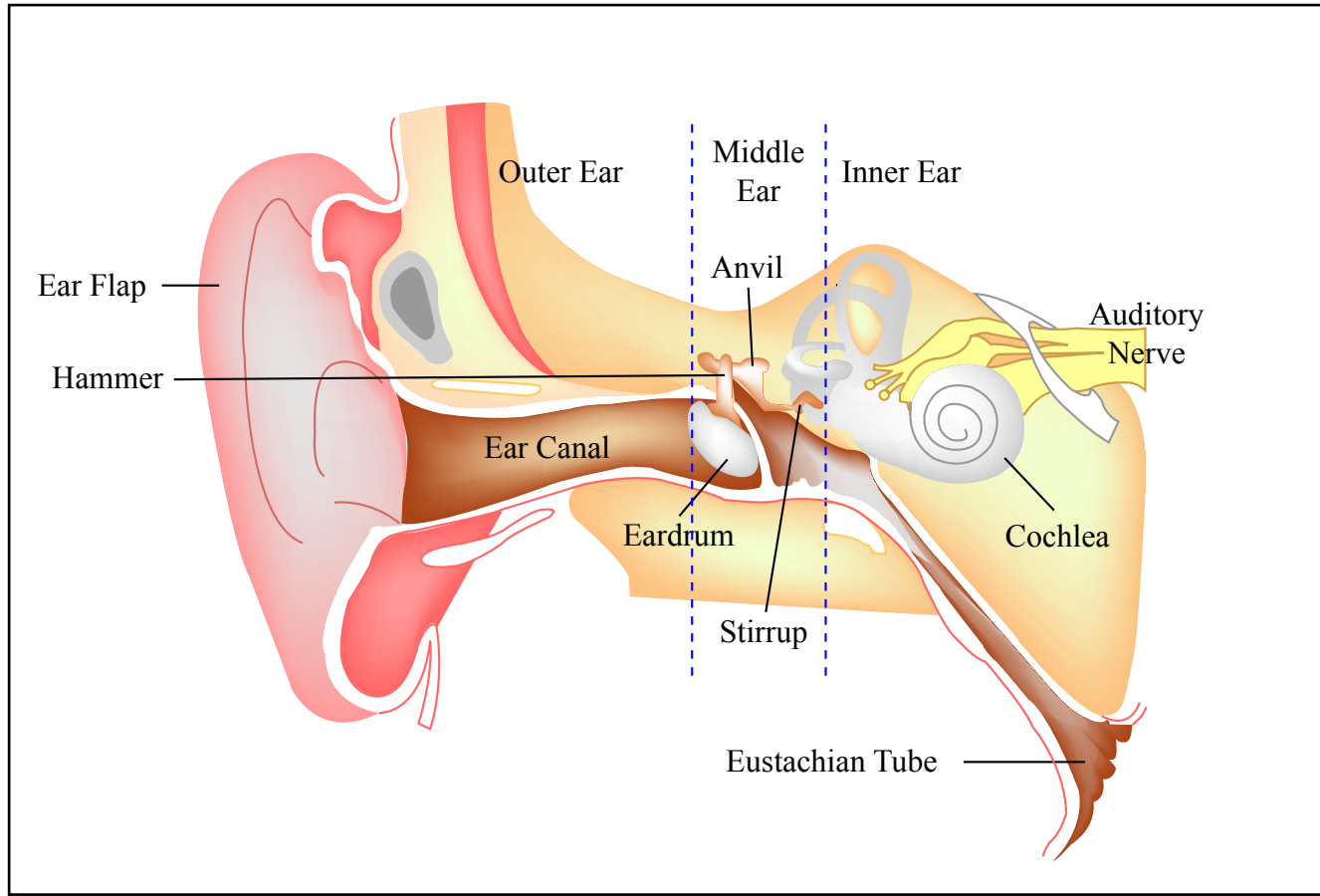


Figure by MIT OpenCourseWare.

Anatomy

Audition

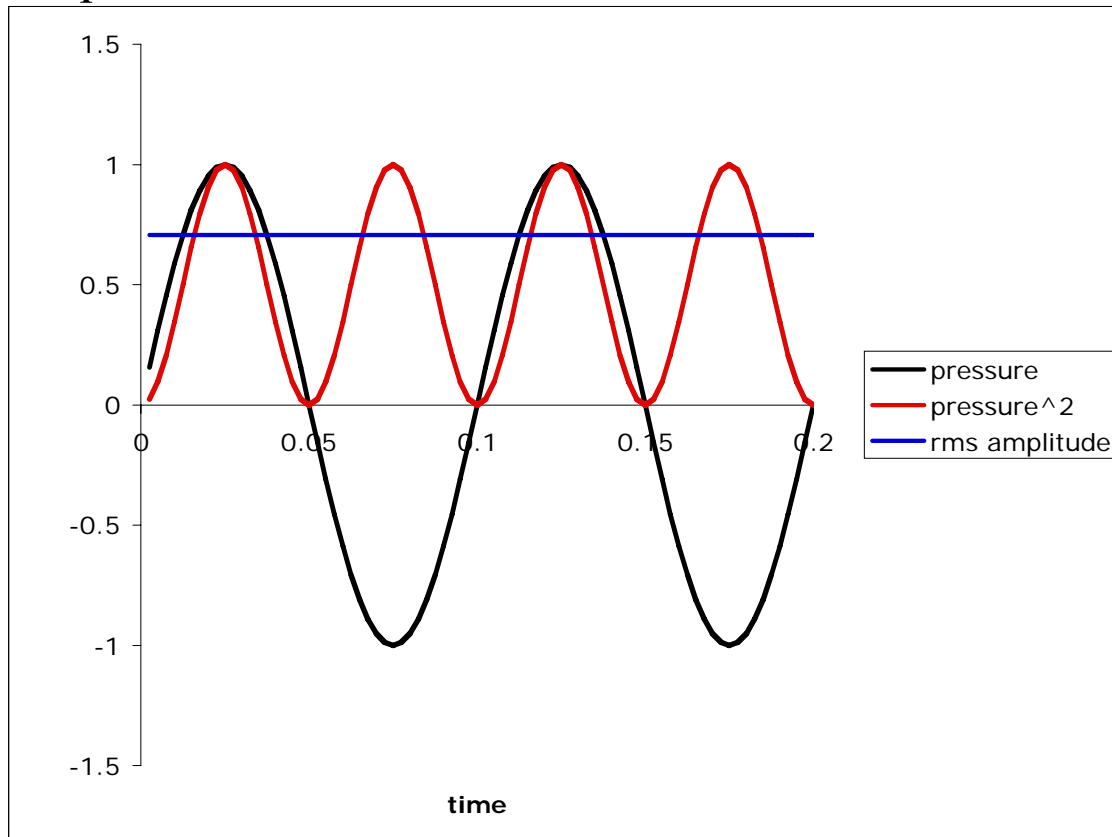
- Loudness
- Pitch
- ‘Auditory spectrograms’

Loudness

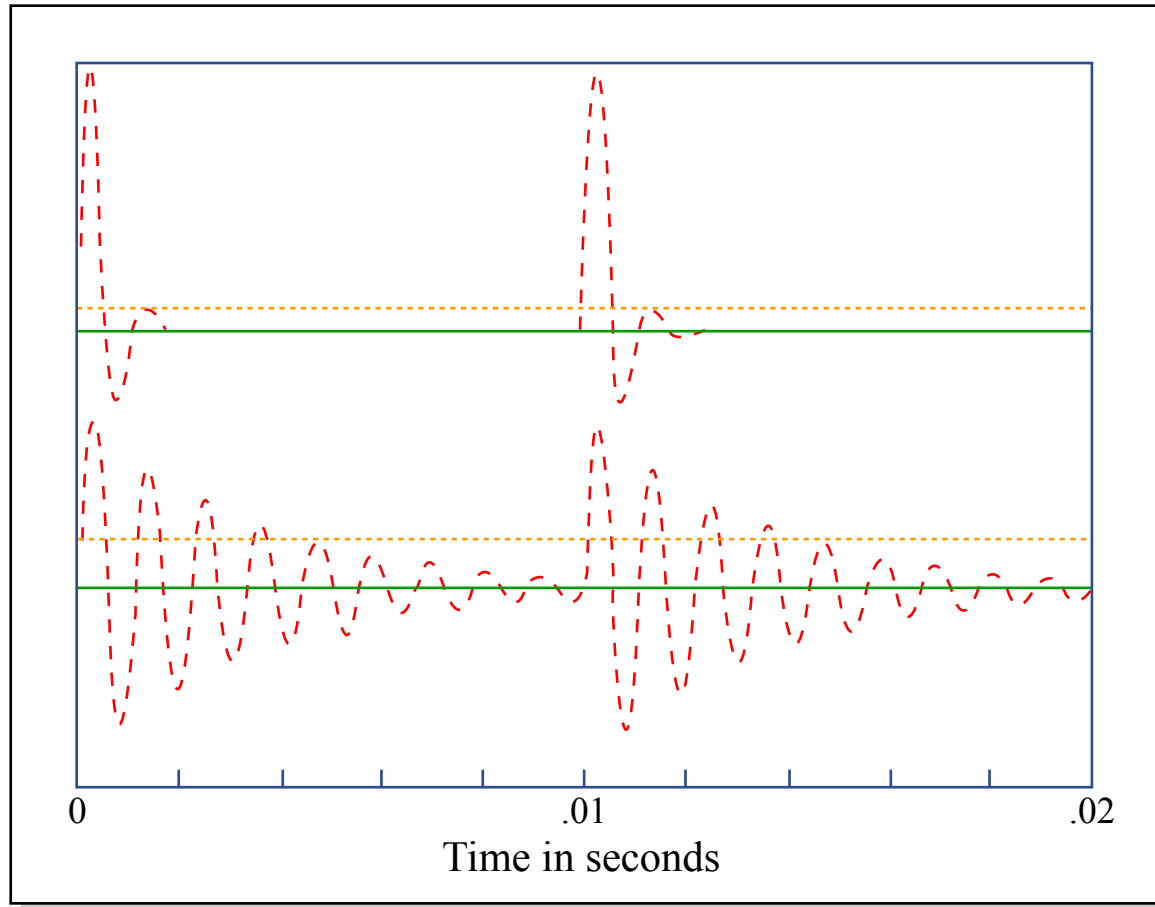
- The perceived loudness of a sound depends on the amplitude of the pressure fluctuations in the sound wave.
- Amplitude is usually measured in terms of root-mean-square (rms amplitude):
 - The square root of the mean of the squared amplitude over some time window.

