CERN Academic Training 2023

The Physics of Music from Pythagoras to Microtones

Isabella Masina (Univ. of Ferrara and INFN, Italy)
Giuseppe Lo Presti (IT Dept., CERN)

Context and Motivation

Origin: PhD / BSc Course at University of Ferrara, Italy

Multidisciplinary, with contributions from Theoretical Physics, Neurosciences, Acoustics, and Digital Signal Processing

The Physics of Music represents a vast corpus of knowledge: we chose the topics that we deemed most interesting for (us and) you here

...with some original touch (especially C&D) based on our scientific background and skills

Context and Motivation

- What are the physical foundations of Western music?
- Why are there 12 notes inside an octave?
- Why is the equal temperament so successful?
- What about microtones?

We will try to answer these and many other questions, that might have come to the mind of scientists playing some instrument, from amateur to semi-professional level.

At the end of the series, the scientist-musician will hopefully acquire a deeper awareness of the art of playing and composing music.

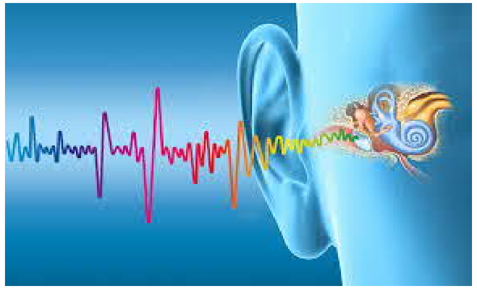
CONTENTS

- Our Detector: the ear
- Case study: vibrating strings and Fourier Analysis
- Psychoacoustics: notion of “roughness”, beatings, etc
- Consonance and Dissonance: a test for dyads (intervals)

OUR DETECTOR: THE EAR

The ear actually functions as a Fourier analysis device

Ohm's law of hearing: the perception of the tone of a sound is a function of the frequencies and amplitudes of the harmonics and not of the phase relationships between them



https://www.odyo.ca/anatomy-of-the-ear-and-the-hearing-system-works/

Signal associated to pressure waves in air, time domain $s(t)$ → $S(f)$ (S)FT in frequency domain

Fourier Analysis

The Fourier Transform $S(f)$ of a time-dependent function (signal) $s(t)$ is defined as:

$$S(f) := \int_{-\infty}^{+\infty} s(t) e^{-i2\pi f t}\, dt$$

Typically, the *Power Spectral Density* $|S(f)|^2$ is analyzed, which relates to the *autocorrelation*

To capture transients, time-frequency analysis through the *Short-Time* Fourier Transform is common:

$$S_{st}(t, f) = \int_{-\infty}^{+\infty} s(\tau) w(\tau - t) e^{-i2\pi f \tau}\, d\tau$$

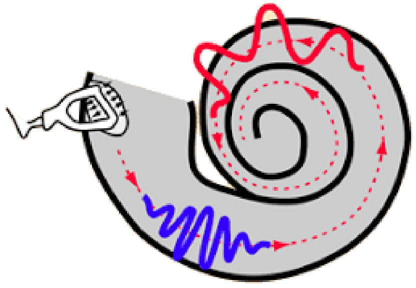
where $w(t)$ is a *windowing* function (typically a Gaussian). For the ear its width is about 0.05s

The *Spectrogram* is defined as $|S_{st}(t, f)|^2$. Easy to produce with your smartphone!

OUR DETECTOR: THE EAR

The ear actually functions as a Fourier analysis device

This is consistent with the *place theory* of hearing, which correlates the observed pitch with the position along the basilar membrane of the inner ear that is stimulated by the corresponding frequency



http://hyperphysics.phy-astr.gsu.edu

Non linearities are nevertheless relevant and represent a current subject of research, see e.g.

Jacob N. Oppenheim and Marcelo O. Magnasco
Human Time-Frequency Acuity Beats the Fourier Uncertainty Principle
PRL 110, 044301 (2013)

OUR DETECTOR: THE EAR

The ear actually functions as a Fourier analysis device

fundamental
frequency of the tone

We are able to establish the PITCH of a sound
if its harmonics (or partials) are «harmonic», i.e.

