COMP 546

Lecture 19

Sound 2: frequency analysis

Tues. March 27, 2018

Speed of Sound

Sound travels at about 340 m/s, or 34 cm/ ms.

(This depends on temperature and other factors)

Wave equation

$$Pressure = \ I_{atm} + I(X, Y, Z, t)$$

$I(X, Y, Z, t)$ is not an arbitrary function.

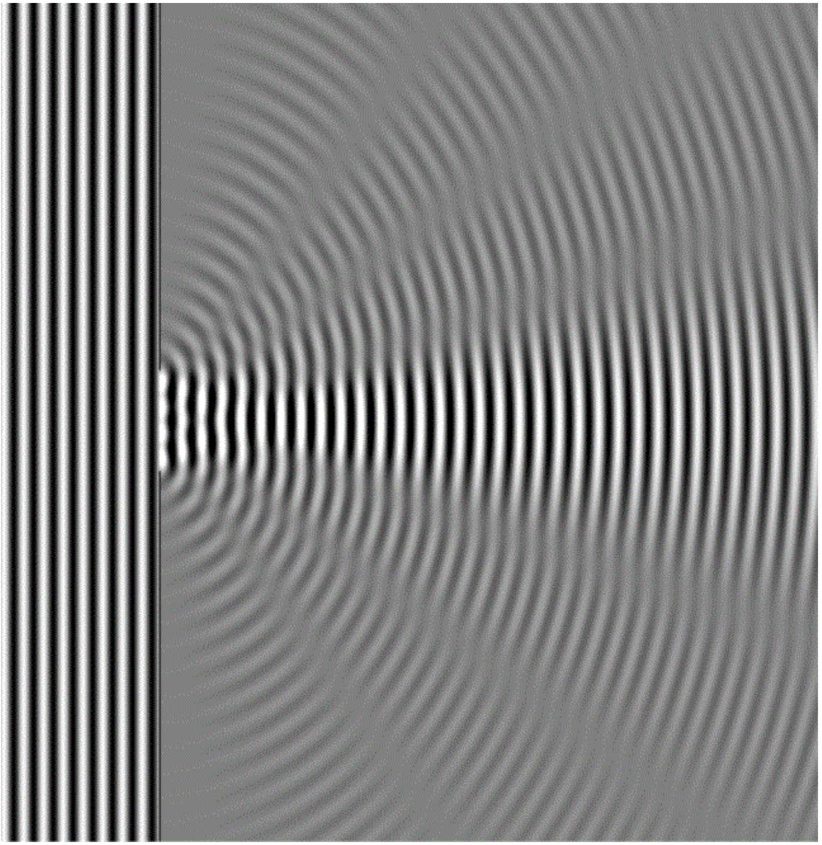
Rather:

$$\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2} + \frac{\partial^2}{\partial Z^2}\right) I(X, Y, Z, t) \ = \ \frac{1}{v^2}\frac{\partial^2}{\partial t^2} I(X, Y, Z, t)$$

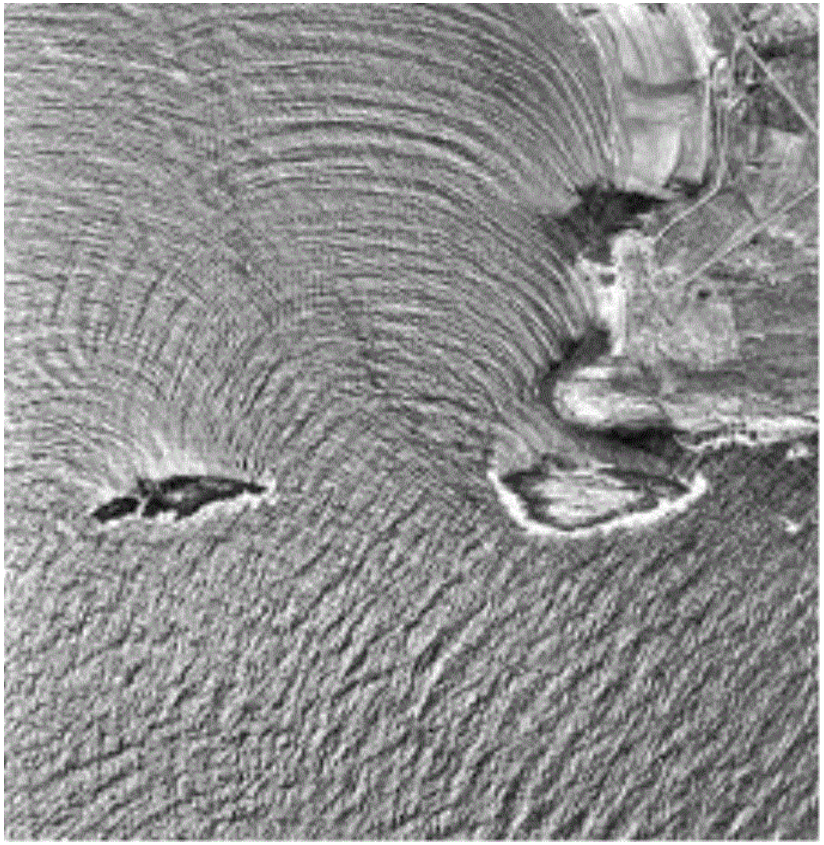
v = 340 m/s

The wave equation + boundary conditions give complicated shadow and reflection effects.

What happens when sound enters the ear ?



plane wave + single slit



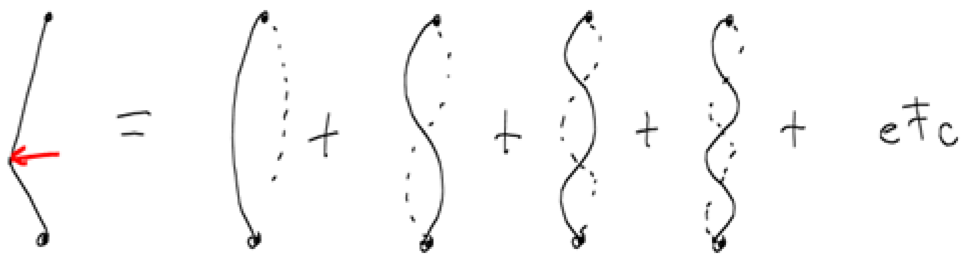
sea waves + islands

Musical sounds

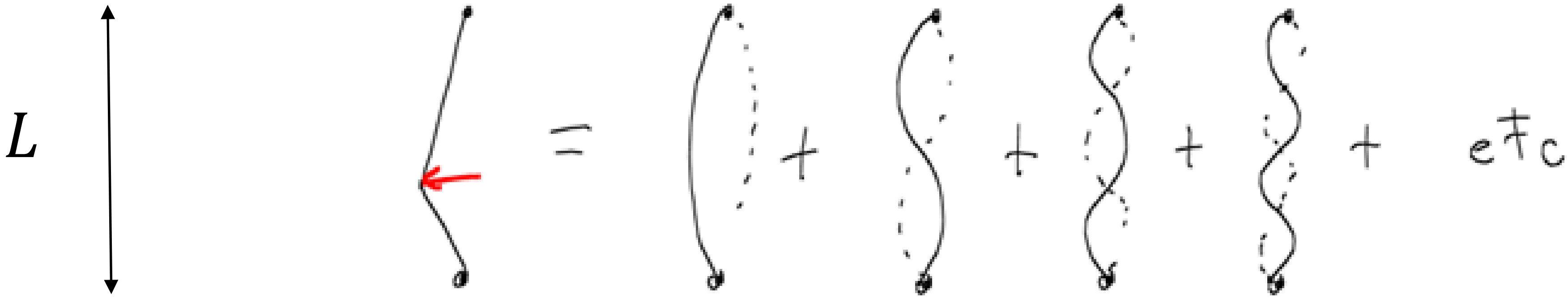
(brief introduction)

Example: guitar

Write one string displacement at t = 0 as sum of sines.



Modes are $\sin(\frac{\pi}{L}\, jx)$ where L is the length of the string, j is an integer.