rms amplitude

- Square each sample in the analysis window.
- Calculate the mean value of the squared waveform:
 - Sum the values of the samples and divide by the number of samples.
- Take the square root of the mean.



rms amplitude

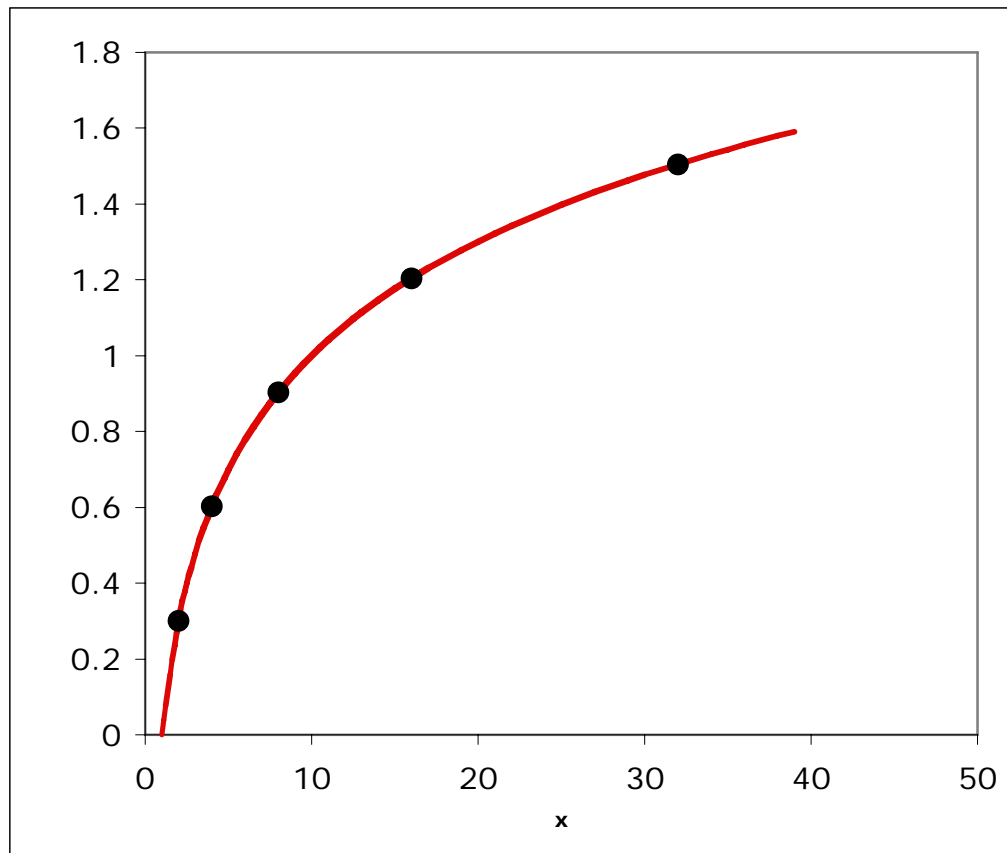


Intensity

- Perceived loudness is more closely related to intensity (power per unit area), which is proportional to the square of the amplitude.
- relative intensity in Bels = $\log_{10}(x^2/r^2)$
- relative intensity in dB = $10 \log_{10}(x^2/r^2)$
= $20 \log_{10}(x/r)$
- In absolute intensity measurements, the comparison amplitude is usually $20\mu\text{Pa}$, the lowest audible pressure fluctuation of a 1000 Hz tone (dB SPL).

logarithmic scales

- $\log x^n = n \log x$



Loudness

- The relationship between intensity and perceived loudness is not exactly logarithmic.

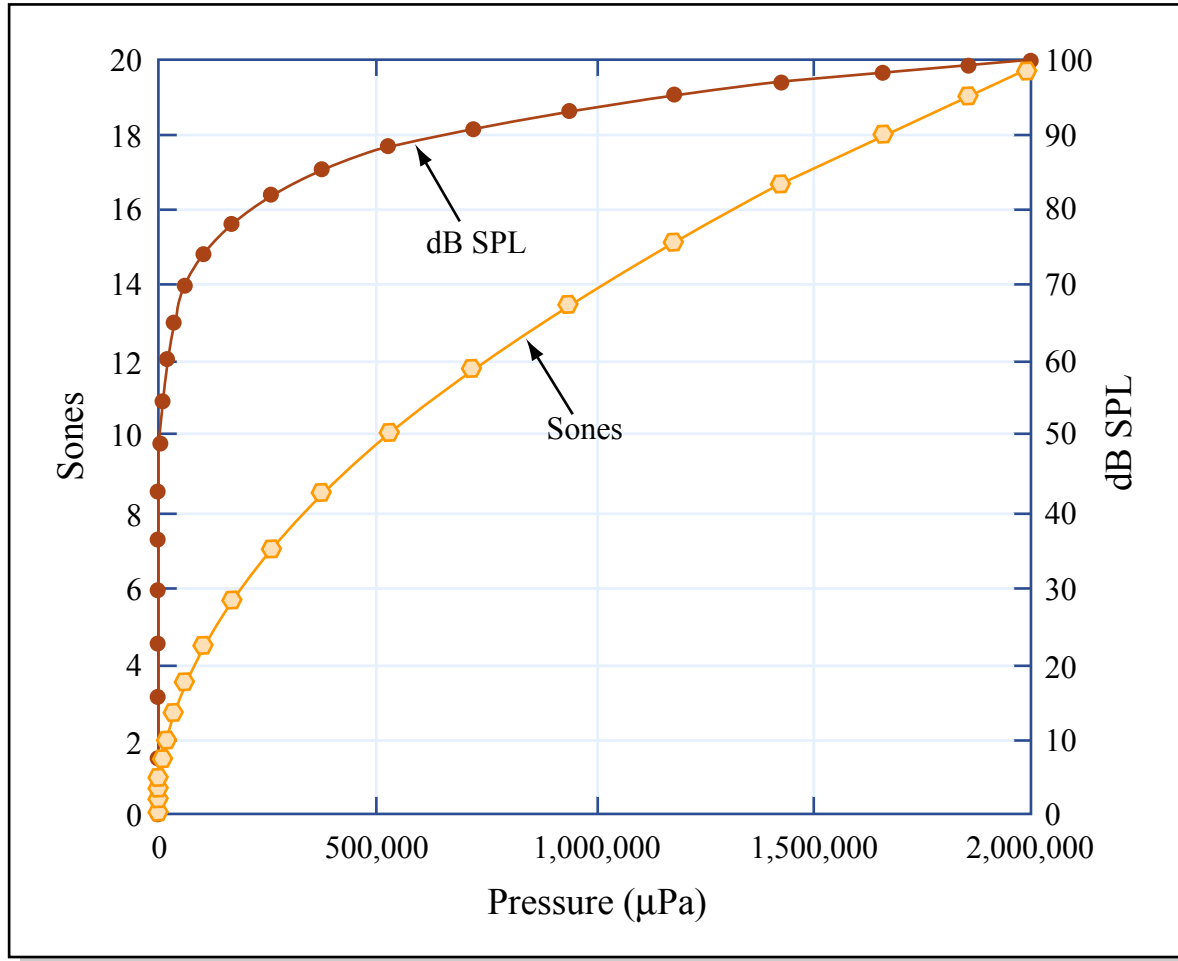
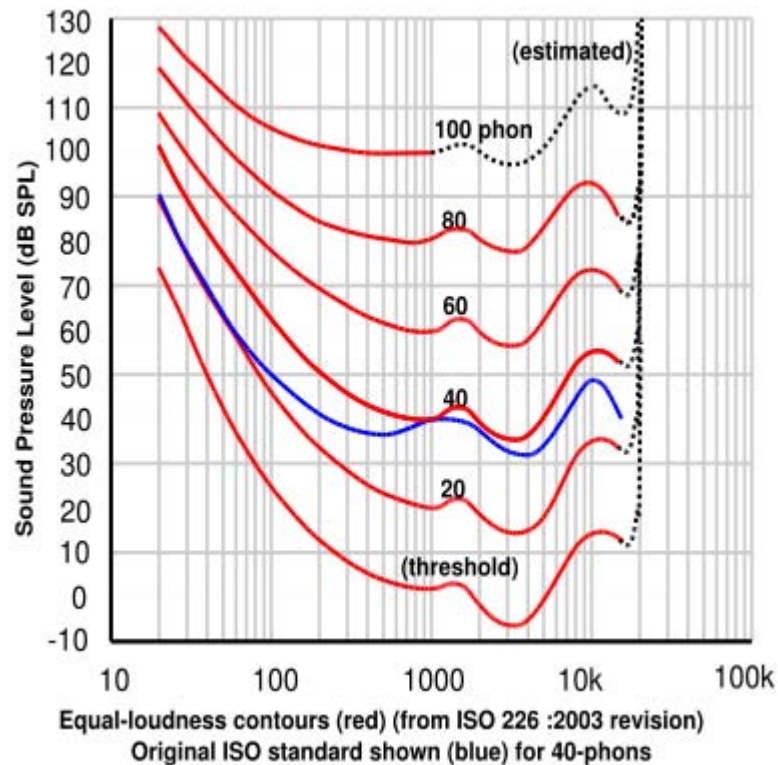


Figure by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997. ISBN: 9780631188483.

