Perception & Multimedia Computing

Week 11 - Signals as Sinusoids

Michael Zbyszyński
Lecturer, Department of Computing
Goldsmiths University of London

https://goo.gl/Zfb9NK

Please fill in the survey at the link above

Key challenges in computing with audio

- How do we experience (see, hear) a piece of media (audio, image, video, etc.)?
- How can we efficiently create media to have a desired effect?
- How can we efficiently process media to have a desired effect?
- How can we efficiently analyze media to understand its contents?
- How can we represent media in a computer for storage, transmission, processing, analysis, etc.?

Becoming a sound ninja...

https://www.scientificamerican.com/article/a-learning-secret-don-t-take-notes-with-a-laptop/

SUBSCRIBE

SCIENTIFIC AMERICAN

English Cart 0 Sign In | Register

THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS STORE

MIND

A Learning Secret: Don't Take Notes with a Laptop

Students who used longhand remembered more and had a deeper understanding of the material

By Cindi May on June 3, 2014 26 **Véalo en español**

READ THIS NEXT

The Science of Education: Back to School

Signals as Sinusoids
&
Implications for Audio Perception

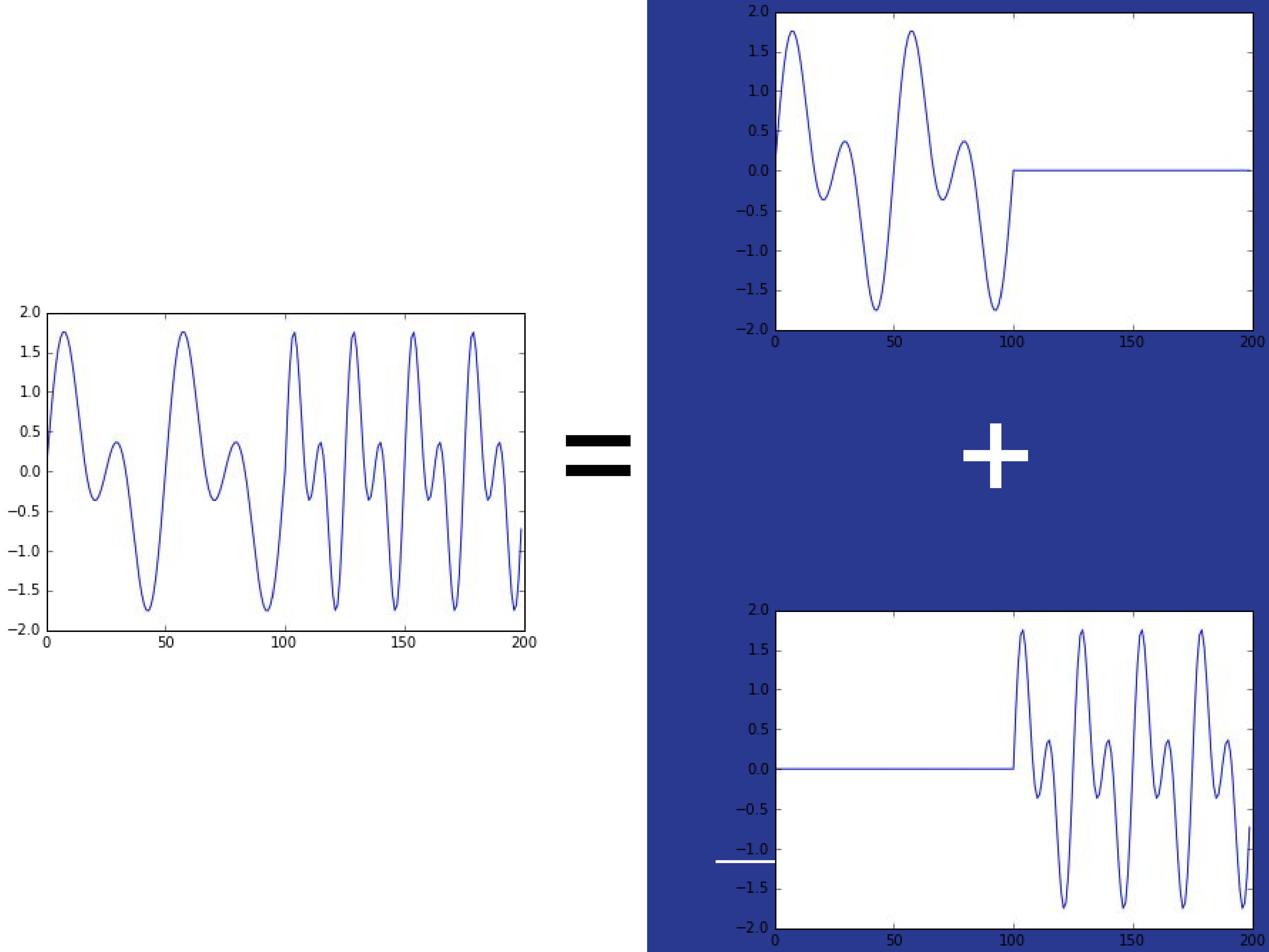
Key challenges in computing with audio

- How do we experience (see, hear) a piece of media (audio, image, video, etc.)?
- How can we efficiently create media to have a desired effect?
- How can we efficiently process media to have a desired effect?
- How can we efficiently analyze media to understand its contents?
- How can we represent media in a computer for storage, transmission, processing, analysis, etc.?

Two insights for answering these questions

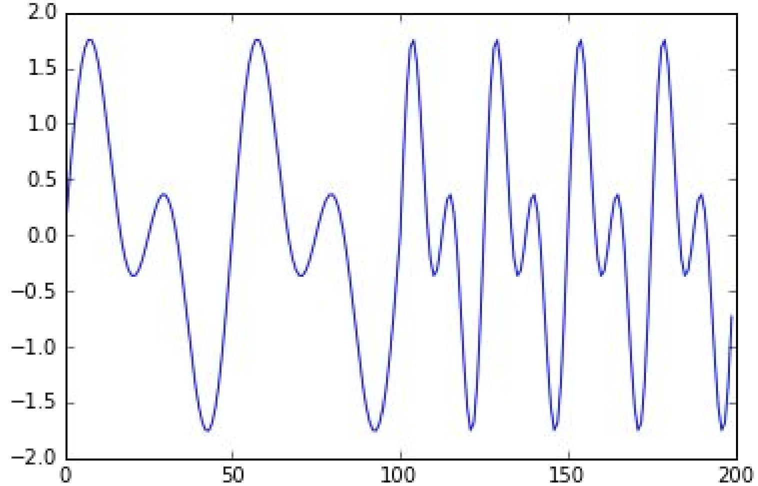
1. Sound, image, video, etc. are *functions*
2. All functions can be represented as *sums of simpler functions*

Sometimes called “basis functions”

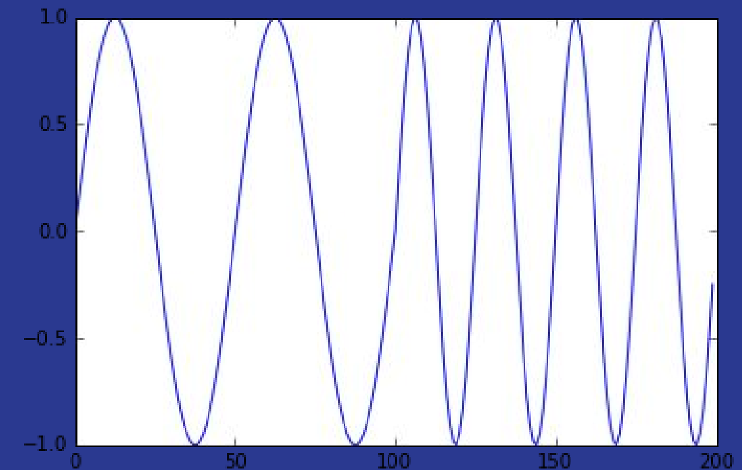


2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
50
100
150
200
=
+

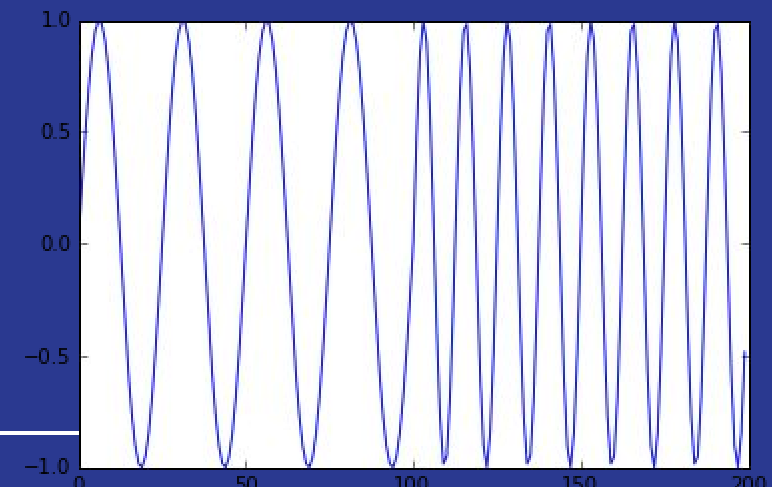
or...

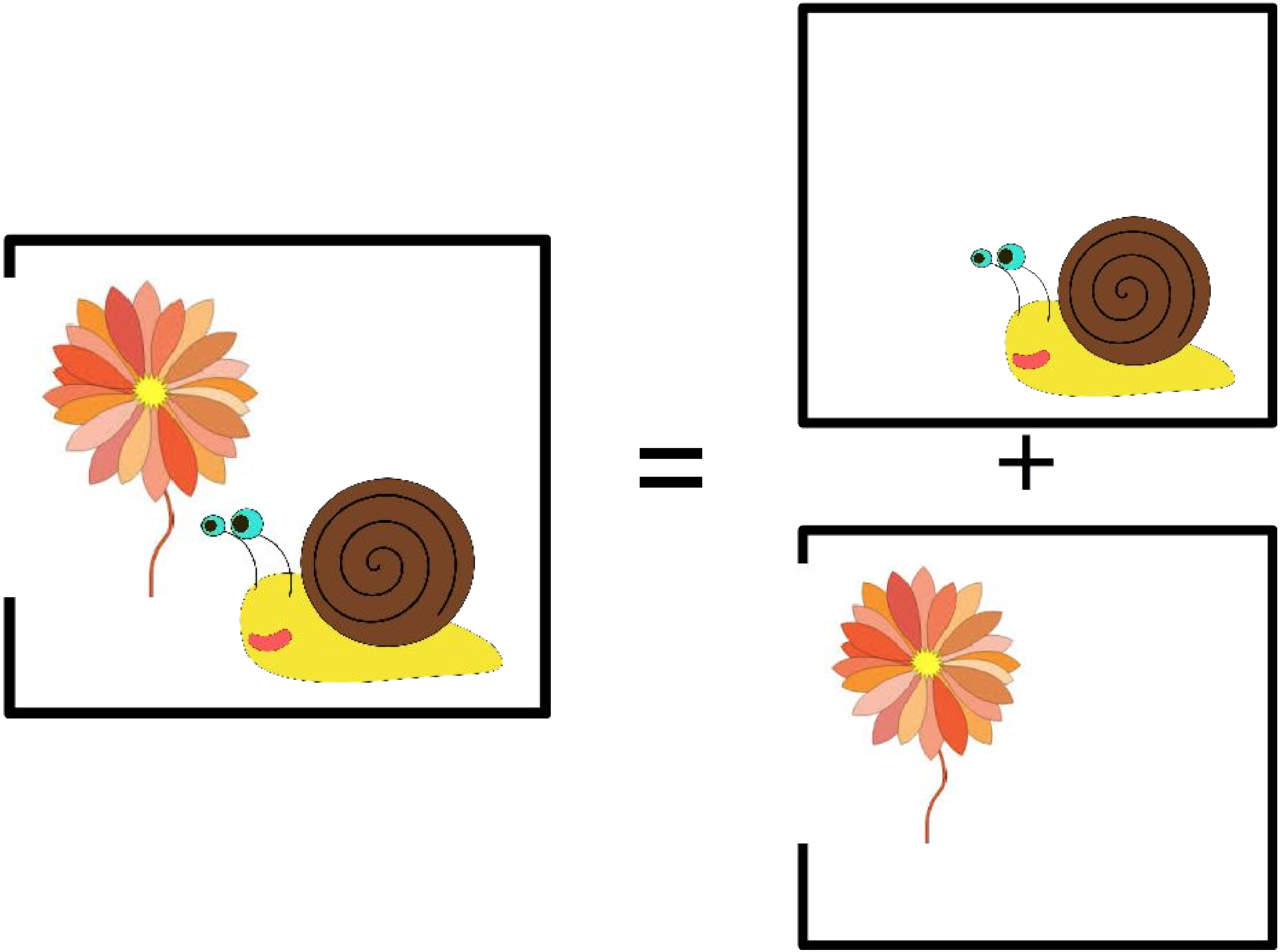

2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
50
100
150
200

=

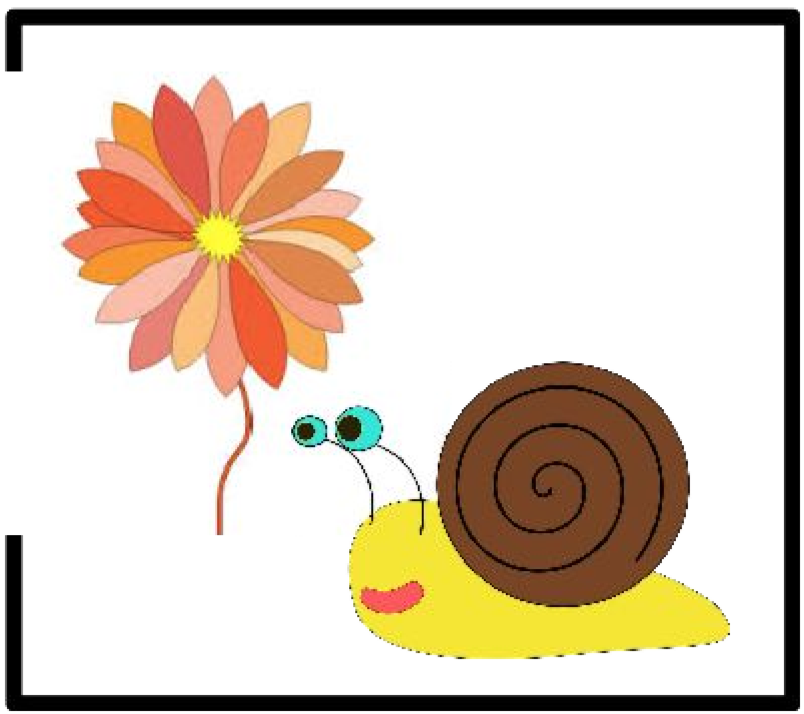

1.0
0.5
0.0
−0.5
−1.0
0
50
100
150
200

+

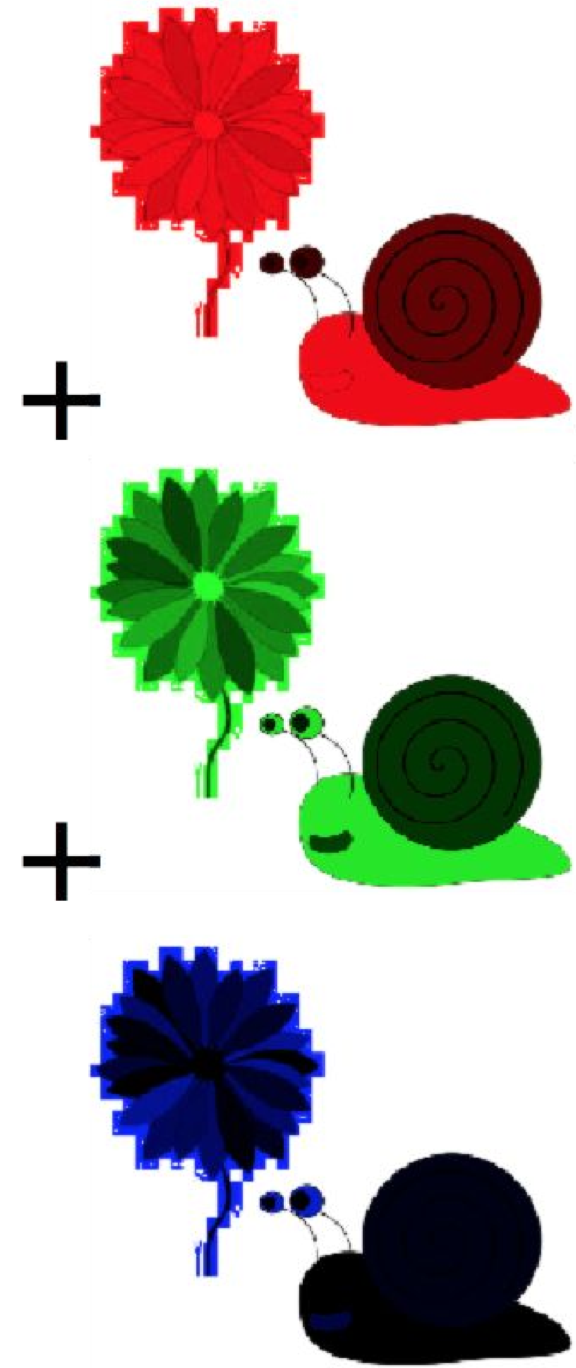

1.0
0.5
0.0
−0.5
−1.0
0
50
100
150
200



or...



=



=


COLAVITA
Aged
White
WINE
VINEGAR

Today’s lecture

We can represent any function as a sum of sinusoids.

Why sinusoids?

1. They’re perceptually & physically relevant

e.g. sound frequency, volume, timbre, location, vowel, …; image contour shape, edges, texture, …

2. We have mathematical & computational tools that make working with sines convenient & efficient

Fourier analysis, Fast Fourier Transform, filters

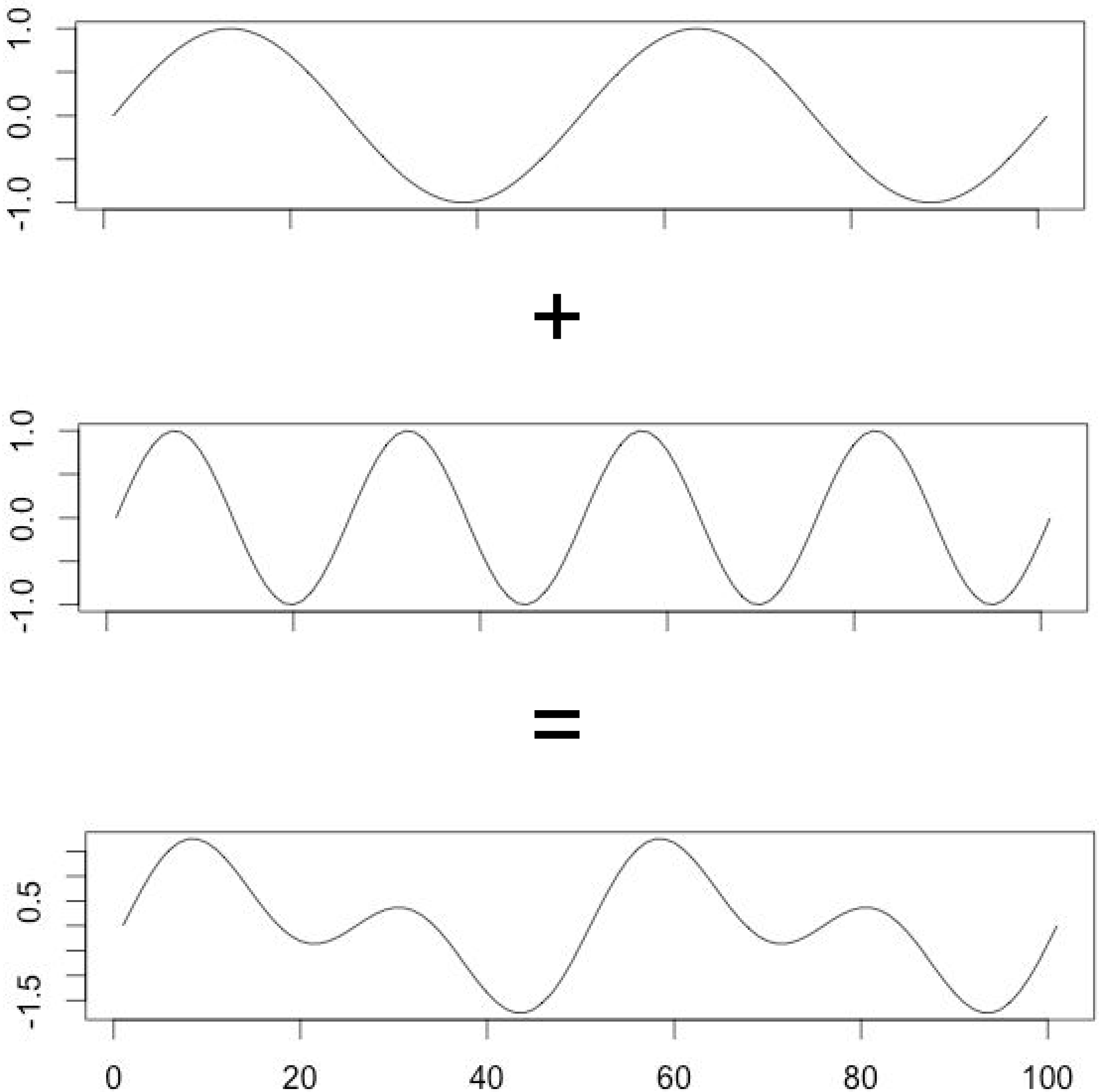
Last term

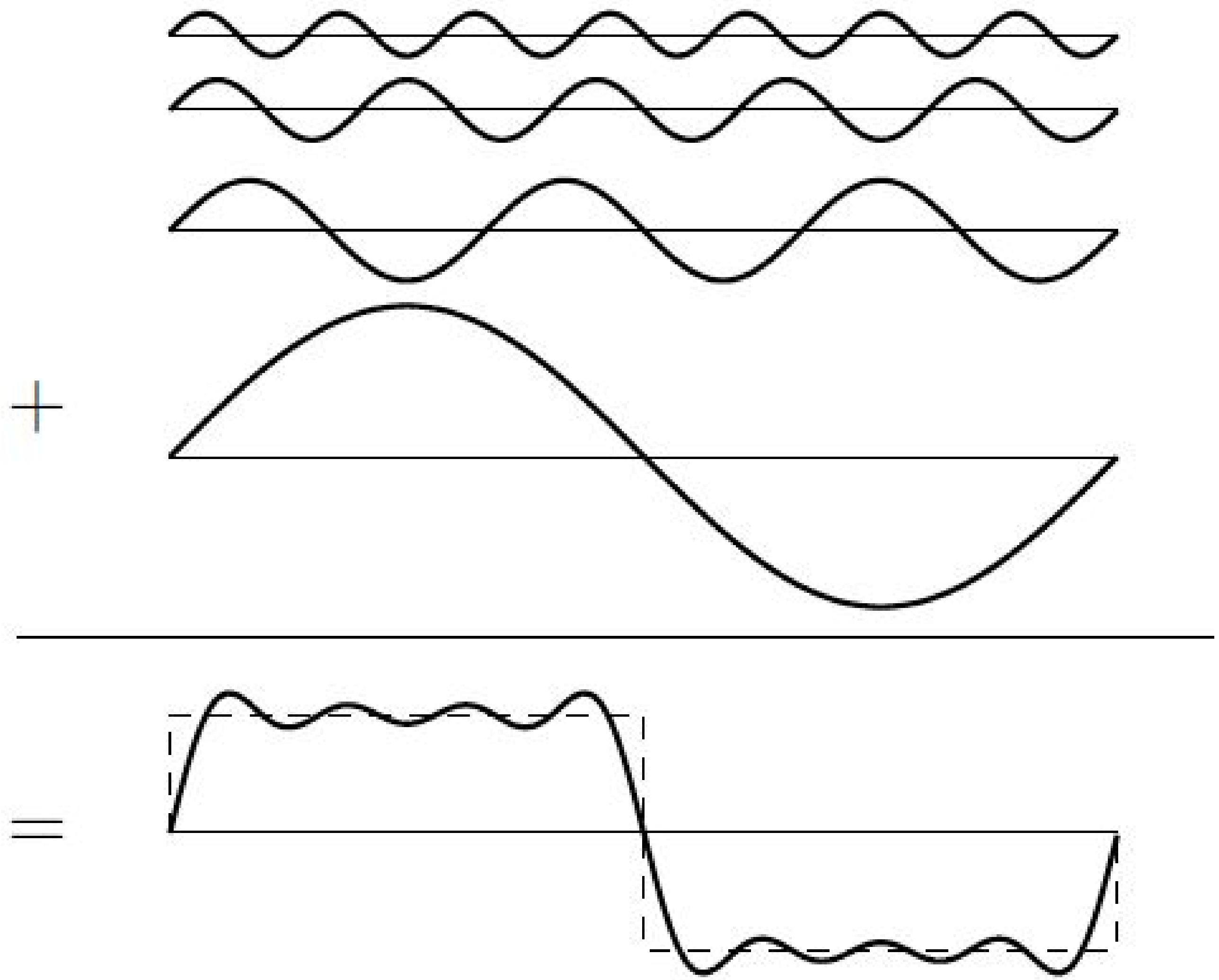
- Wave propagation
- Intro to audio perception: pitch, volume, location
- Basic analog-to-digital conversion: Sampling & quantising
- Intro to perception & and synthesis of complex waveforms