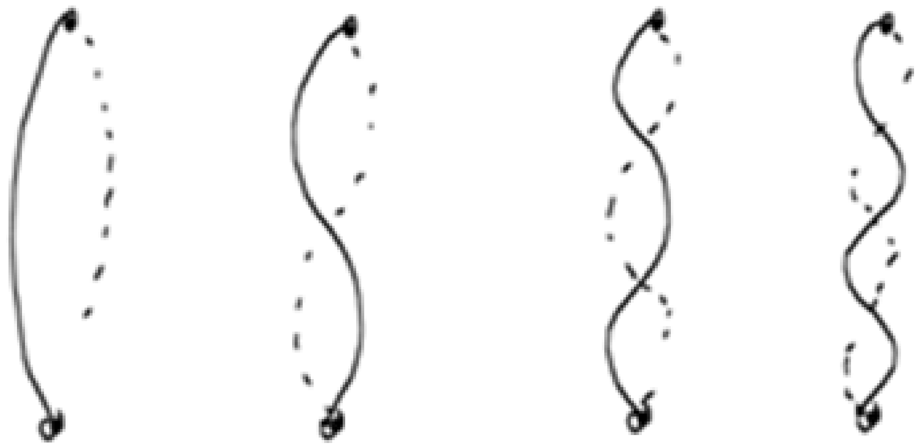


Physics says:

$$\omega = \frac{c}{L}$$

where constant c depends on physical properties of string (mass density, tension)

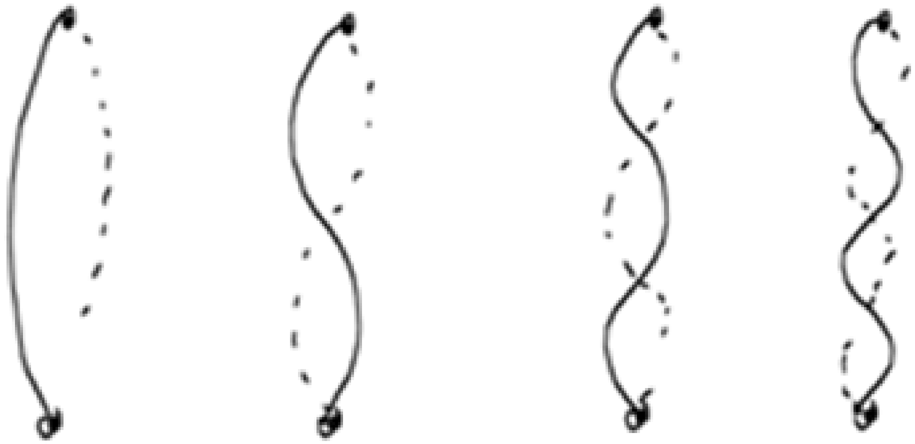
Modes of a vibrating string each have fixed points which reduce the effective length.



$L \qquad \frac{L}{2} \qquad \frac{L}{3} \qquad \frac{L}{4}$

Physics says:

$$\omega = \frac{c}{L} \qquad \frac{2c}{L} \qquad \frac{3c}{L} \qquad \frac{4c}{L}$$



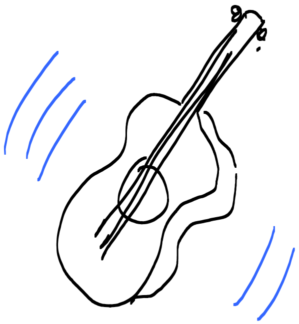
$$\omega = \frac{c}{L} \qquad \frac{2c}{L} \qquad \frac{3c}{L} \qquad \frac{4c}{L}$$

ω_0

“fundamental”
(1st harmonic)

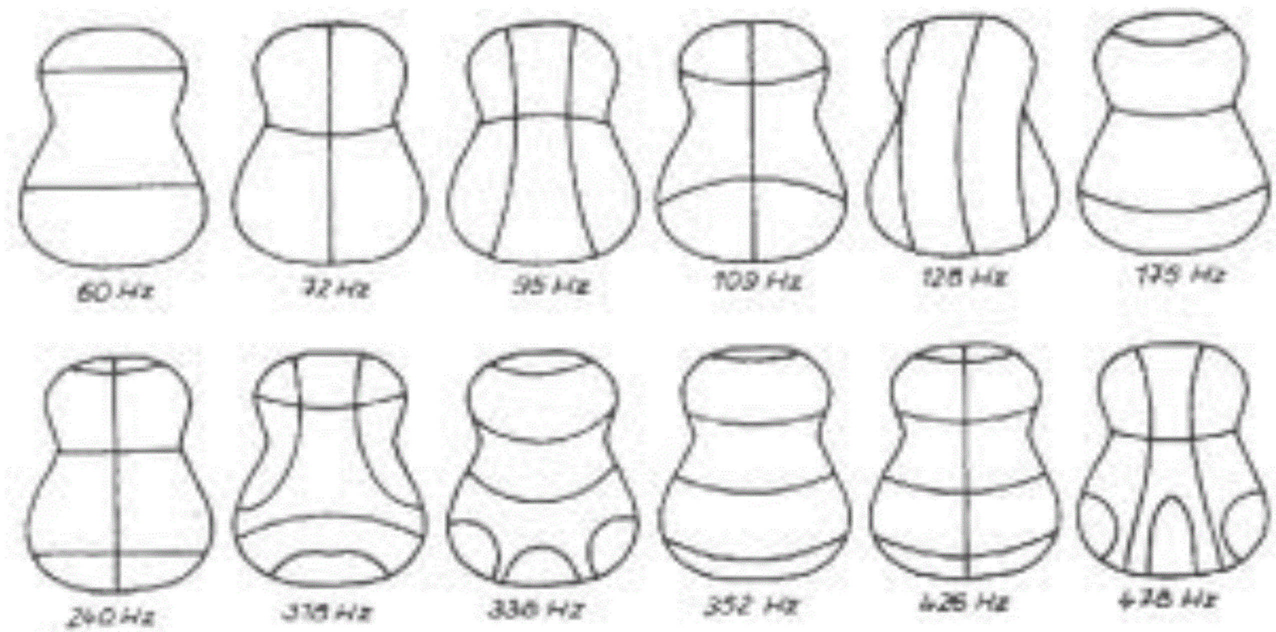
“overtones”

The temporal frequency $m\,\omega_0$ is called the m-th harmonic.



For stringed instruments, most of the sound is produced by vibrations of the instrument body (neck, front and back plates).
http://www.acs.psu.edu/drussell/guitars/hummingbird.html

The lines in the sketches below are the nodal points. They don't move.



These are vibration *modes*, not harmonics.   The guitar sound is a sum of these modes.

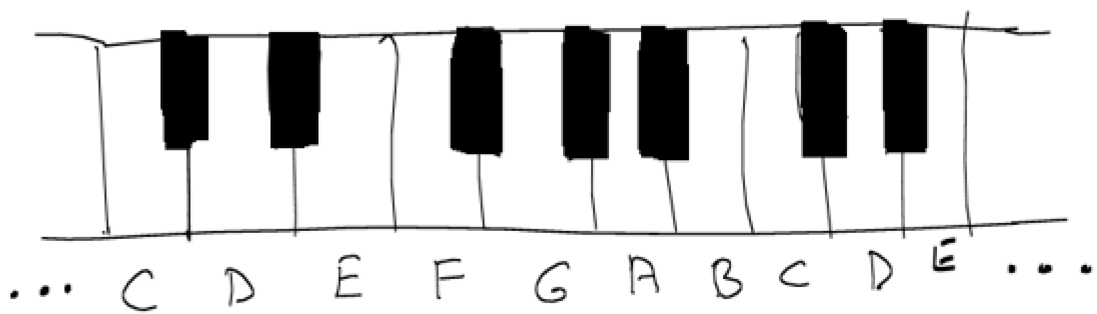
Difference of two frequencies ω_1 and ω_2 :

$$log_2 \frac{\omega_2}{\omega_1} \quad \text{octaves.}$$

e.g. 1 octave is a doubling of frequency.

(Western) Musical Notes

Each “octave” ABCDEFGA is divided into 12 “semitones”, separated into 1/12 octave.



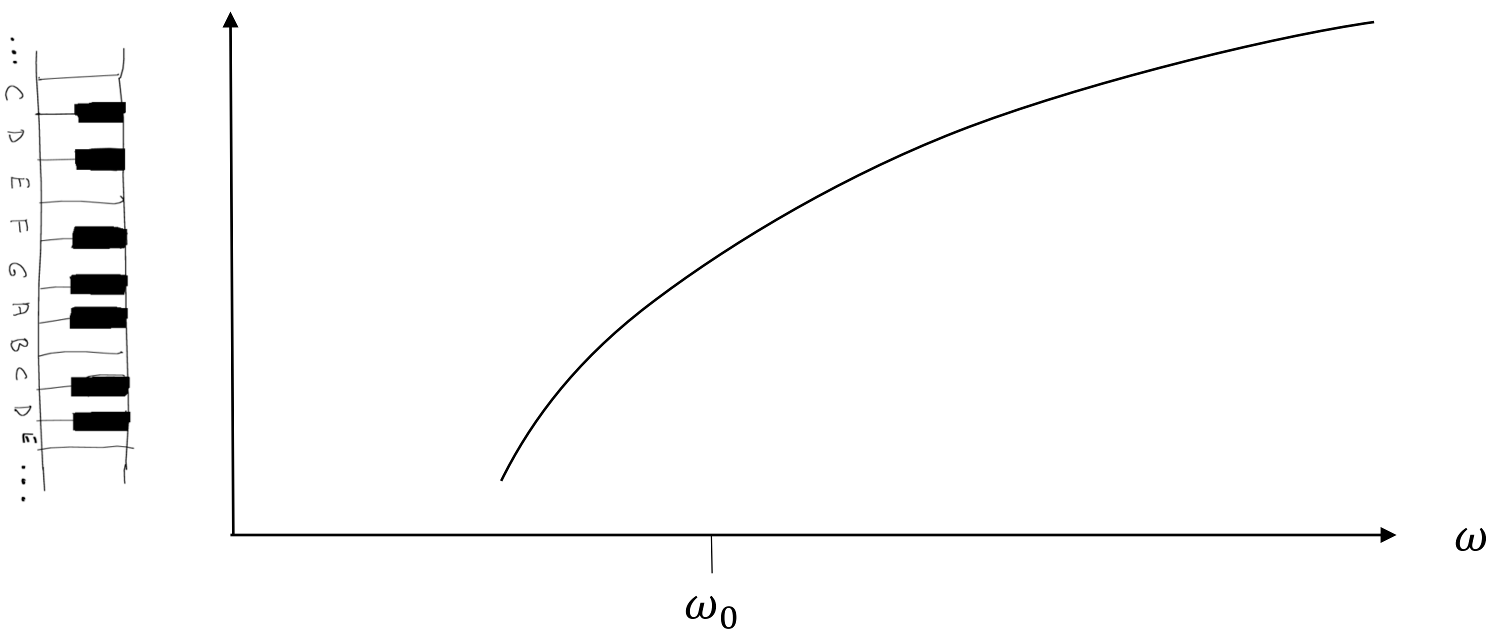
C-D, D-E, F-G, G-A, A-B are two semitones each

E-F, B-C are one semitone each.