Loudness

- Loudness also depends on frequency.
- equal loudness contours for pure tones:



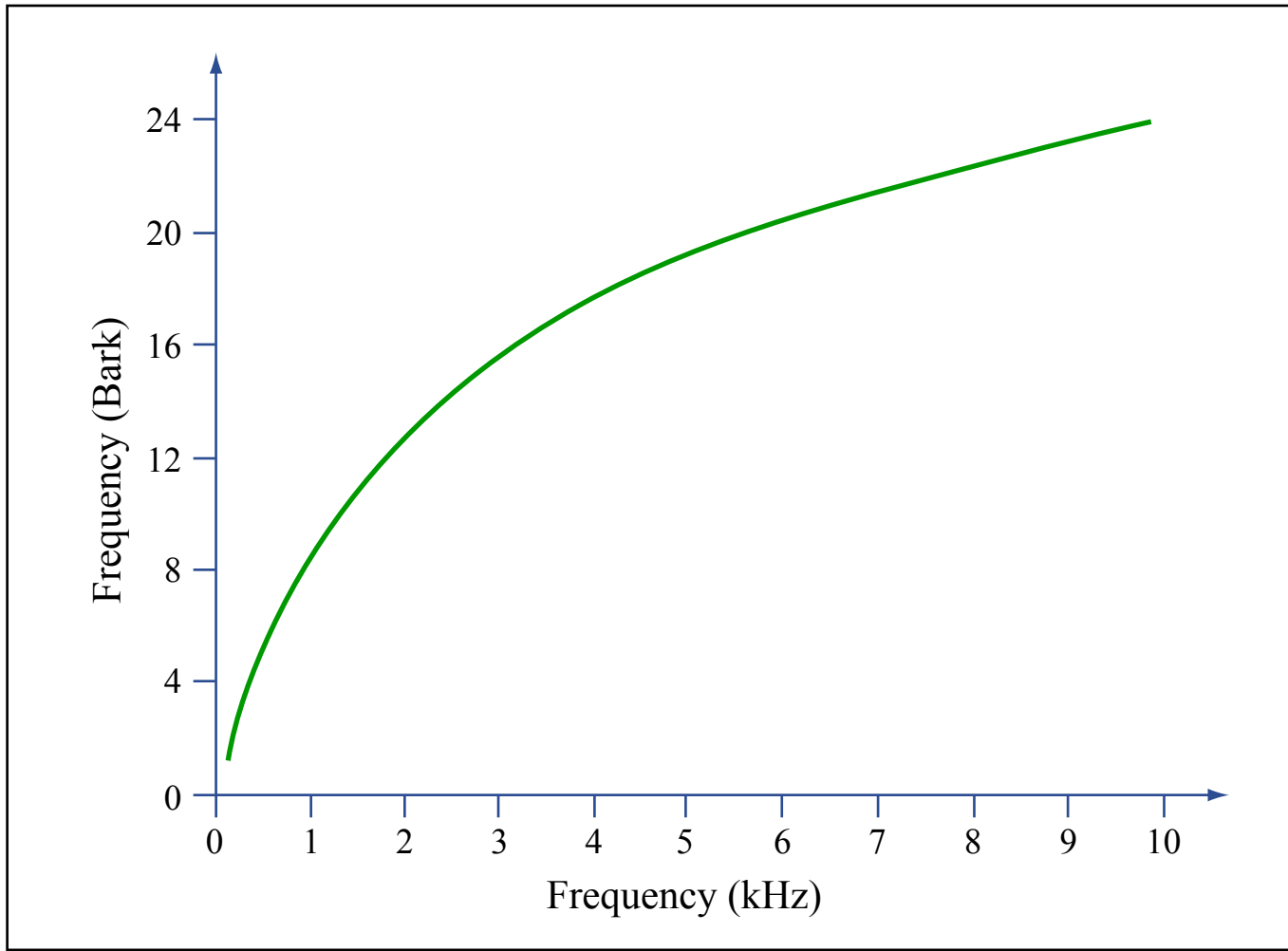
Source: Wikimedia Commons.

Loudness

- At short durations, loudness also depends on duration.
- Temporal integration: loudness depends on energy in the signal, integrated over a time window.
- Duration of integration is often said to be about 200ms, i.e. relevant to the perceived loudness of vowels.

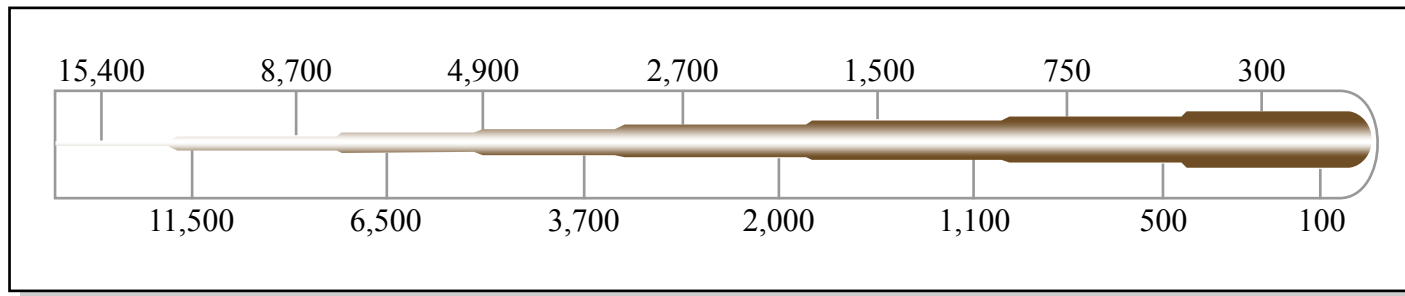
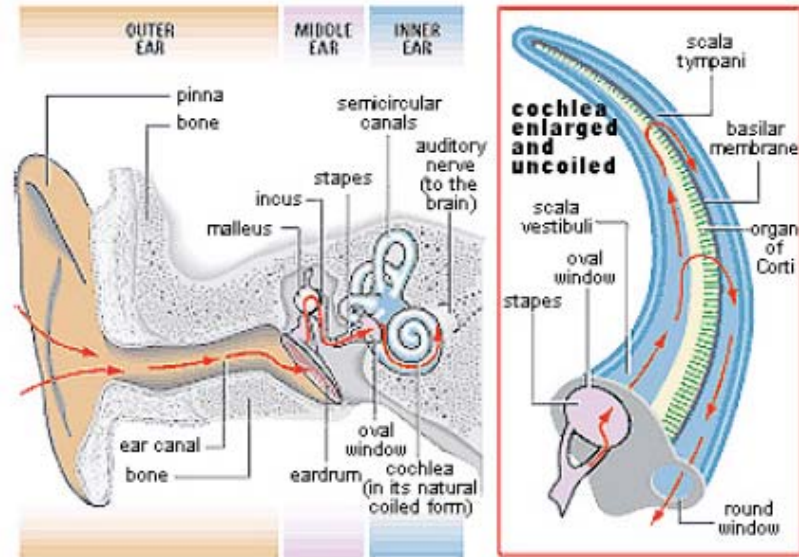
Pitch

- Perceived pitch is approximately linear with respect to frequency from 100-1000 Hz, between 1000-10,000 Hz the relationship is approximately logarithmic.