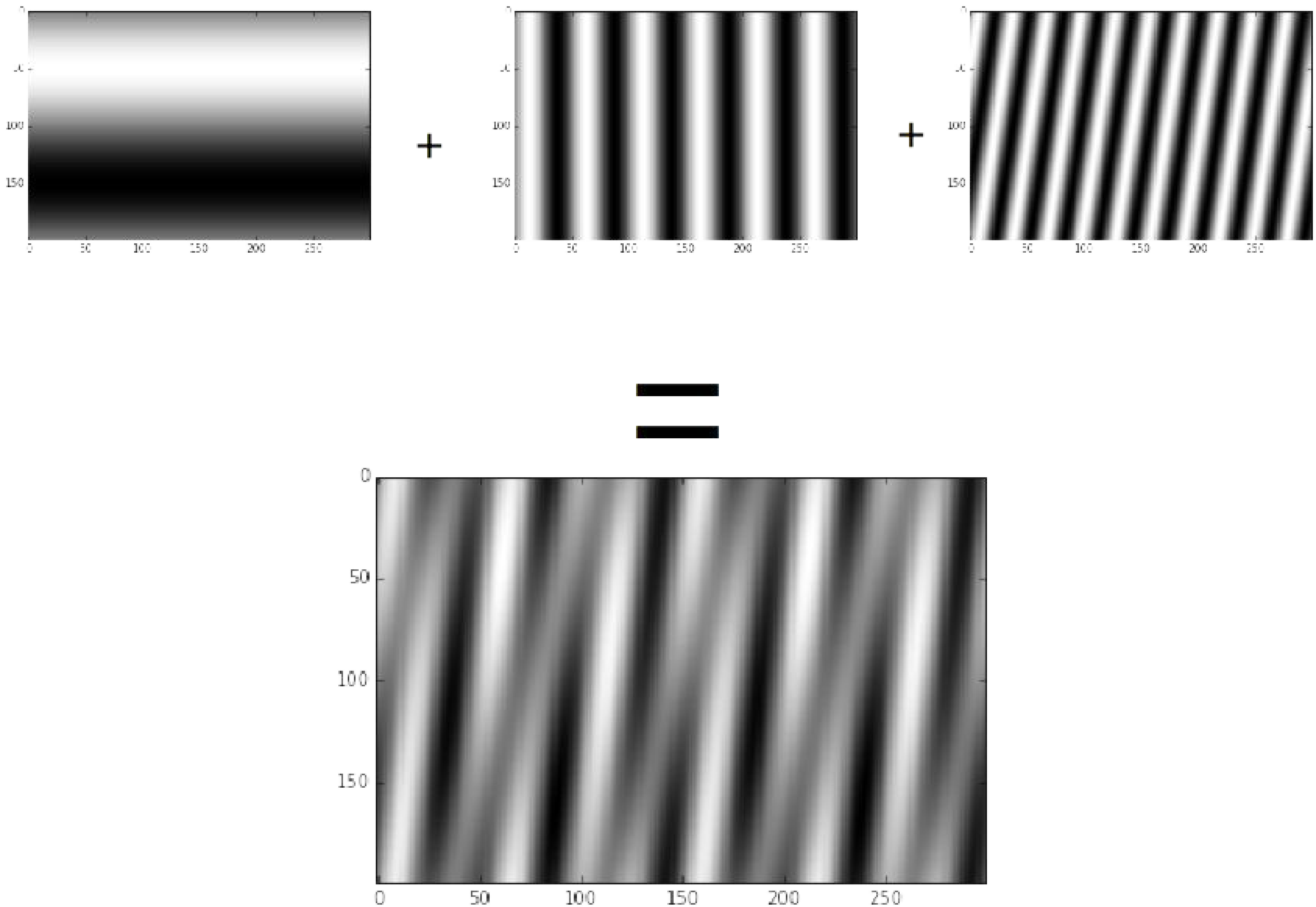
Today

- Understanding signals as sums of sinusoids
- Fourier analysis (i.e., which sinusoids to sum?)
- Using sinusoids to understand perception of complex audio waveforms
  - Pitched & unpitched sounds
  - Consonance and dissonance
  - Human speech
- Implications for synthesis and analysis

Waveforms as sums of sinusoids


1.0
0.0
-1.0
+
1.0
0.0
-1.0
=
0.5
-1.5
0
20
40
60
80
100

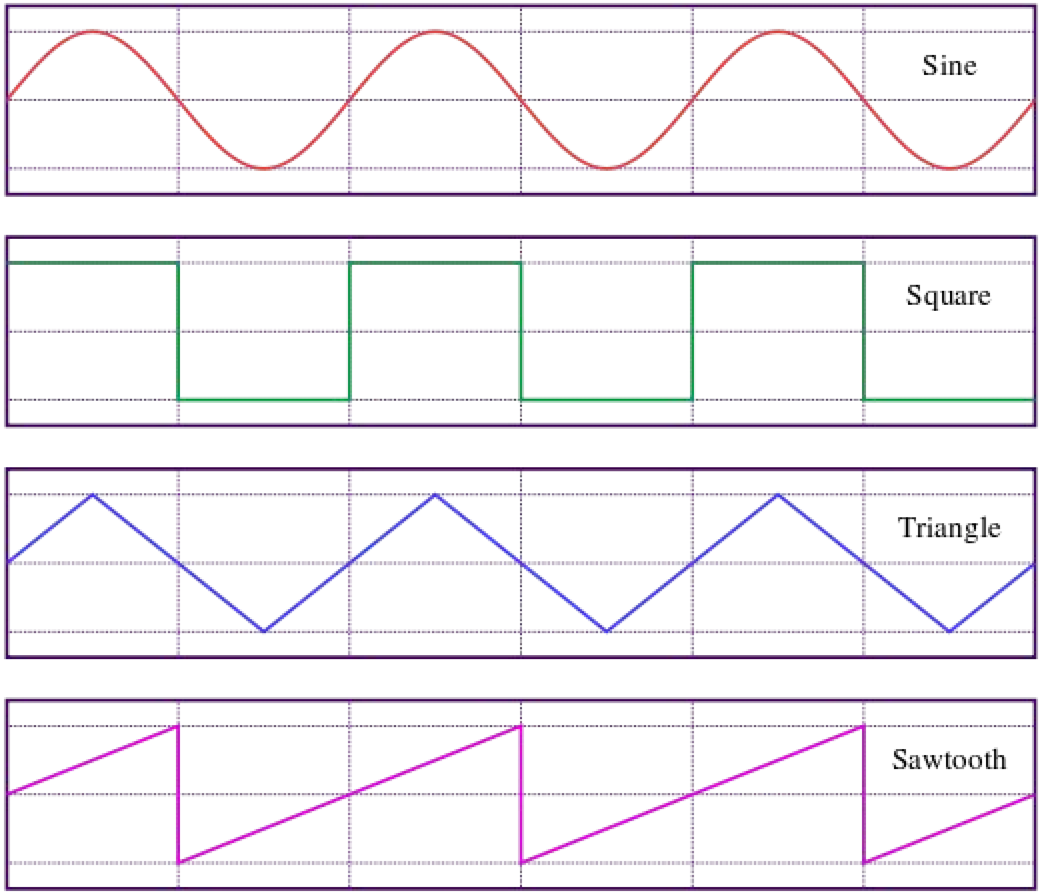



+
+
=
0
50
100
150
0
50
100
150
200
250

How to synthesize more interesting sounds?

1. Generate waveforms with different basic shapes

Other waveforms...



How to synthesize more interesting sounds?

1. Generate waveforms with different basic shapes
2. Add multiple sines together

^^^ Same expressive potential ^^^

We can express any* waveform as a sum of sinusoids

**requires waveform be either infinitely periodic or (more likely) finite in duration*

$$S = A_1\sin(2\pi f_1 t + \Phi_1)$$
$$+ A_2\sin(2\pi f_2 t + \Phi_2)$$
$$+ A_3\sin(2\pi f_3 t + \Phi_3)$$
$$+ \ldots$$

Intro to Fourier Analysis

Given a waveform, what are its sinusoidal components?

If $s(t) = A_1\sin(2\pi f_1 t + \Phi_1)$

$+ A_2\sin(2\pi f_2 t + \Phi_2)$

$+ A_3\sin(2\pi f_3 t + \Phi_3)$

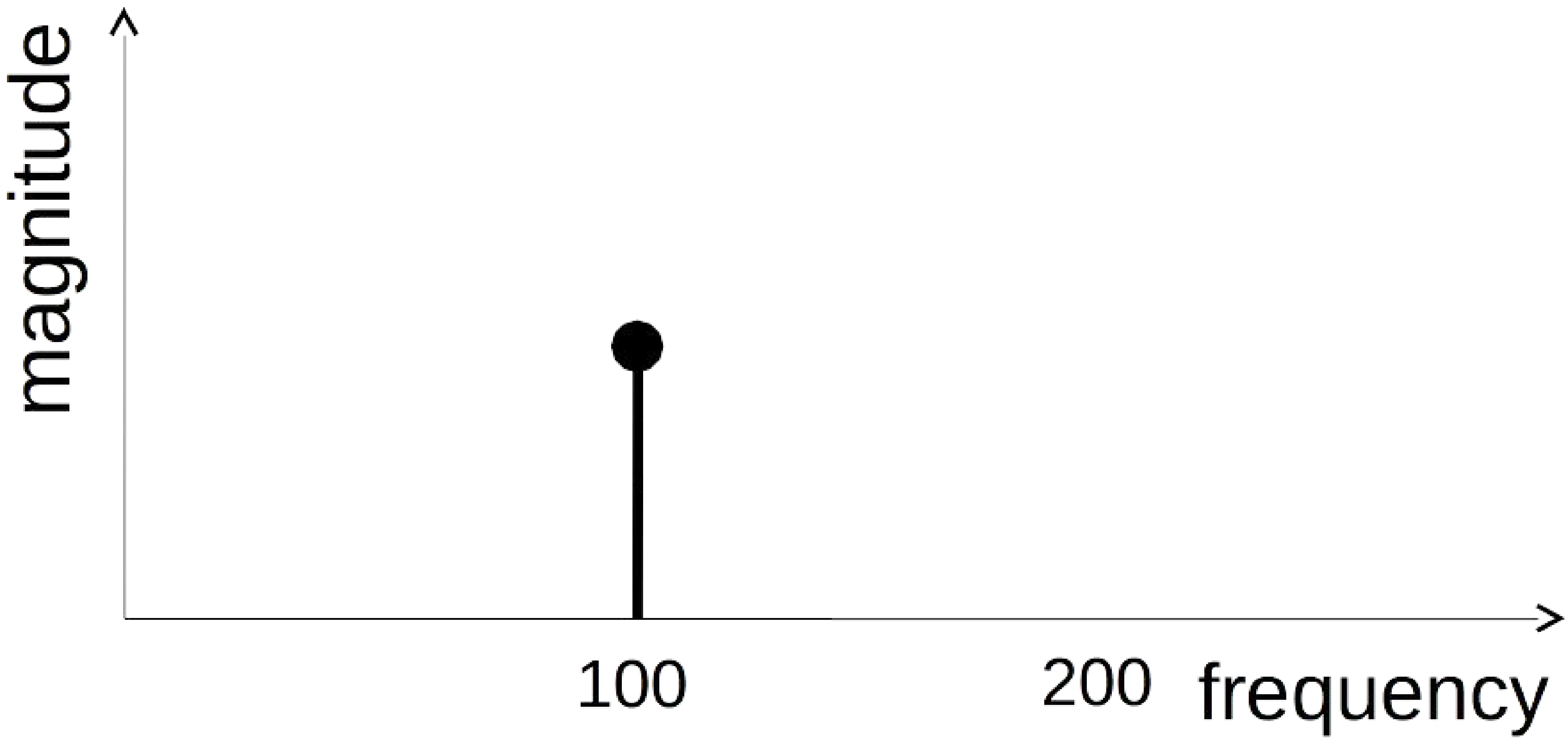
$+ \ldots$

What are A_n, f_n, Φ_n for all n?

The spectrum of s

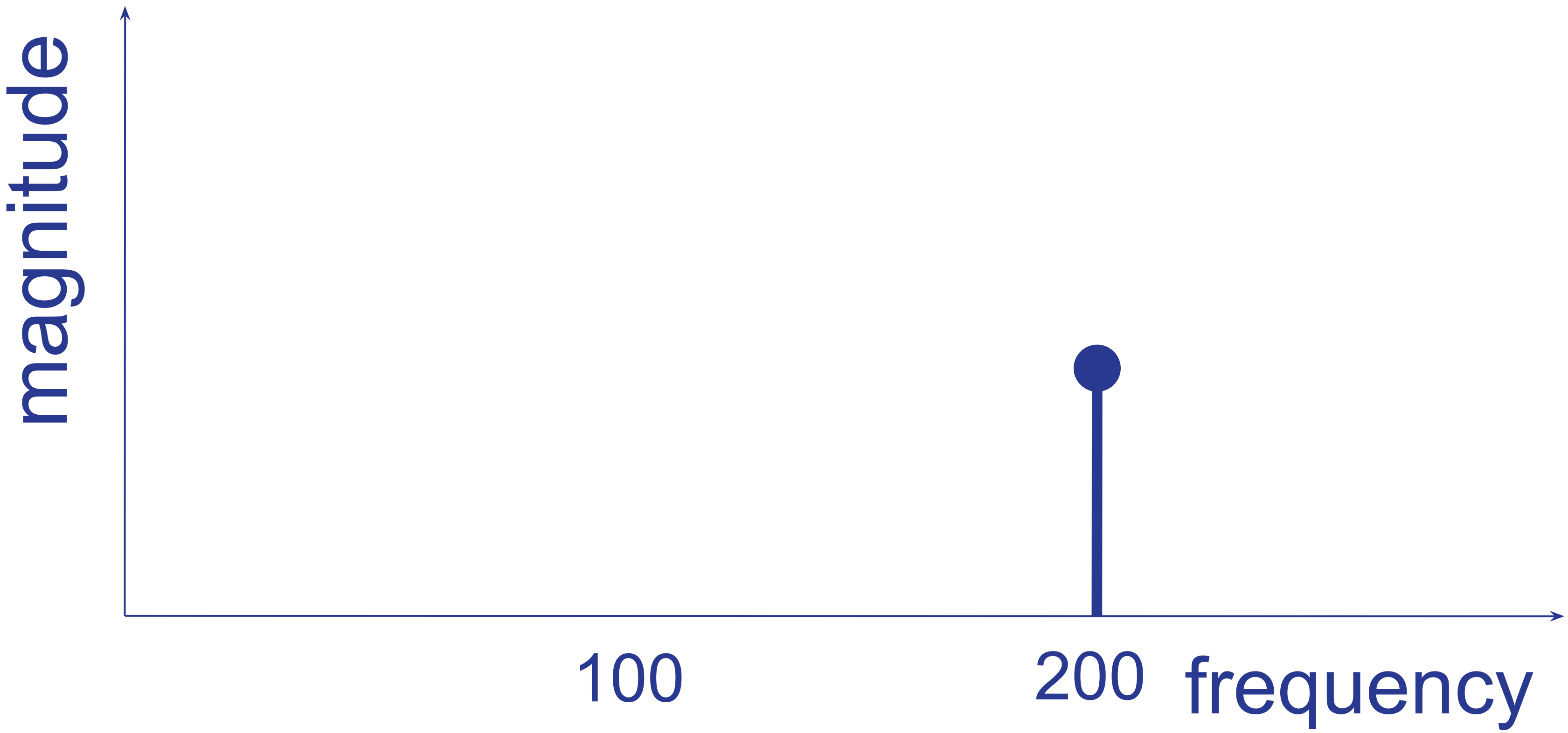
Plotting magnitude spectra
(amplitude & frequency, ignoring phase)

$$s = A \sin(2\pi * (100) * t)$$



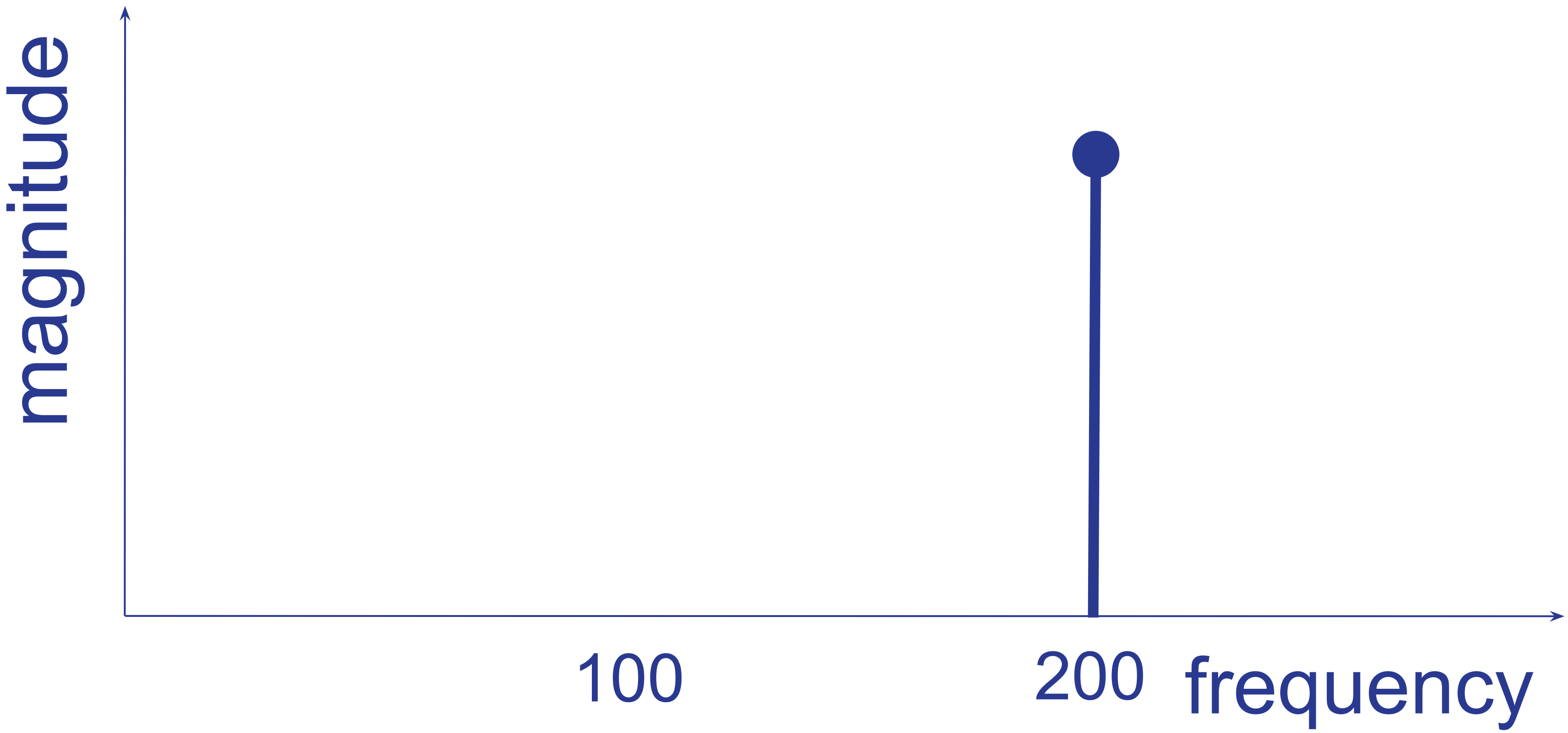
Plotting magnitude spectra
(amplitude & frequency, ignoring phase)

