Perception & Multimedia Computing

Week 13 – Moar Fourier; spectral perception

Michael Zbyszyński
Lecturer, Department of Computing
Goldsmiths University of London

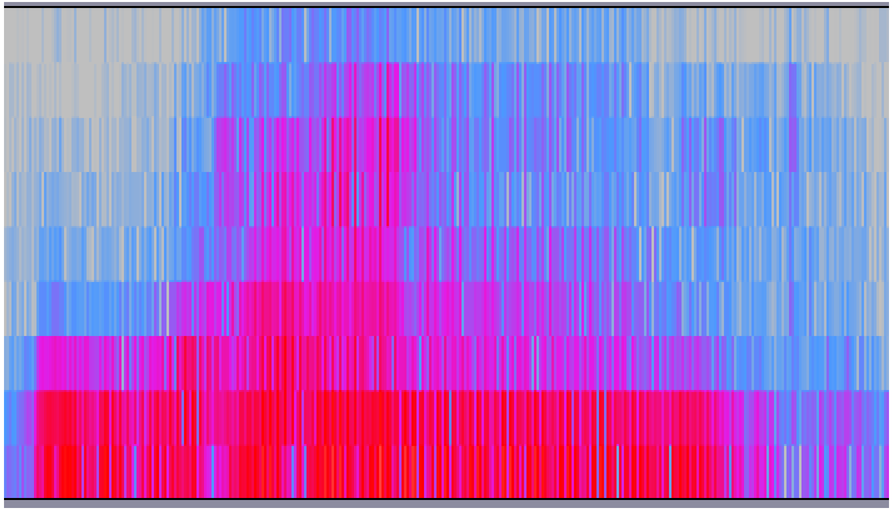
Fourier Analysis, review & conclusion

Why Fourier analysis?

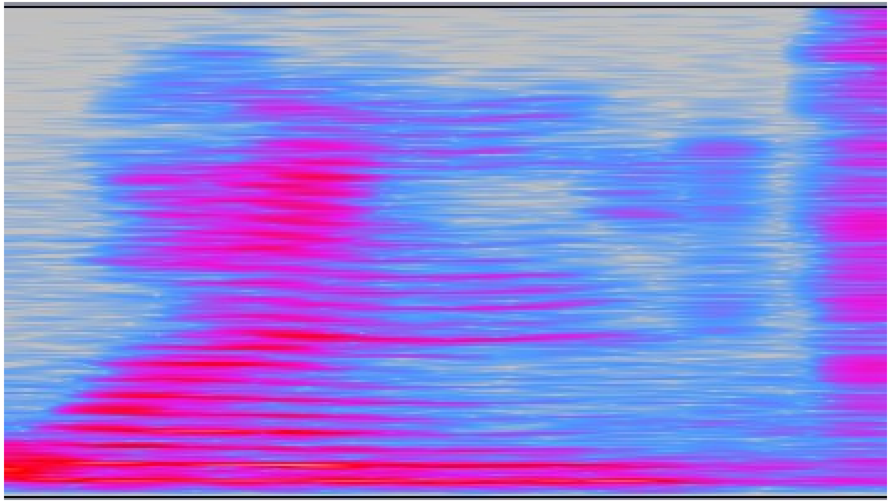
Audio:

- Tells us about pitch, timbre, instrumentation, mastering; speech/speaker; recording environment; …
- Re-synthesize and process sounds (e.g. time stretch, pitch shift)
- Reason about how filters, reverb, EQ, etc. will affect a sound
- *Design* filters, reverb, EQ, etc.

Time/Frequency tradeoff



N=64



N=4096

https://live.codecircle.com/d/2uqCQkLbNn6GB5AH9

How many bins to use? (What should *N* be?)

More bins?

- Better frequency resolution
- Worse time resolution (FFT can't detect changes within the analysis frame)

Fewer bins?

- Worse frequency resolution
- Better time resolution

The Discrete Fourier Transform

$$X_k = \sum_{n=1}^{N} x[n] \times e^{-i2\pi(k/N)n}$$

x is our input signal (e.g. our audio signal)

```
void DFT::idft1()
{
    double pi2 = 2.0 * M_PI;
    double angleTerm,cosineA,sineA;
    double invs = 1.0 / size;
    for(unsigned int y = 0;y < size;y++) {
        output_seq[y] = 0;
        for(unsigned int x = 0;x < size;x++) {
            angleTerm = pi2 * y * x * invs;
            cosineA = cos(angleTerm);
            sineA = sin(angleTerm);
            output_seq[y].real += input_seq[x].real * cosineA -
input_seq[x].imag * sineA;
            output_seq[y].imag += input_seq[x].real * sineA +
input_seq[x].imag * cosineA;
        }
        output_seq[y] *= invs;
    }
}
```

Discussion: Computation time

Recall: Summation is like a for-loop in code

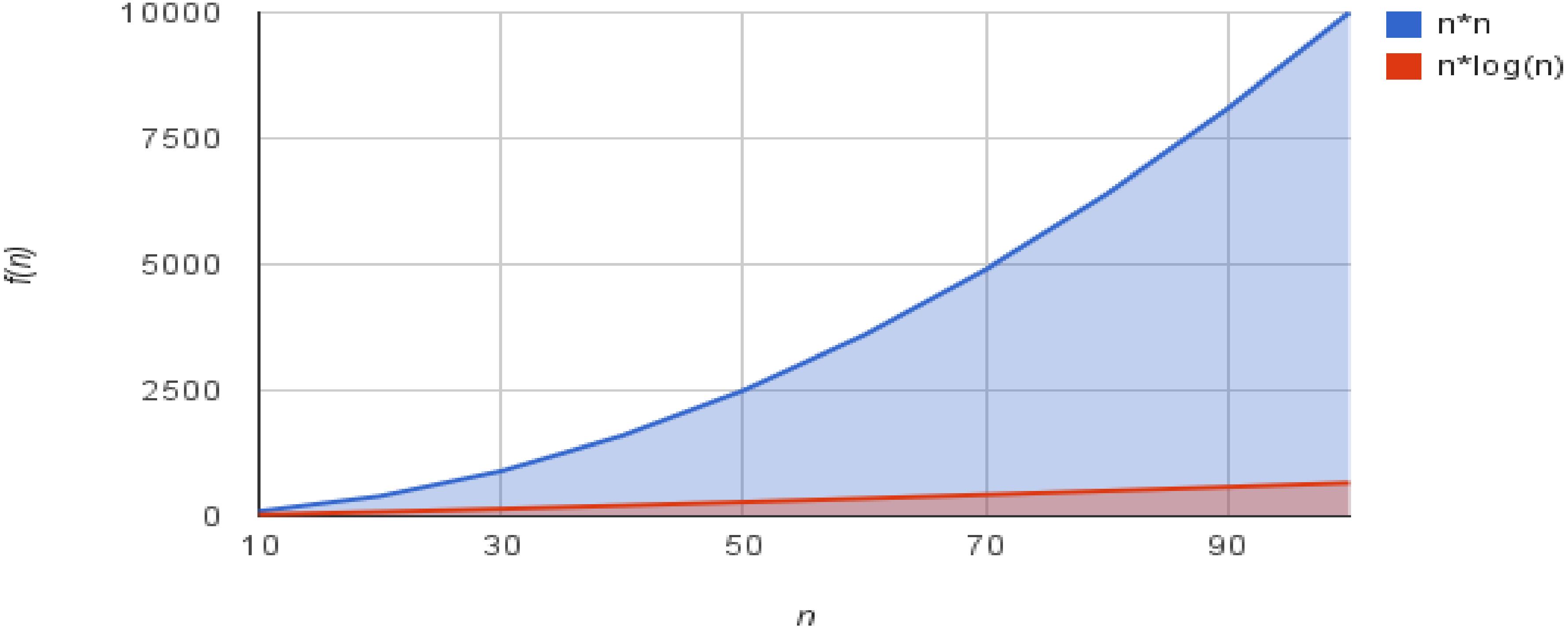
We do *N* summations, each with *N* iterations

A faster way

The FFT :

- FAST Fourier transform ☺
- Computes the same results, but using fewer computations
- $O(N \log N)$ rather than $O(N^2)$
- Fastest when N is a power of 2

$O(N^2)$ vs $O(N \log N)$



From http://java.dzone.com/articles/algorithm-week-insertion-sort

Using frequency content to reason about audio perception

Rule of Thumb #1

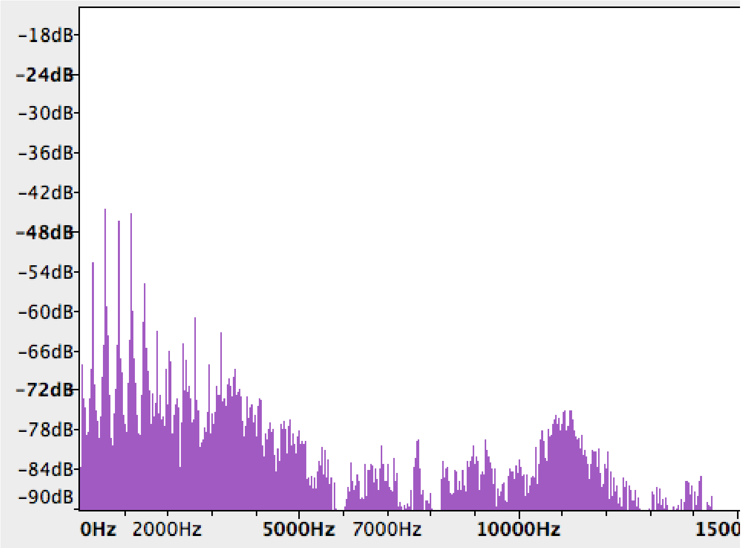
Pitched sounds have sinusoidal components that are harmonically related.