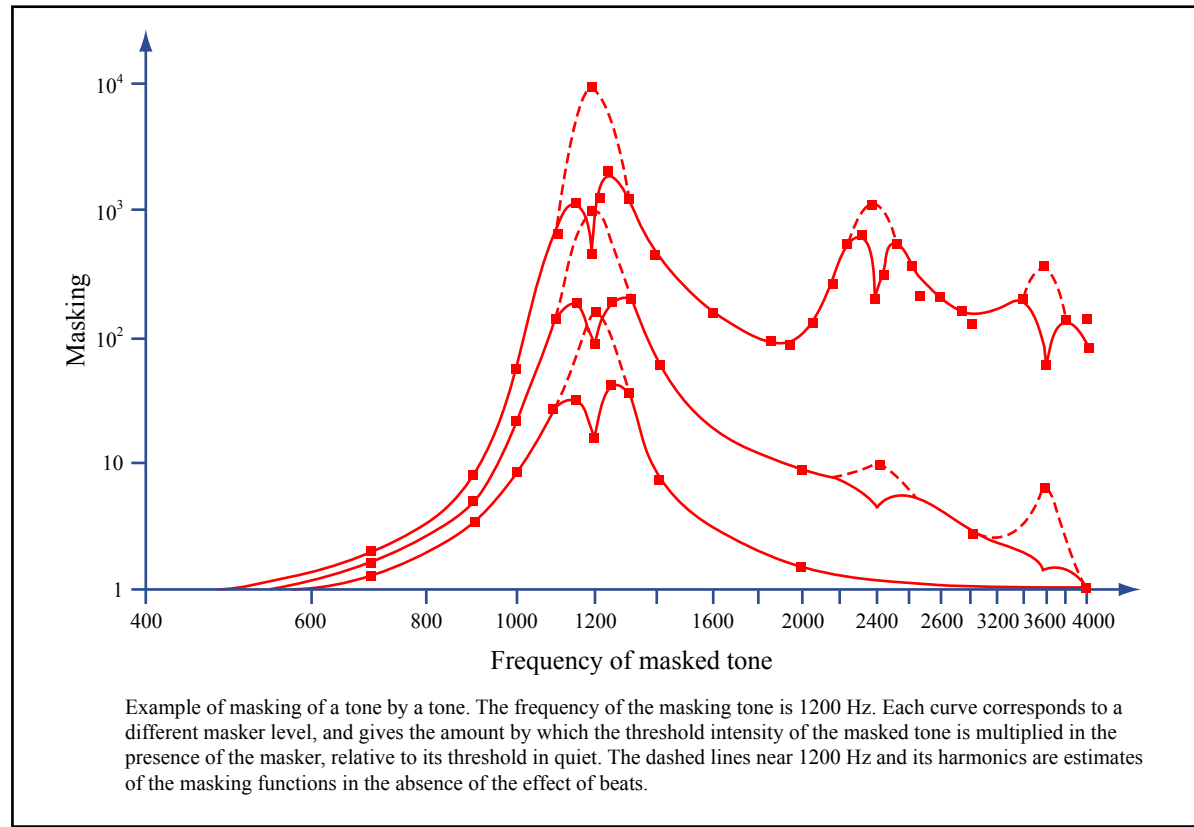
Pitch

- The non-linear frequency response of the auditory system is related to the physical structure of the basilar membrane.
- basilar membrane ‘uncoiled’:



Masking - simultaneous

- Energy at one frequency can reduce audibility of simultaneous energy at another frequency (masking).
- One sound can also mask a preceding or following sound.



Time course of auditory nerve response

Response to a noise burst:

- Strong initial response
- Rapid adaptation (~ 5 ms)
- Slow adaptation (> 100 ms)
- After tone offset, firing rate only gradually returns to spontaneous level.

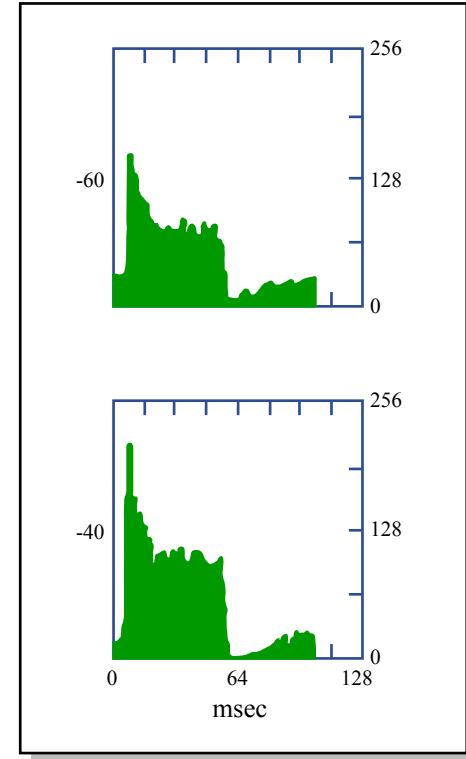


Figure by MIT OpenCourseWare. Adapted from Kiang et al. (1965)

Interactions between sequential sounds

- A preceding sound can affect the auditory nerve response to a following tone (Delgutte 1980).

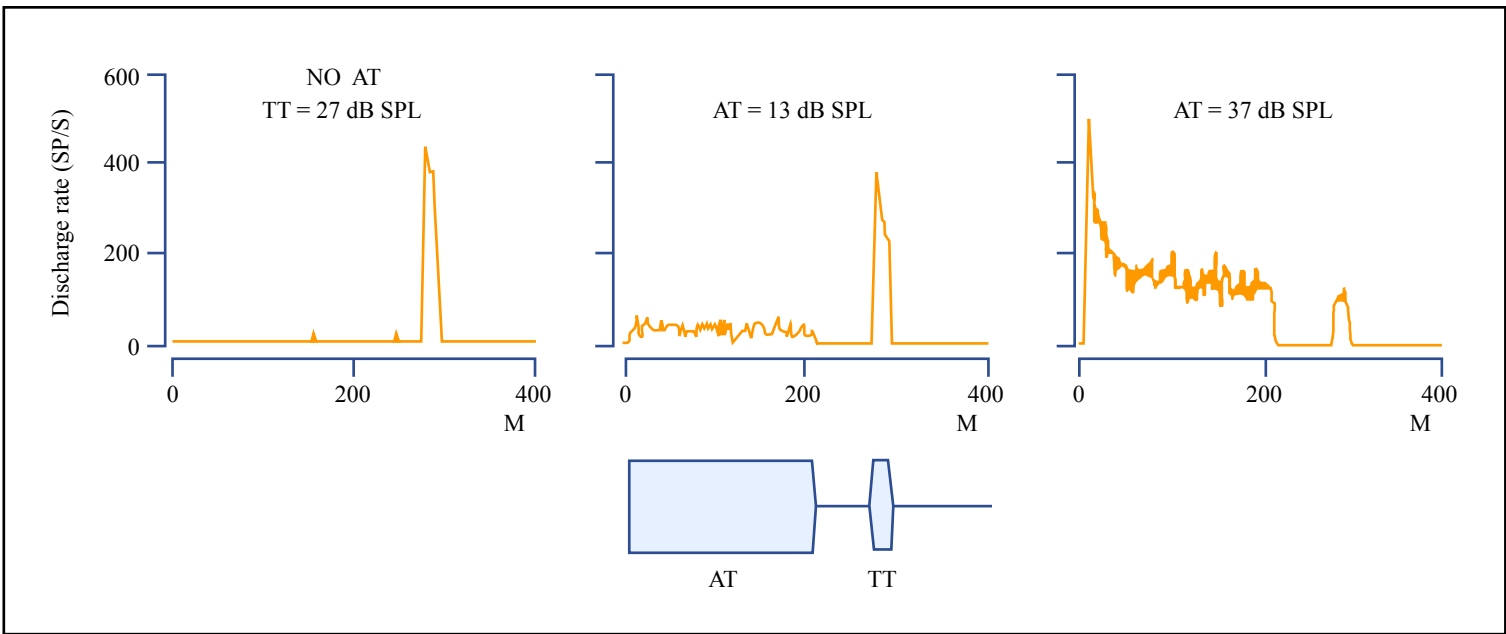


Figure by MIT OpenCourseWare. Adapted from Stevens, Kenneth N. *Acoustic Phonetics*. Cambridge, MA: MIT Press, 1999. ISBN: 9780262194044, after Delgutte, B. "Representation of Speech-like Sounds in the Discharge Patterns of Auditory-nerve Fibers." *Journal of the Acoustical Society of America* 68, no. 3 (1980): 843-857.

Auditory ‘spectrograms’

The auditory system performs a running frequency analysis of acoustic signals - cf. spectrogram.

- A regular spectrogram analyzes frequency of equal widths, but the peripheral auditory system analyzes frequency bands that are wider at higher frequencies.
- Further disparities are introduced by the non-linearities of the peripheral auditory system, e.g.
 - loudness is non-linearly related to intensity
 - masking(simultaneous and nonsimultaneous)

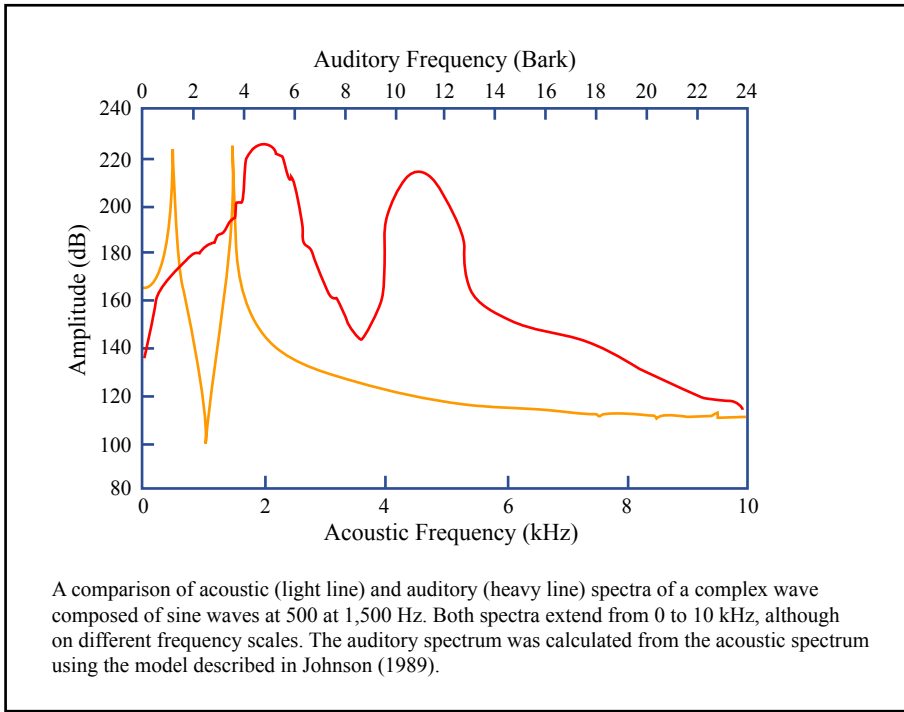


Image by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997. ISBN: 9780631188483.