Q: How many semi-tones are there from ω_0 to ω ?

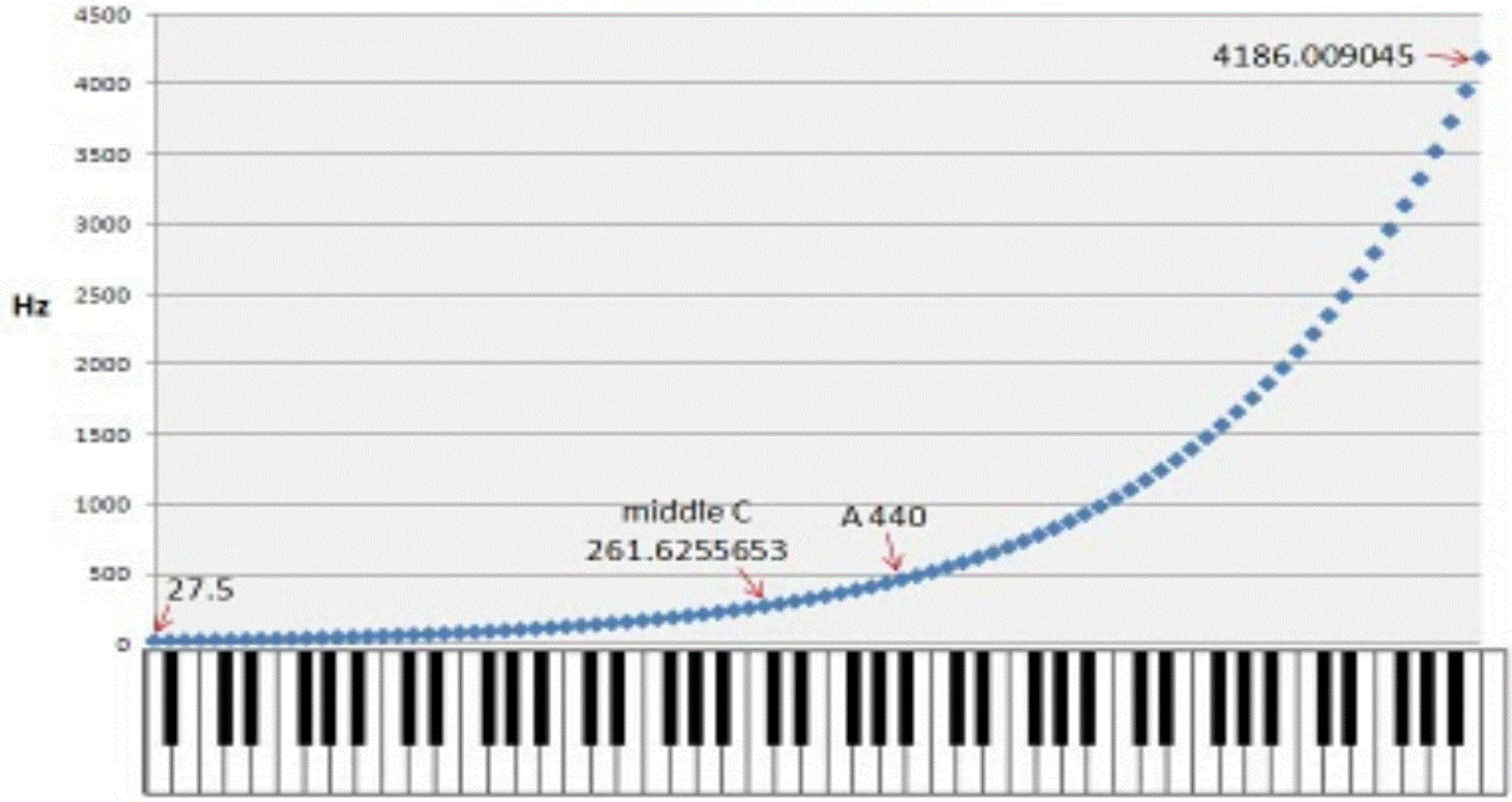
Q: How many semi-tones are there from ω_0 to ω ?

A: $$12\ log_2 \frac{\omega}{\omega_0}$$



Fundamental frequency of note

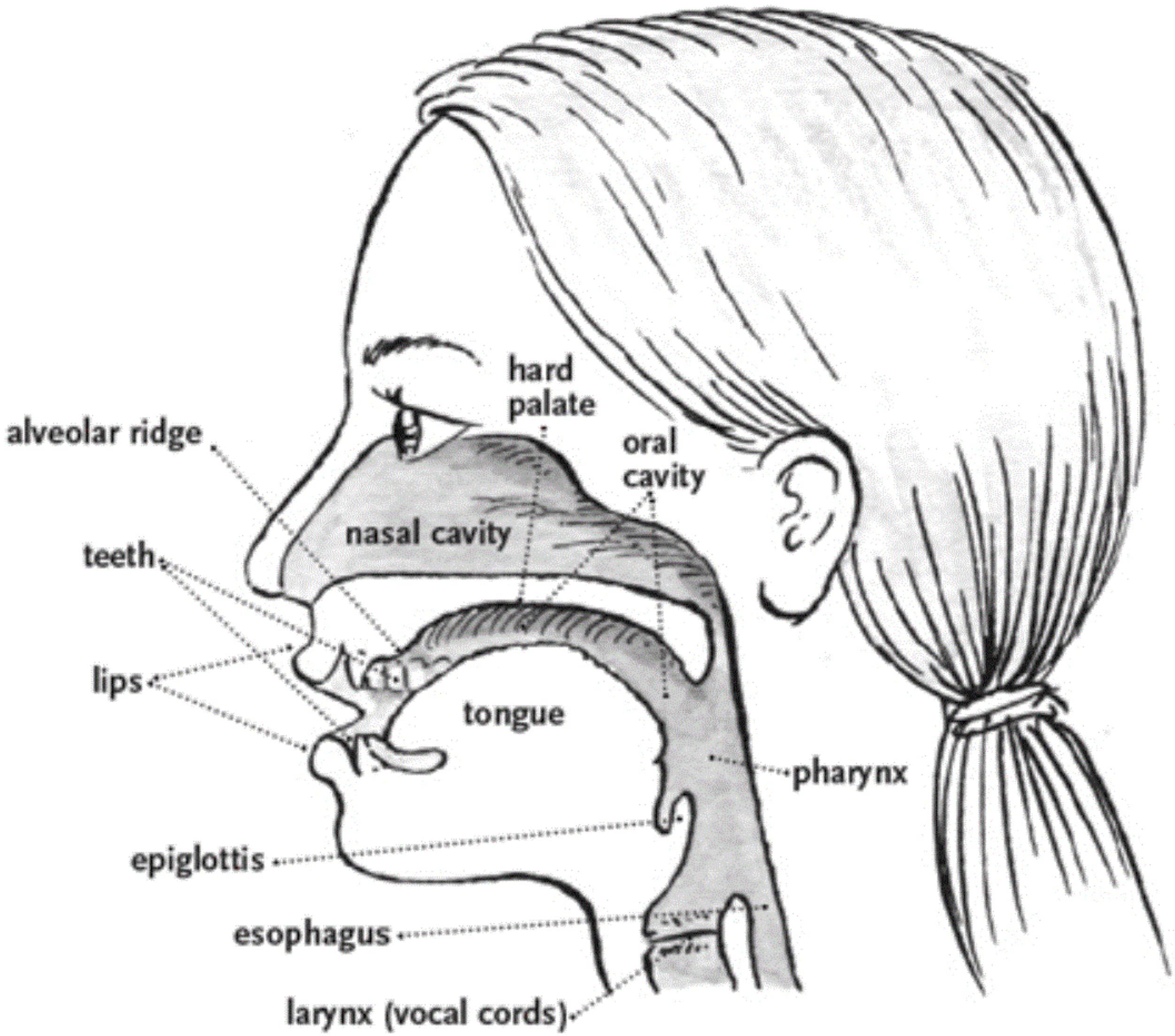
88 fundamental frequencies (Hz) on a keyboard



The fundamental frequencies of successive notes define a *geometric progression.*

This is different from the harmonics of a vibrating string which define an *arithmetic progression*.

Speech Sounds



What determines speech sounds?