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.

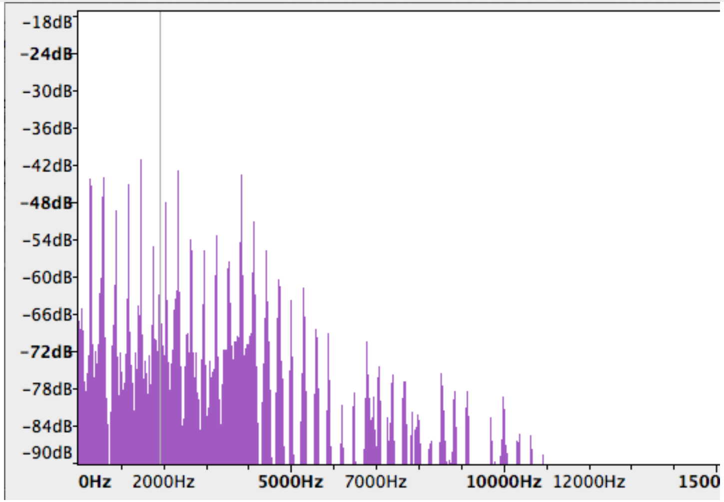
Rule of Thumb #2

The pitch we hear is determined by the fundamental frequency

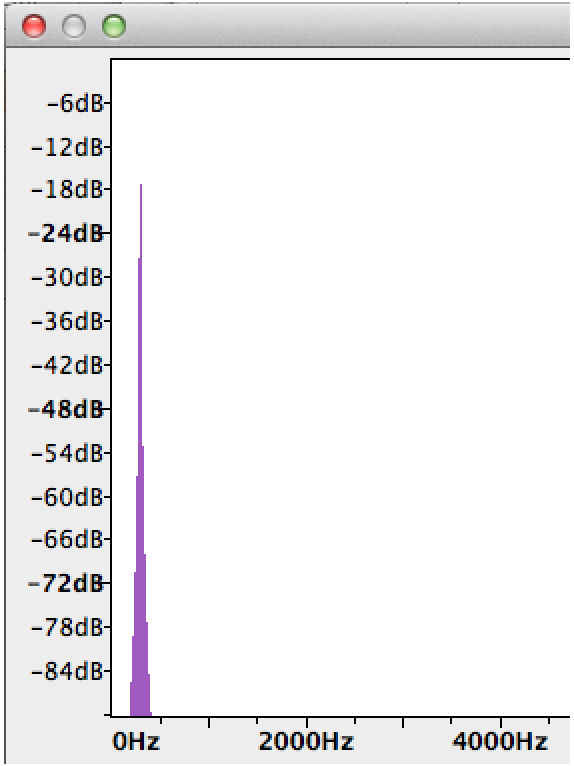
Flute playing "D" above middle C



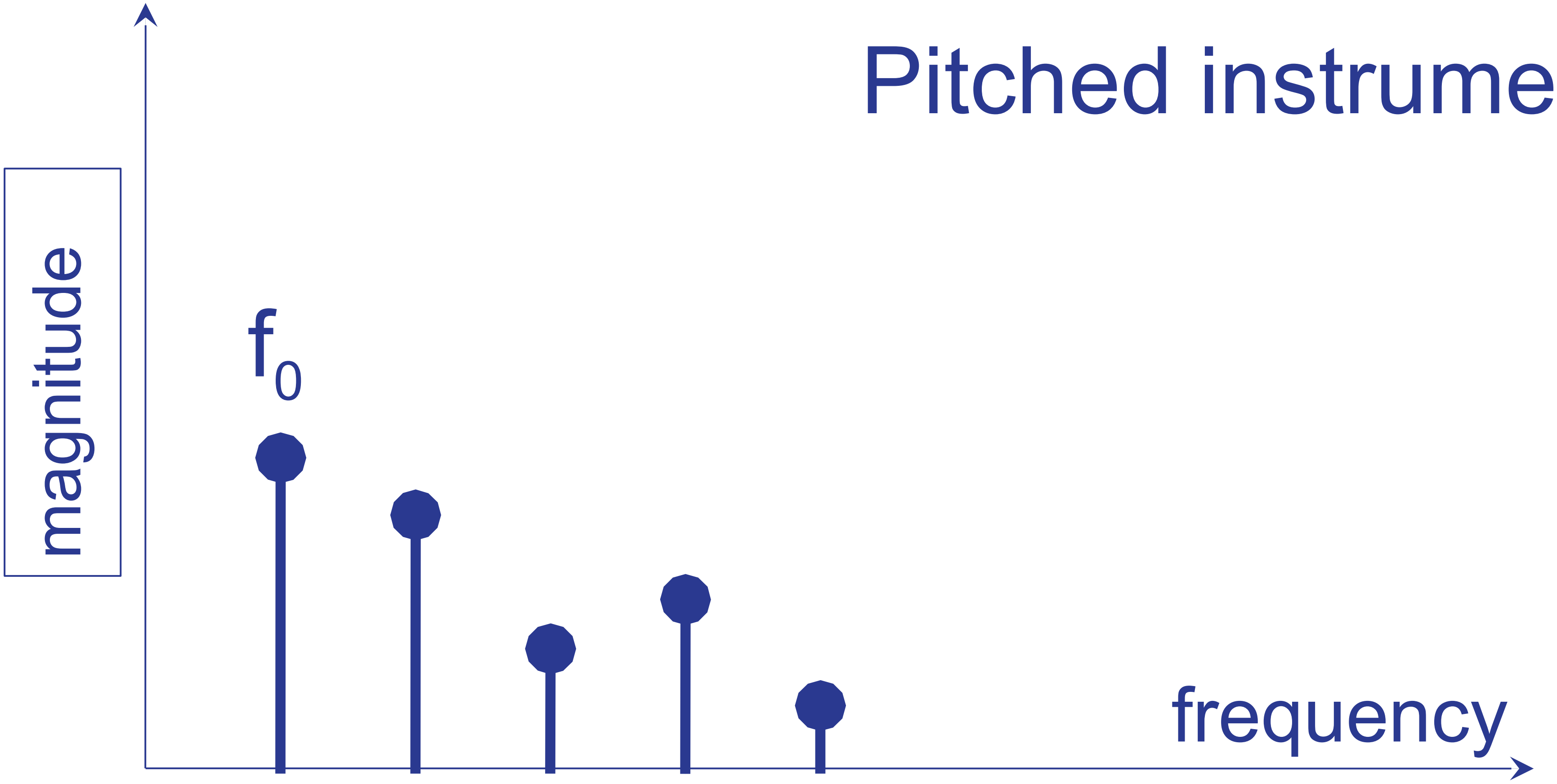
Violin playing same note



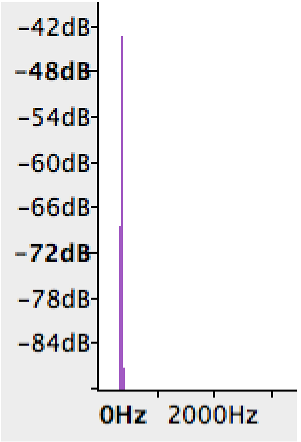