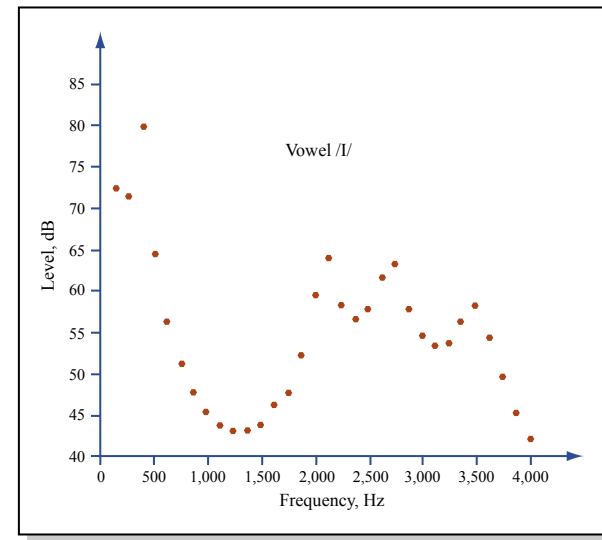


Image by MIT OpenCourseWare. Adapted from Moore, Brian. *The Handbook of Phonetic Science*. Edited by William J. Hardcastle and John Laver. Malden, MA: Blackwell, 1997. ISBN: 9780631188483.

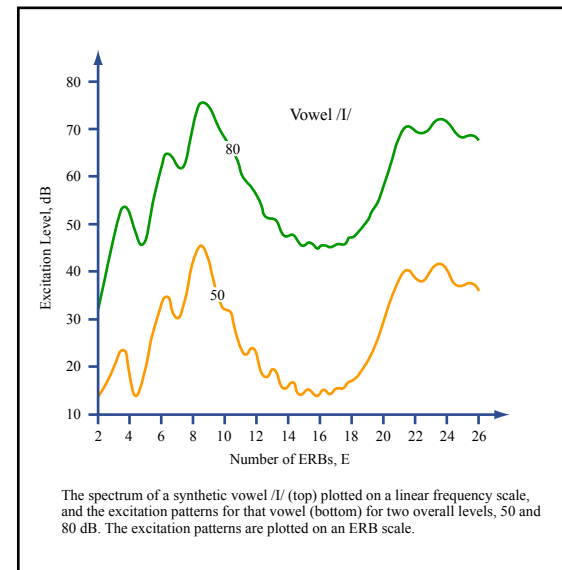
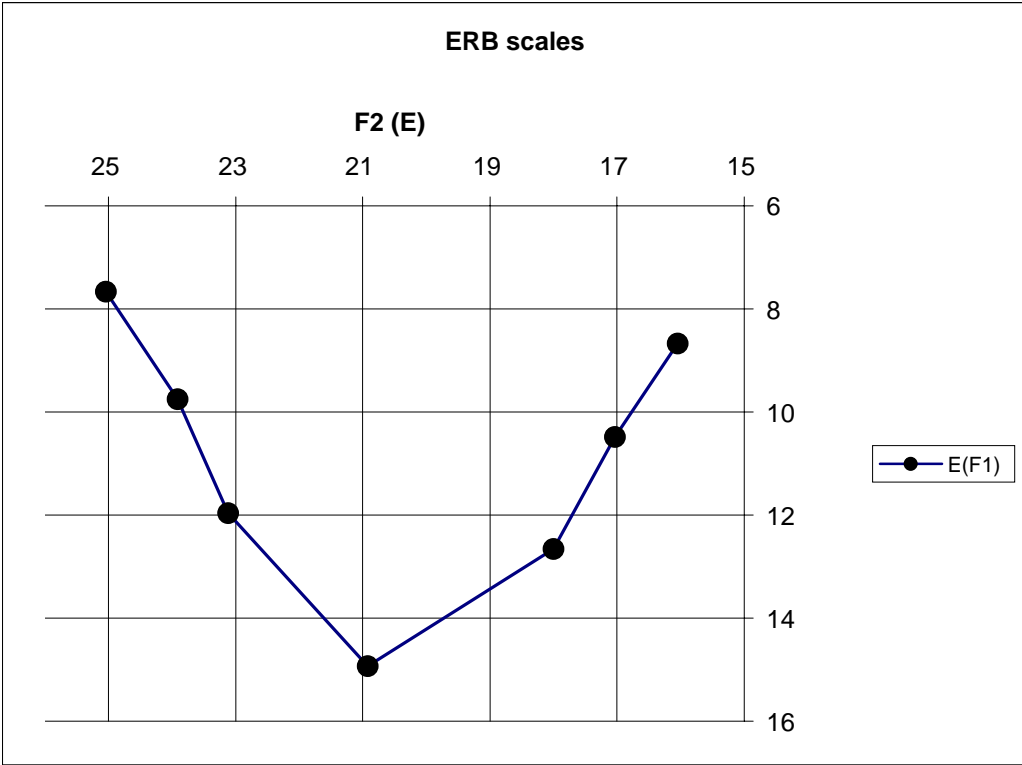
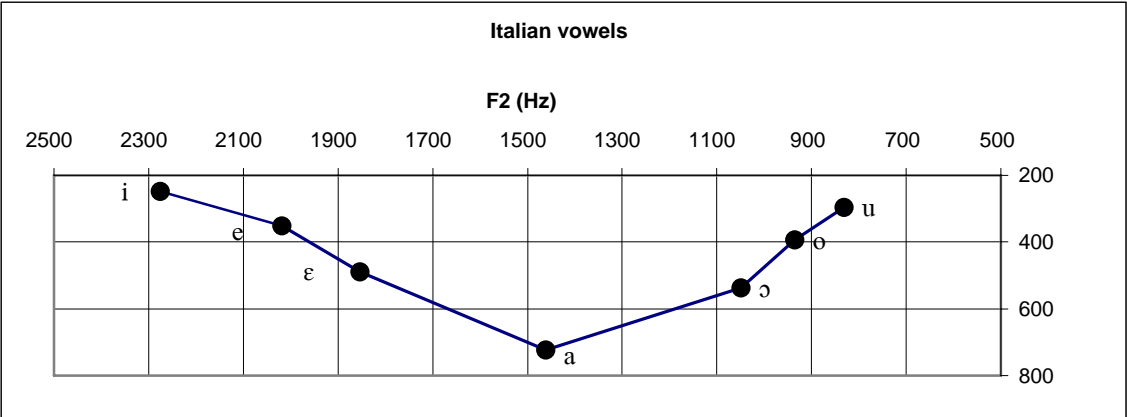


Image by MIT OpenCourseWare. Adapted from Moore, Brian. *The Handbook of Phonetic Science*. Edited by William J. Hardcastle and John Laver. Malden, MA: Blackwell, 1997. ISBN: 9780631188483.

Spectrogram images removed due to copyright restrictions.

Figure 3.8 in Johnson, Keith. "Comparison of Normal Acoustic Spectrogram and Auditory Spectrogram or Cochleagram."
Acoustic and Auditory Phonetics. Malden, MA: Blackwell Publishers, 1997. ISBN: 9780631188483.