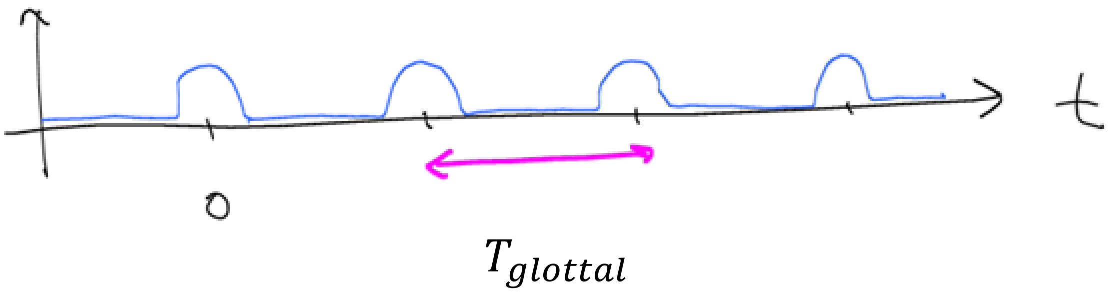
- voiced vs. unvoiced

  ‘zzzz’ vs. ‘ssss’,   ‘vvvv’ vs. ‘ffff’

- articulators   (jaw, tongue, lips)

  ‘aaaa’, ‘eeee’, ‘oooo’, …

Voiced sounds are produced by "glottal pulses".

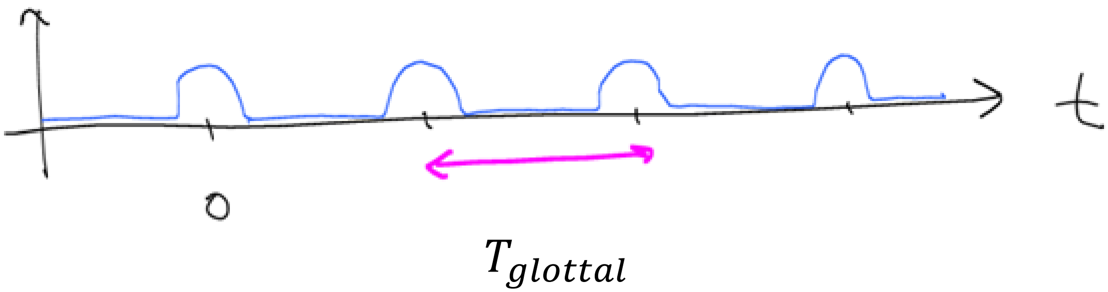


$$\sum_{j=0}^{n_{glottal}} g\left(t - j\, T_{glottal}\right)$$

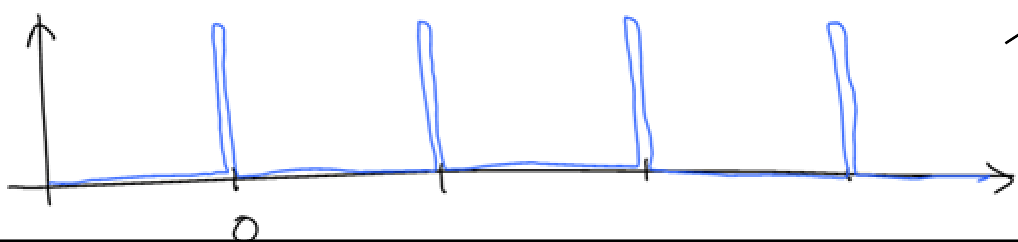
Exercise 16 Q7.

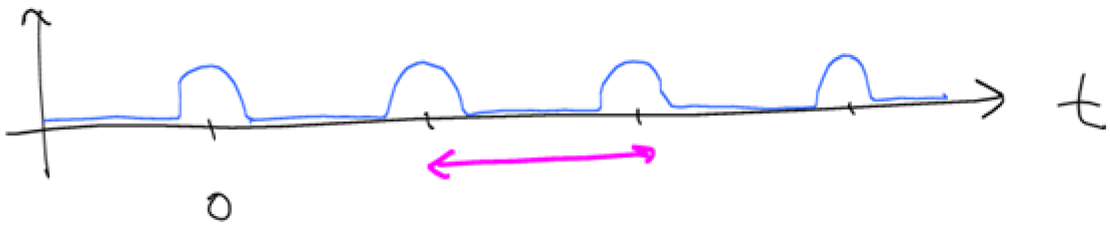
$$g(t - t_0) = g(t) * \delta(t - t_0)$$

Voiced sounds are produced by “glottal pulses”.



$$\sum_{j=0}^{n_{glottal}} g(t - j\,T_{glottal}) = g(t) * \sum_{j=0}^{n_{glottal}} \delta(t - j\,T_{glottal})$$





$$\sum_{j=0}^{n_{glottal}} g(t - j\,T_{glottal})$$

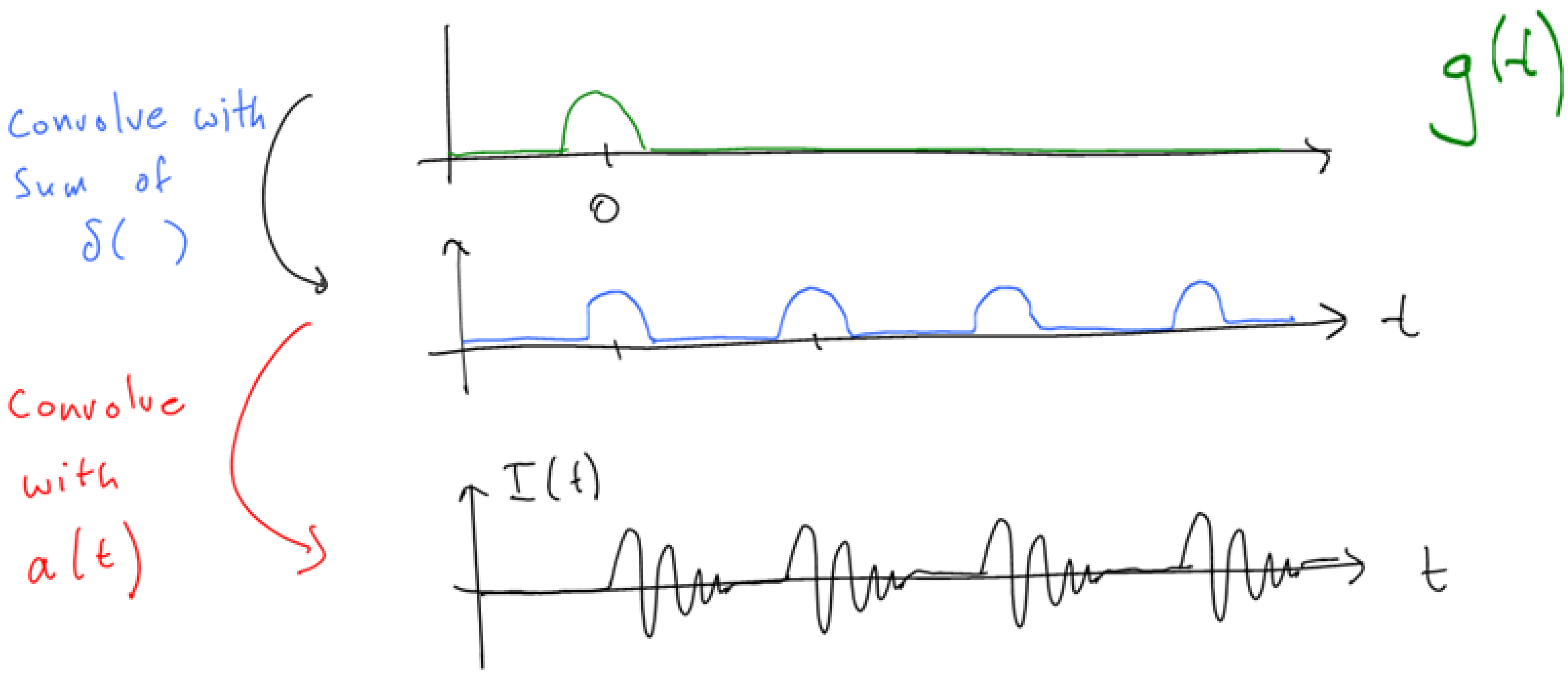
decrease $T_{glottal}$ by increasing tension in vocal cords

≡ increase frequency of pulses

Let $a(t)$ be the impulse response function of the articulators.