$$s = A \sin(2\pi*(200)*t)$$



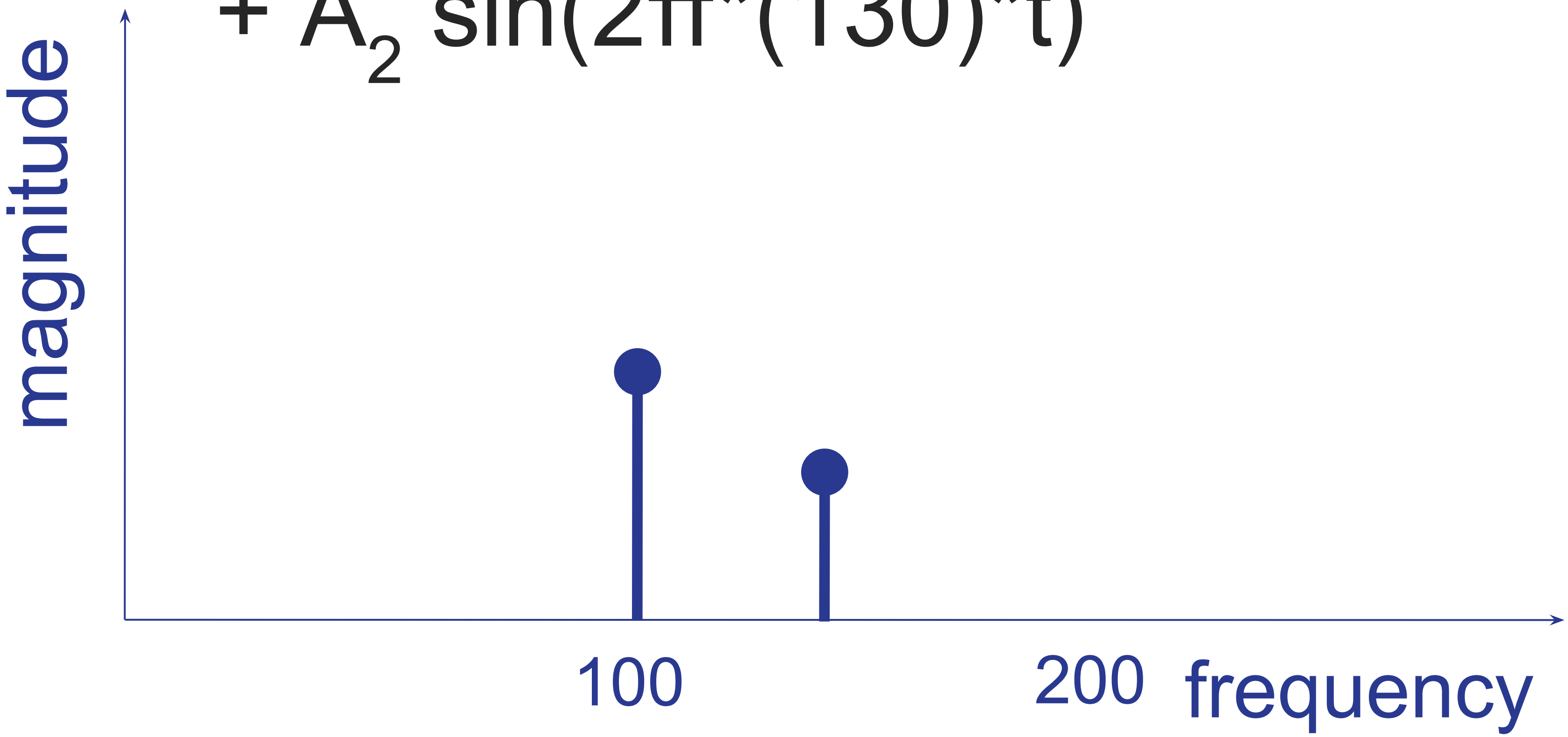
Plotting magnitude spectra
(amplitude & frequency, ignoring phase)

$$s = 2A \sin(2\pi * (200) * t)$$



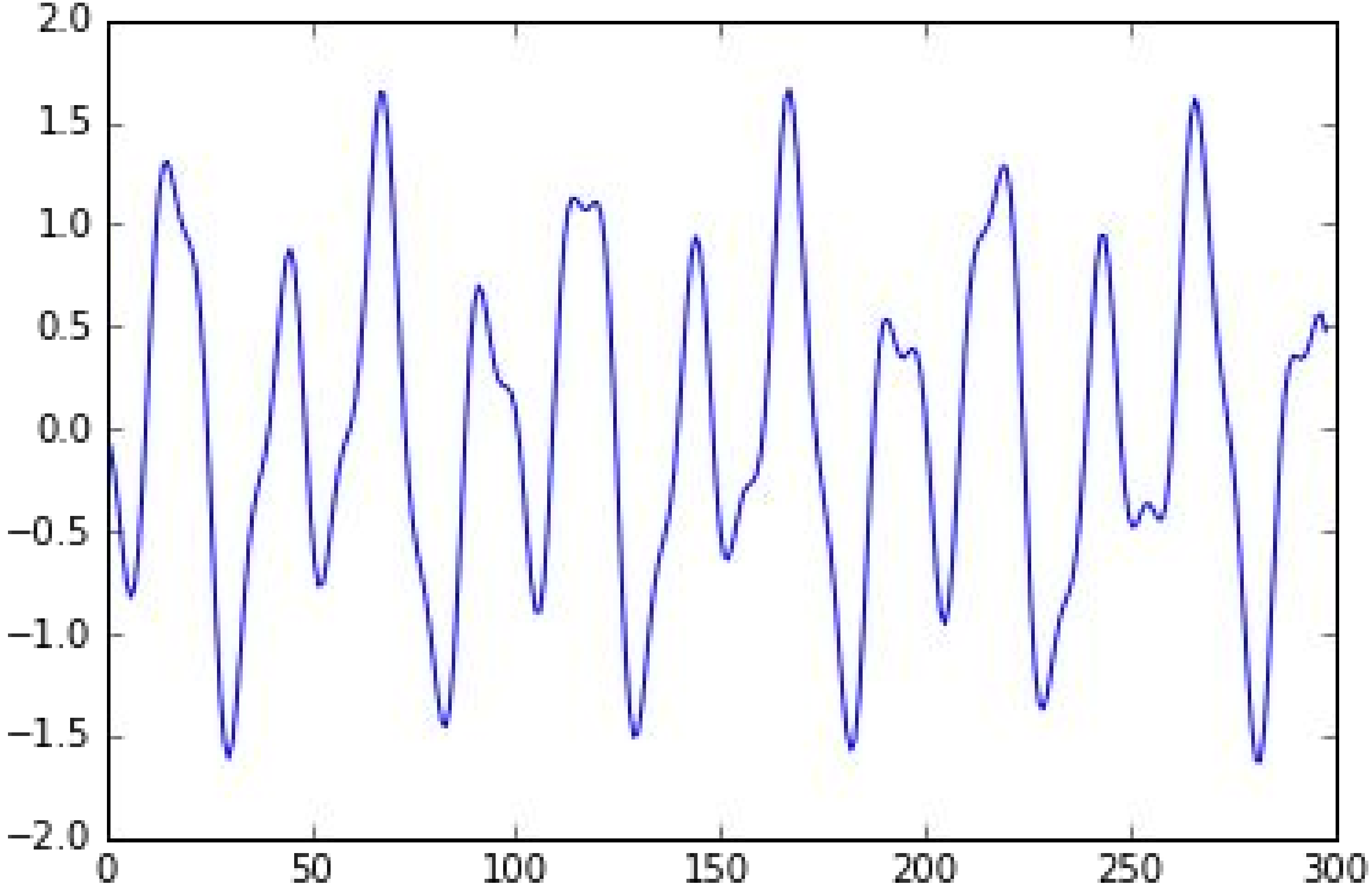
Plotting magnitude spectra
(amplitude & frequency, ignoring phase)

$$s = A_1 \sin(2\pi^*(100)^*t)$$
$$+ A_2 \sin(2\pi^*(130)^*t)$$

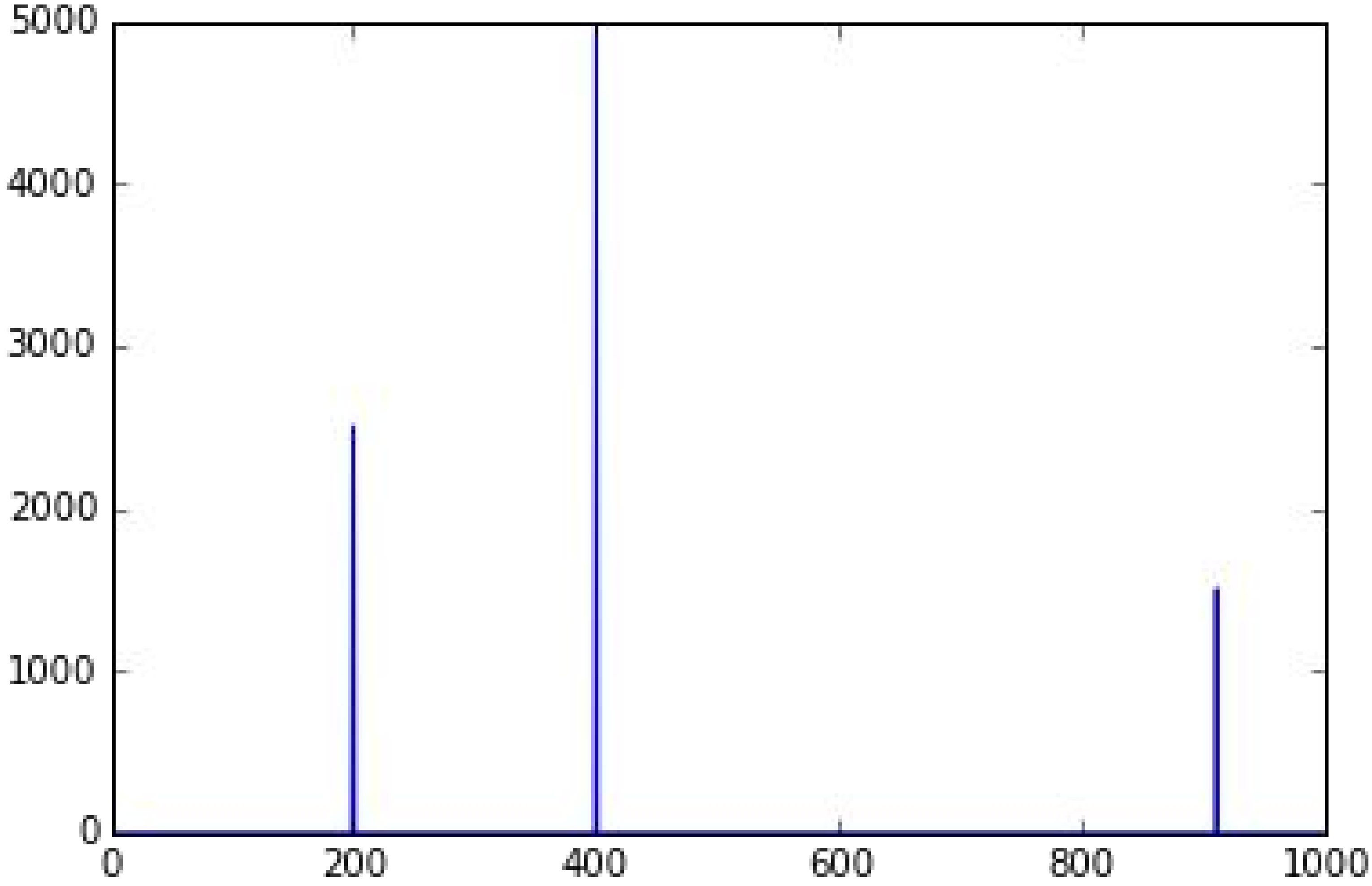


Why plot magnitude spectra?

Audio waveform



Spectrum

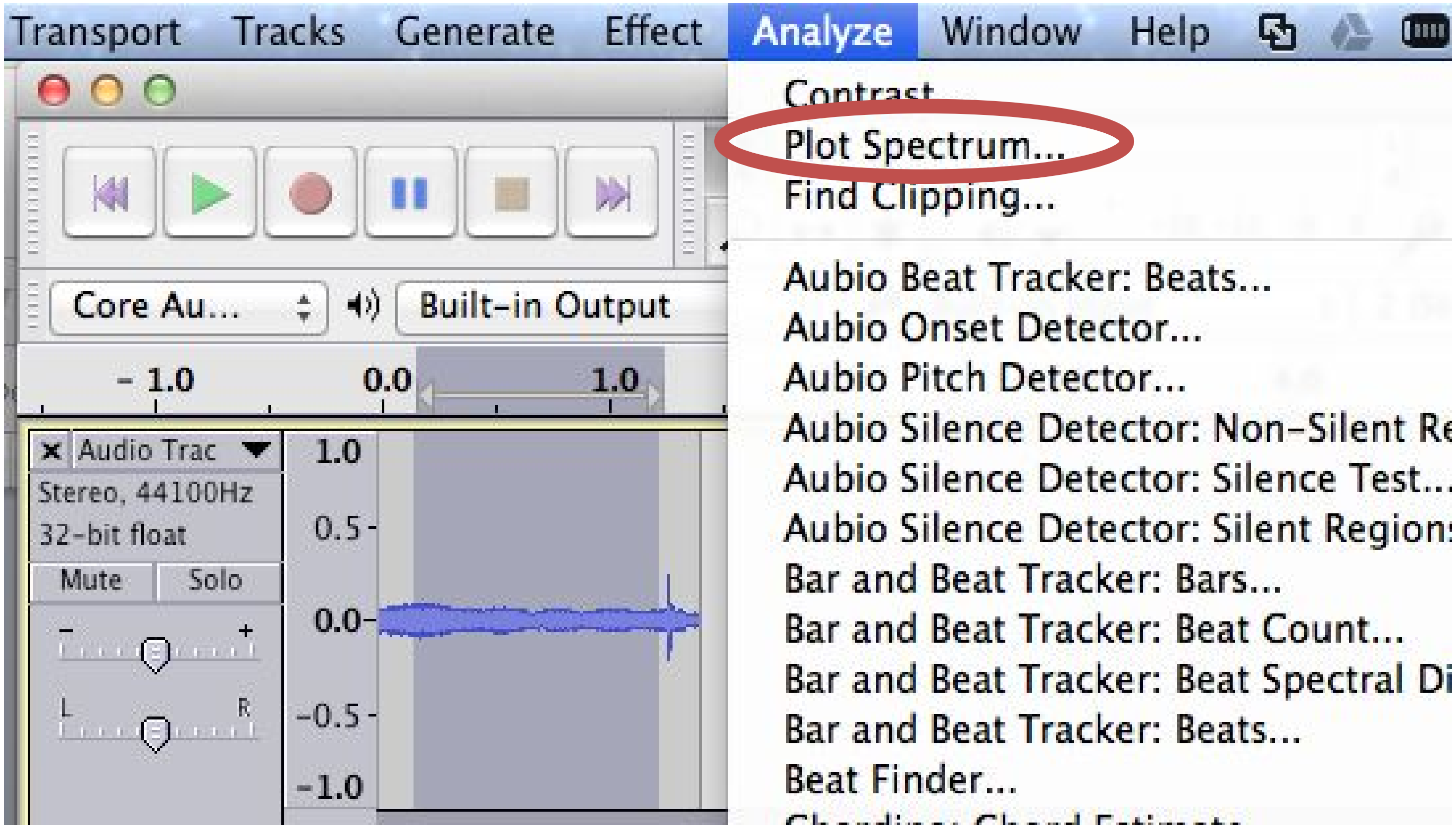


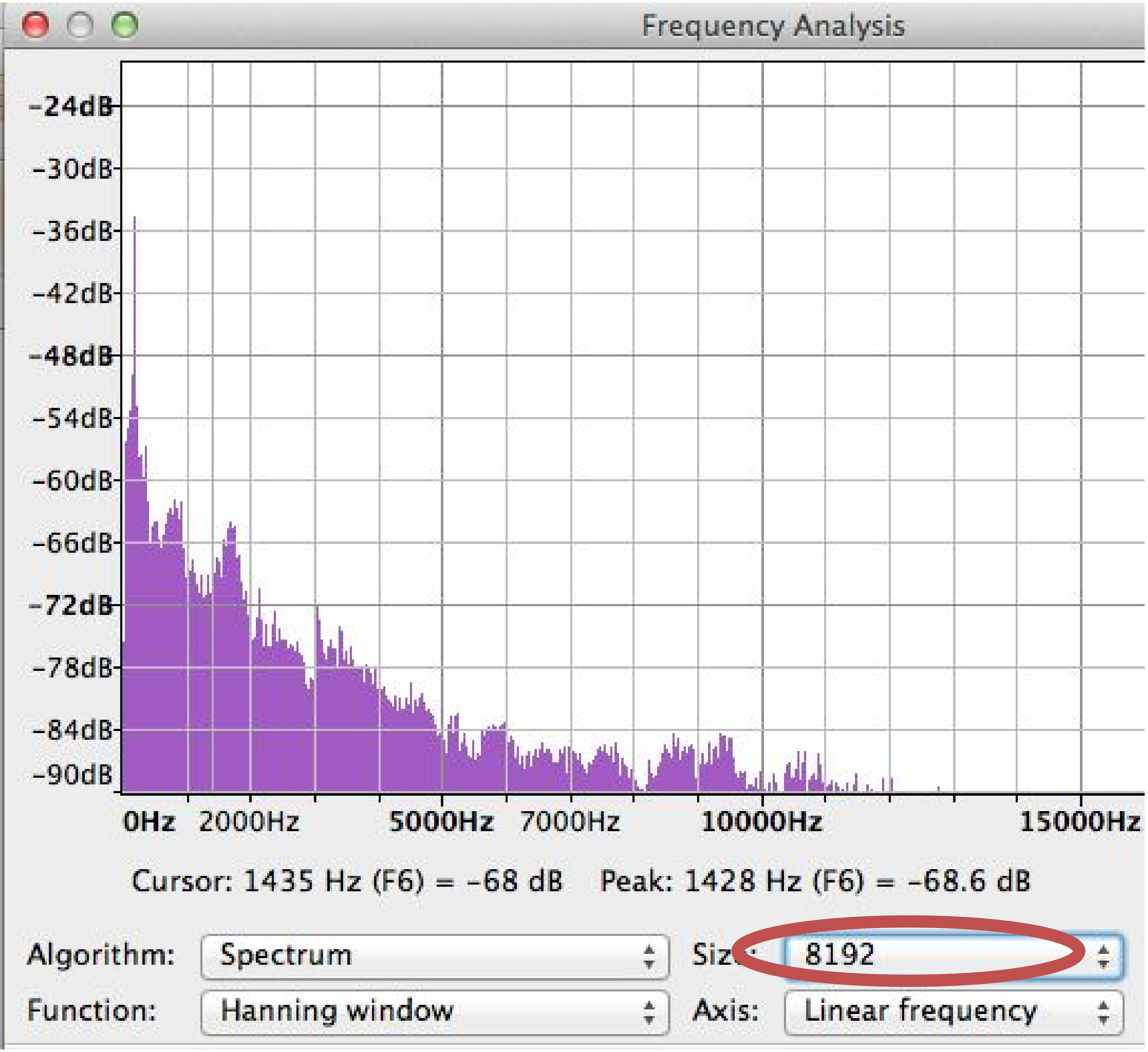
How to obtain a magnitude spectrum?