24.910

Linguistic Phonetics

Analog-to-digital conversion of speech signals

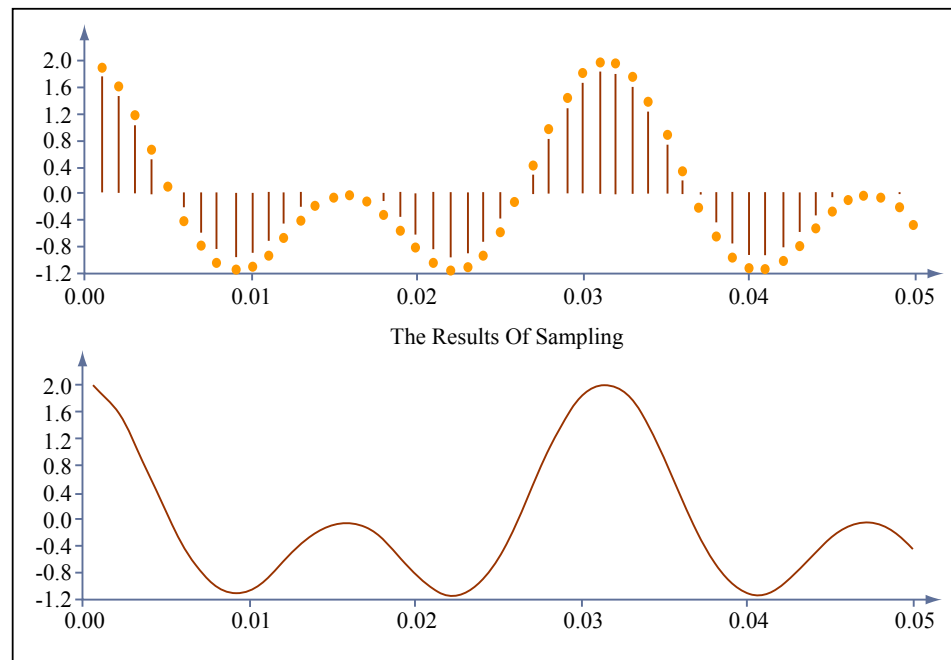
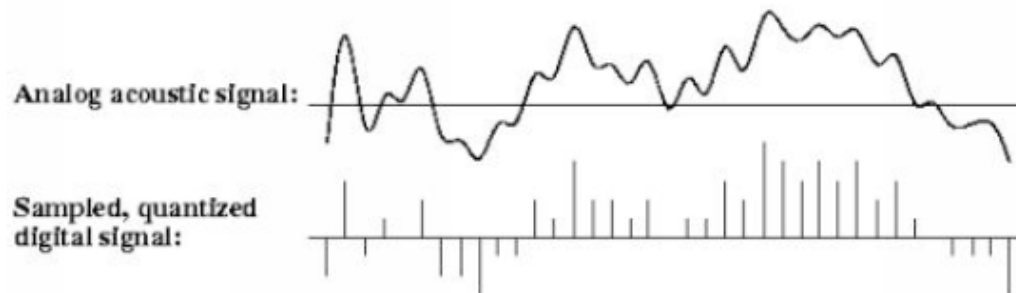


Figure by MIT OpenCourseWare.

Analog-to-digital conversion

- Almost all acoustic analysis is now computer-based.
- Sound waves are analog (or continuous) signals, but digital computers require a digital representation - i.e. a series of numbers, each with a finite number of digits.
- There are two continuous scales that must be divided into discrete steps in analog-to-digital conversion of speech: time and pressure (or voltage).
 - Dividing time into discrete chunks is called **sampling**.
 - Dividing the amplitude scale into discrete steps is called **quantization**.



Sampling

- The amplitude of the analog signal is sampled at regular intervals.
- The sampling rate is measured in Hz (samples per second).
- The higher the sampling rate, the more accurate the digital representation will be.

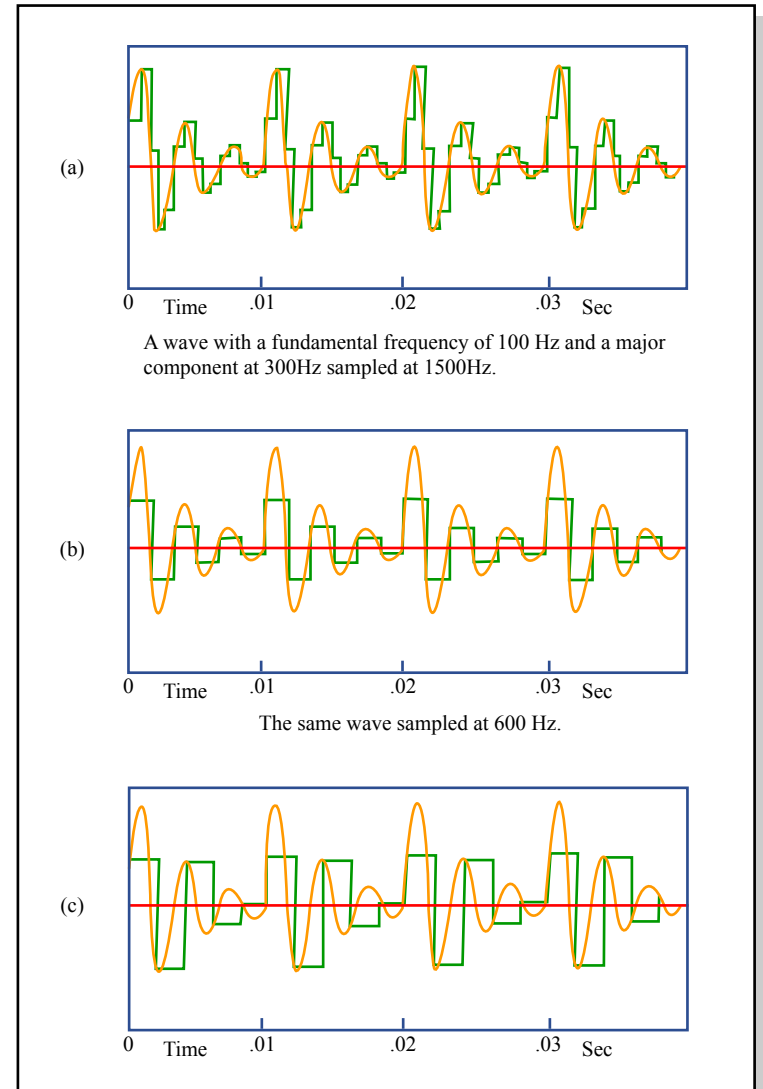


Figure by MIT OpenCourseWare. Adapted from Ladefoged, Peter.
L104/204 Phonetic Theory lecture notes, University of California, Los Angeles.

Sampling

- In order to represent a wave component of a given frequency, it is necessary to sample the signal with at least twice that frequency (the Nyquist Theorem).
- The highest frequency that can be represented at a given sampling rate is called the Nyquist frequency.
- The wave at right has a significant harmonic at 300 Hz
 - (a) sampling rate 1500 Hz
 - (b) sampling rate 600 Hz
 - (c) sampling rate 500 Hz

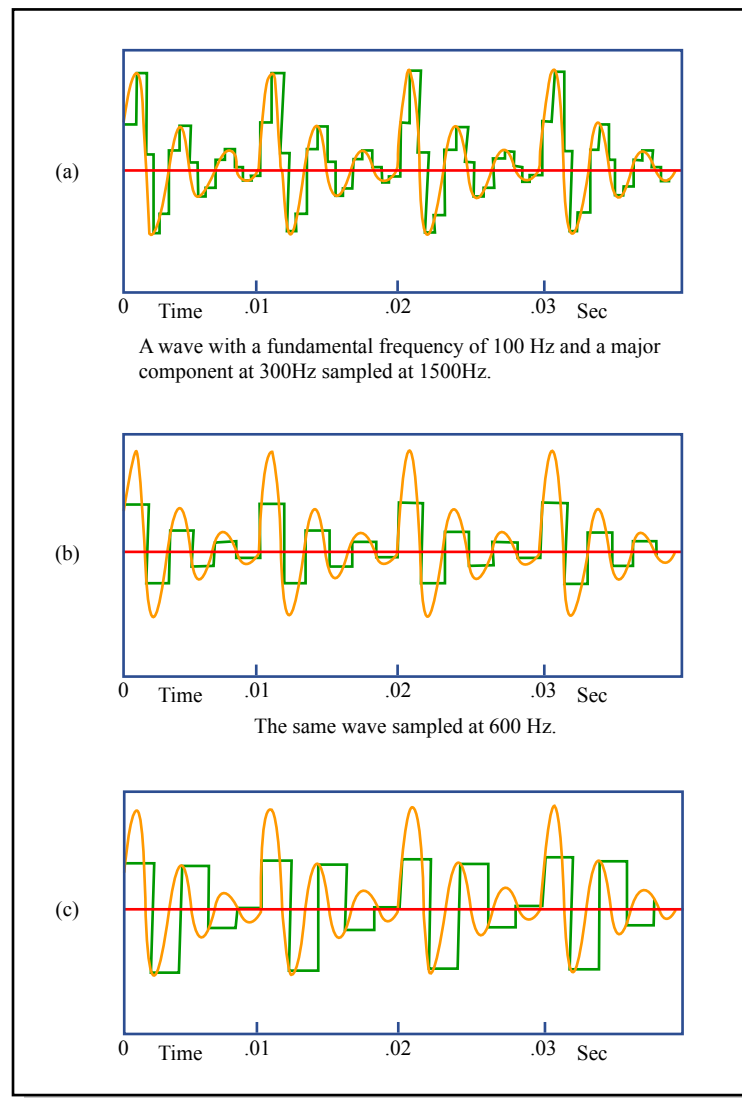


Figure by MIT OpenCourseWare. Adapted from Ladefoged, Peter. L104/204 Phonetic Theory lecture notes, University of California, Los Angeles.

What sampling rate should you use?

- The highest frequency that (young, undamaged) ears can perceive is about 20 kHz, so to ensure that all audible frequencies are represented we must sample at $2 \times 20 \text{ kHz} = 40 \text{ kHz}$.
- The ear is relatively insensitive to frequencies above 10 kHz, and almost all of the information relevant to speech sounds is below 10 kHz, so high quality sound is still obtained at a sampling rate of 20 kHz.
- There is a practical trade-off between fidelity of the signal and memory, but memory is getting cheaper all the time.