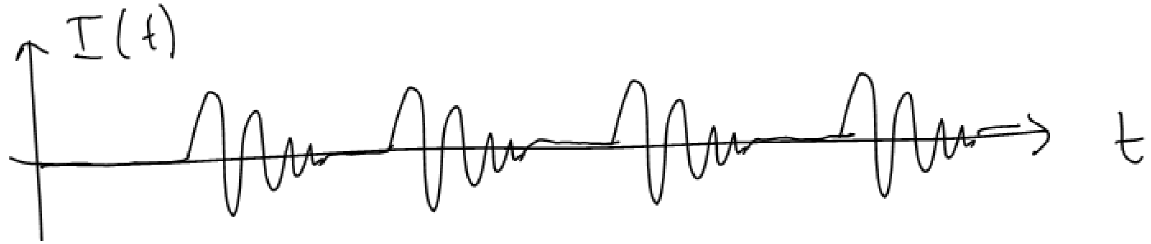
(jaw, tongue,lips)

$$I(t) = a(t) * g(t) * \sum_{j=0}^{n_{glottal}} \delta(t - j\,T_{glottal})$$



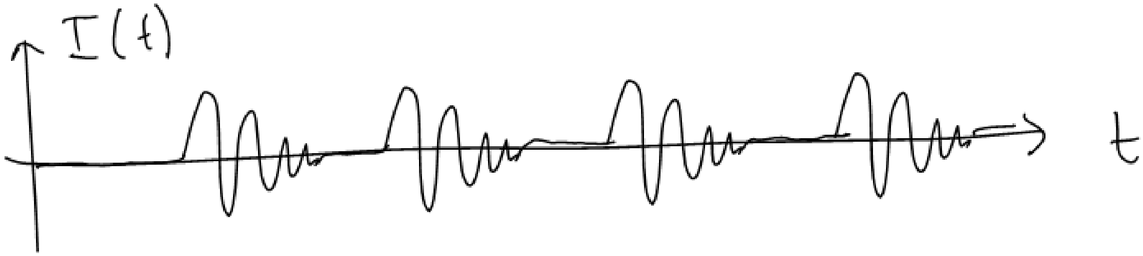
Q: What is the Fourier transform of

$$I(t) = a(t) * g(t) * \sum_{j=0}^{n_{pulse}-1} \delta(t - jT_g) \; ?$$



Q: What is the Fourier transform of

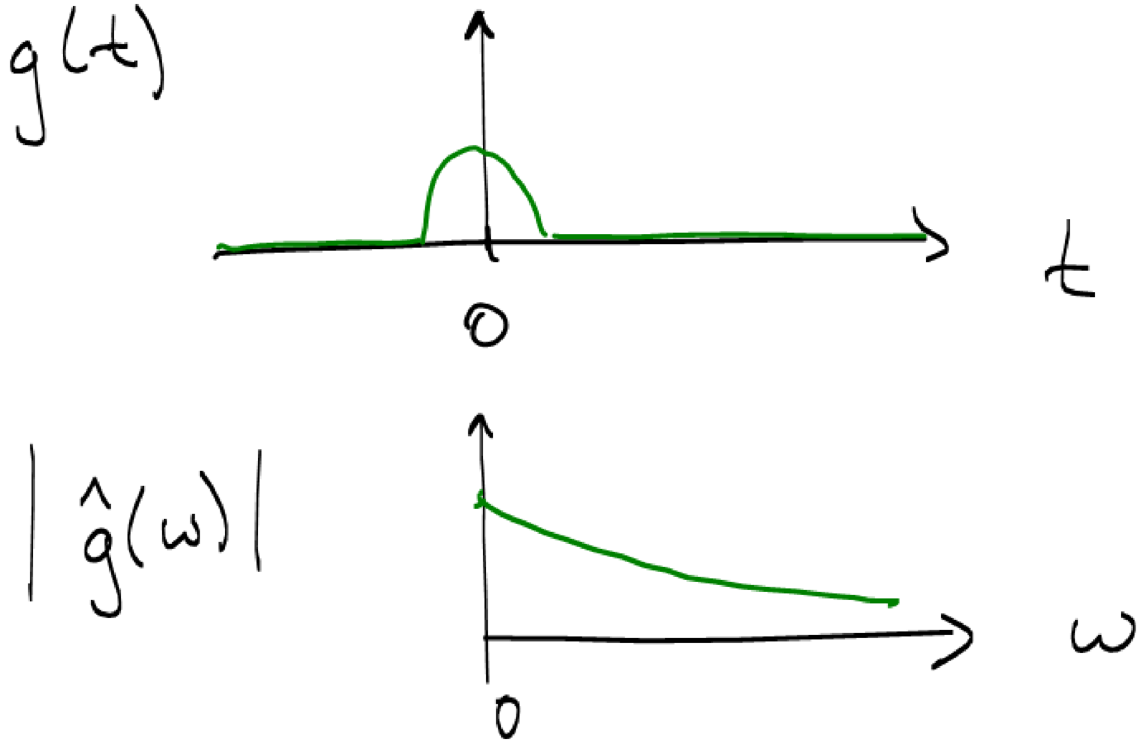
$$I(t) = a(t) * g(t) * \sum_{j=0}^{n_{pulse}-1} \delta(t - jT_g) \ ?$$



A:

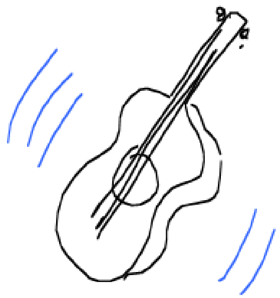
$$\hat{I}(\omega) = \hat{a}(\omega) \cdot \hat{g}(\omega) \cdot \underbrace{F \sum_{j=0}^{n_{pulse}-1} \delta(t - jT_g)}_{?}$$

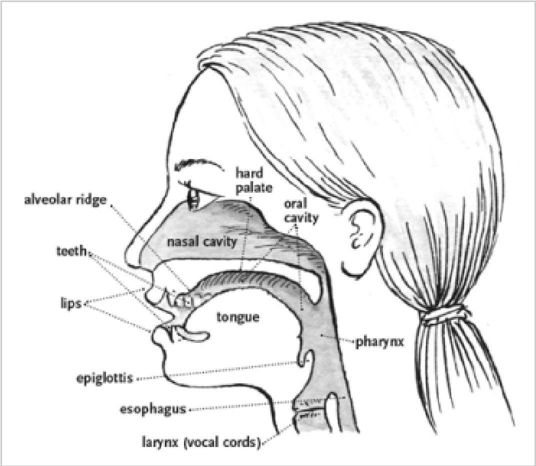
Glottal Pulse shape



Glottal pulse $g(t)$ has a shape roughly between a Gaussian and impulse.

What about $a(t)$?

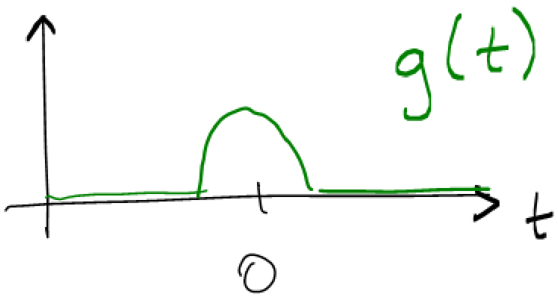


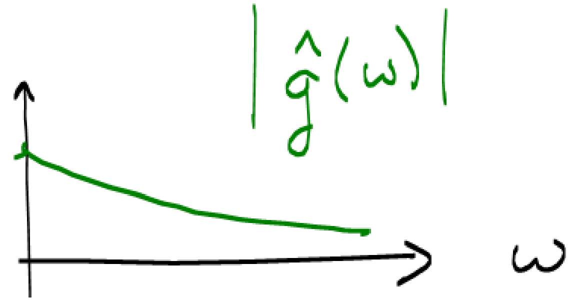


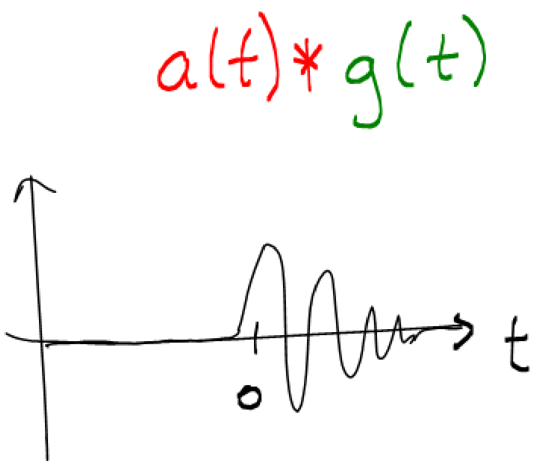
Oral and nasal cavity have resonant modes of vibration, like air cavity in guitar does.

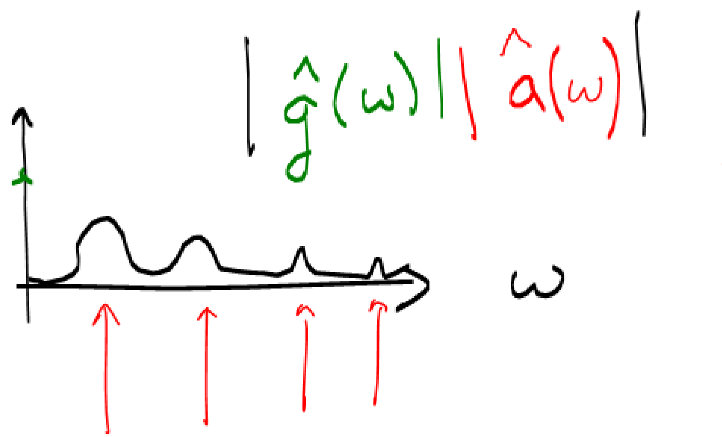
Time domain

Temporal frequency domain



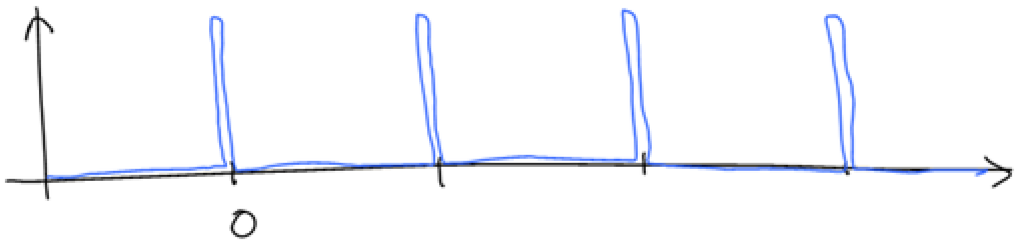






Peaks are called “formants”

$$\mathbf{F} \sum_{j=0}^{n_{glottal}} \delta(t - j\, T_{glottal}) \;=\; ?$$



T_g is the period of the glottal pulse train.