Sine wave at 294 Hz



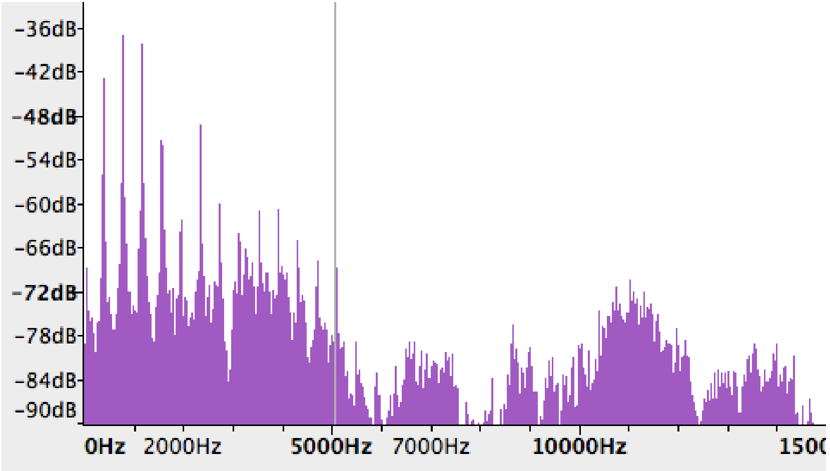
Pitched instrument



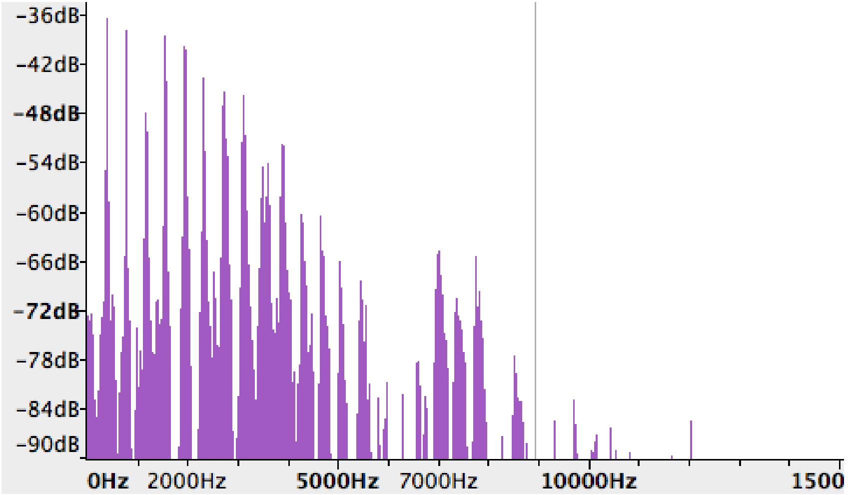
Sine
at 392 Hz



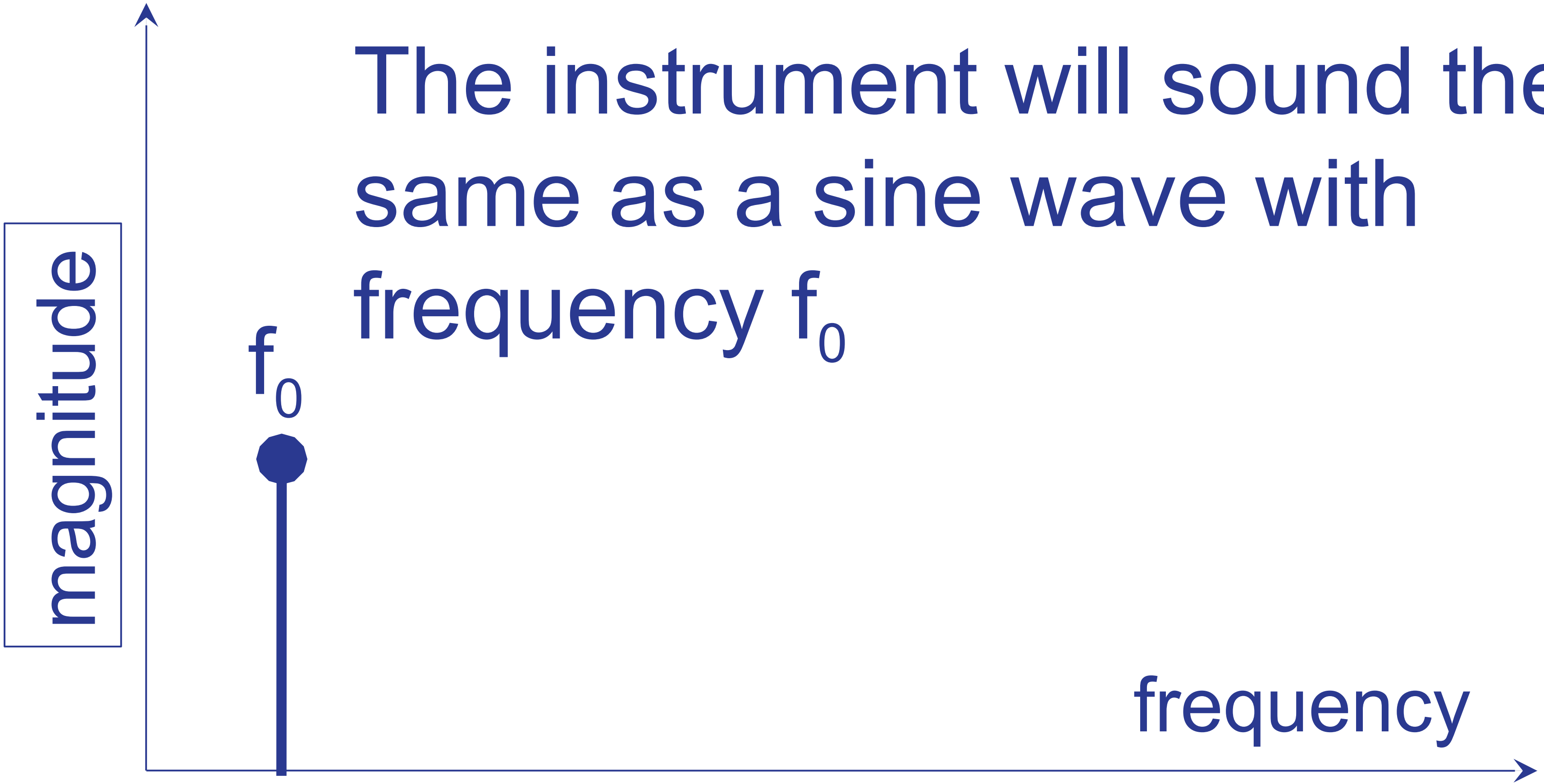
Flute



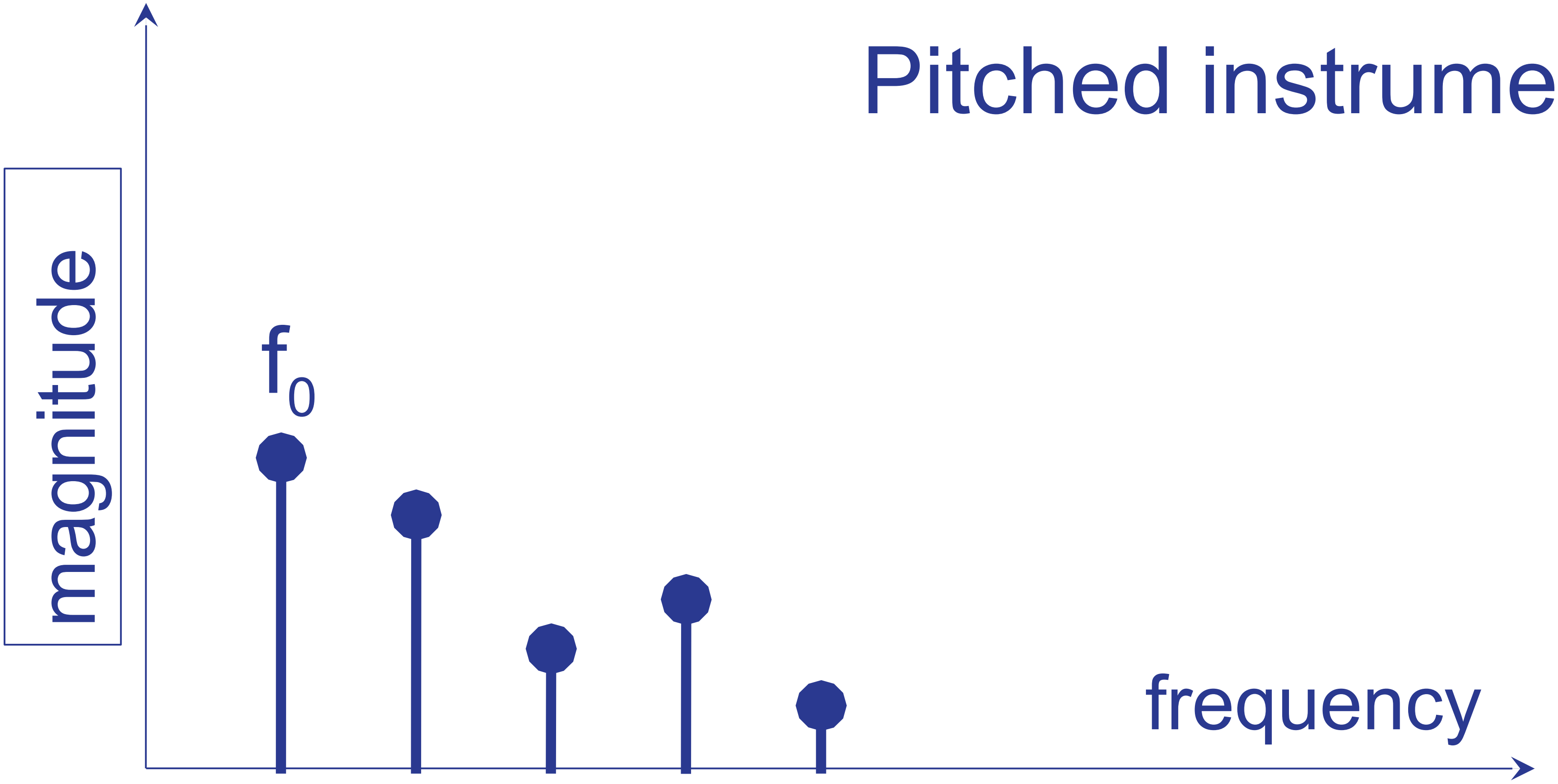
Violin



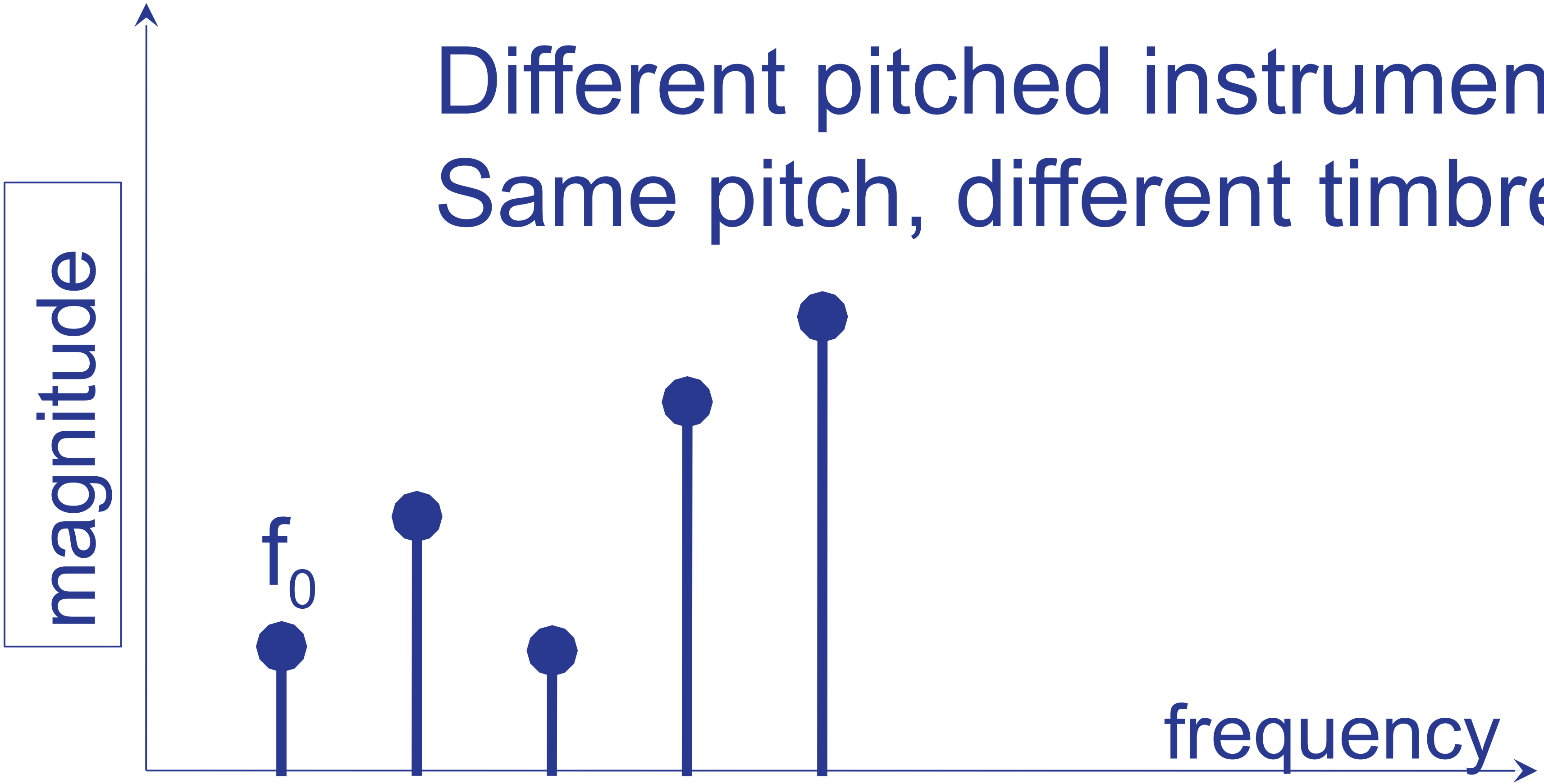
The instrument will sound the same as a sine wave with frequency f_0