1. Choose some finite segment of the waveform to analyze
2. Compute Fast Fourier Transform (FFT), which will tell us magnitude & phase for sinusoids at different frequencies
3. Plot, with frequency on x-axis and magnitude on y-axis (don’t plot phase)

Result shows frequency content throughout the analyzed segment (no way to tell if/how things change over time within the segment)

Example: Spectra in Audacity





Try changing size

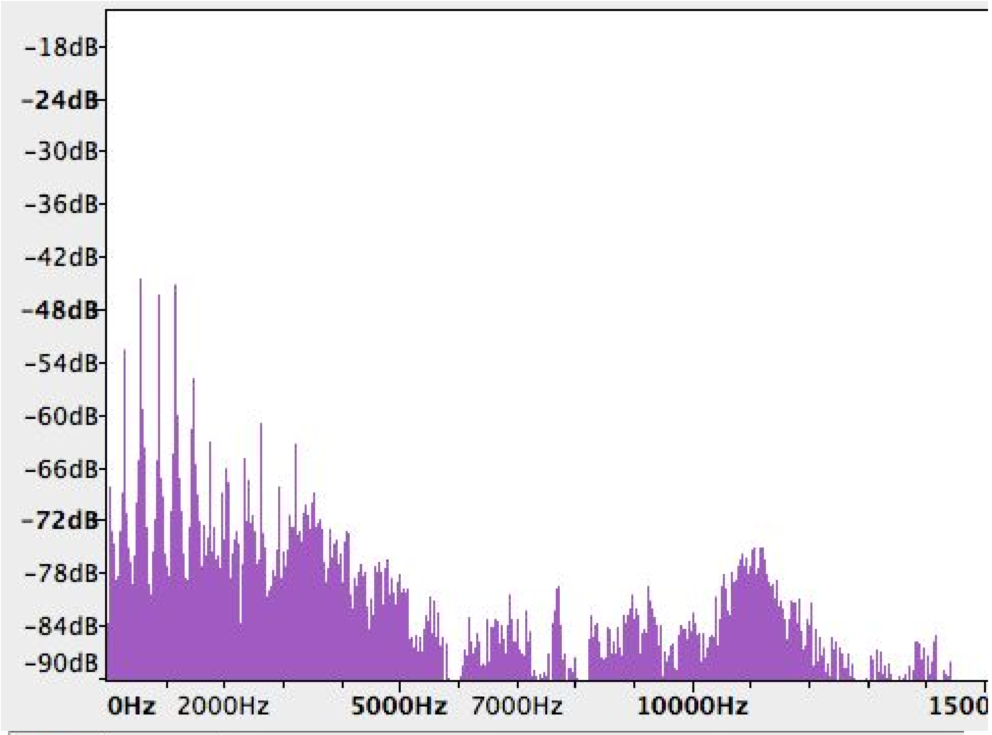
Demo

Examining Audio Spectra using Audacity

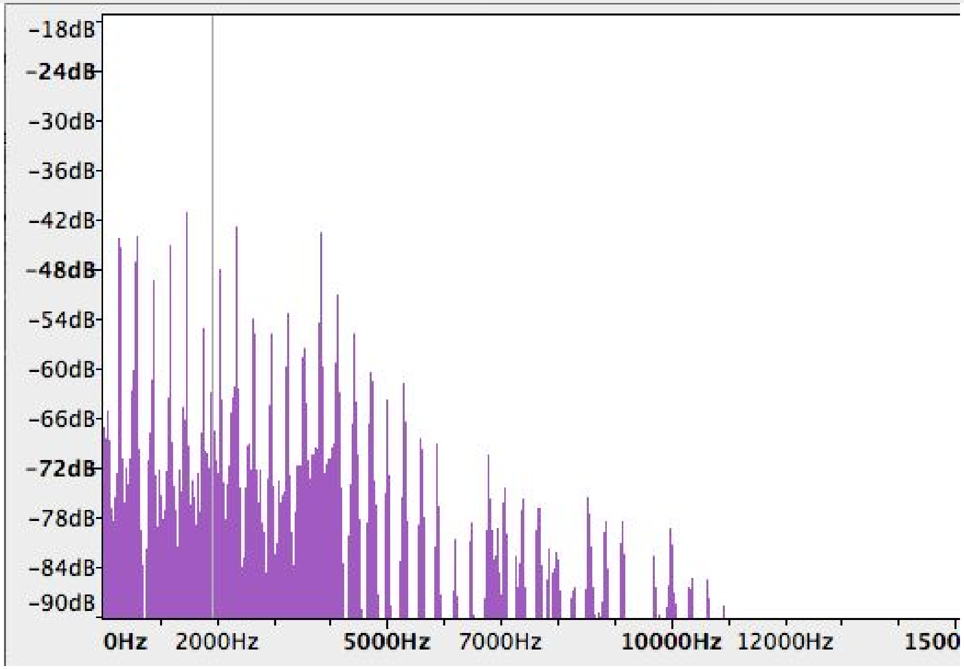
- Practical tip: try different window sizes & types
- Magnitude typically expressed in dB
  - 0 dB is a reference point here, corresponding to amplitude of 1 ($dBy = 20 \log_{10}(y)$)
  - *0dB doesn't mean same thing as when measuring sound volume!*
  - *Frequencies typically present with "negative" decibels*
  - *BUT 6dB rule of thumb still applies*

Using frequency content to reason about audio perception

Flute playing "D" above middle C



Violin playing same note



Rule of Thumb #1

Pitched sounds have sinusoidal components that are harmonically related.

$$s(t) = A_1\sin(2\pi f_1 t + \Phi_1) + A_2\sin(2\pi f_2 t + \Phi_2) + A_3\sin(2\pi f_3 t + \Phi_3) + \ldots$$

If $s(t) = A_1\sin(2\pi f_1 t + \Phi_1)$
$+ A_2\sin(2\pi f_2 t + \Phi_2)$
$+ A_3\sin(2\pi f_3 t + \Phi_3)$
$+ \ldots$

Then $f_2 = 2 \times f_1$

$f_3 = 3 \times f_1$

$f_4 = 4 \times f_1$

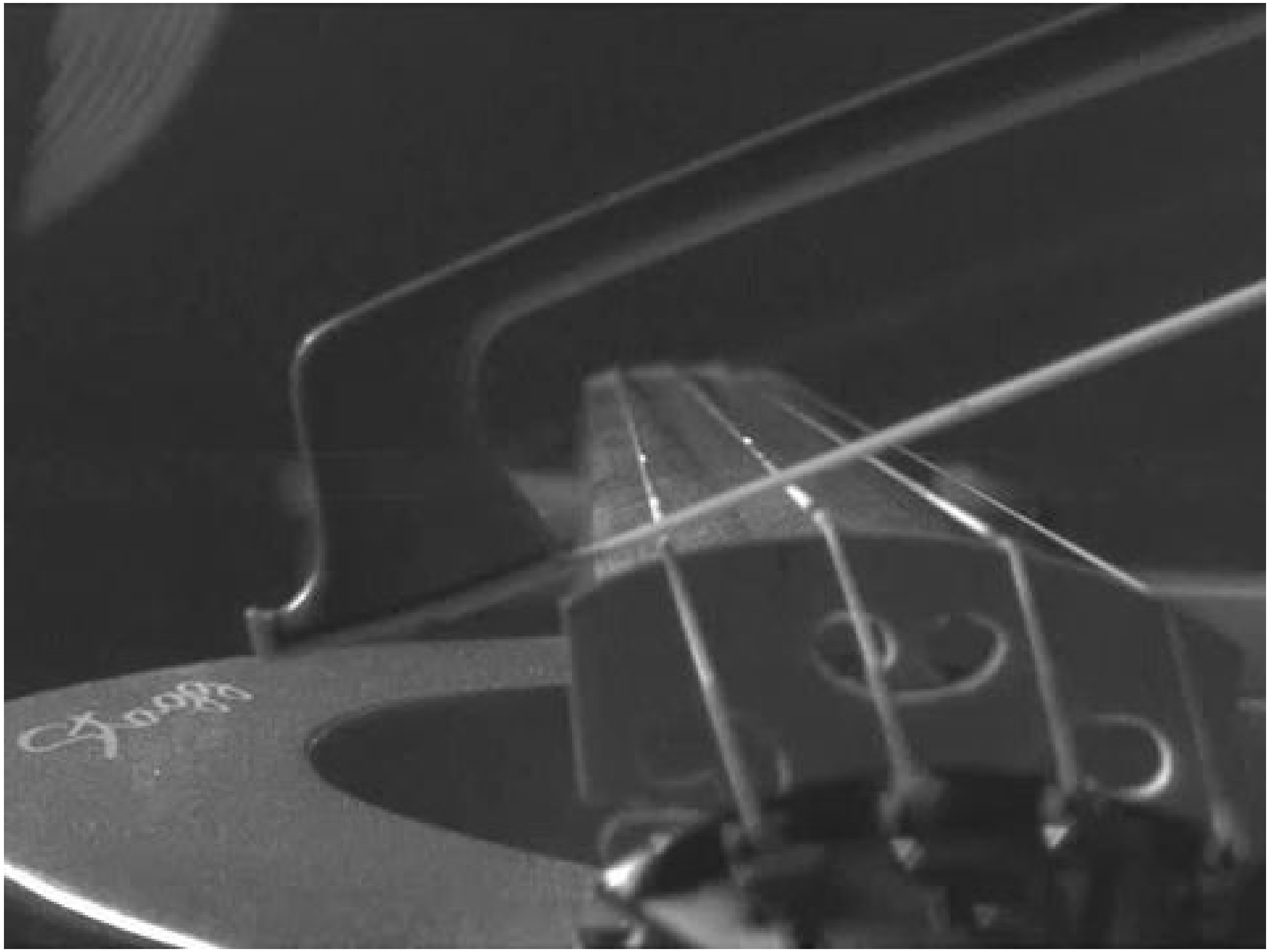
$f_5 = 5 \times f_1$

etc.

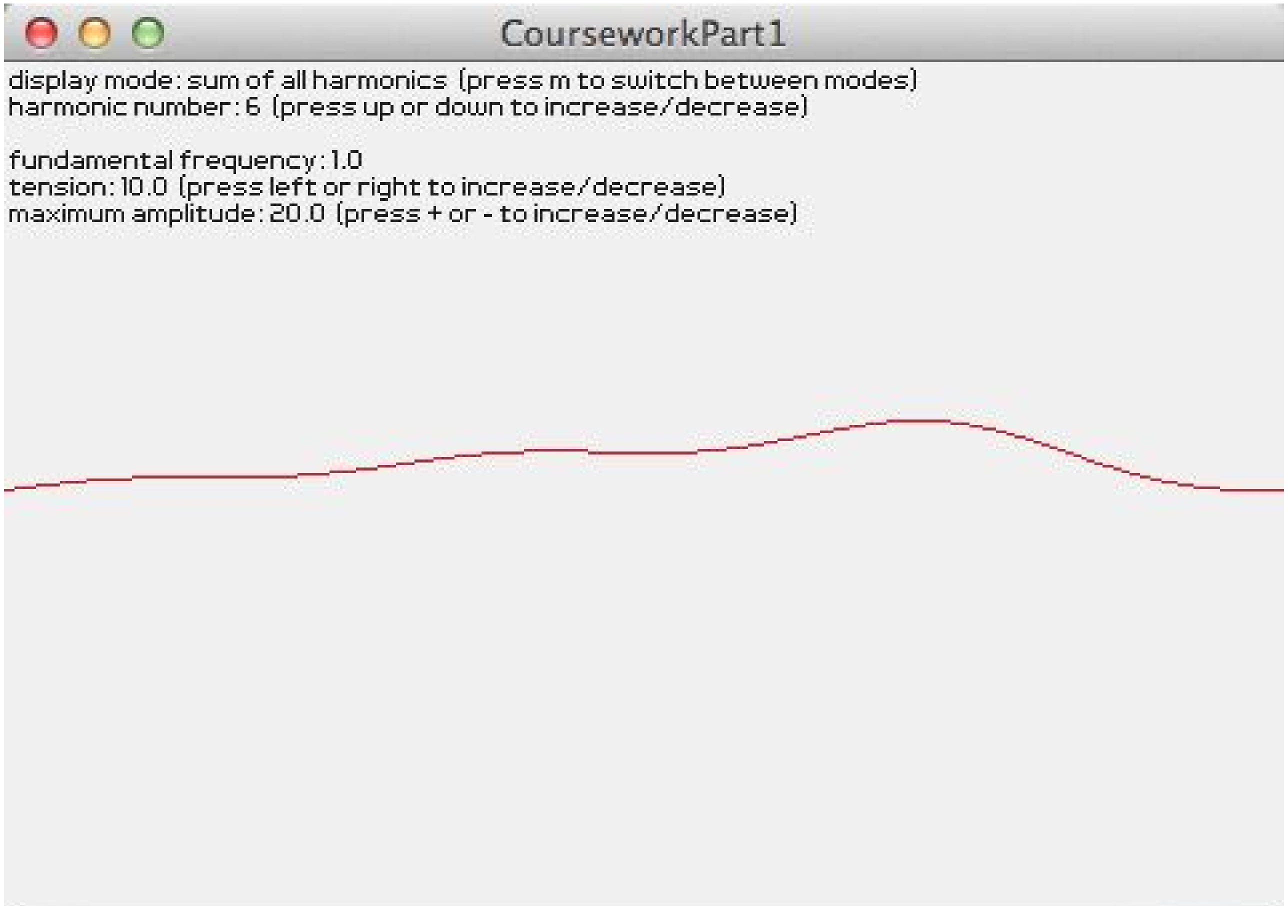
e.g., {100, 200, 300, 400}

Or {250, 500, 750, 1000}

etc.



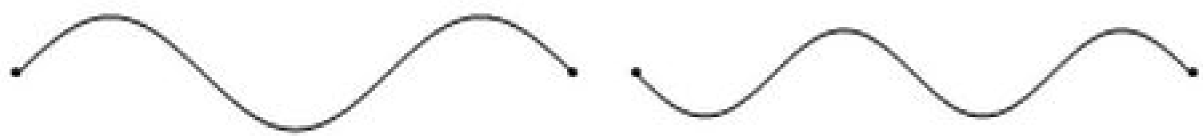
http://www.youtube.com/watch?v=6JeyiM0YNo4


CourseworkPart1
display mode: sum of all harmonics (press m to switch between modes)
harmonic number: 6 (press up or down to increase/decrease)
fundamental frequency: 1.0
tension: 10.0 (press left or right to increase/decrease)
maximum amplitude: 20.0 (press + or - to increase/decrease)

Harmonic Modes of Vibration

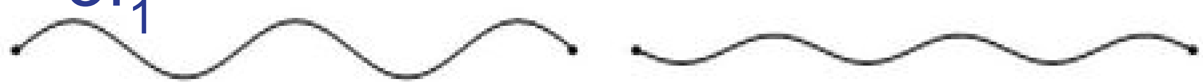
λ_1 =L, f_1 = v/L λ_2 =L/2, f_2 = 2v/L = $2f_1$

Recall:
λ = v/f
and v (speed) is constant



λ_3 =L/3, f_3 = 3v/L

= $3f_1$



http://en.wikipedia.org/wiki/Harmonic

http://www.youtube.com/watch?v=Ut7gy_7NDRI

Rule of Thumb #1

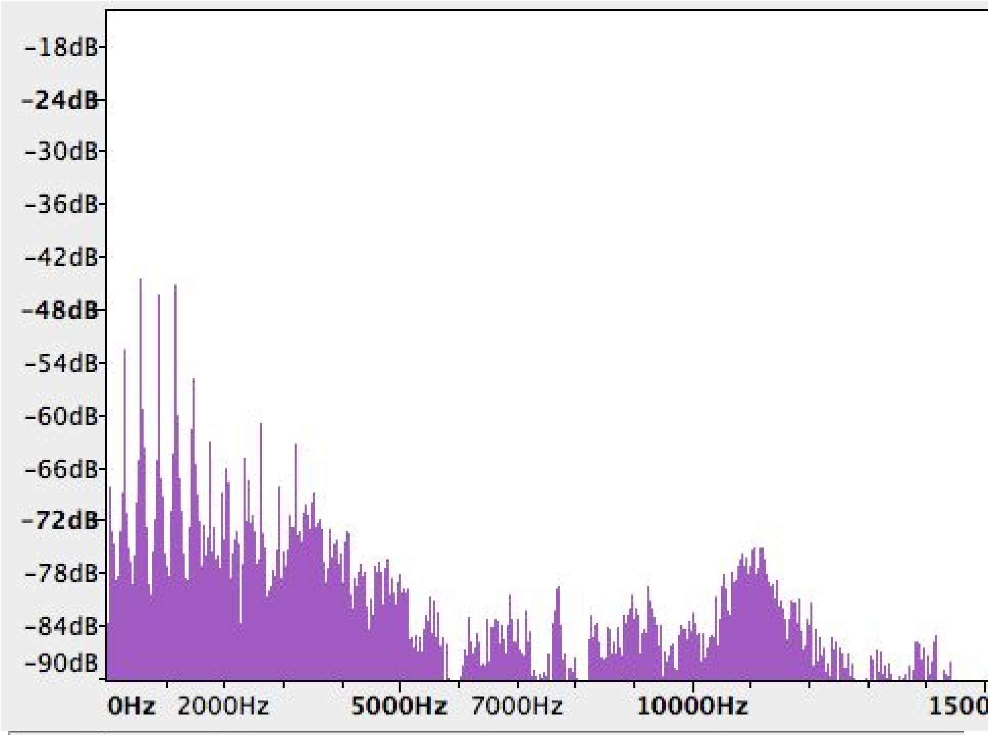
Pitched sounds have sinusoidal components that are harmonically related.

This is due to the physics of strings and air columns. When bowed / plucked / blown / etc., they will vibrate in certain way and not others.