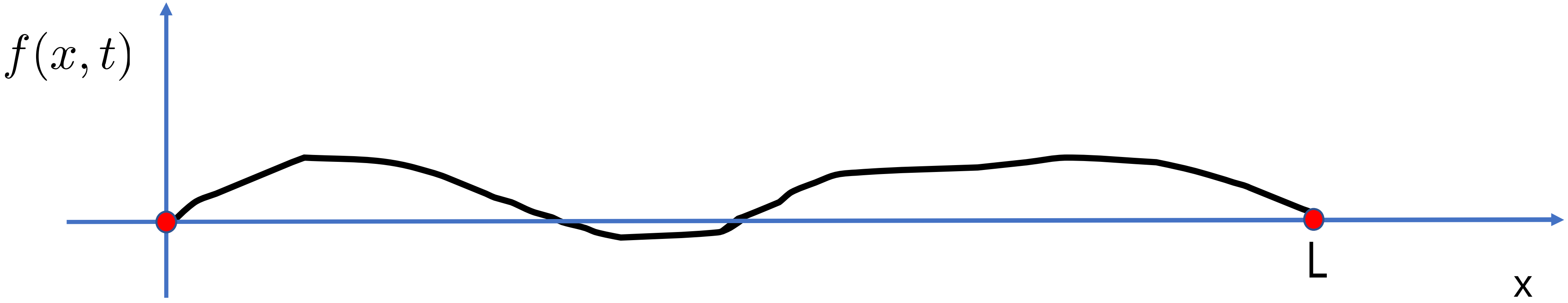
$$\boxed{f_n = n f_1}$$

where n=1,2,3,...

Q: Is a vibrating string harmonic? What are its amplitudes and frequencies?

$$s(t) = \sum_{n=1,2,\ldots} A_n \sin(2\pi f_n t)$$

The vibrating string with fixed endpoints



The displacement f(x,t) satisfies the d'Alembert wave equation (small displacements approx)

$$\boxed{\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f(x,t)}{\partial t^2}}$$

velocity of propagation

$$v = \sqrt{\frac{T}{\mu}}$$

tension

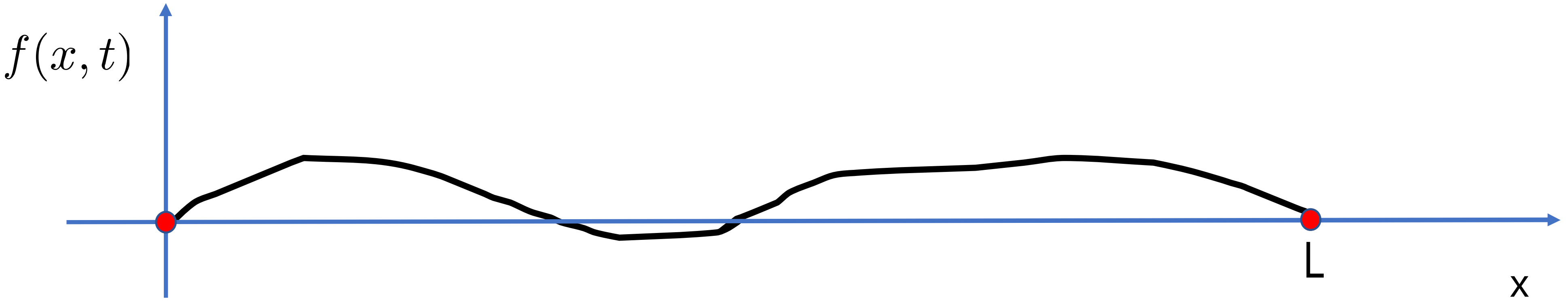
linear mass density

General solution:

$$f(x,t) = \sum_{n=1,2,\ldots} \sin(n\pi \frac{x}{L})\,(C_n^s \sin(2\pi f_n t) + C_n^c \cos(2\pi f_n t))$$

$$f_n = \frac{n}{2L}\sqrt{\frac{T}{\mu}}$$

The vibrating string with fixed endpoints



The ear detects the (S)FT of the pressure wave generated in air, that is of a signal that can be written as

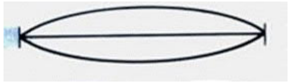
$$s(t) = \sum_{n=1,2,\ldots} (C_n^s + C_n^c) \sin(2\pi f_n t)$$