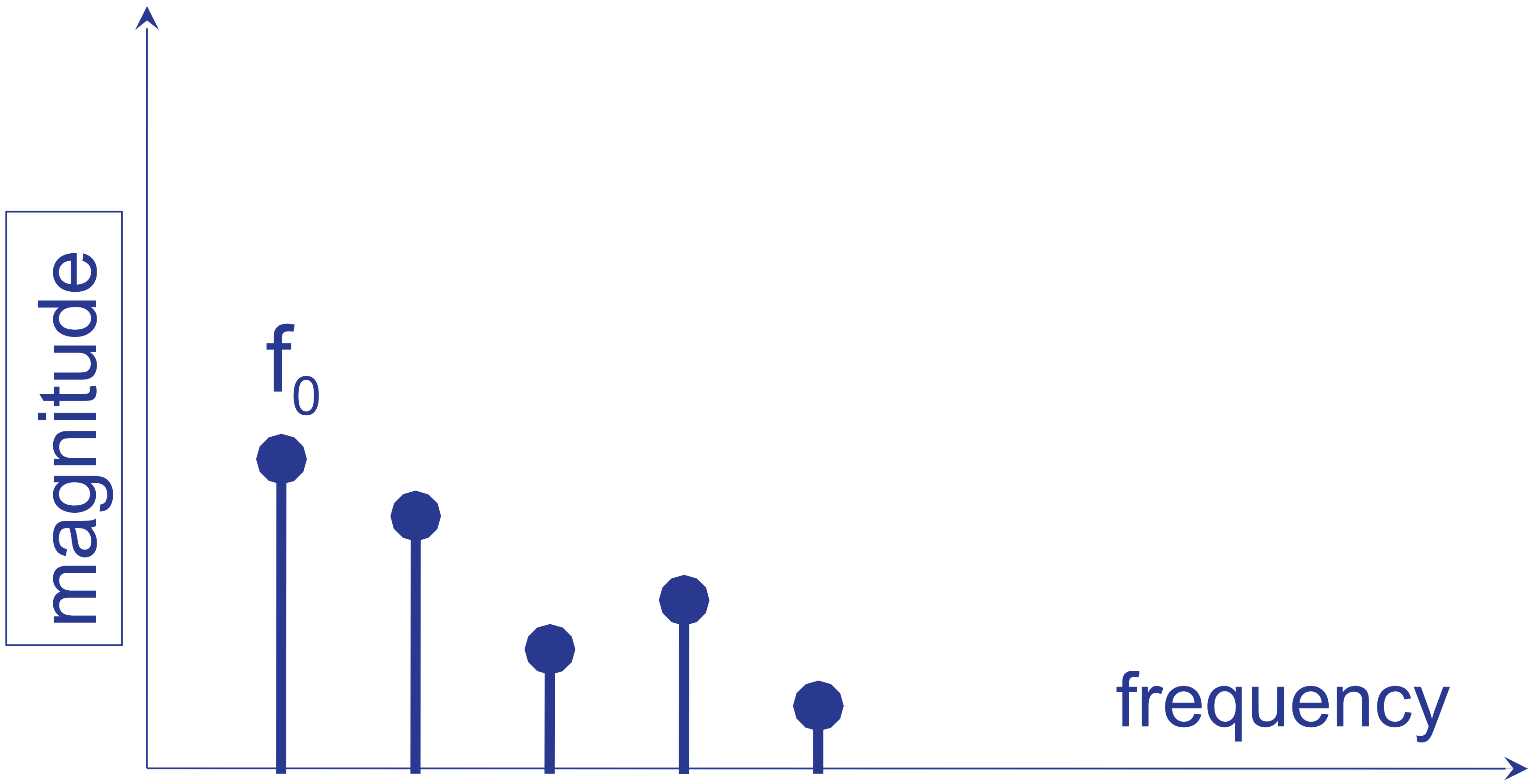


Pitched instrument

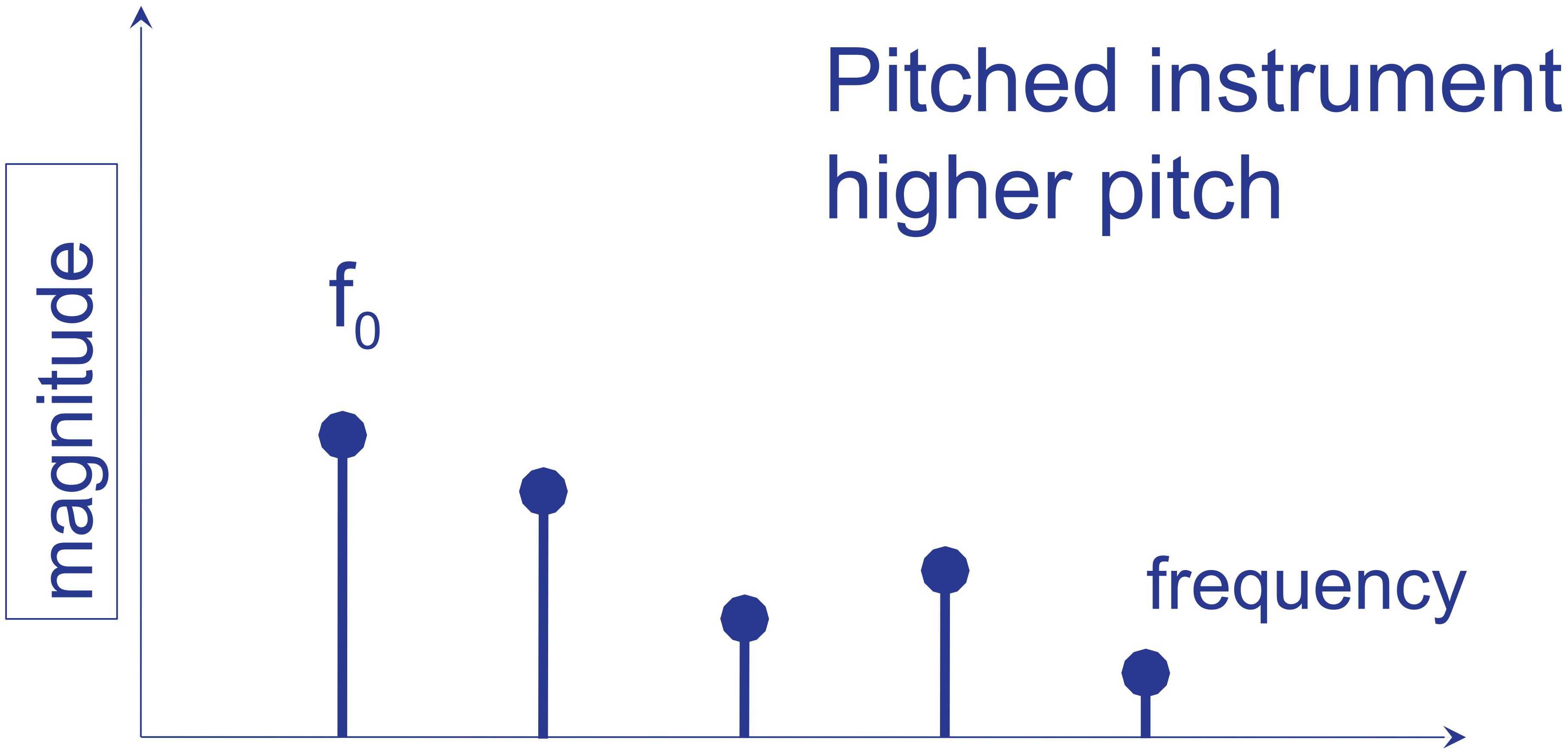


Different pitched instrument
Same pitch, different timbre





magnitude
f_0
frequency



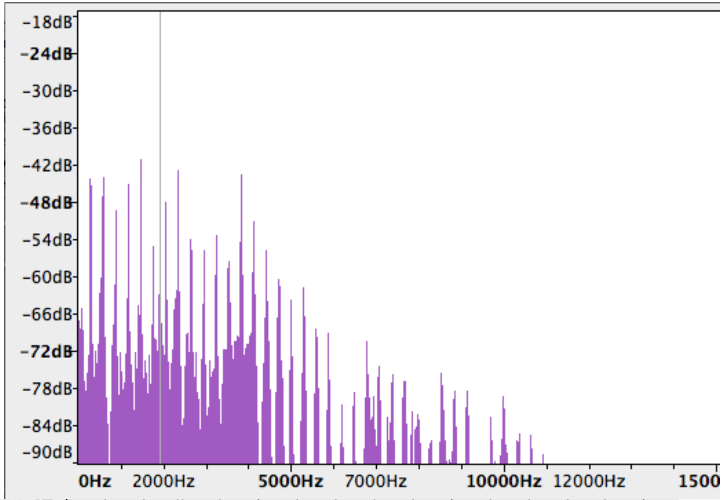
Pitched instrument
higher pitch
magnitude
f0
frequency

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."

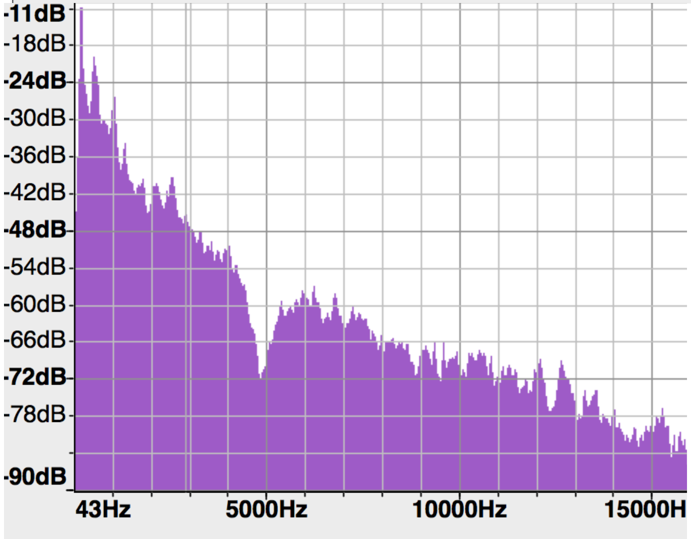