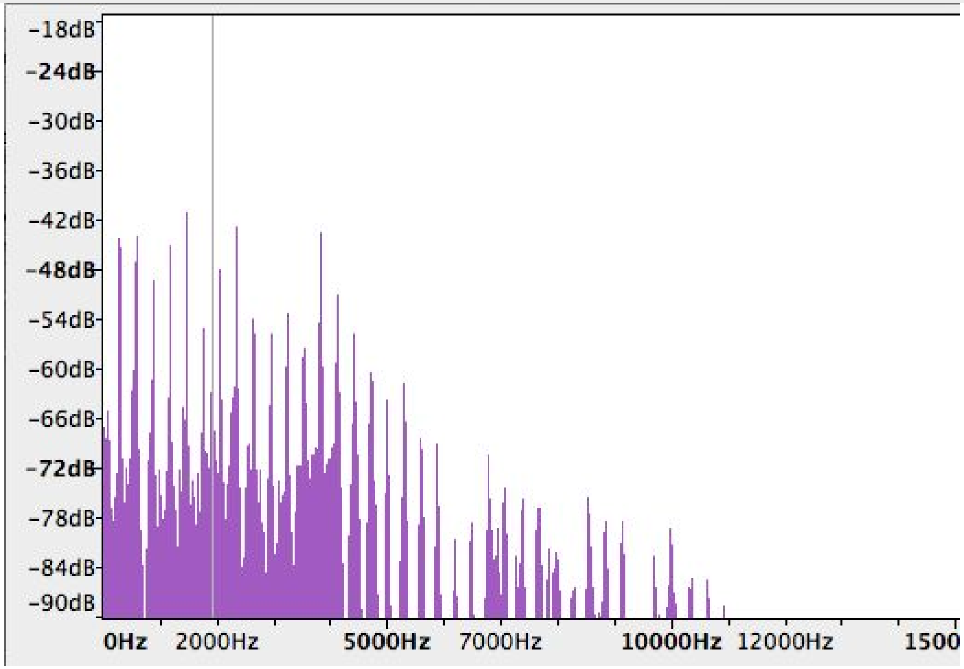
Rule of Thumb #2

The pitch we hear is determined by the fundamental frequency

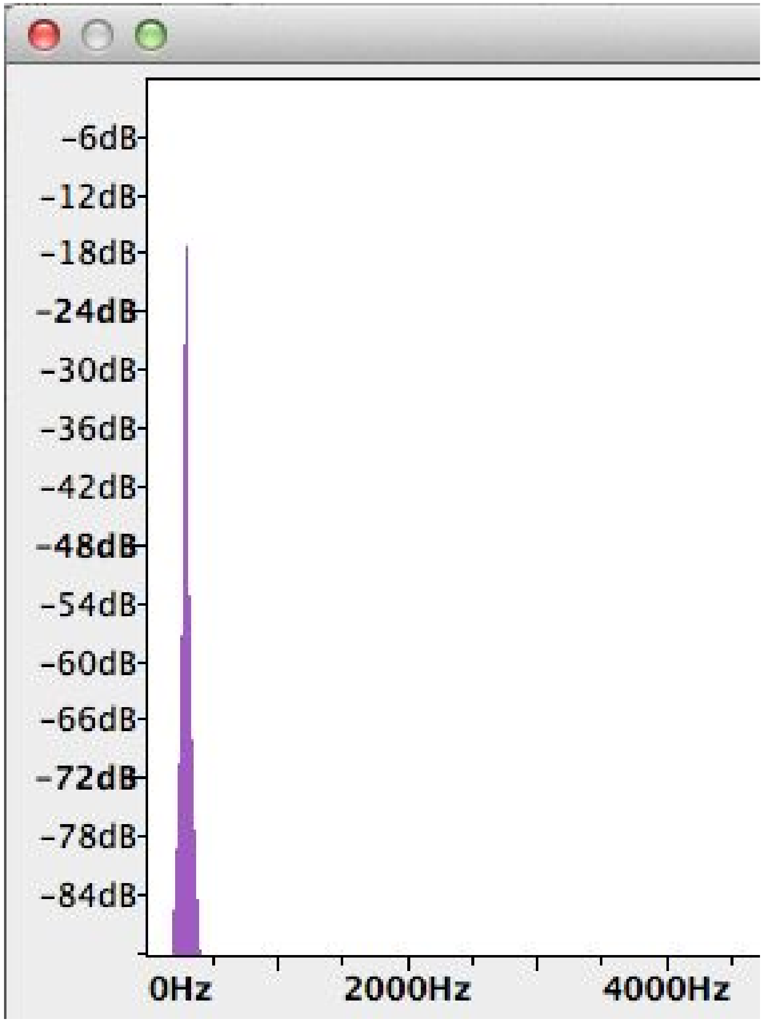
Flute playing "D" above middle C



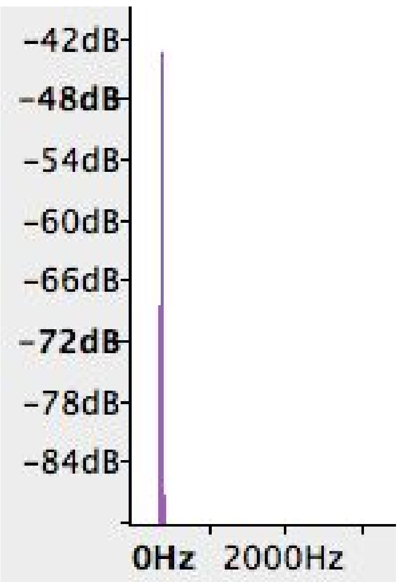
Violin playing same note



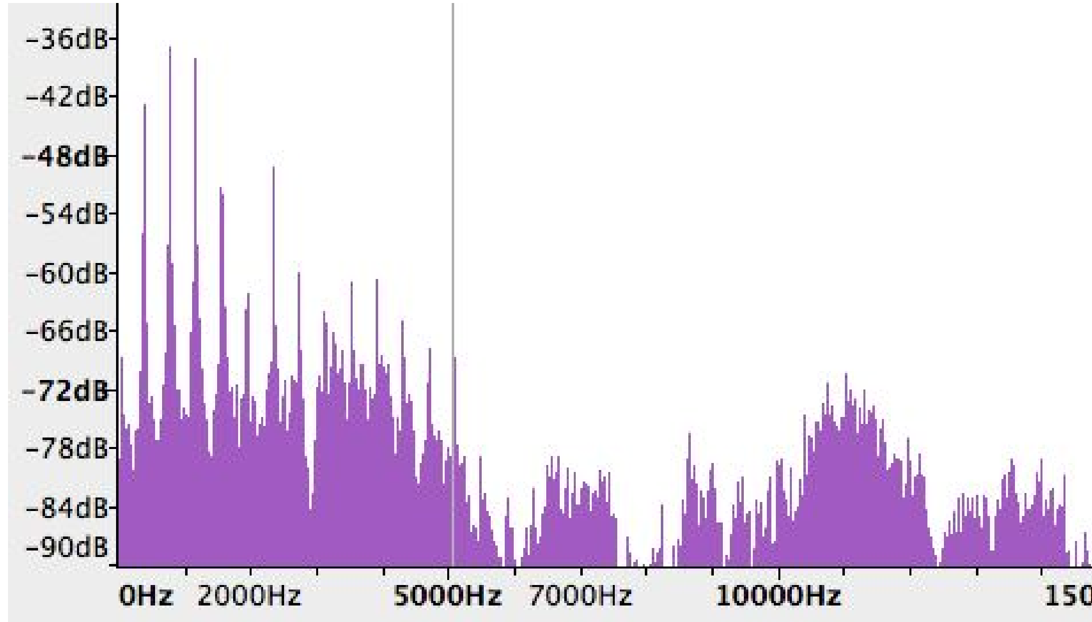
Sine wave at
294 Hz



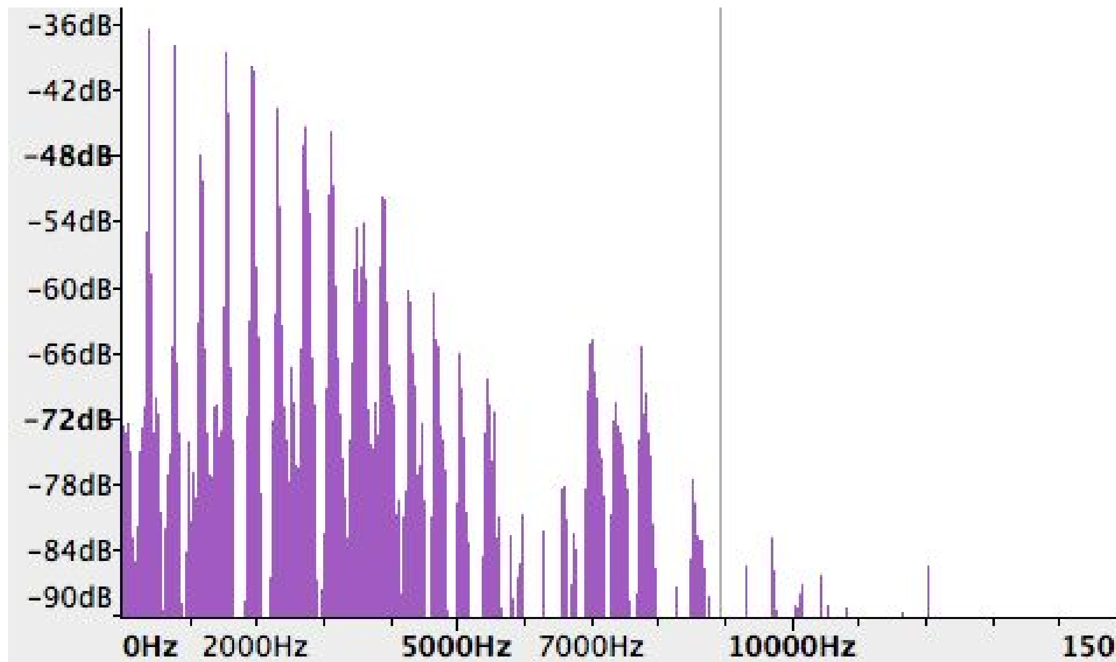
Sine
at 392 Hz

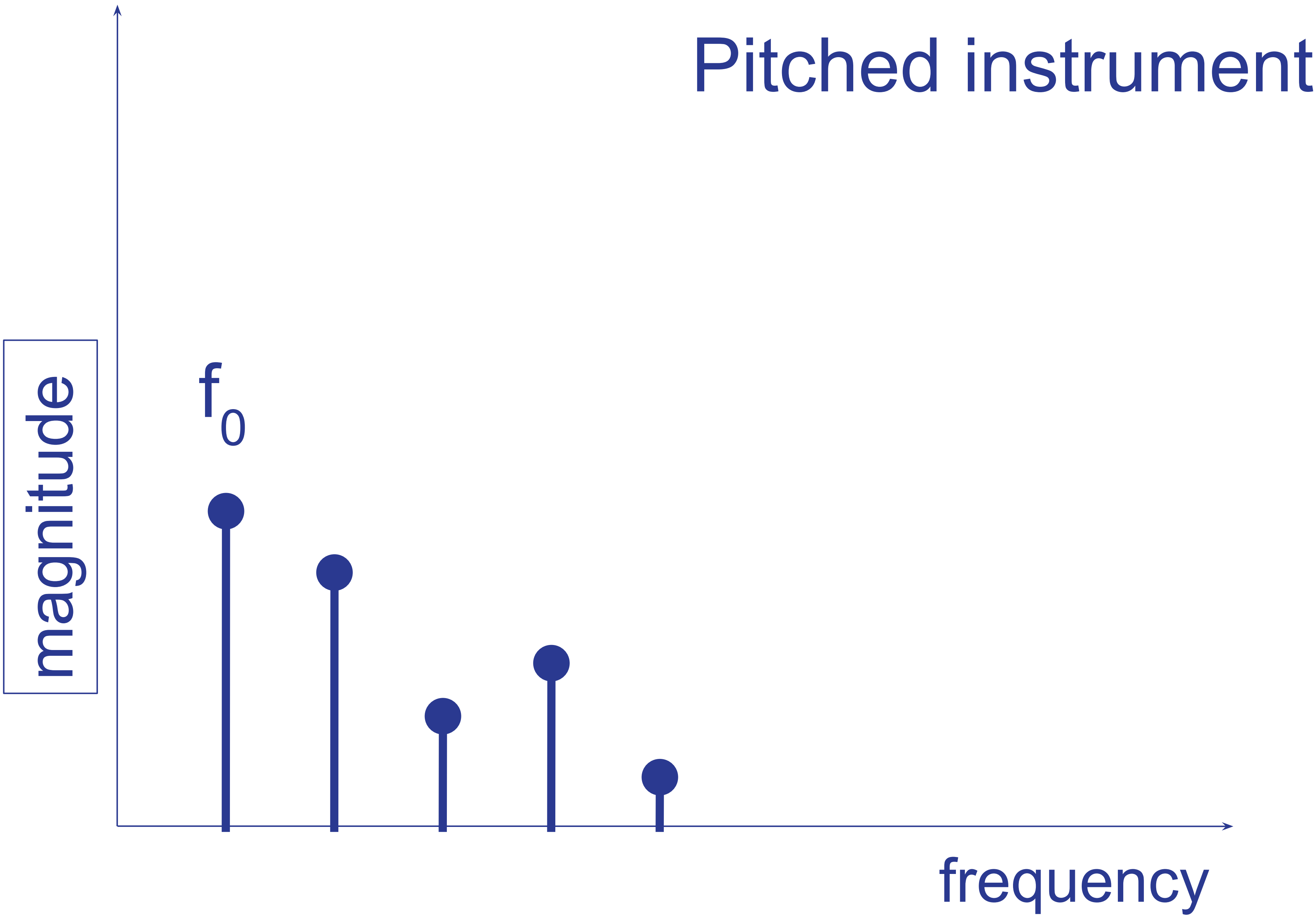


Flute

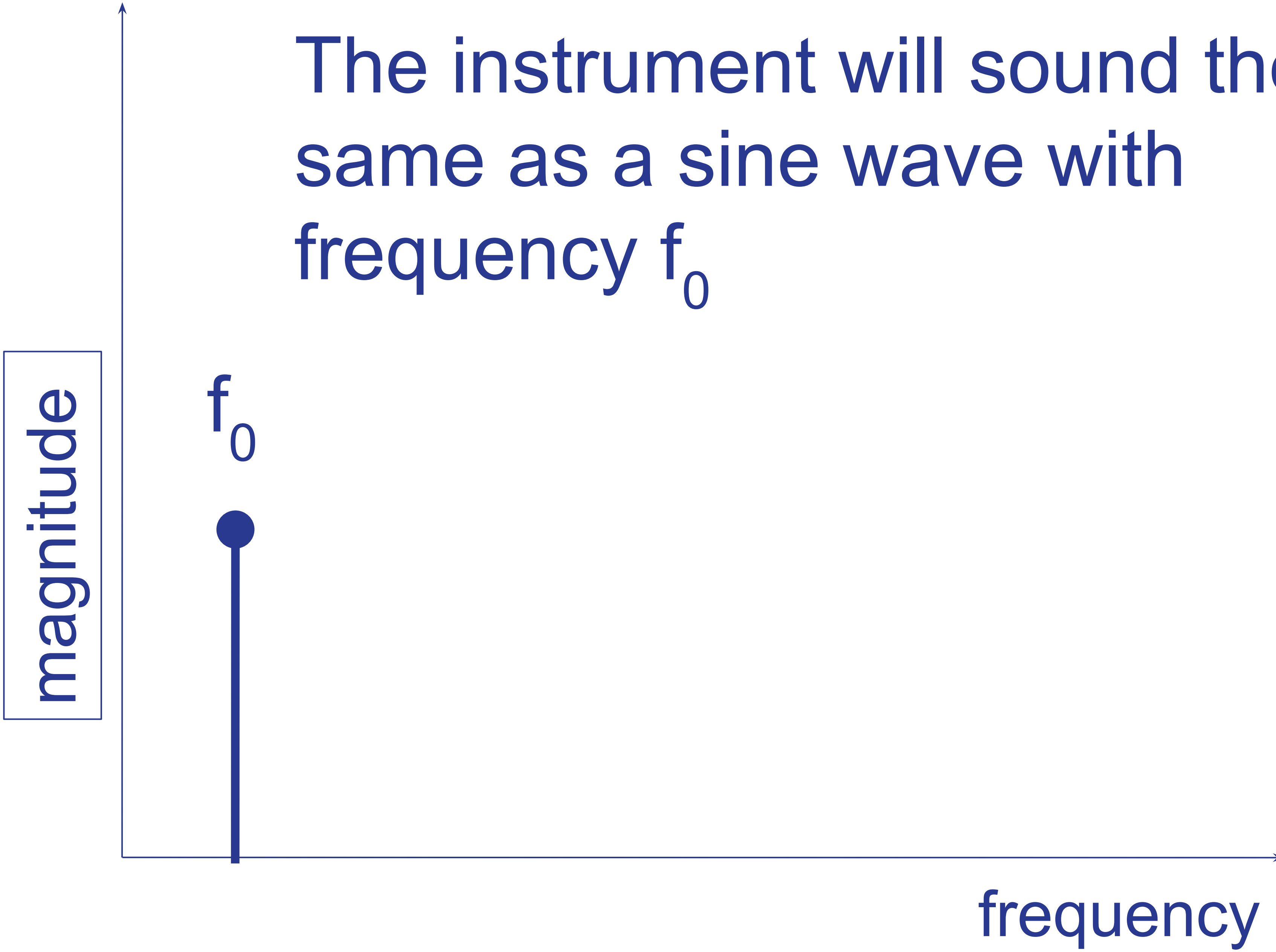


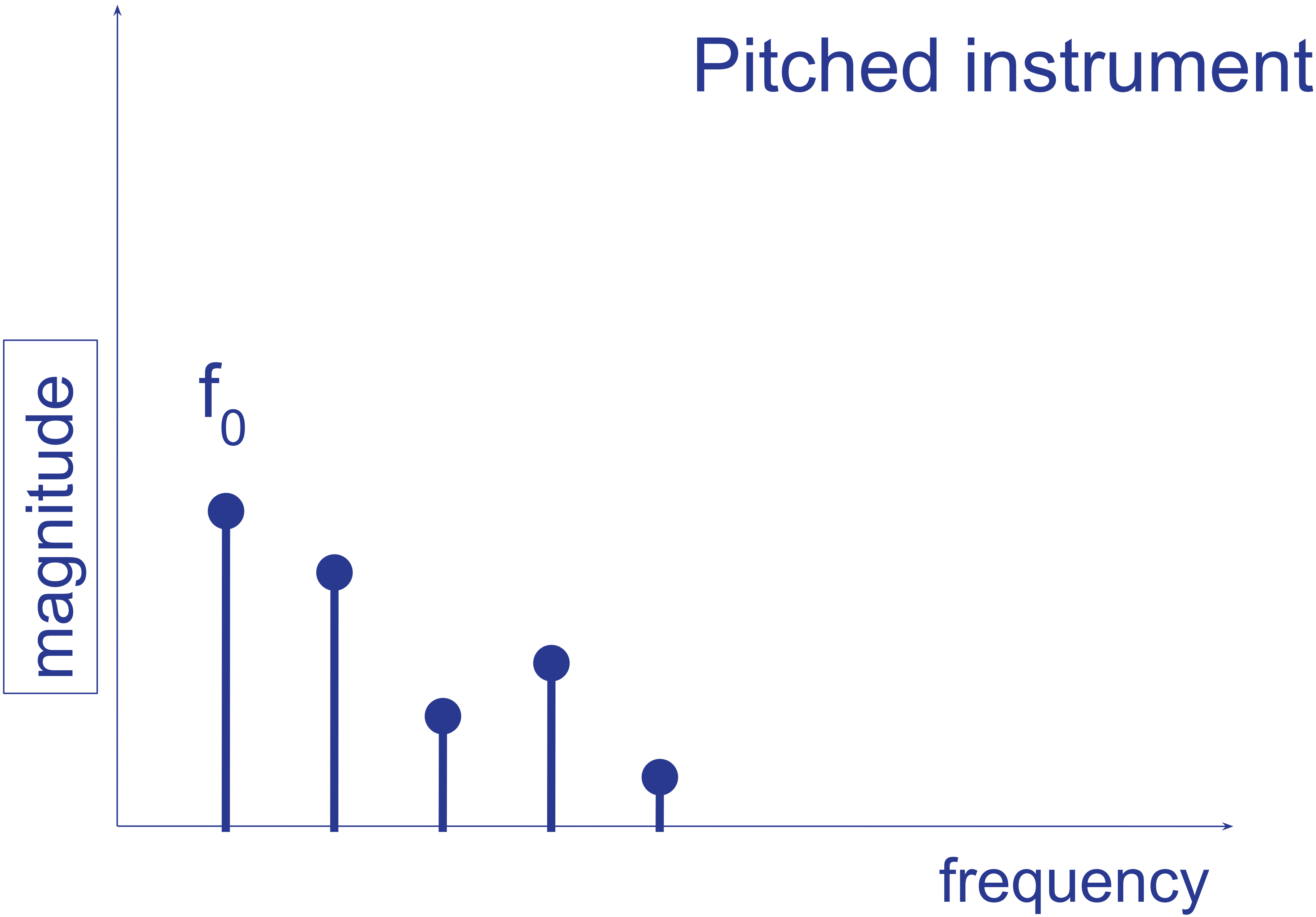
Violin



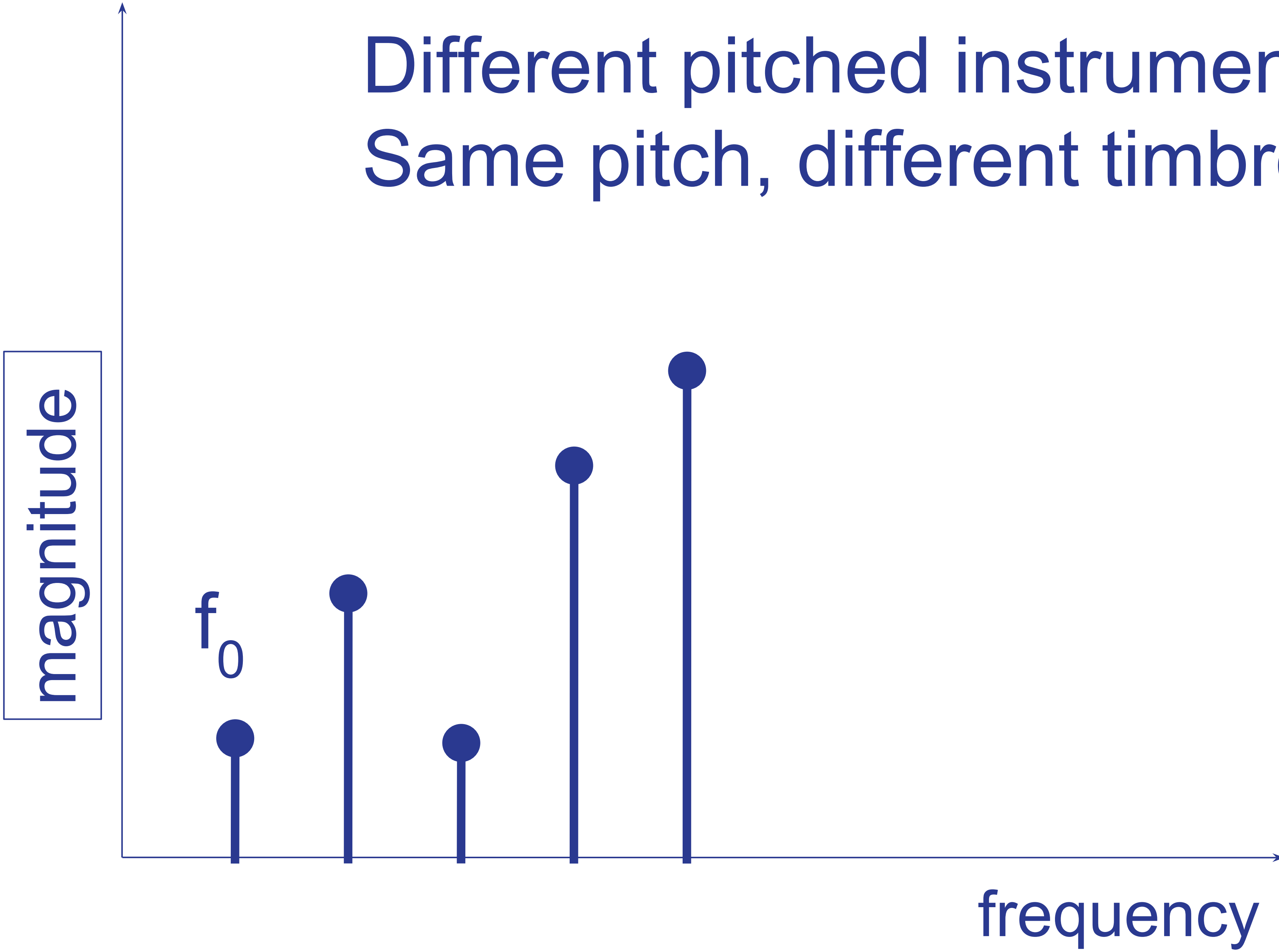

Pitched instrument
magnitude
f
0
frequency

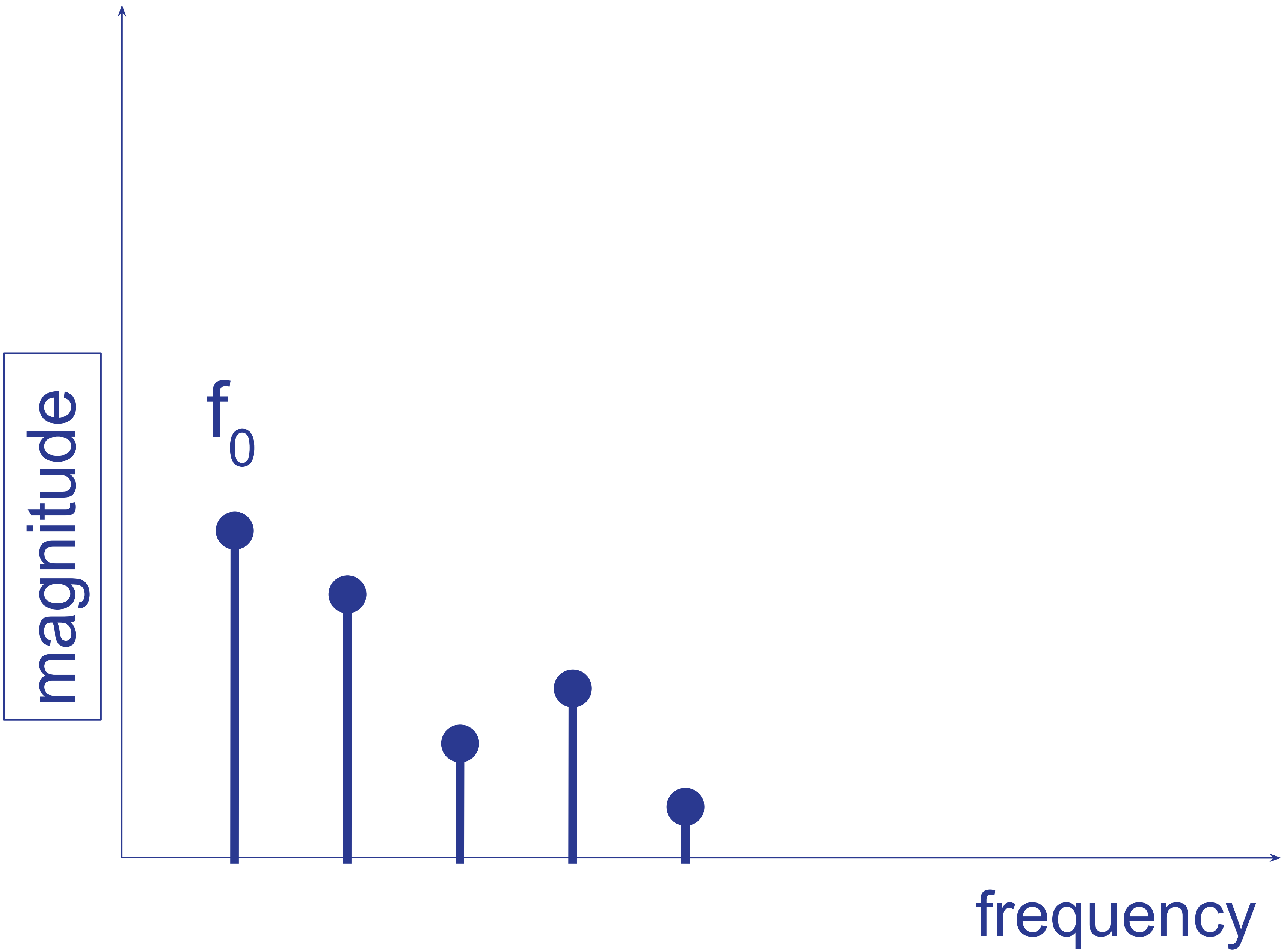
The instrument will sound the same as a sine wave with frequency f_0