What sampling rate should you use?

- For some purposes (e.g. measuring vowel formants), a high sampling rate can be a liability, but it is always possible to **downsample** before performing an analysis.
- Audio CD uses a sampling rate of 44.1 kHz.
- Many A-to-D systems only operate at fractions of this rate (22050 Hz, 11025 Hz).

Aliasing

- Components of a signal which are above the Nyquist frequency are misrepresented as lower frequency components (**aliasing**).
- To avoid aliasing, a signal must be filtered to eliminate frequencies above the Nyquist frequency.
- Since practical filters are not infinitely sharp, this will attenuate energy near to the Nyquist frequency also.

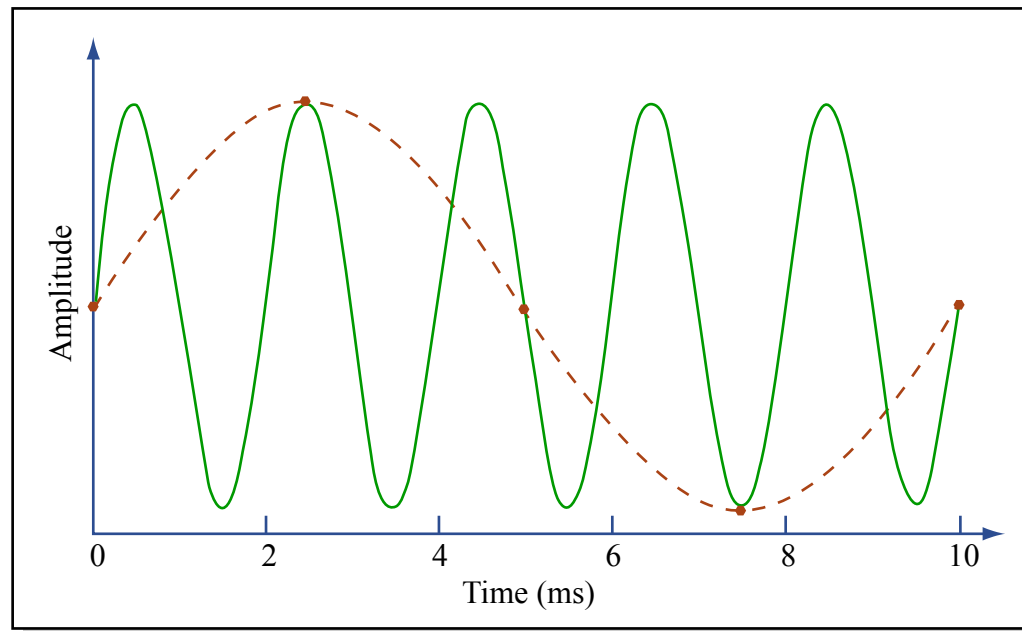


Figure by MIT OpenCourseWare. Adapted from Johnson, Keith.
Acoustic and Auditory Phonetics. Malden, MA: Blackwell Publishers, 1997. ISBN: 9780631188483.

Quantization

- The amplitude of the signal at each sampling point must be specified digitally - quantization.
- Divide the continuous amplitude scale into a finite number of steps. The more levels we use, the more accurately we approximate the analog signal.

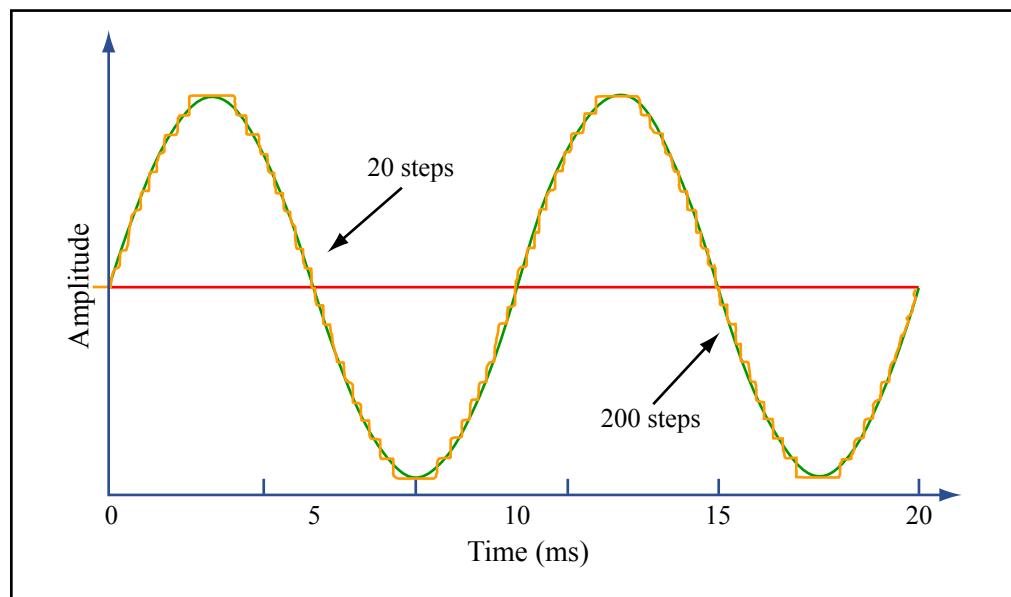


Figure by MIT OpenCourseWare. Adapted from Johnson, Keith.
Acoustic and Auditory Phonetics. Malden, MA: Blackwell Publishers, 1997. ISBN: 9780631188483.

Quantization

- The number of levels is specified in terms of the number of bits used to encode the amplitude at each sample.
 - Using n bits we can distinguish 2^n levels of amplitude.
 - e.g. 8 bits, 256 levels.
 - 16 bits, 65536 levels.
- Now that memory is cheap, speech is almost always digitized at 16 bits (the CD standard).

Quantization

- Quantizing an analog signal necessarily introduces quantization errors.
- If the signal level is lower, the degradation in signal-to-noise ratio introduced by quantization noise will be greater, so digitize recordings at as high a level as possible without exceeding the maximum amplitude that can be represented (clipping).
- On the other hand, it is essential to avoid clipping.

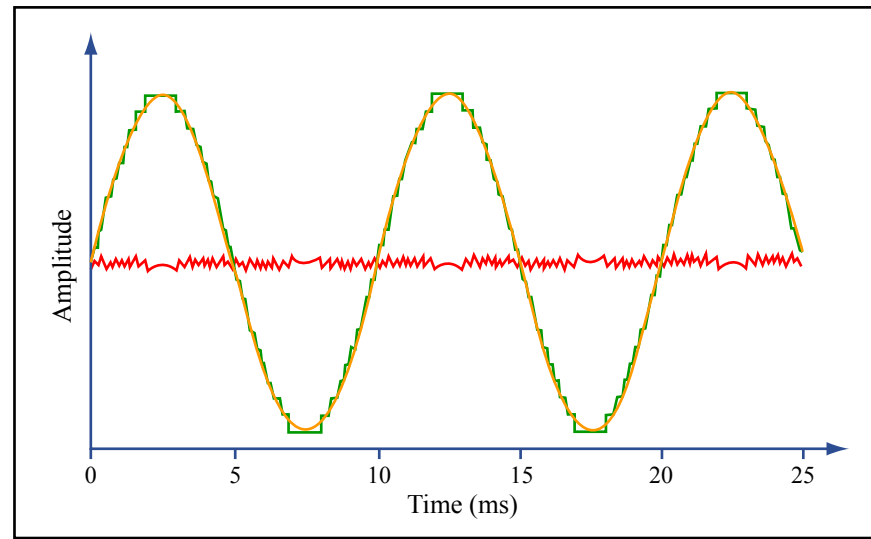





Figure by MIT OpenCourseWare. Adapted from Johnson, Keith.
Acoustic and Auditory Phonetics. Malden, MA: Blackwell Publishers, 1997. ISBN: 9780631188483.

Voicing and aspiration

- Many languages make a contrast between two sets of stops with different laryngeal properties, loosely referred to as ‘voiced’ and ‘voiceless’.
- The precise details of these laryngeal contrasts differ from language to language.
- Some broad distinctions:
 - voiced [b]: vocal fold vibration during closure
 - **bal** (‘hair’) 
 - voiceless unaspirated [p]: no vibration of the vocal folds, short VOT
 - **pal** (‘take care of’) 
 - voiceless aspirated [p^h]: no vibration of the vocal folds, long VOT (high airflow after release)
 - **p^hal** (‘knife blade’) 

Listen to all three sound files [here](#).