(less harmonically related = less pitched)

Violin



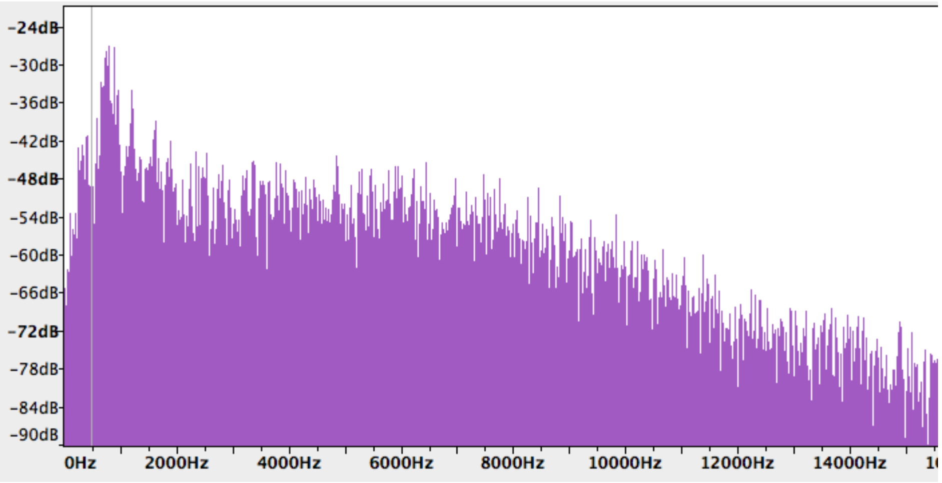
Peaks at approx. 294Hz, 589Hz, 884, …

Ciblon



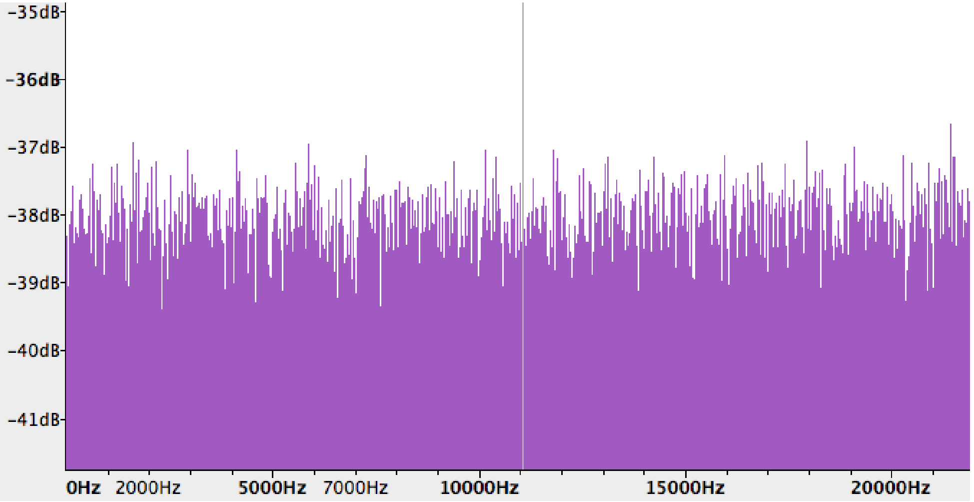
Peaks at approx. 195Hz, 522Hz, 1320Hz, 2560Hz…

Snare



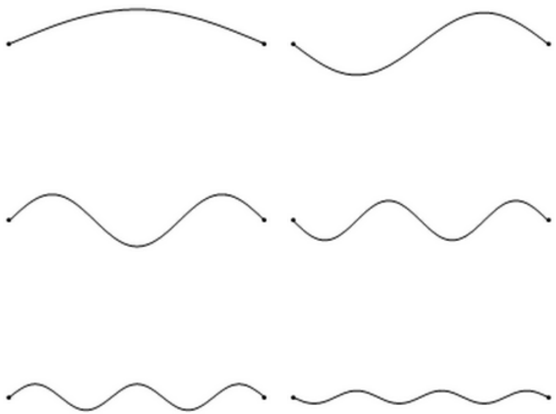
White

noise



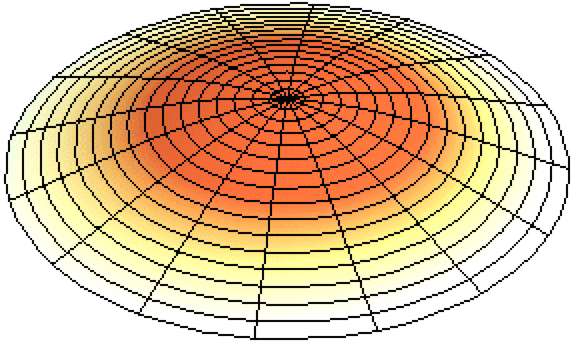
Why aren't drums pitched?