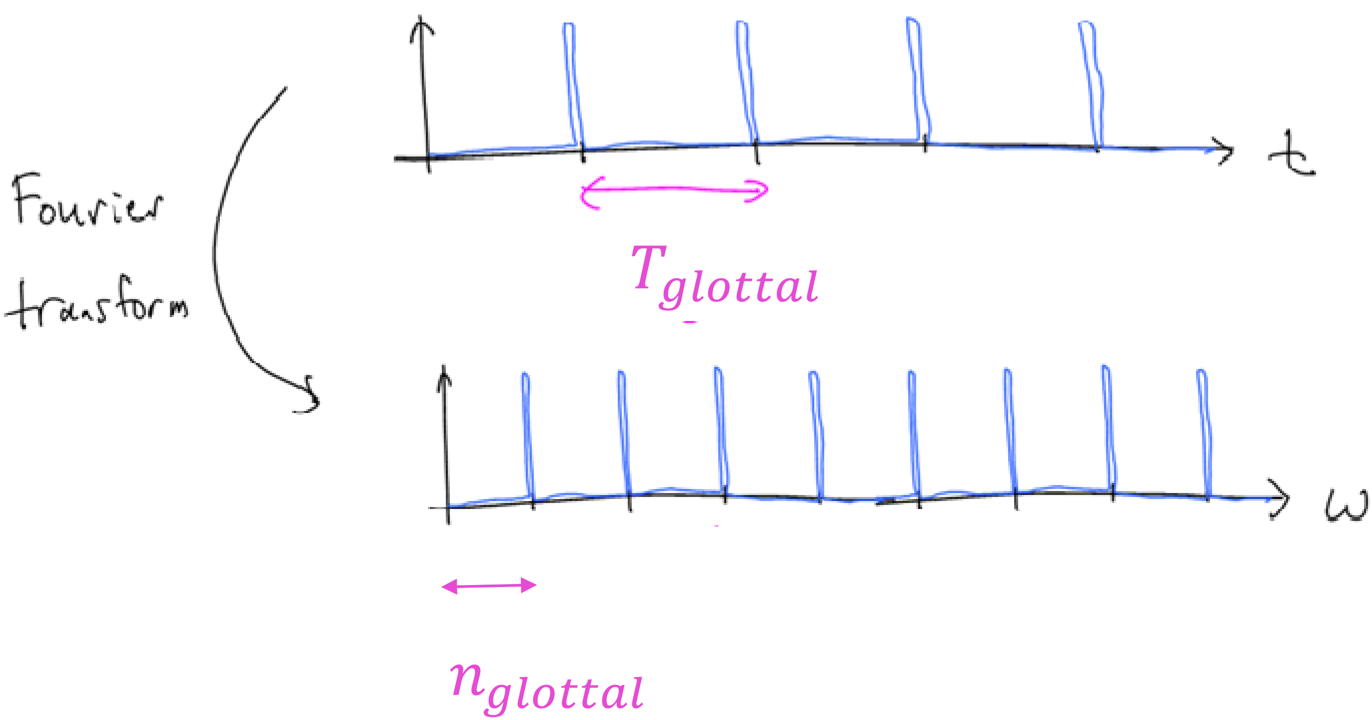
The pulse train has $n_{glottal}$ pulses in T time steps, i.e.

$$T_{glottal}\ n_{glottal} = T.$$

Assume that the Fourier transform is taken over T samples.

Assignment 3: Show

$$\mathbf{F} \sum_{j=0}^{n_{glottal}-1} \delta(t - j\,T_{glottal}) = n_{glottal} \sum_{m=0}^{T_{glottal}-1} \delta(\omega - m\;n_{glottal})$$



Units of temporal frequency ω

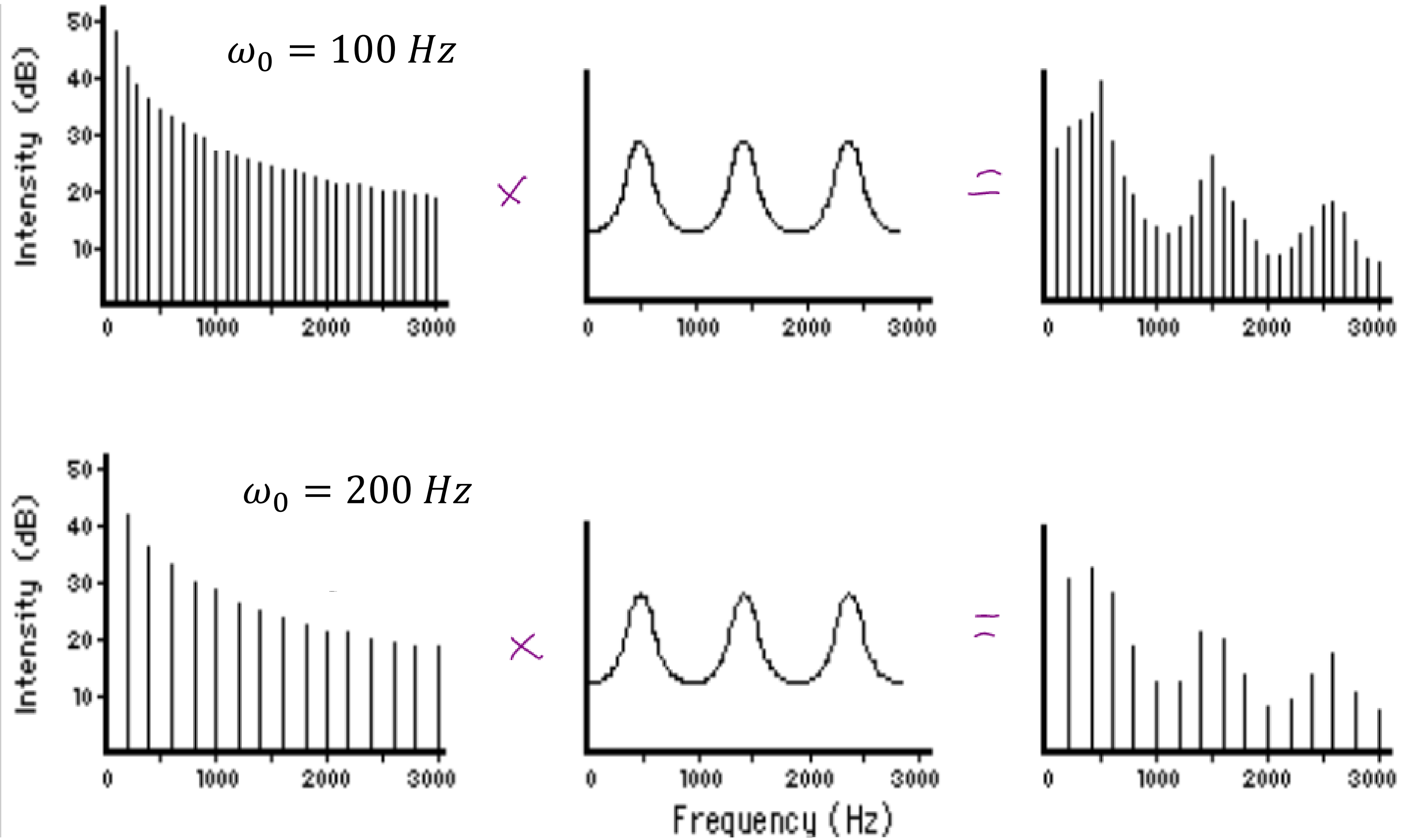
$T_{glottal}$ is the period of the glottal pulse train.
$n_{glottal}$ pulses in T time samples.

To convert $n_{glottal}$ to 'pulses per second', we divide T (to get pulses per sample) and then multiply by 'time samples per second'. High quality audio uses 44,100 samples per second.

$n_{glottal}$ is the fundamental frequency of the voiced sound. It determines the "pitch".