Pitched instrument
magnitude
f_0
frequency

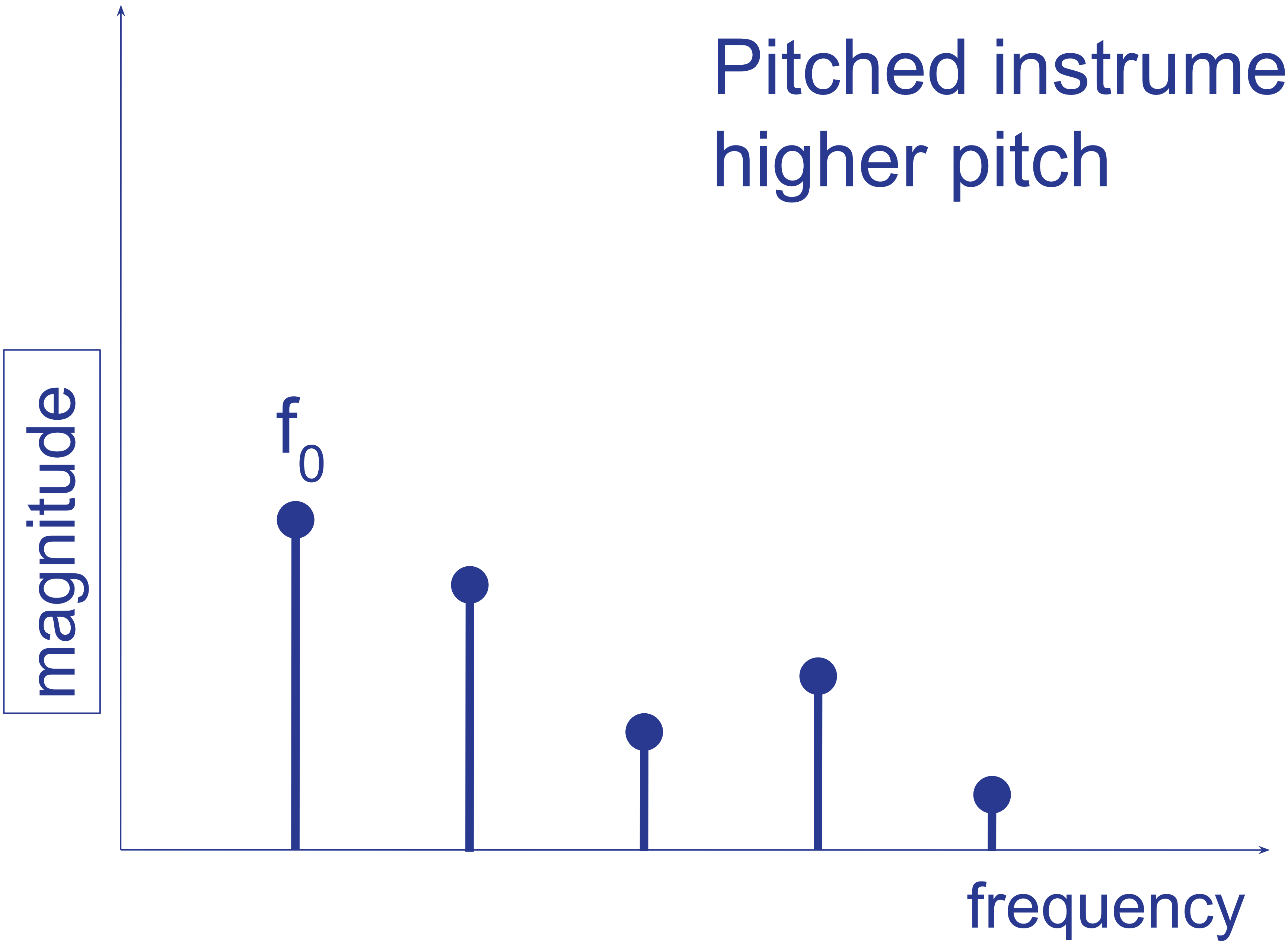
Different pitched instrument
Same pitch, different timbre





magnitude
f0
frequency

Pitched instrument
higher pitch

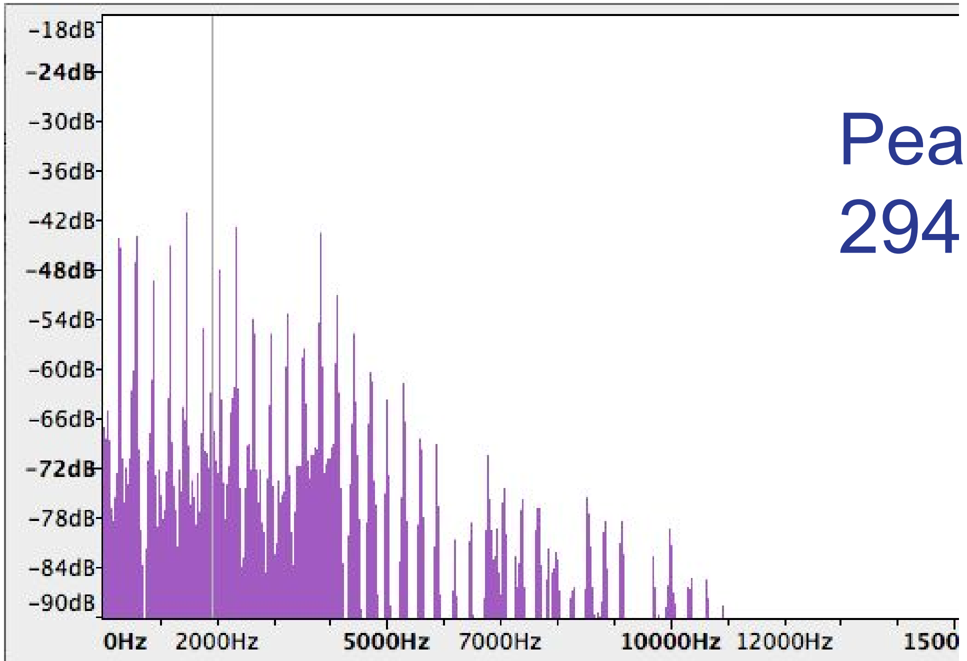


Rule of Thumb #3

The degree to which a sound's frequencies are harmonically related influences the degree to which we hear it as "pitched."

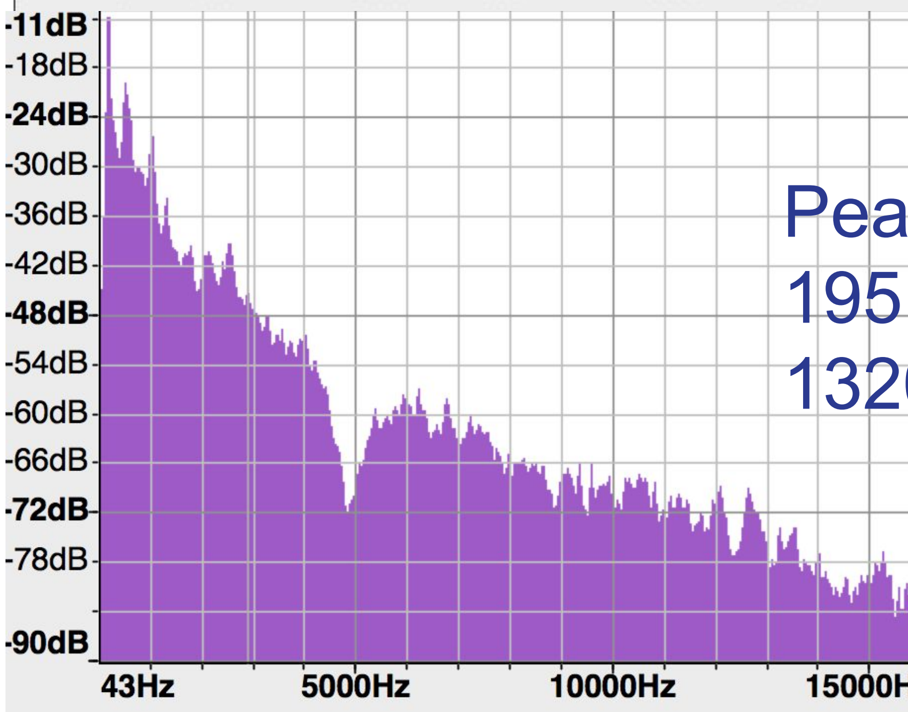
(less harmonically related = less pitched)

Violin



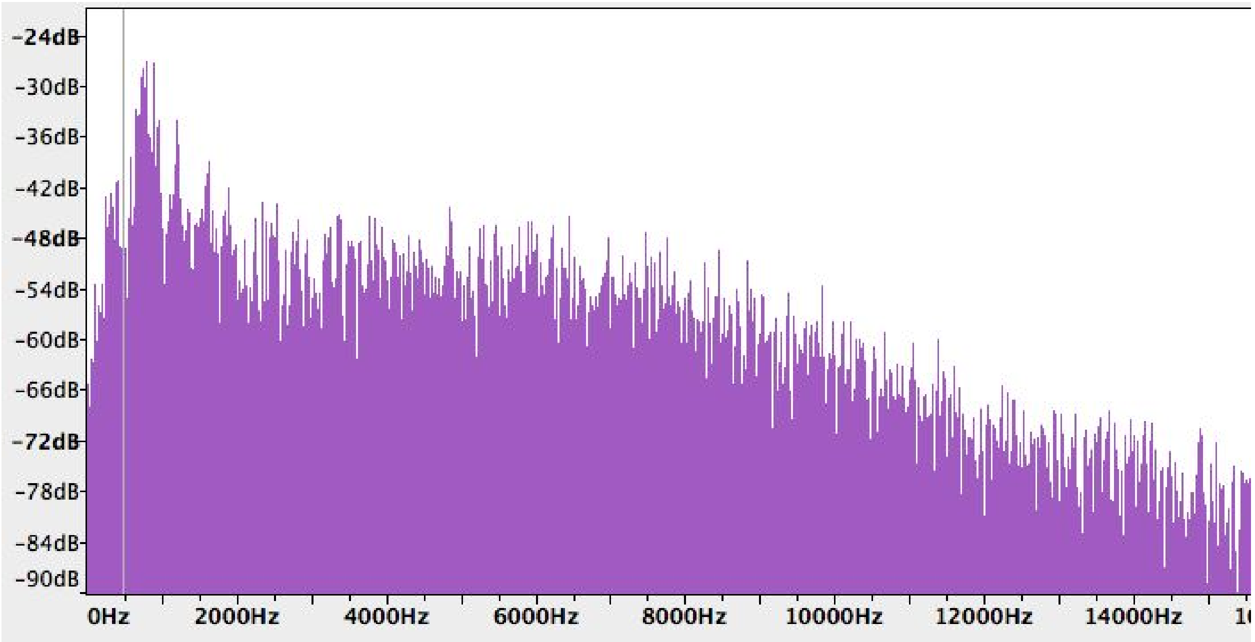
Peaks at approx.
294Hz, 589Hz, 884, …

Ciblon

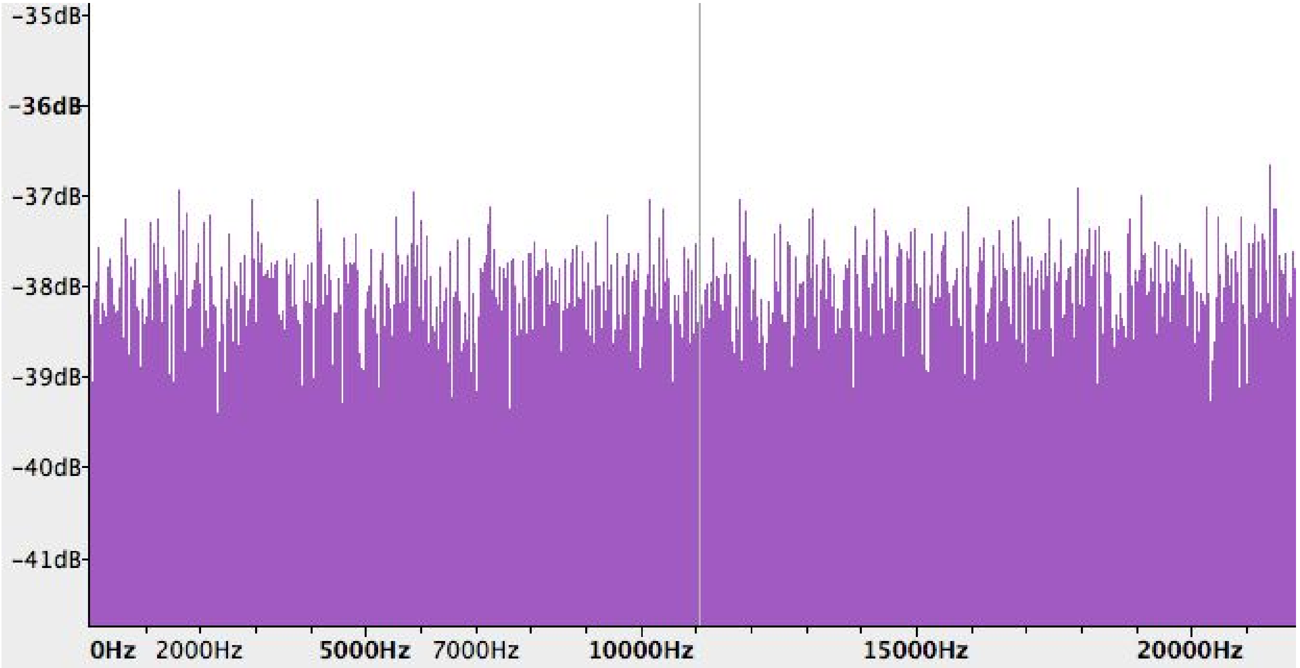


Peaks at approx.
195Hz, 522Hz,
1320Hz, 2560Hz…

Snare

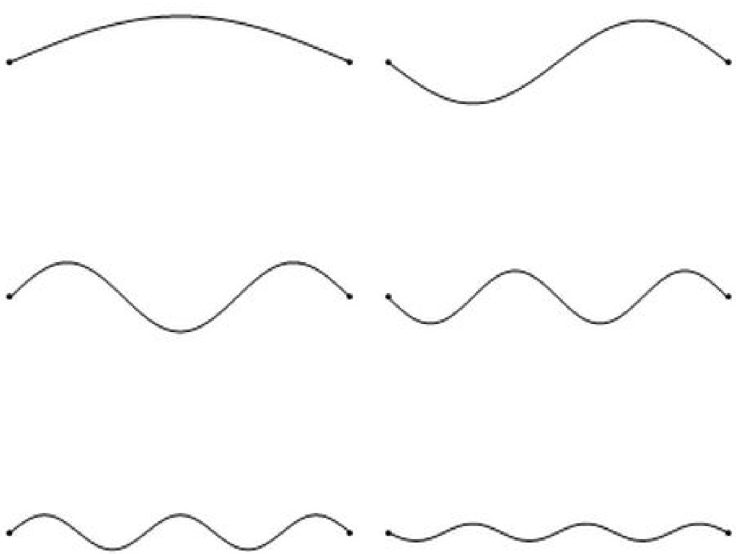


White
noise



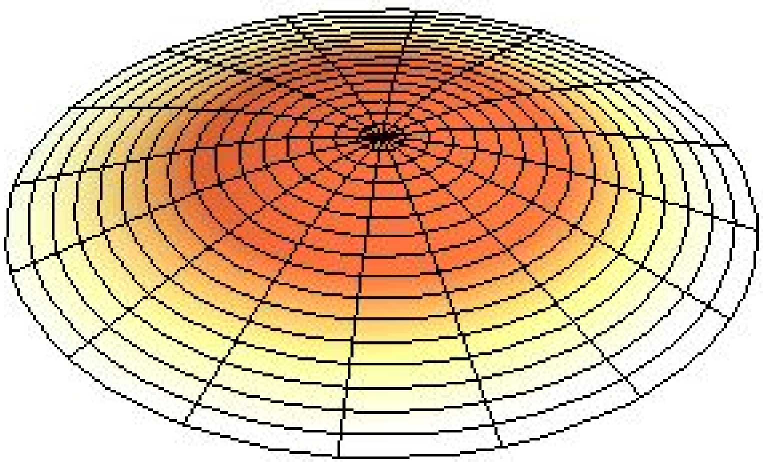
Why aren’t drums pitched?

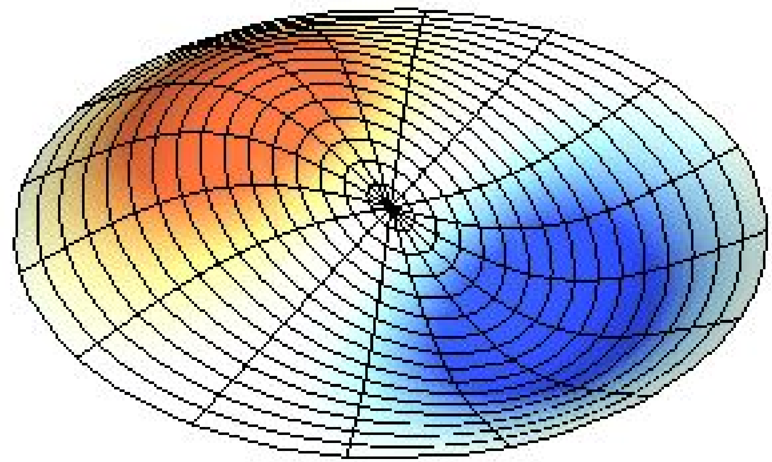
- Strings, air columns vibrate at harmonics:



- 2D surfaces do not (*inharmonicity)*

Modes on a drum head



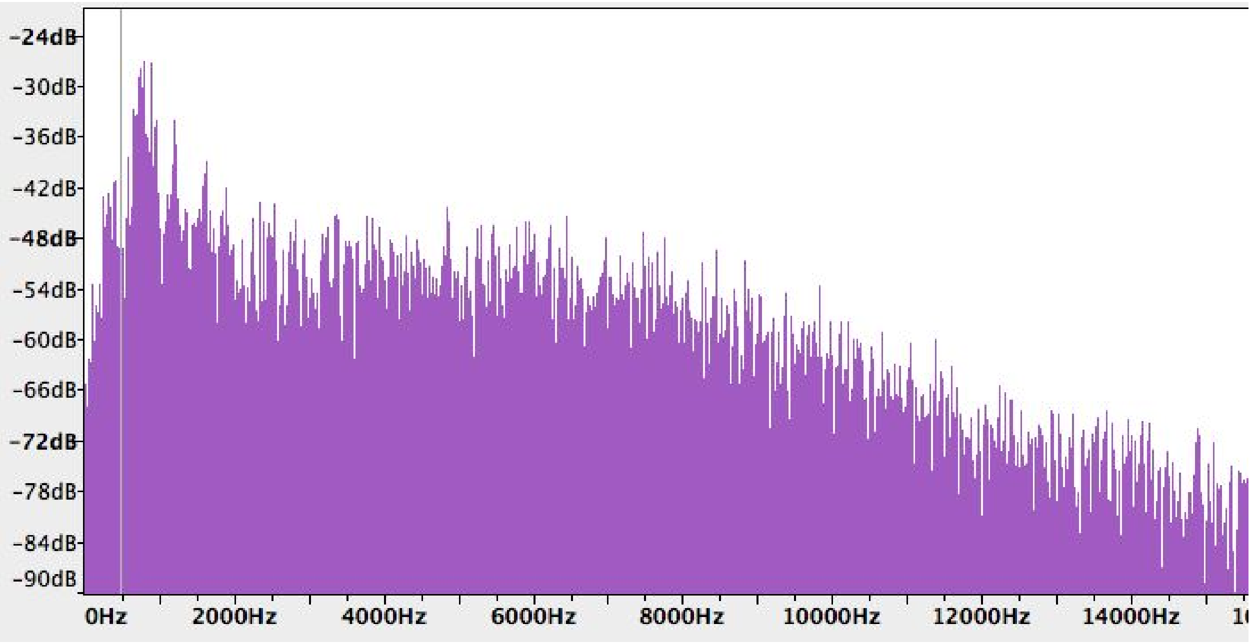


(and many more)

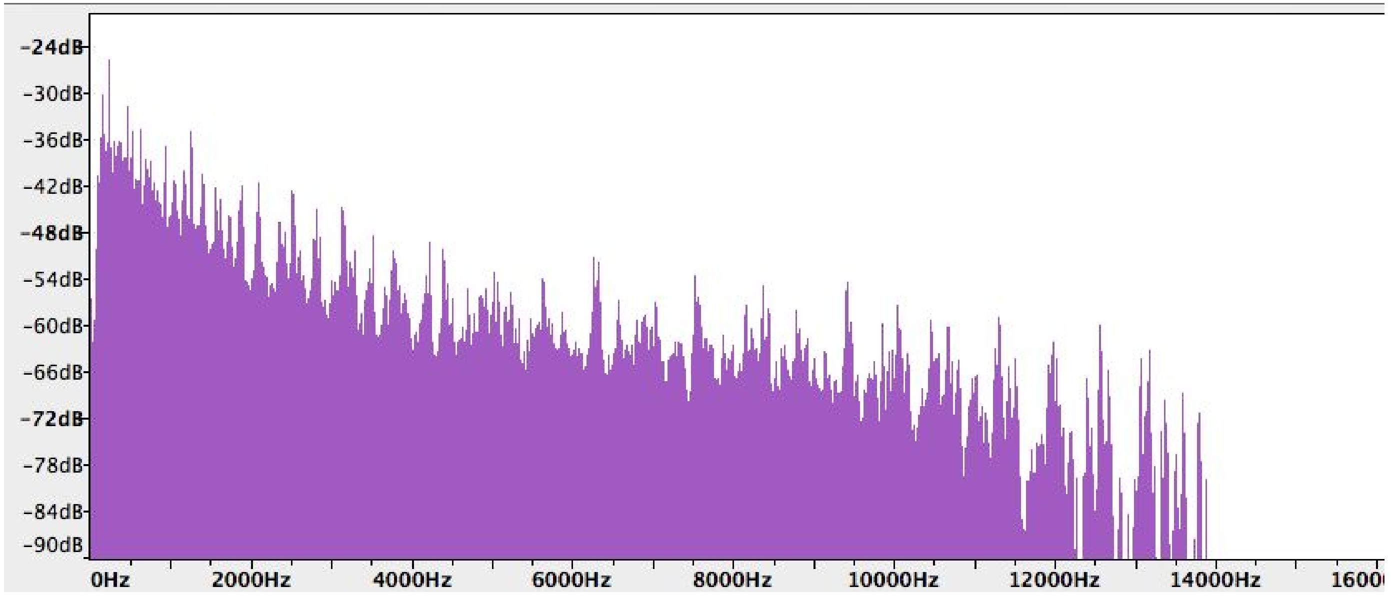
http://www.acs.psu.edu/drussell/demos/membranecircle/circle.html

One sound or many?

Snare



Orchestra



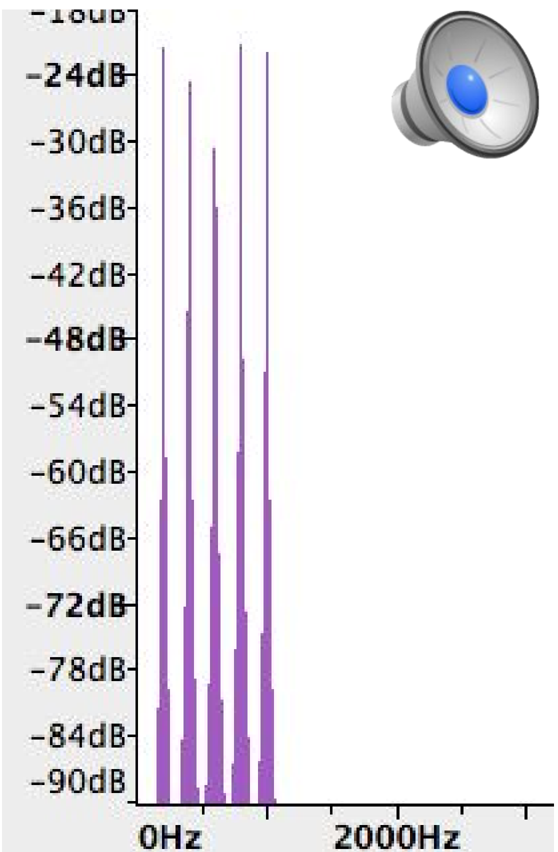
Single or multiple sound sources?