Amplitudes to be calculted by specifying boundary conditions

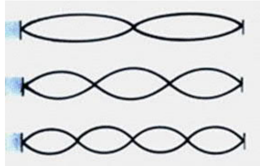
$$f_n = \frac{n}{2L}\sqrt{\frac{T}{\mu}}$$

HARMONIC!
(like voice)

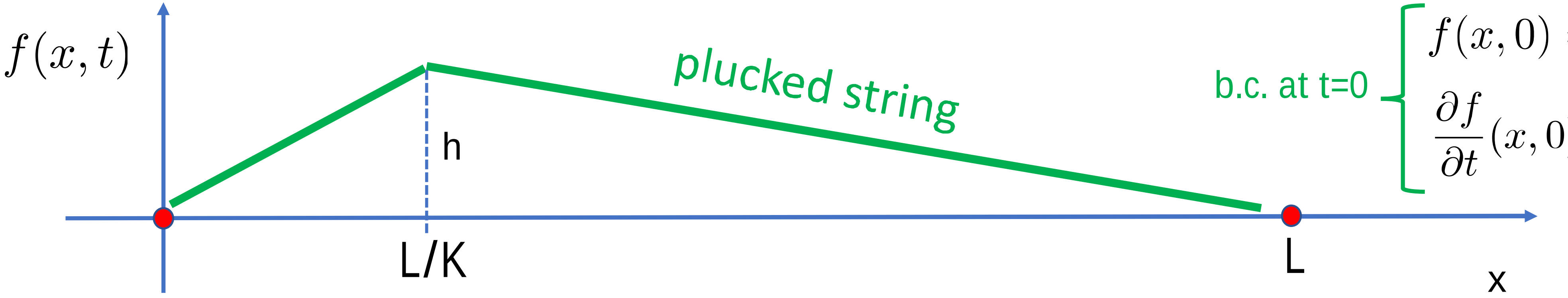
n=1 → fundamental



n=2,3,... → (higher) harmonics



The vibrating string with fixed endpoints



b.c. at t=0
$$\begin{cases} f(x,0) = \text{triangle} \\ \dfrac{\partial f}{\partial t}(x,0) = 0 \end{cases}$$

The ear detects the (S)FT of the pressure wave generated in air, that is of a signal that can be written as

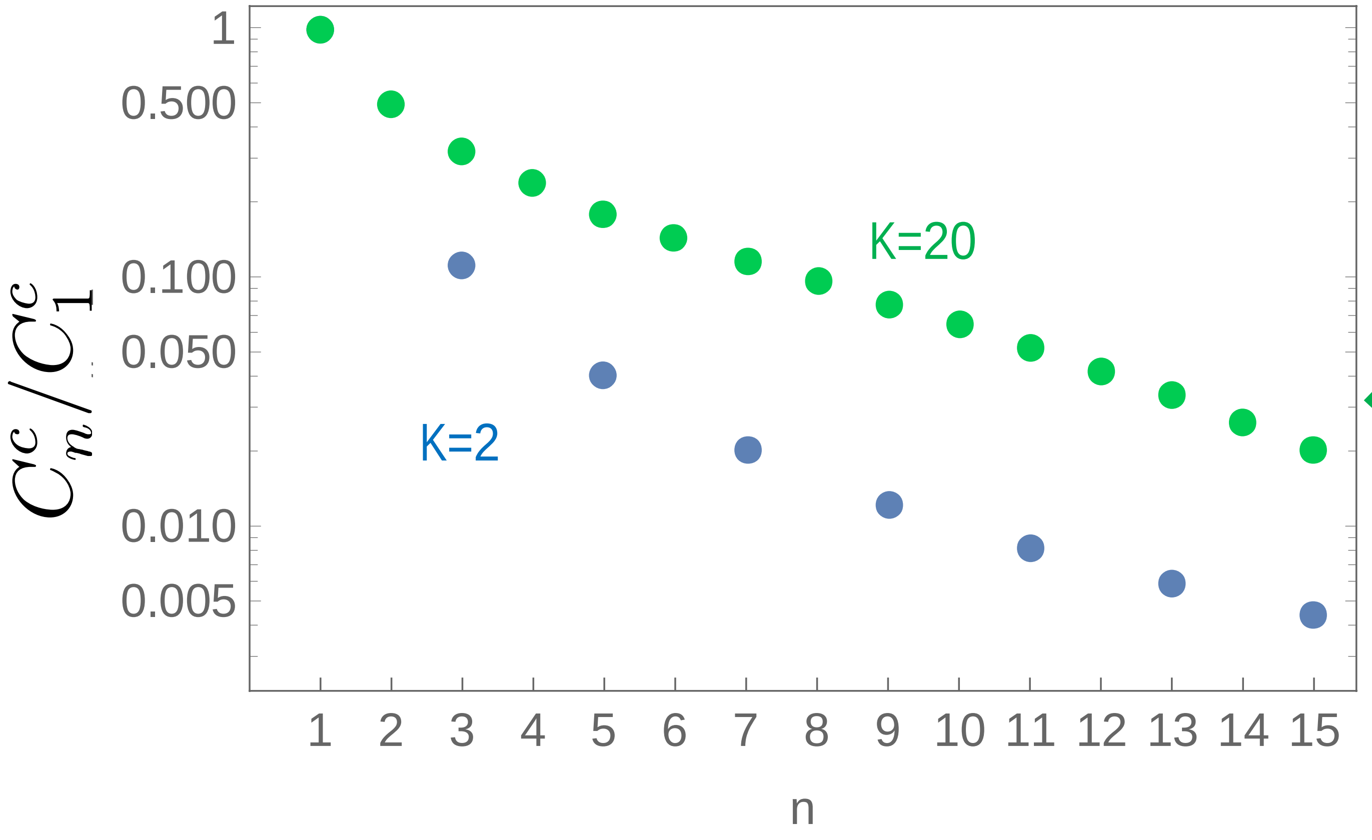
$$s(t) = \sum_{n=1,2,\dots} (\cancel{C_n^s} + C_n^c) \sin(2\pi f_n t)$$