Adult males : 100-150
Adult females : 150-250 Hz
Children: over 250 Hz

glottal pulse spectrum | “formants” | sound spectrum



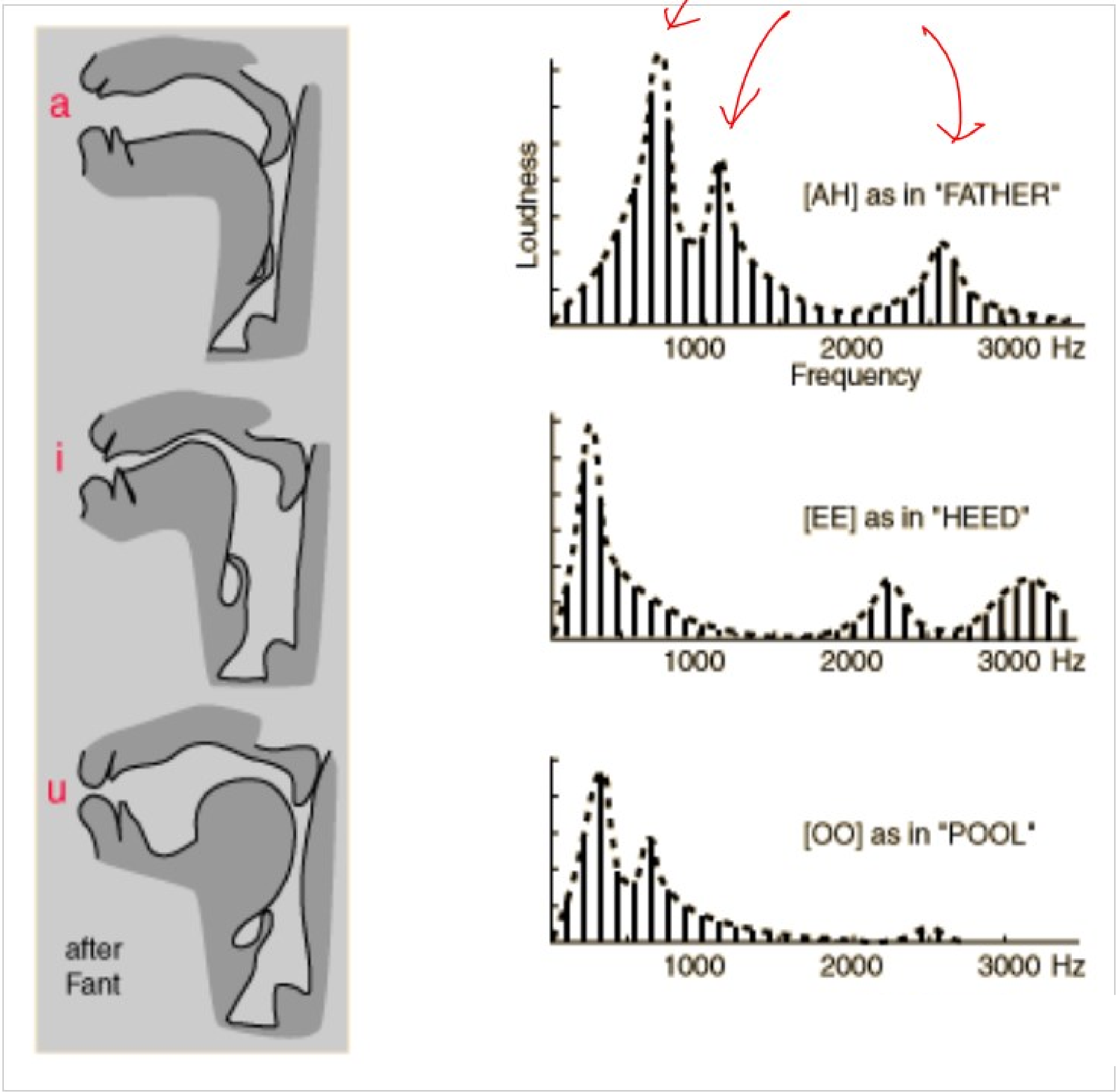
Voiced vowel sounds

formants

a

i

u



Unvoiced sounds

noise instead of glottal pulses

$$I(t) = a(t) * n(t)$$

Unvoiced sounds

noise instead of glottal pulses

$$I(t) = a(t) * n(t)$$

$$\hat{I}(\omega) = \hat{a}(\omega)\,\hat{n}(\omega)$$

Flat amplitude spectrum on average ('white noise')

Consonants

Restrict flow of air by moving tongue, lips into contact with the teeth & palate.

Fricatives

- voiced   z, v,   zh, th (the)
- unvoiced   ?

Stops

- voiced   b, d, g
- unvoiced   ?

Nasals (closed mouth)

-      m, n, ng

Consonants

Restrict flow of air by moving tongue, lips into contact with the teeth & palate.

Fricatives

- voiced z, v, zh, th (the)
- unvoiced s, f, sh, th (theta)

Stops

- voiced b, d, g
- unvoiced p, t, k

Nasals (closed mouth)

- m, n, ng

I did not have time to cover the following slides properly.

I will present them again in lecture 22.

Spectrogram

Partition a sound signal into B blocks of T samples each (i.e. the sound has BT samples in total).

Take the Fourier transform of each block.

Spectrogram

Partition a sound signal into B blocks of T samples each (i.e. the sound has BT samples in total).

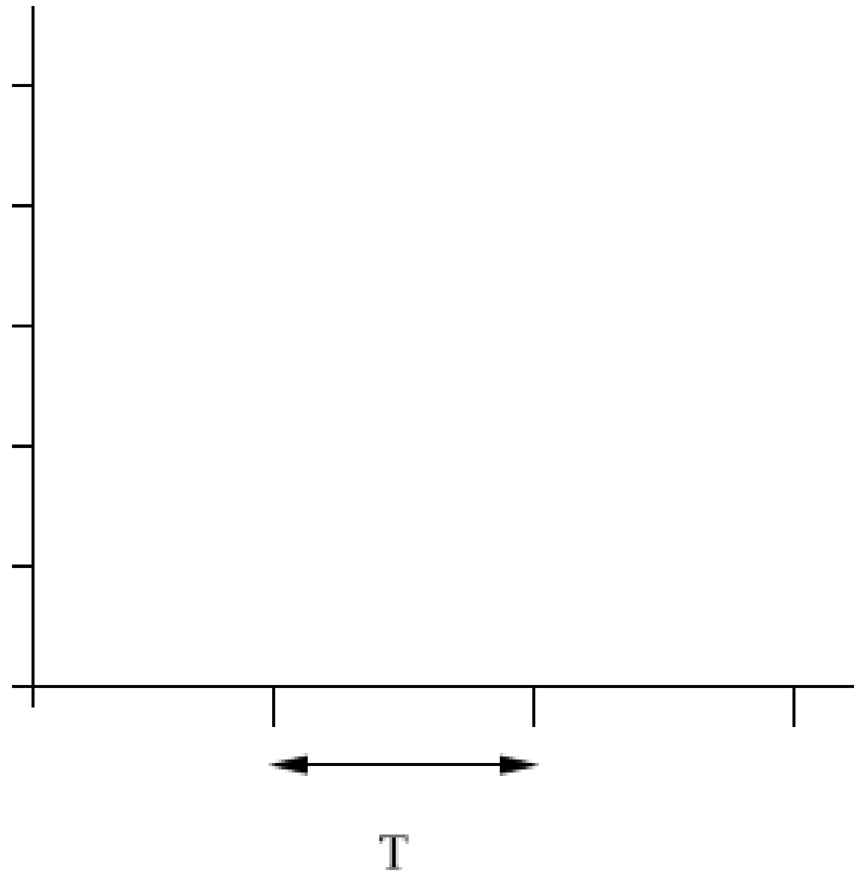
Take the Fourier transform of each block.

Let b be the block number, and ω units be cycles per block.

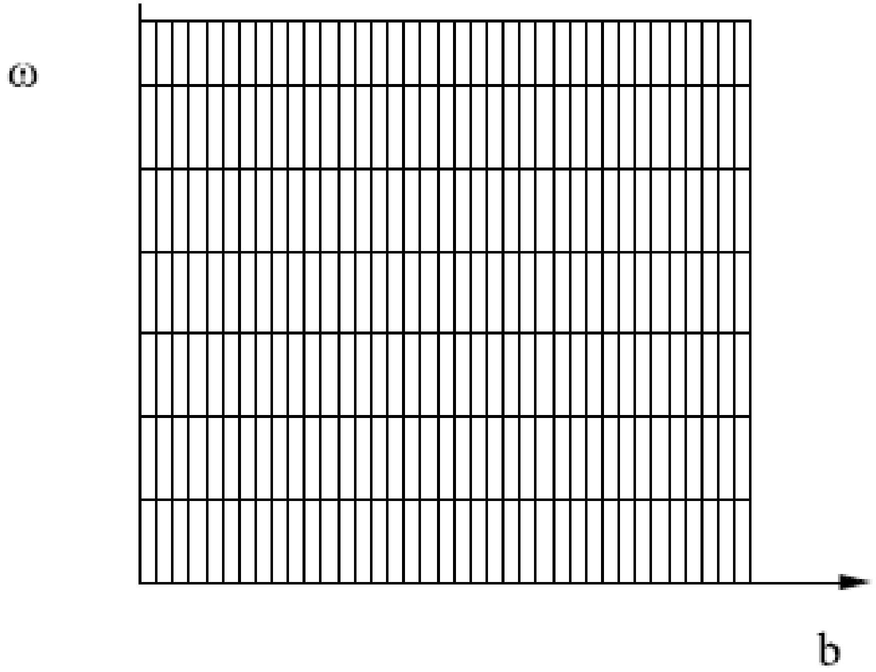
$$\hat{I}(b, \omega) = \sum_{t=0}^{T-1} I(\, b\, T + t)\; e^{-i \frac{2\pi}{T} \omega t}$$