These make it more likely to hear a single sound ->

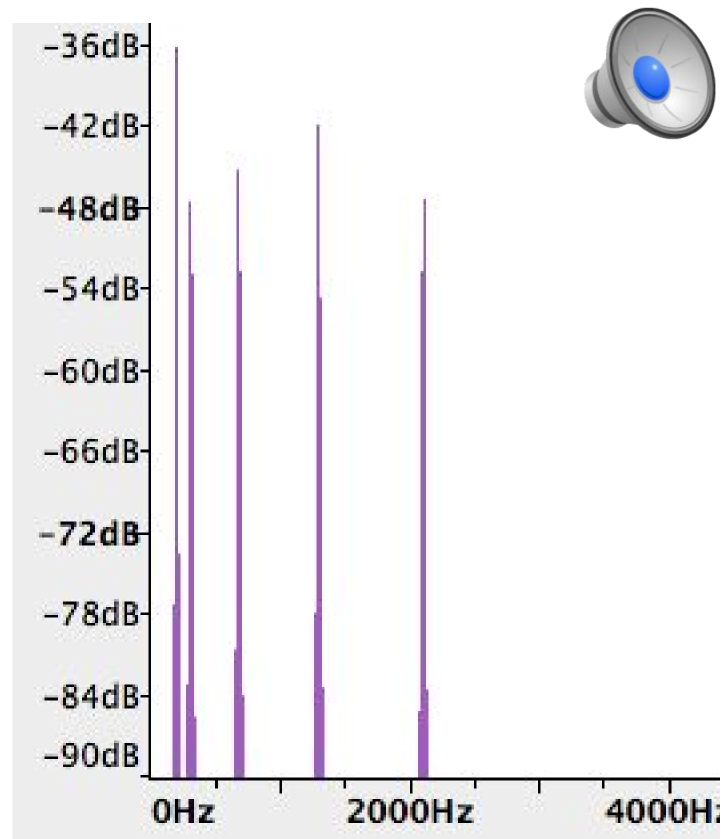
- Harmonic relationship
- Shared onset time
- Shared location
- Shared changes in amplitude (envelope)
- Shared changes in frequency (vibrato)

Harmonic relationship

Sound with 5
harmonically-related partials
(200, 400, 600, 800, 1000Hz)



Sound with 5
inharmonically-related partials
(200, 311, 682, 1300, 2109Hz)



Onset time (when does the sound begin?)

8 harmonics, shared onset time

Same 8 harmonics, different onset times

Shared location

8 Harmonics, 4 panned left and 4 panned right

4 moving left→ right, other 4 moving right → left

Same 4, all centre

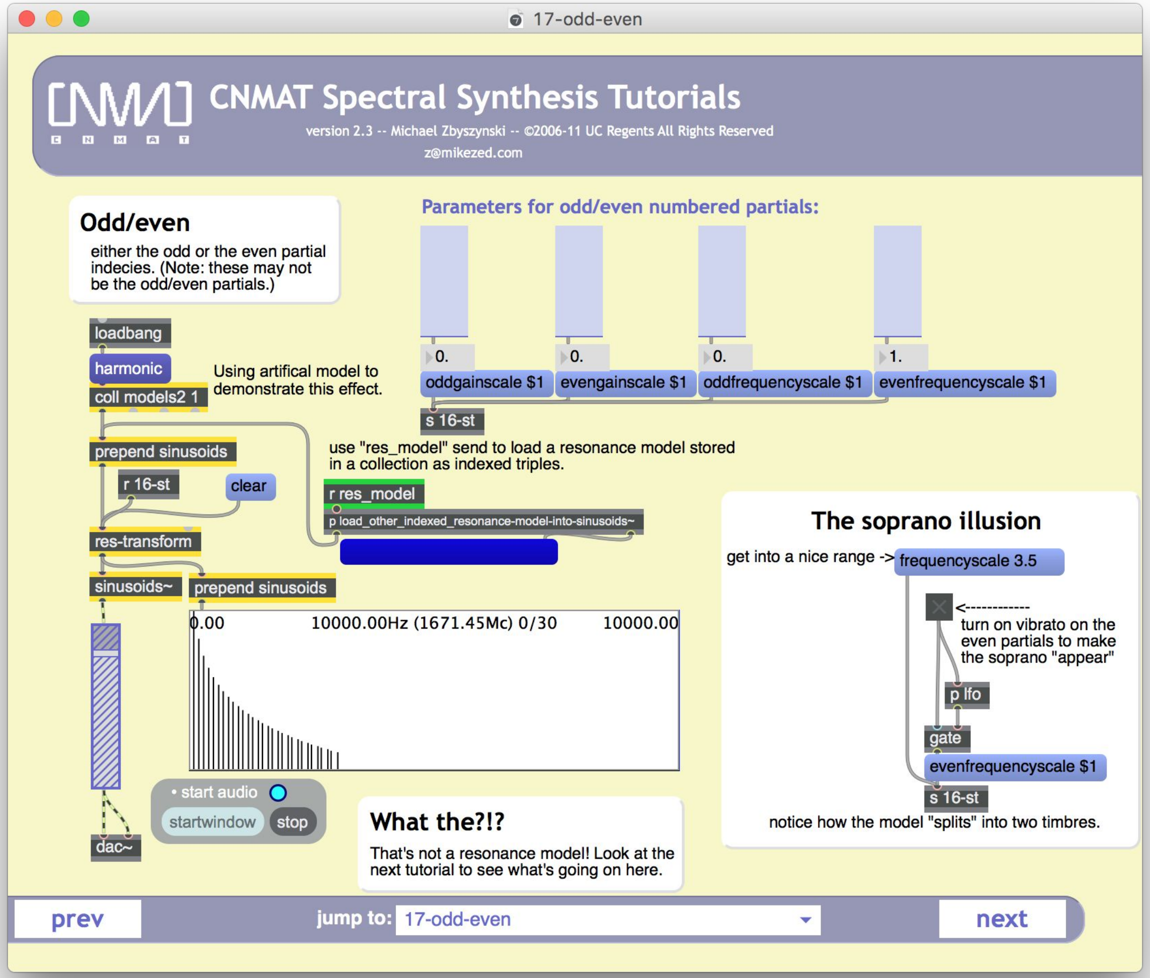
Shared changes in amplitude

Envelope:

Describes changes in the overall amplitude of a signal over time:

Envelope demos

Shared changes in frequency: vibrato demos



Vibrato demo: the soprano illusion

Single or multiple sound sources?

The following make it more likely to hear a single sound:

- Harmonic relationship
- Shared onset time
- Shared location
- Shared changes in amplitude (envelope)
- Shared changes in frequency (vibrato)

Compare to Gestalt principles of visual perception

Consonance & Dissonance

Rule of Thumb #4

Dissonance is caused by simultaneous frequencies that are close together

Two pitched sounds (sinusoids or complex waveforms) played simultaneously can be perceived as consonant or dissonant.

(not absolute binary, also has cultural dimensions)

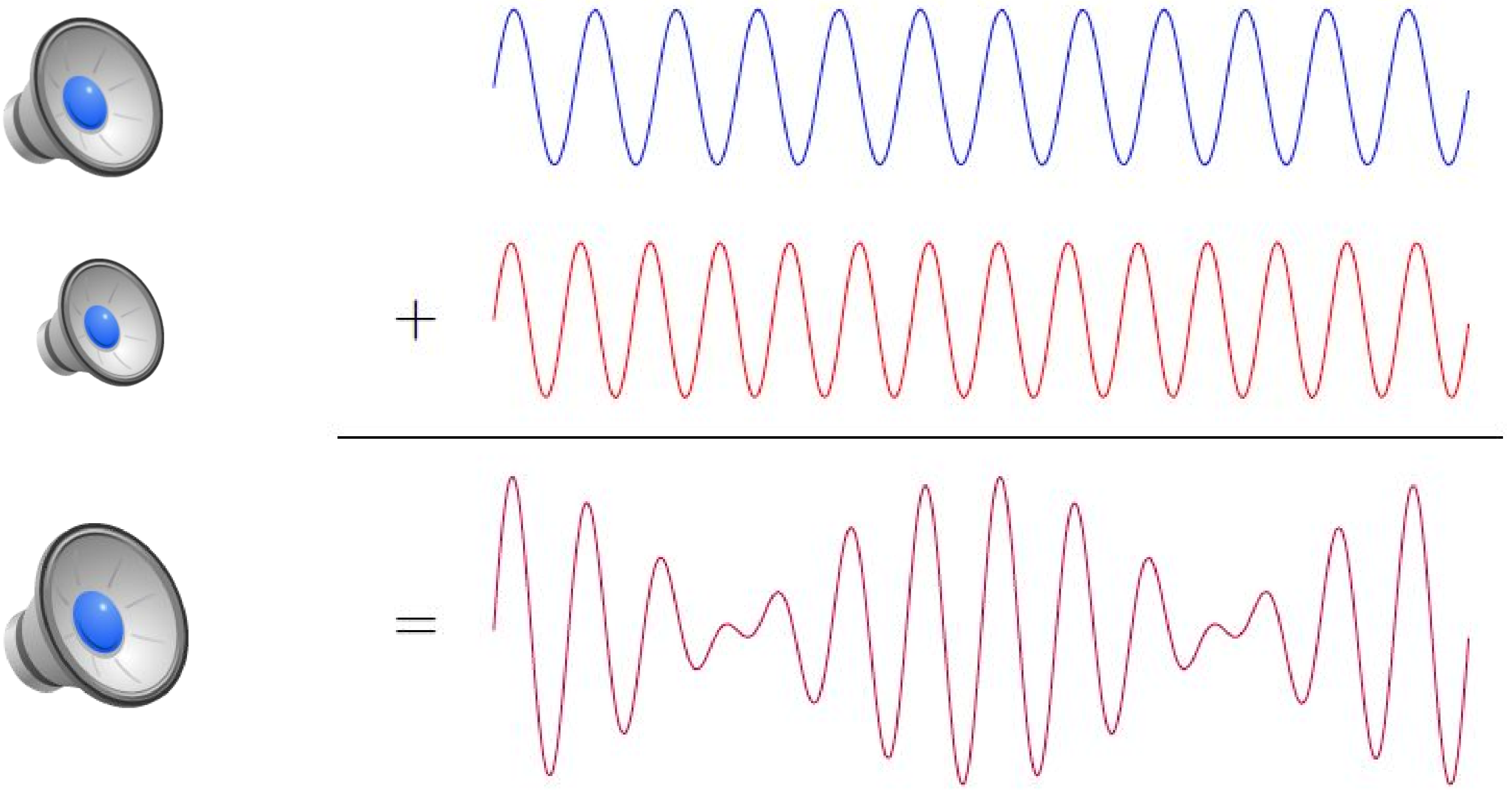
Perception impacted by:

- Relative pitch of sounds
- Absolute pitch of sounds
- Timbre of sounds

When frequencies are relatively close:

- Beating
- “Roughness”

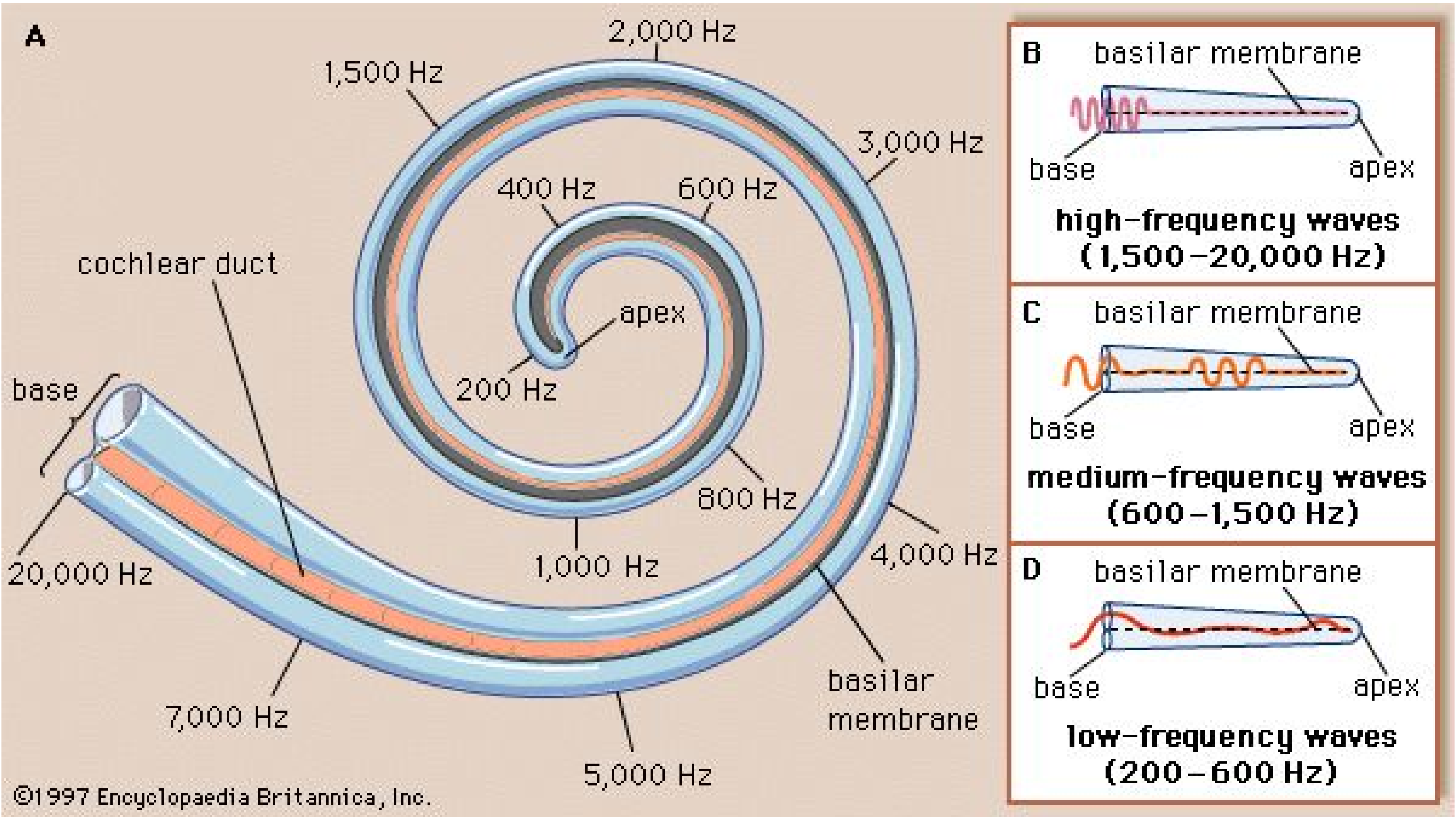
Beating (2 waves close in frequency)



$$\sin(A) + \sin(B) = 2\sin\left[\frac{A+B}{2}\right]\cos\left[\frac{A-B}{2}\right]$$

Audio examples

Basilar membrane



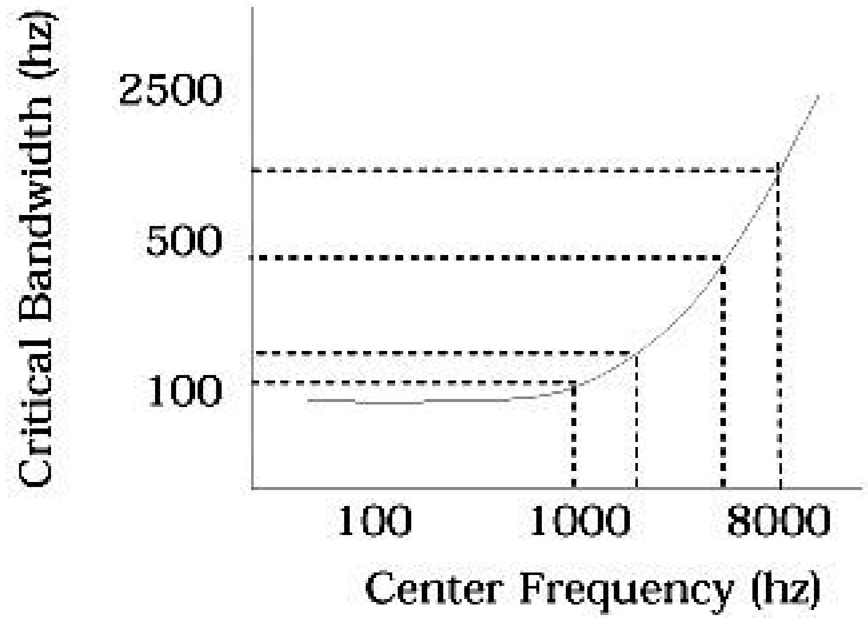
When two tones are close in frequency, they excite nearby locations on basilar membrane.

Critical band

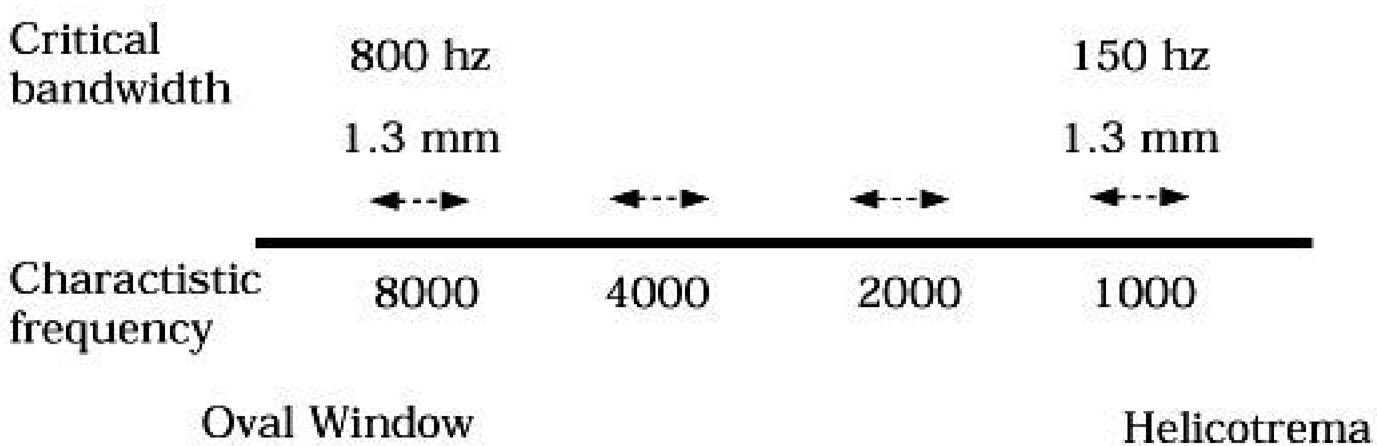
A range of frequencies around a given tone within which addition of a second tone will interfere with accurate perception of the original tone.

Two simultaneous tones with different frequencies but within same critical band will sound “dissonant” or “rough.”

Critical bands and the basilar membrane



Demo



Two pitched sounds (sinusoids or complex waveforms) played simultaneously can be perceived as consonant or dissonant.

(not absolute binary, also has cultural dimensions)

Perception impacted by:

- Relative pitch of sounds
- Absolute pitch of sounds
- Timbre of sounds

When sounds aren’t just sinusoids

- Do harmonics/partials line up?
- Or do they fall within same critical bands, without lining up exactly?

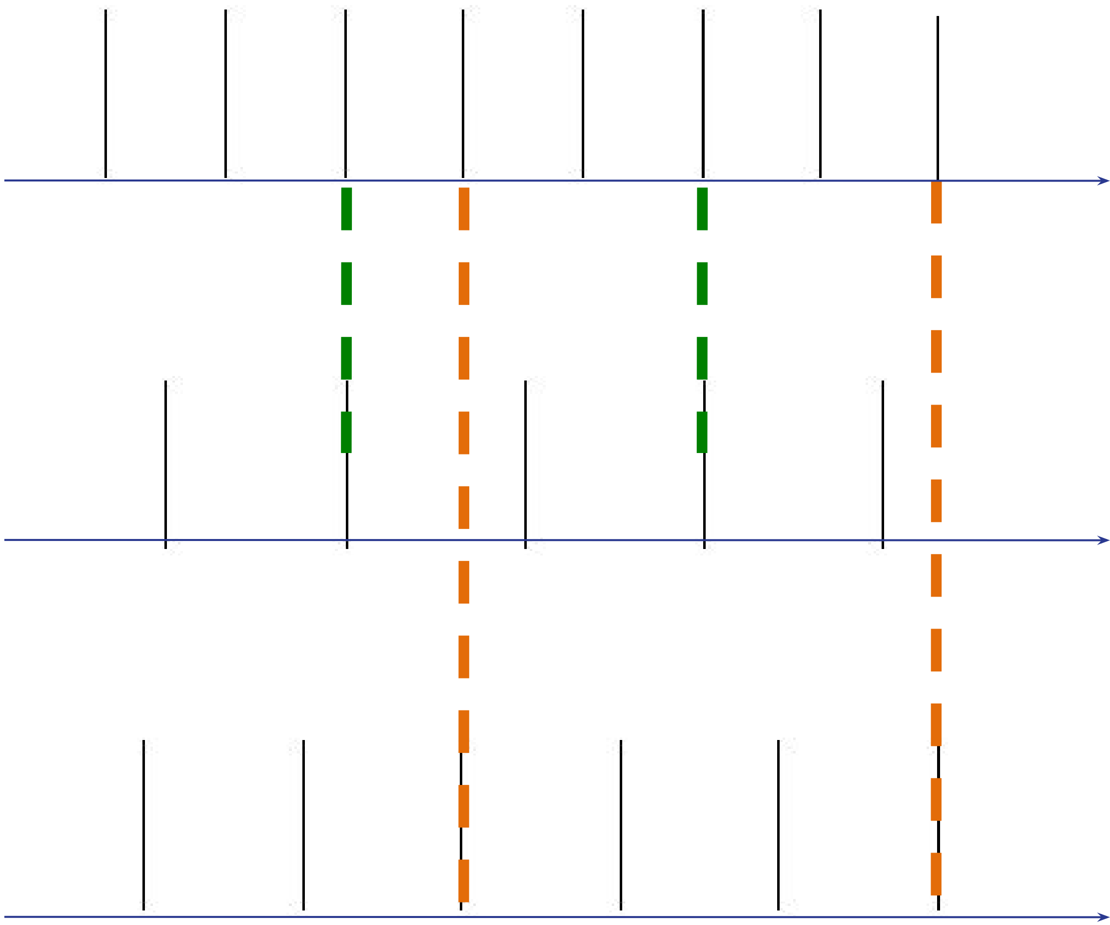
Demo

Consonant intervals reinforce each other

Fundamental f0 + harmonics

Fundamental (3/2*f0) + its harmonics

Fundamental (4/3*f0) + its harmonics



IMPORTANT FOR EXAM

This is a text
the exam, do
signal percep
decompositio
you will do or
exam copies
exams. For ex
perception w

u take
about
oidal
etter
at past
those
d

Basic principle

We perceive it as unpleasant when our ability to accurately sense something is interfered with!