$$= \underbrace{\frac{2hK^2}{\pi^2(K-1)}}_{C_1^c} \frac{1}{n^2} \sin\frac{n\pi}{K}$$

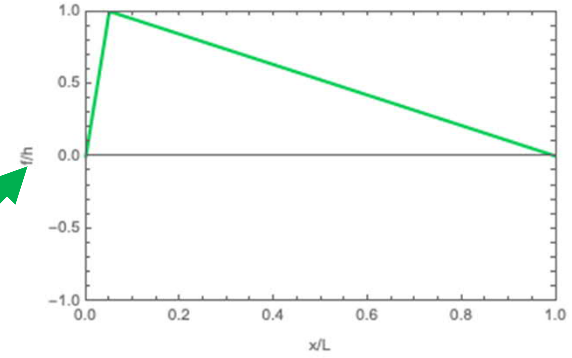
Amplitudes to be calculted by specifying boundary conditions

where do you pluck your string?

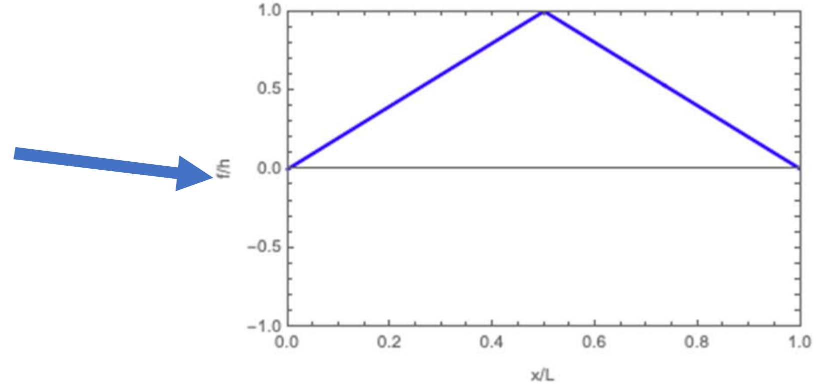
Plucked string: harp (K=2) vs guitar (K=20)