Cycles per second (Hz)

$$\omega_0 = \frac{44000}{T}$$

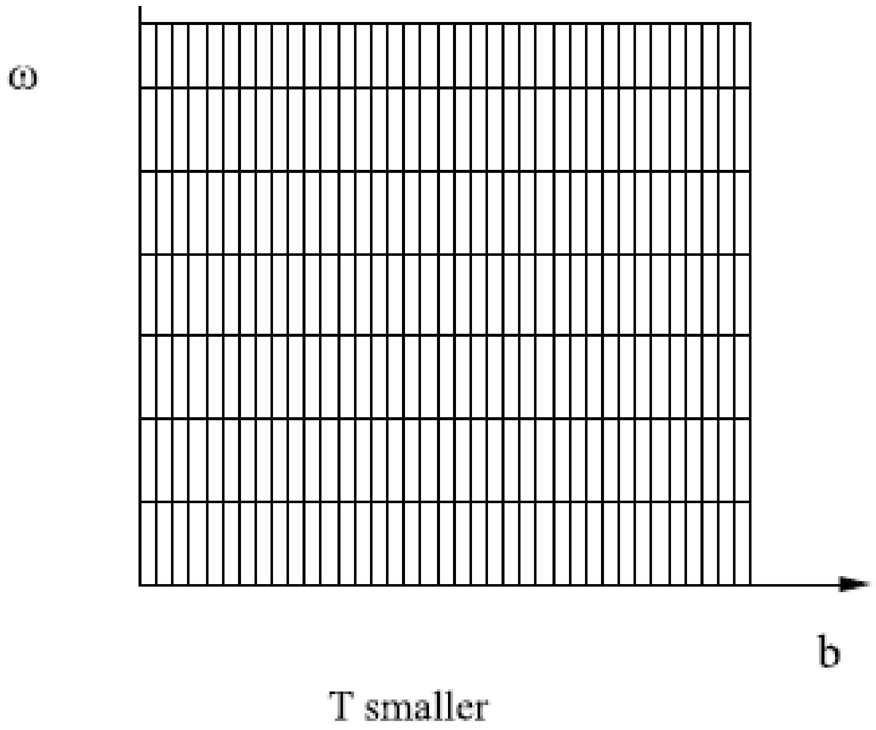


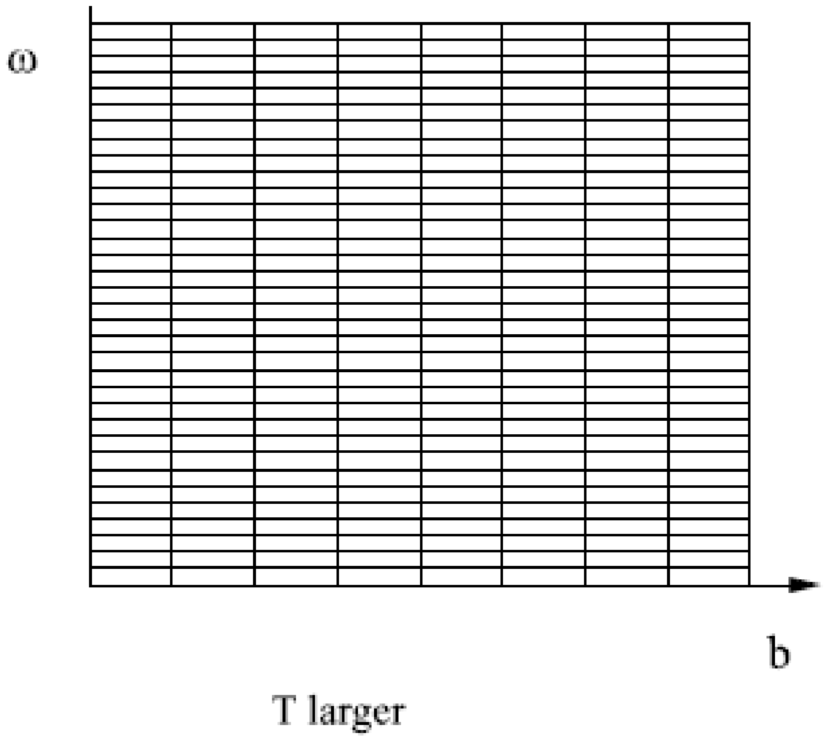
Time (samples)



T smaller

e.g. T = 512 samples (12 ms), ω_0 = 86 Hz

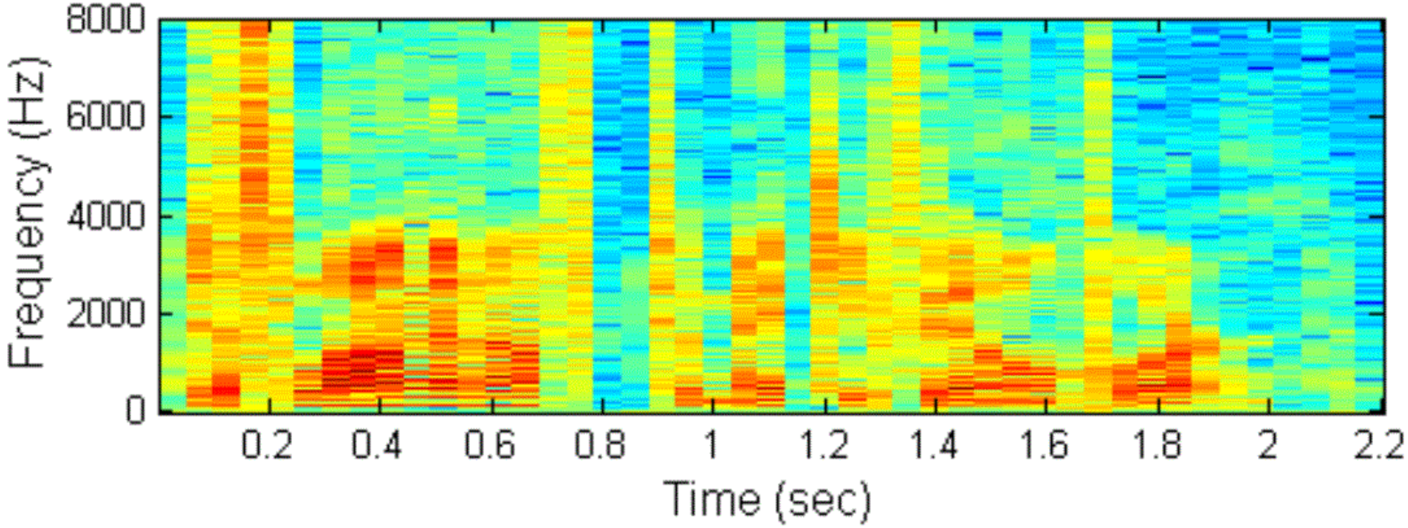




e.g. T = 512 samples (12 ms), ω_0 = 86 Hz

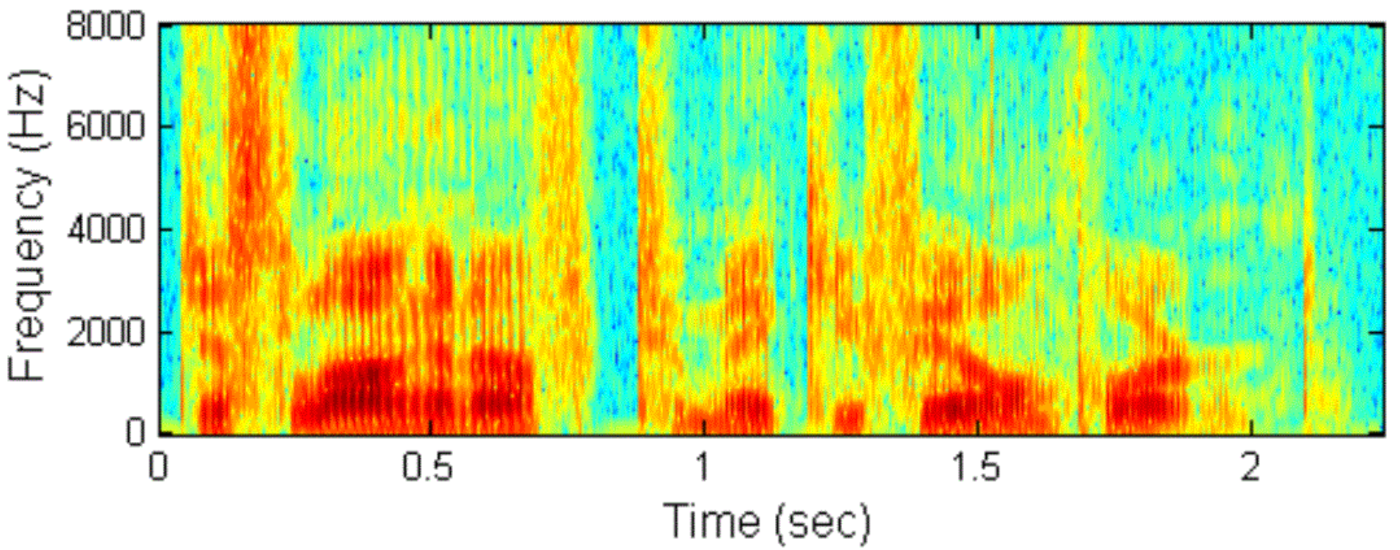
T = 2048 samples (48 ms), ω_0 = 21 Hz

You cannot simultaneously localize the frequency and the time. This is a fundamental tradeoff. We have seen it before (recall the Gaussian).



Narrowband

(good frequency resolution, poor temporal resolution … ~50ms)



Wideband

(poor frequency resolution, good temporal resolution)

Examples: Spectrograms of 10 vowel sounds