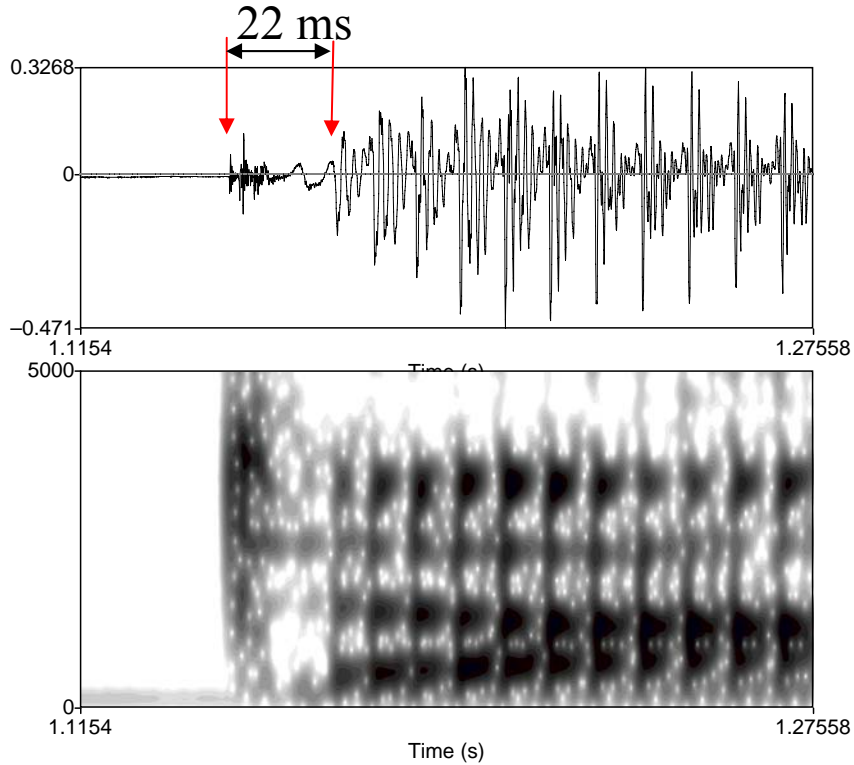
Voicing and aspiration

- Voiced vs. voiceless [b vs. p]
 - Russian, French, Dutch
- Unaspirated vs. aspirated [p vs. p^h]
 - Mandarin, Cantonese
- Voiced vs. voiceless unaspirated vs. aspirated [b vs. p vs. p^h]
 - Hindi, Thai
- English shows contextual variation between voicing and aspiration.

Voice Onset Time

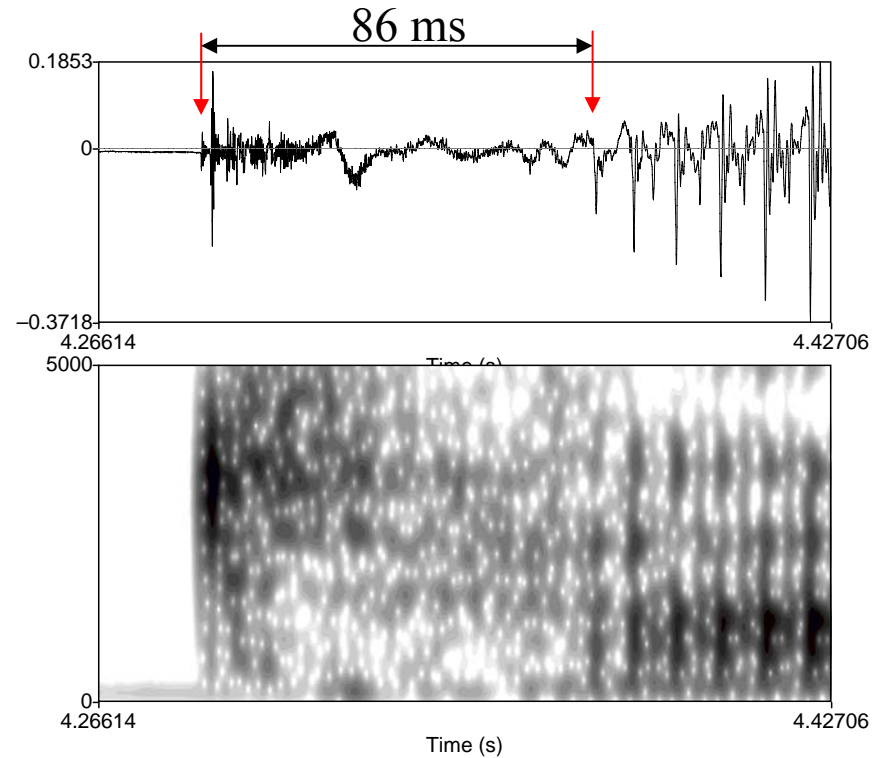
- English utterance-initial stops

Voiceless unaspirated



die

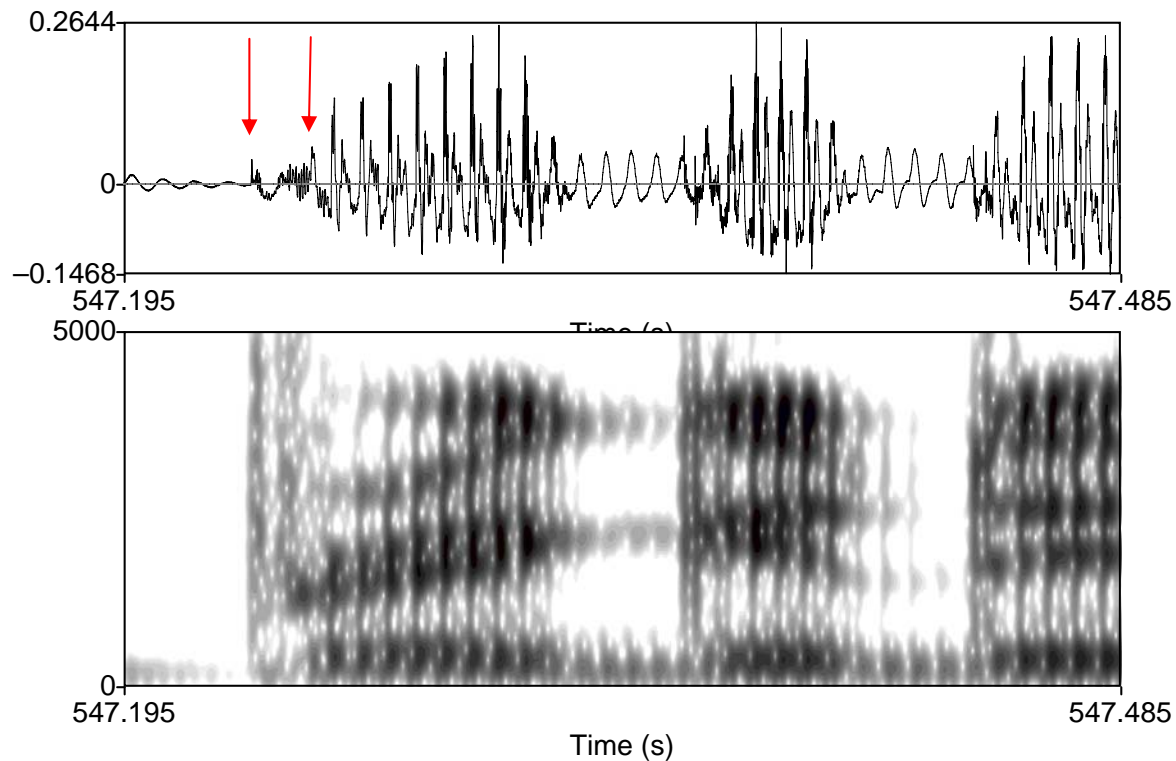
Voiceless aspirated



tie

VOT, closure voicing

- English intervocalic stops can be fully voiced
 - VOT is 0 ms in 2nd and 3rd stops



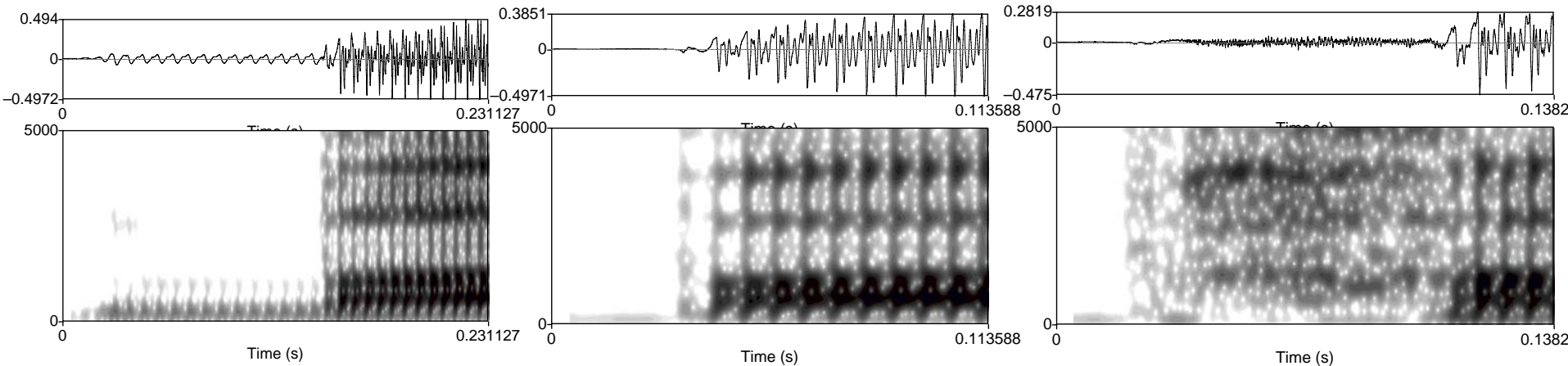
brigadoo(n)

VOT, closure voicing

- Hindi - three-way contrast

- recordings from Ladefoged

- <http://www.phonetics.ucla.edu/vowels/chapter12/hindi.html>



bal

pal

p^hal

‘hair’

‘take care of’

‘knife blade’



Listen to all three sound files [here](#).