Speech analysis & perception

Human speech

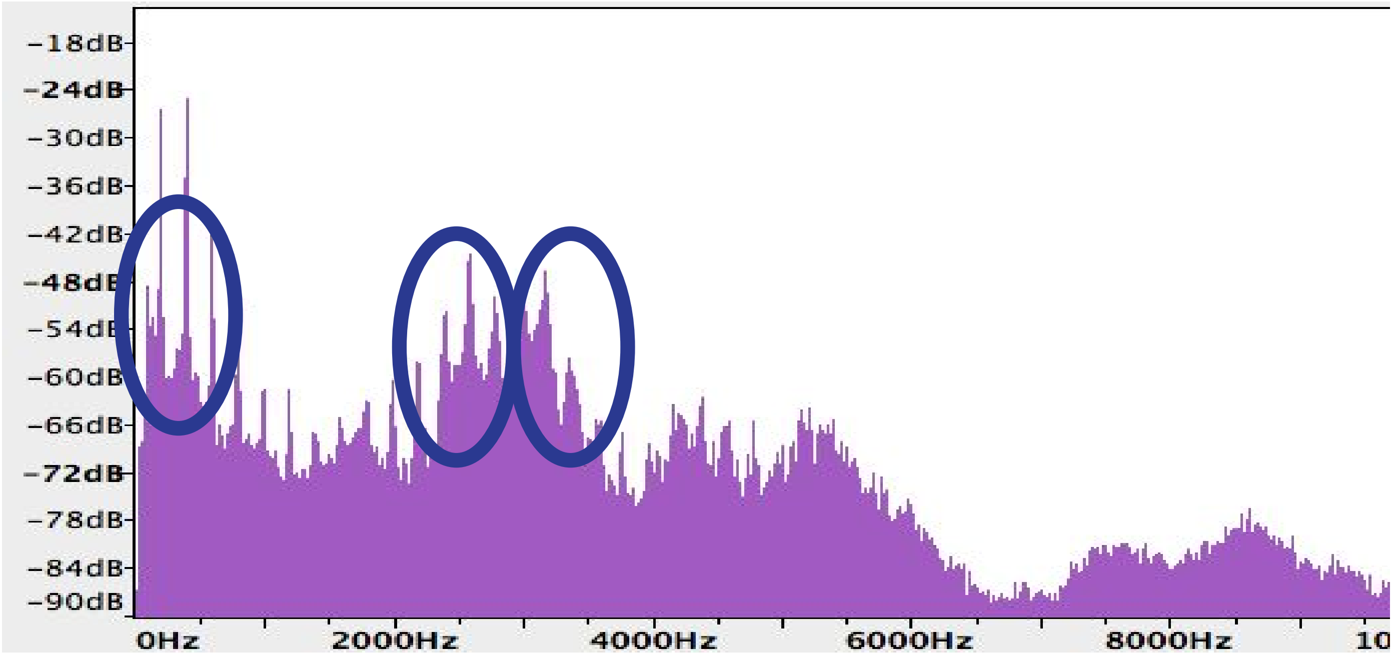
Listen to vowels: What do you hear?

- Constant pitch, volume
- Changing “tone quality”

Rule of Thumb #5

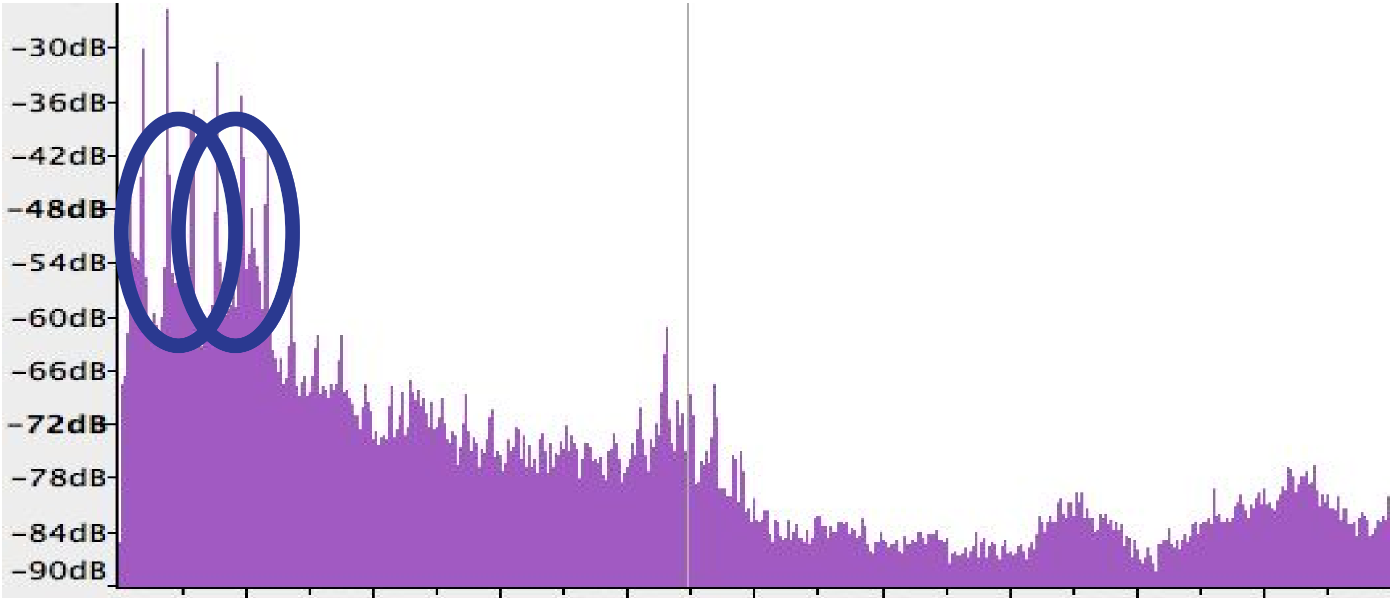
Different vowels are distinguished by relative strengths of particular frequency ranges (“formants”)

EEEE



-18dB
-24dB
-30dB
-36dB
-42dB
-48dB
-54dB
-60dB
-66dB
-72dB
-78dB
-84dB
-90dB
0Hz
2000Hz
4000Hz
6000Hz
8000Hz
10

OOOO



-30dB
-36dB
-42dB
-48dB
-54dB
-60dB
-66dB
-72dB
-78dB
-84dB
-90dB

Formants

Different vowels exhibit greater magnitude in different regions of the frequency spectrum.

Formants

The first two formants are sufficient to distinguish vowel sound.

Consonants

No definite pitch

(unperiodic, inharmonically related partials)

Still distinguishable by frequency content

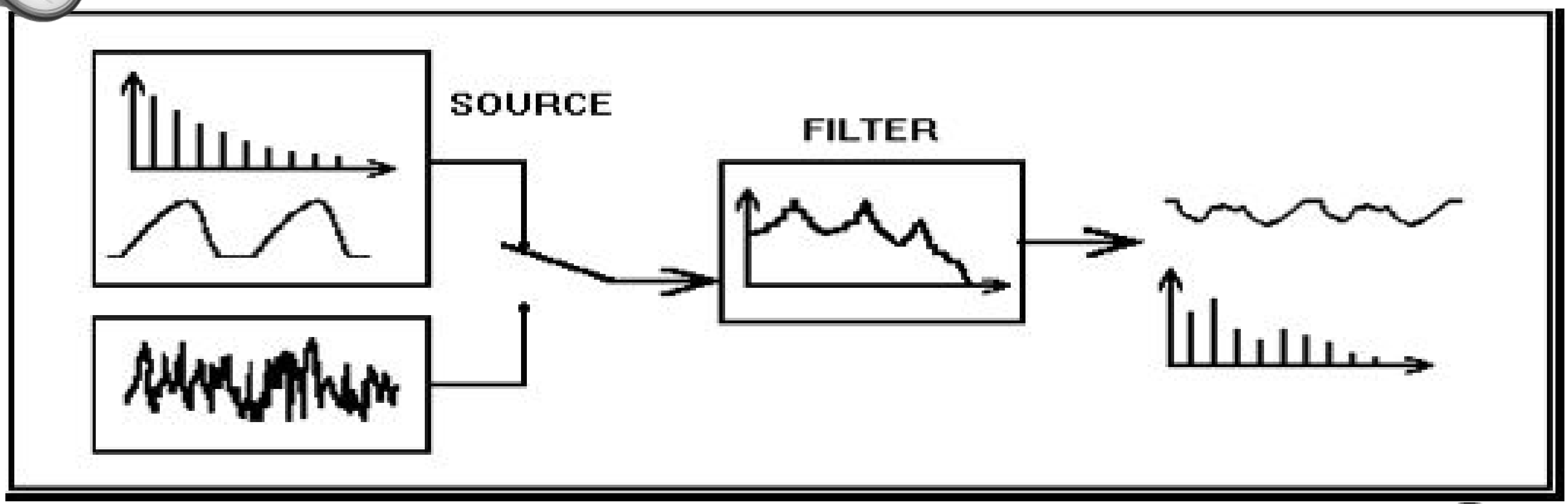
Demo: sndpeek

Singer’s Formant

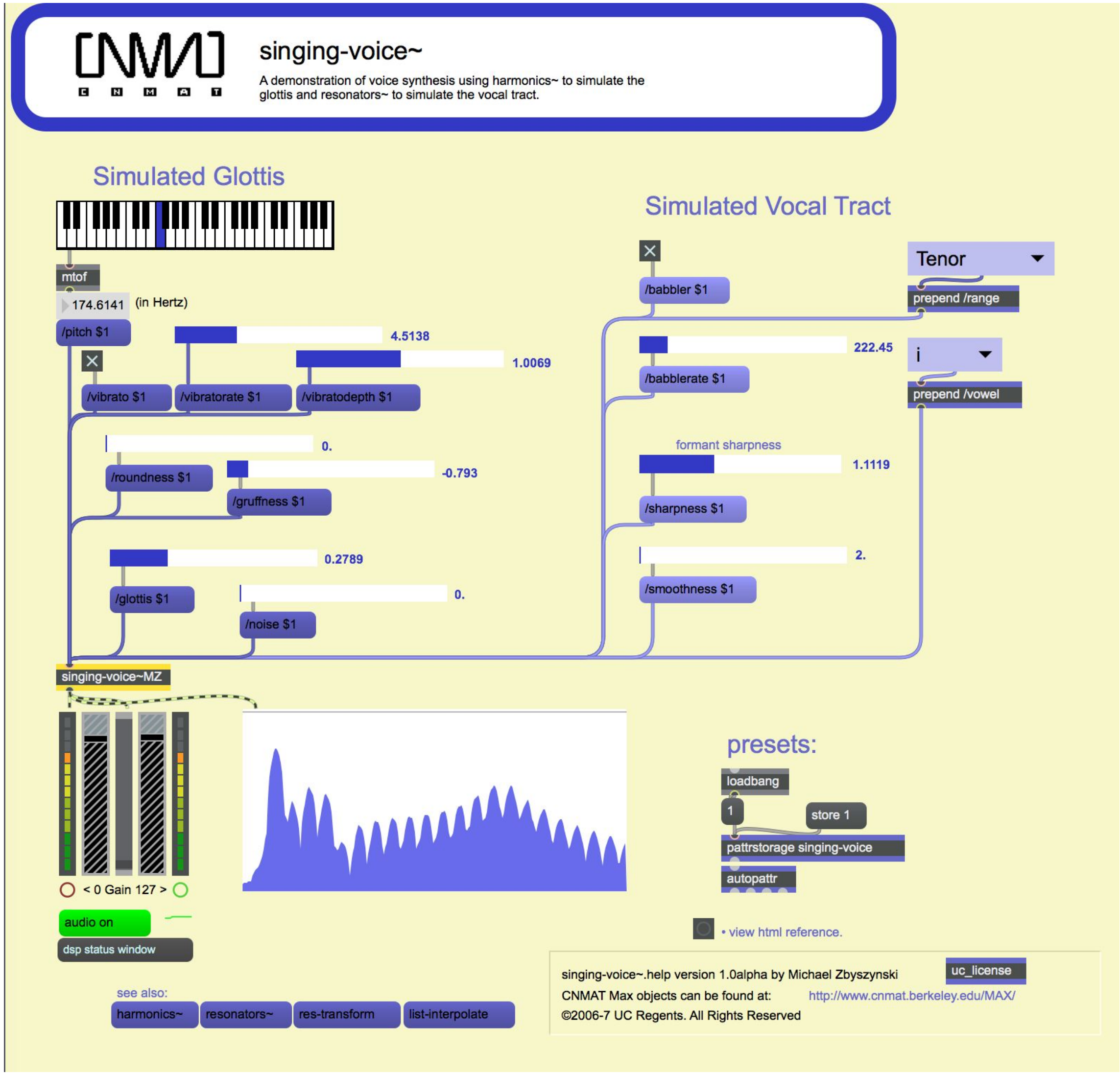
Trained singers have additional formant around 3000Hz.

That allows a singer to be heard above orchestra!

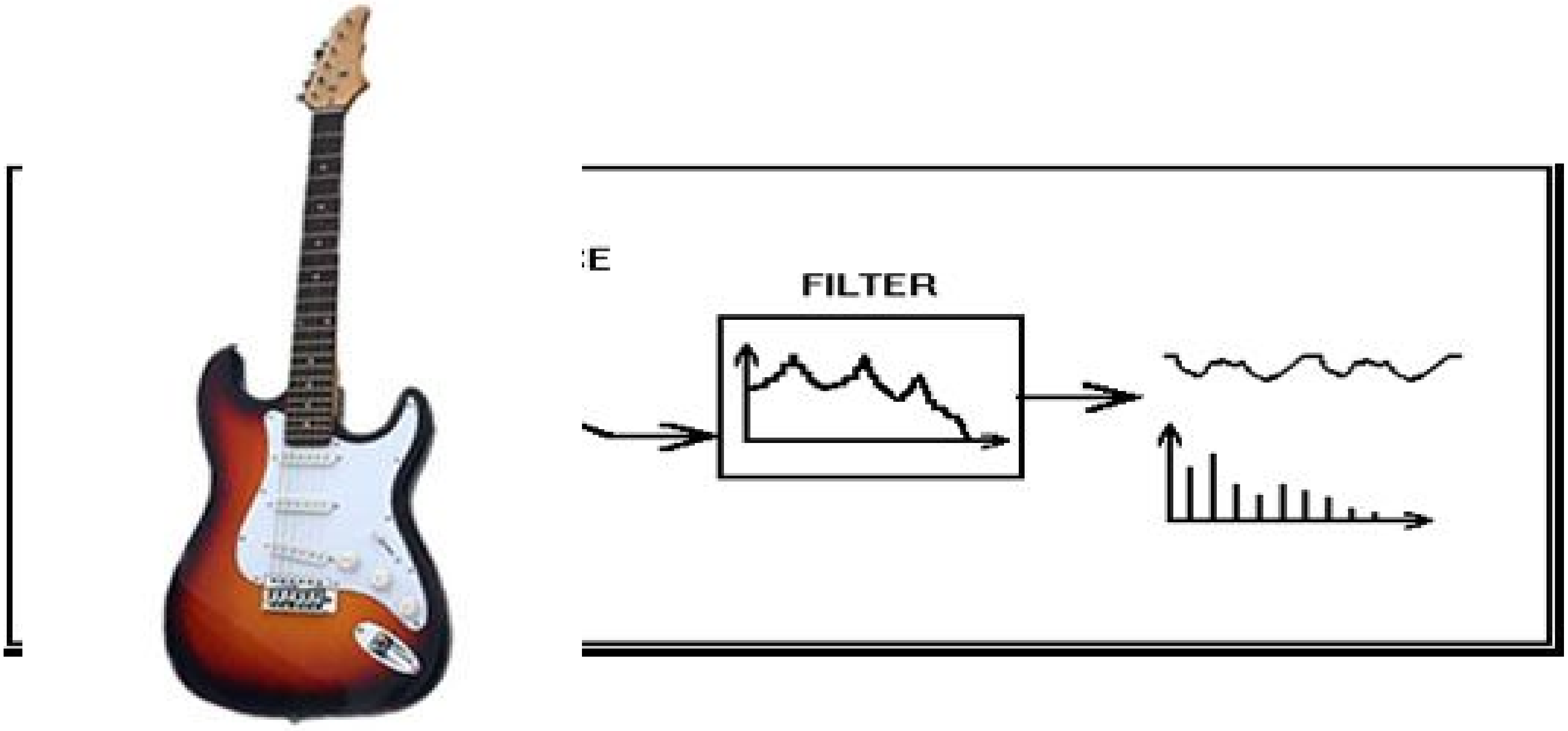
Source-filter model of voice synthesis & analysis



Singing voice demo



Cross-synthesis



Demo:
guitar source spectrum shaped by voice spectrum

Practical applications

- Speech synthesis
- Speech as spectral manipulation
- Compression
- Auto-tune

Speech perception also has a visual component

Demo: McGurk effect

http://www.youtube.com/watch?v=jtsfidRq2tw

Wrap-up

Implications of sinusoidal decomposition

- Synthesis
- Compression
- Processing

Implications of audio percetion principals

- Synthesis
- Compression
- Processing

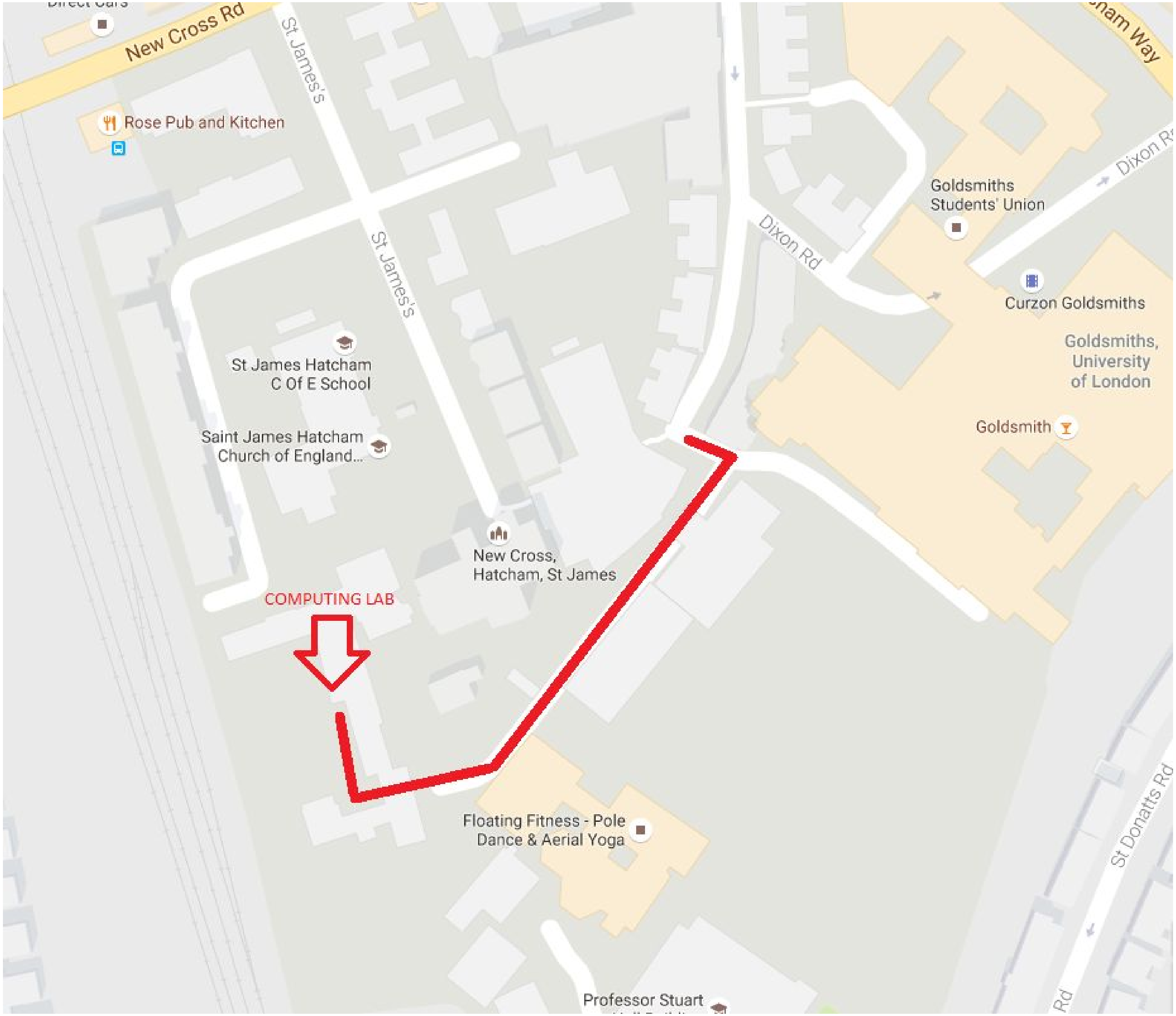
Exam preparation

Can you write an essay on ______ perception?

Can you look at a spectrum and reason about what it will sound like?

Can you talk about how sinusoidal decomposition is useful?

Lab on Monday at 3pm!


New Cross Rd
St James's
Rose Pub and Kitchen
St James's
Dixon Rd
Dixon Rd
Goldsmiths Students' Union
Curzon Goldsmiths
Goldsmiths, University of London
Goldsmith
St James Hatcham C Of E School
Saint James Hatcham Church of England...
New Cross, Hatcham, St James
COMPUTING LAB
Floating Fitness - Pole Dance & Aerial Yoga
Professor Stuart
St Donatts Rd