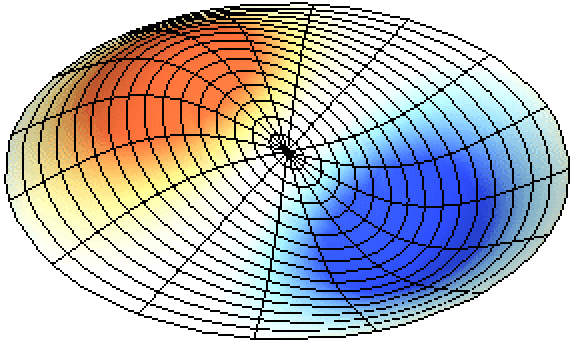
- Strings, air columns vibrate at harmonics:



- 2D surfaces do not (*inharmonicity*)

Modes on a drum head



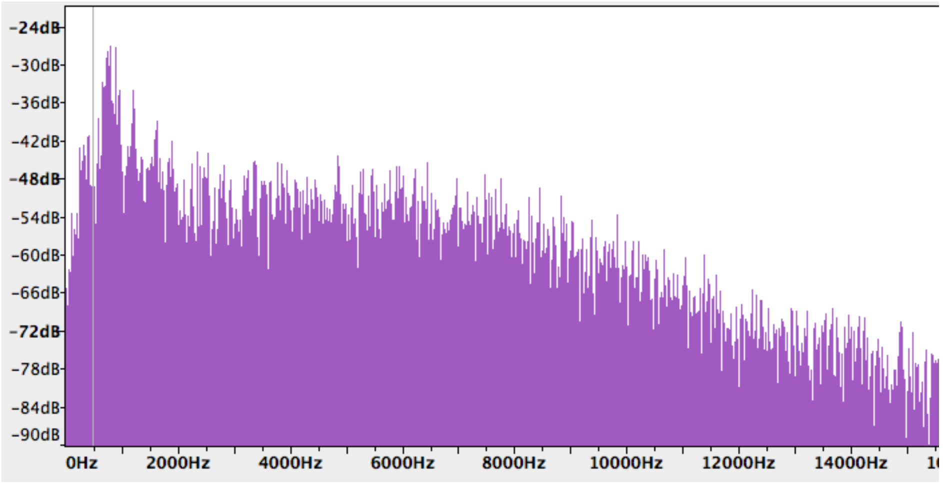


(and many more)

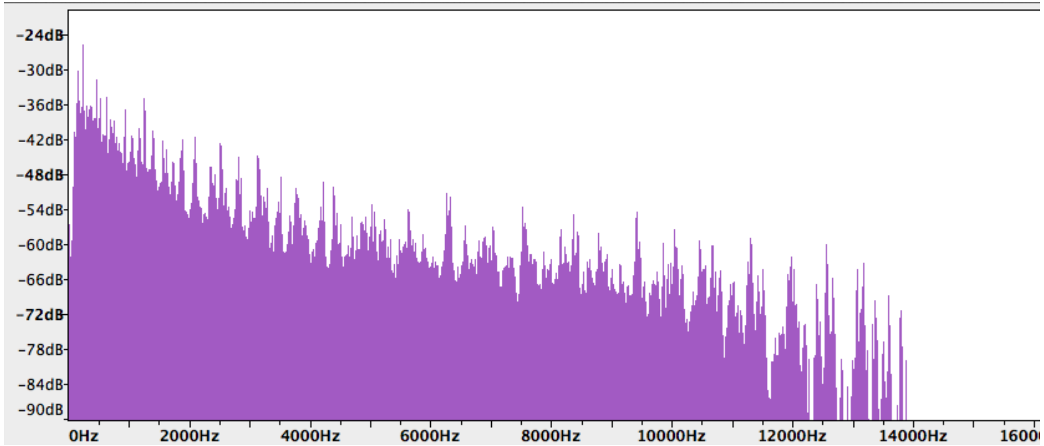
http://www.acs.psu.edu/drussell/demos/membranecircle/circle.html

One sound or many?

Snare



Orchestra



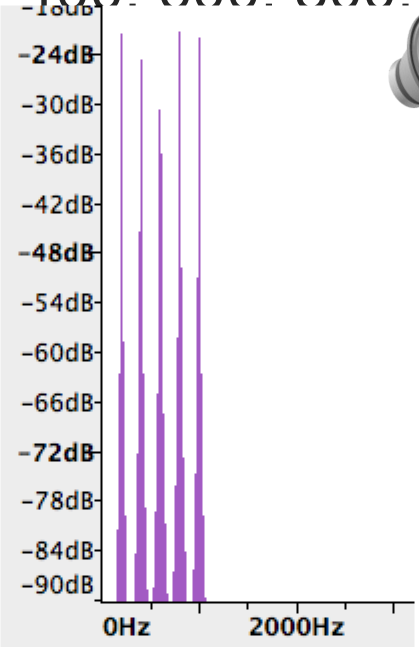
Single or multiple sound sources?

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

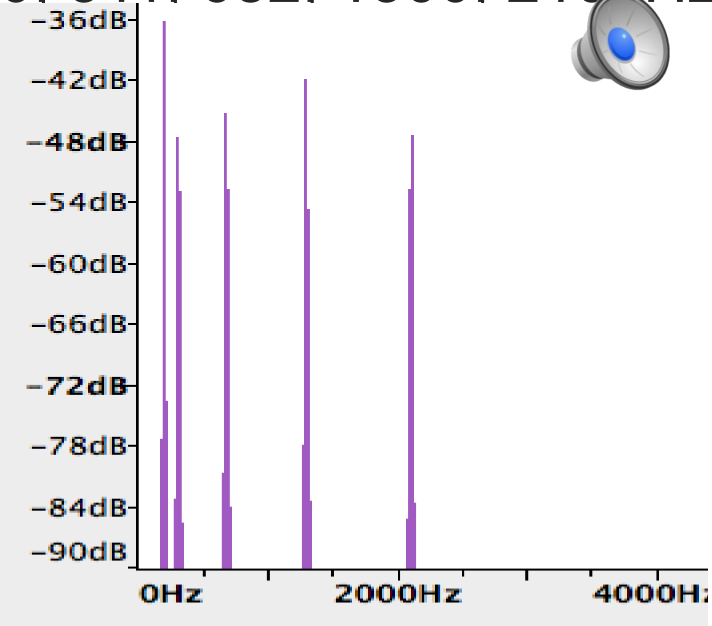
- Harmonic relationship
- Shared onset time
- Shared location
- Shared changes in amplitude (envelope)
- Shared changes in frequency (vibrato)

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

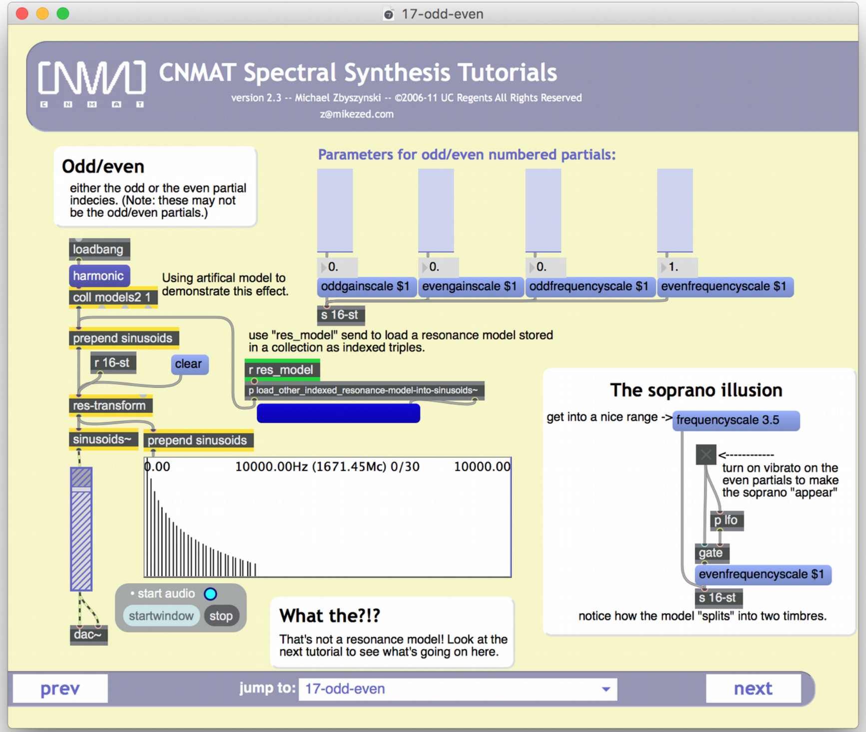
Shared changes in amplitude

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)

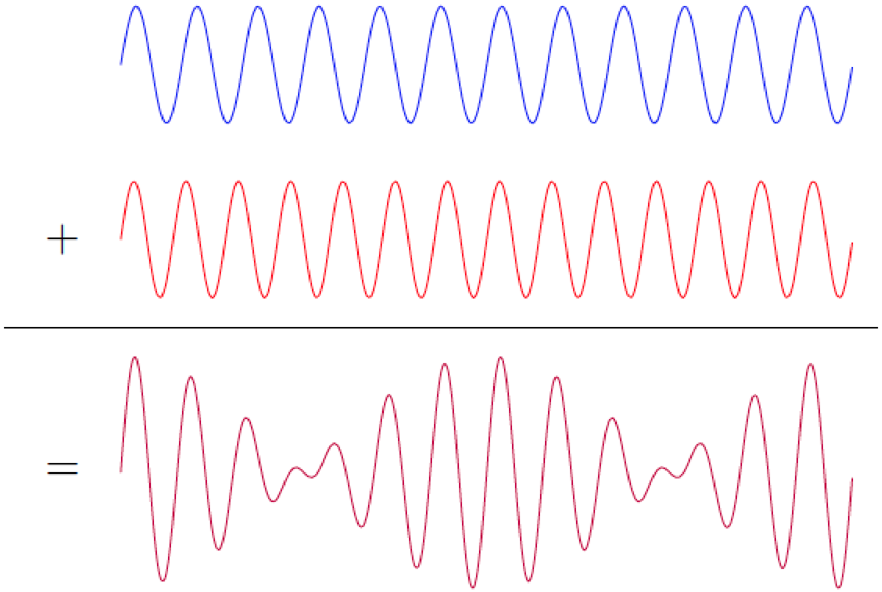
Perception impacted by:

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

When frequencies are relatively close:

- Beating
- “Roughness”

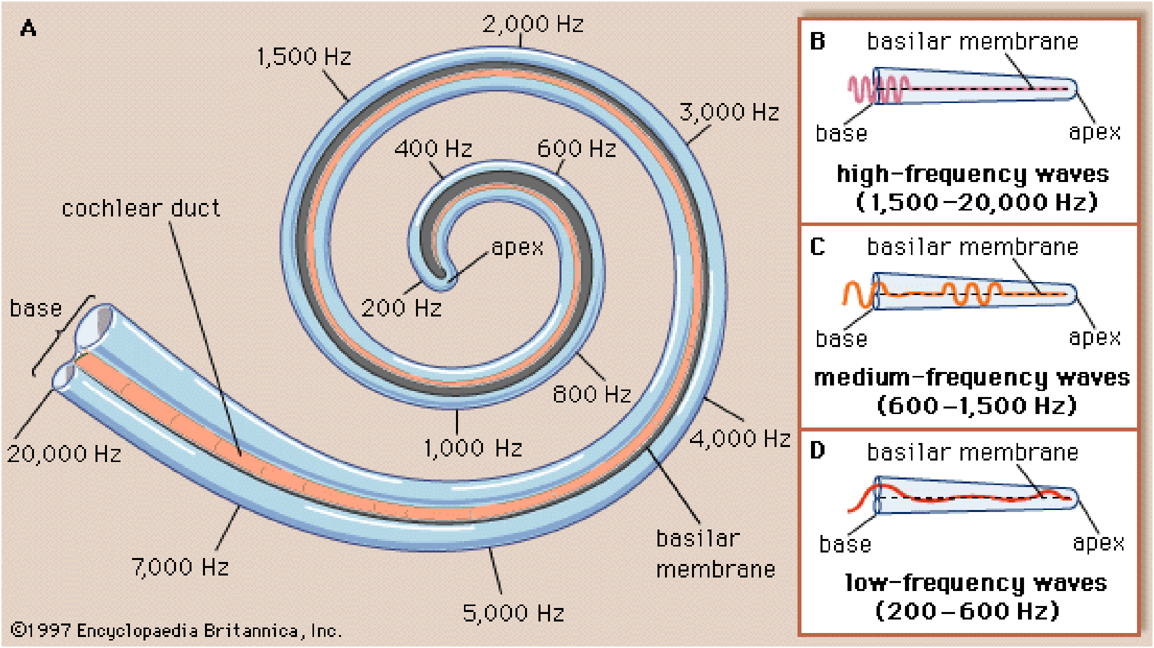
Beating (2 waves close in frequency)



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

Audio examples

Basilar membrane



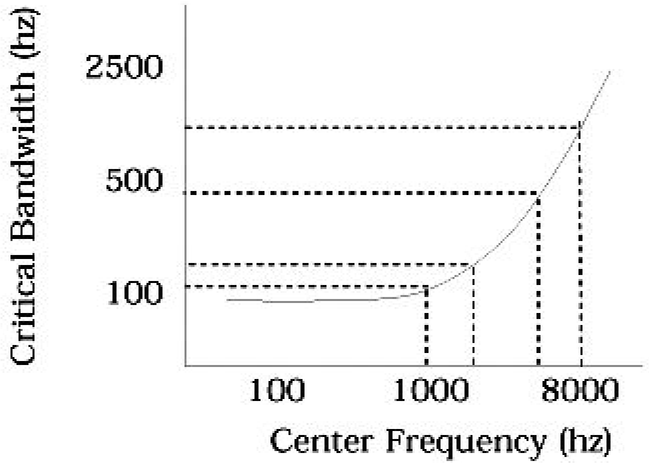
When two tones are close in frequency, they excite nearby locations on basilar membrane.

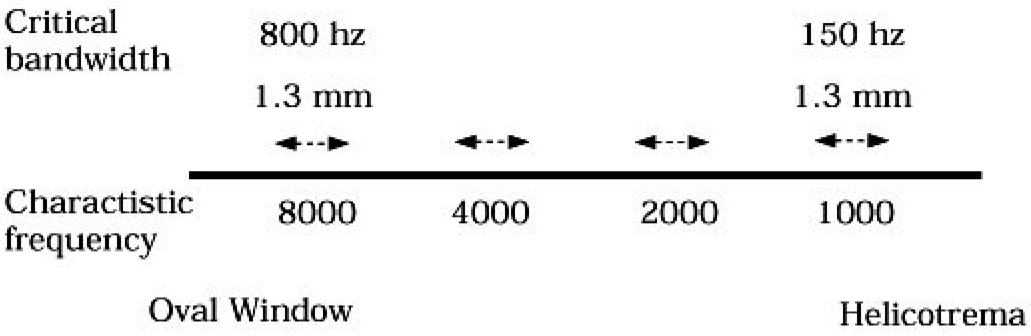
Critical band

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?

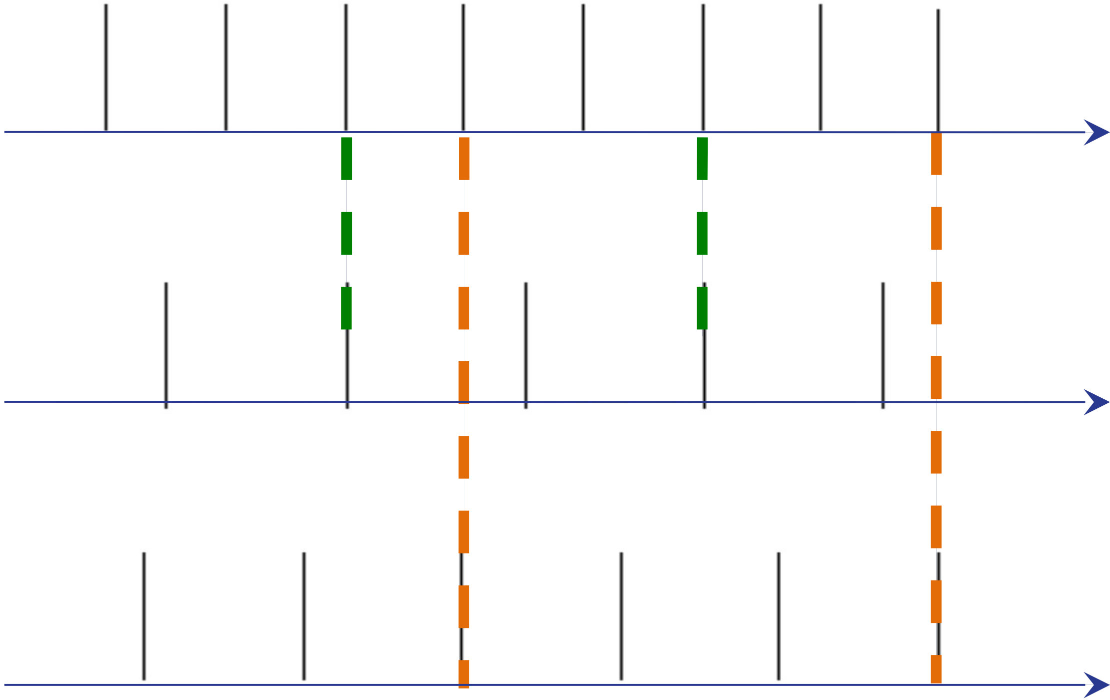
Demo

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 ... u take
the exam, do ... about
signal percep ... oidal
decompositio ... etter you
will do on the ... st exam
copies on the ... exams.
For example, ... otion
which are sin

Basic principle

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

Speech analysis & perception

Human speech

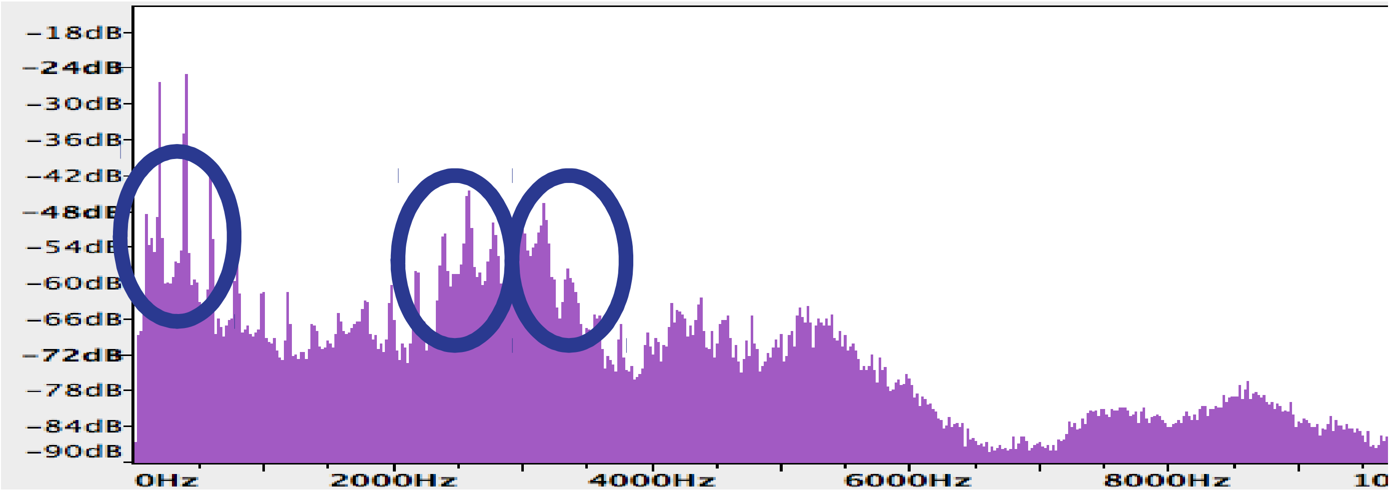
Listen to vowels: What do you hear?

- Constant pitch, volume
- Changing “tone quality”

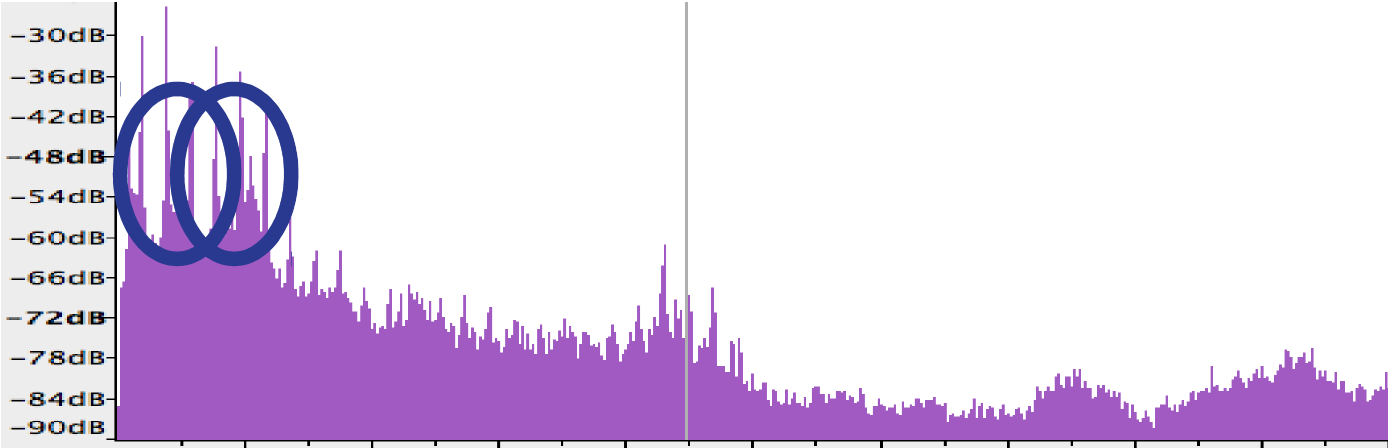
Rule of Thumb #5

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

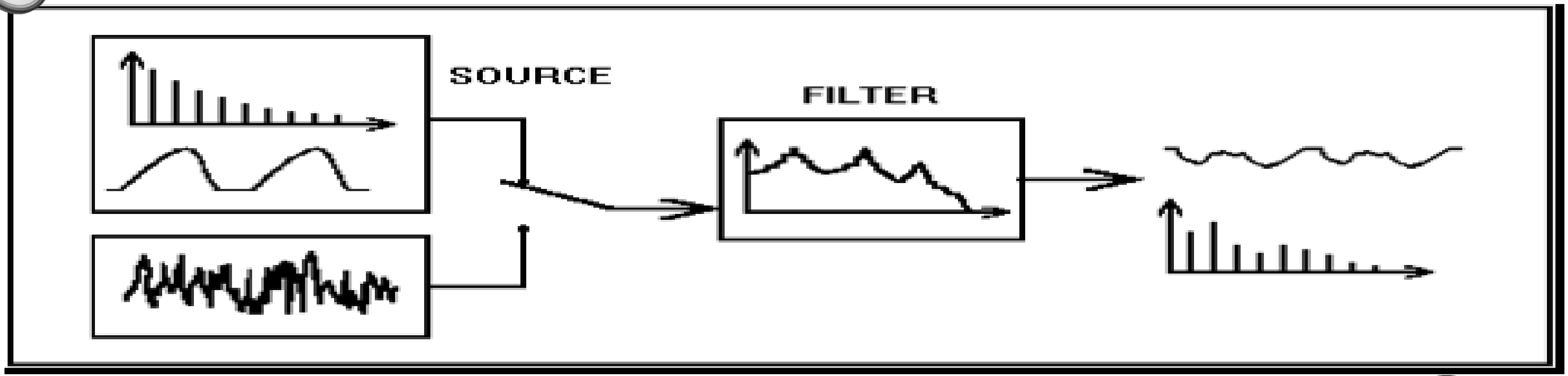
Demo: sndpeek

Singer’s Formant

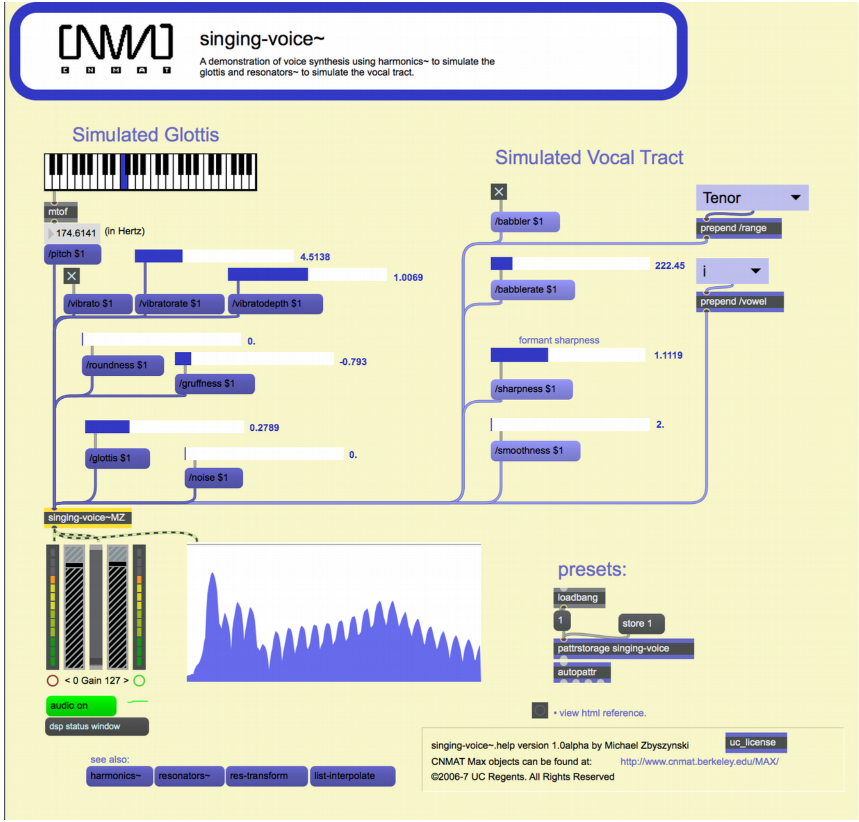
Trained singers have additional formant around 3000Hz.

That allows a singer to be heard above orchestra!

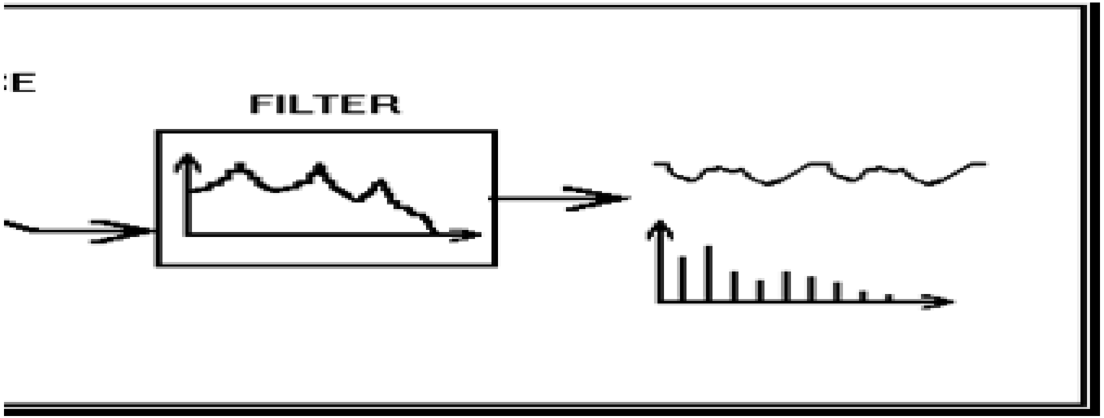
Source-filter model of voice synthesis & analysis



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