guitar



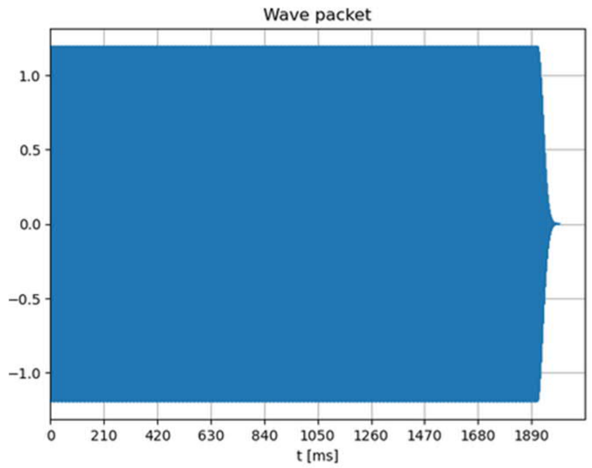
harp



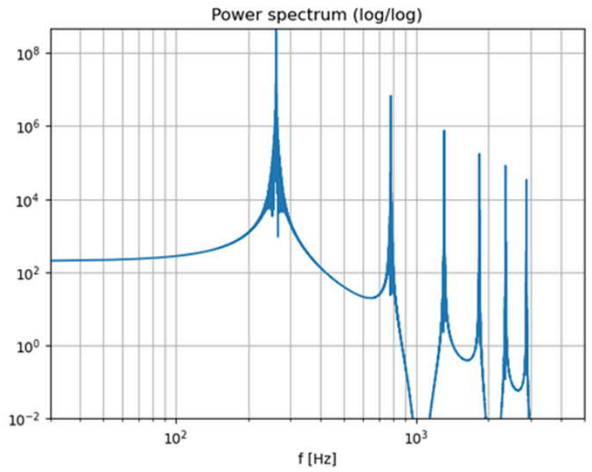
Plucked string?

Let's give it a try! From

$$C_n^c = \underbrace{\frac{2hK^2}{\pi^2(K-1)}}_{C_1^c} \frac{1}{n^2} \sin\frac{n\pi}{K}$$



$s(t)$

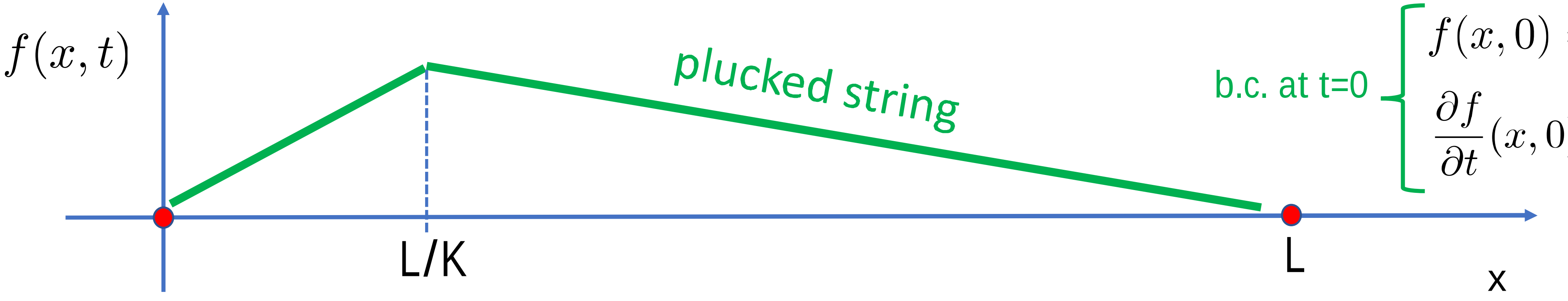


$|S(f)|^2$

We take f=C_4 (central C in a piano), K=2, and generate a waveform with numpy:

```
t = np.linspace(0, self.duration, samples, False)
for n in range(1, self.harmonics):
    Cn = 1 / n**2 * sin(n * pi / self.k_harp)
    s += Cn * np.sin(2 * np.pi * f * n * t)
```

The vibrating string with fixed endpoints + damping



b.c. at t=0 $\begin{cases} f(x,0) = \text{triangle} \\ \dfrac{\partial f}{\partial t}(x,0) = 0 \end{cases}$

The ear detects the FT of the pressure wave generated in air, that is of a signal that can be written as

$$s(t) = \sum_{n=1,2,\dots} (\cancel{C_n^s} + C_n^c) \sin(2\pi f_n t) \times e^{-n\frac{\Gamma}{2}t}$$

$$C_n^c = \underbrace{\frac{2hK^2}{\pi^2(K-1)}}_{C_1^c} \frac{1}{n^2} \sin\frac{n\pi}{K}$$

damping factor:

the higher is the harmonic
the sooner it vanishes

Take e.g. Γ=O(1) s^{-1}

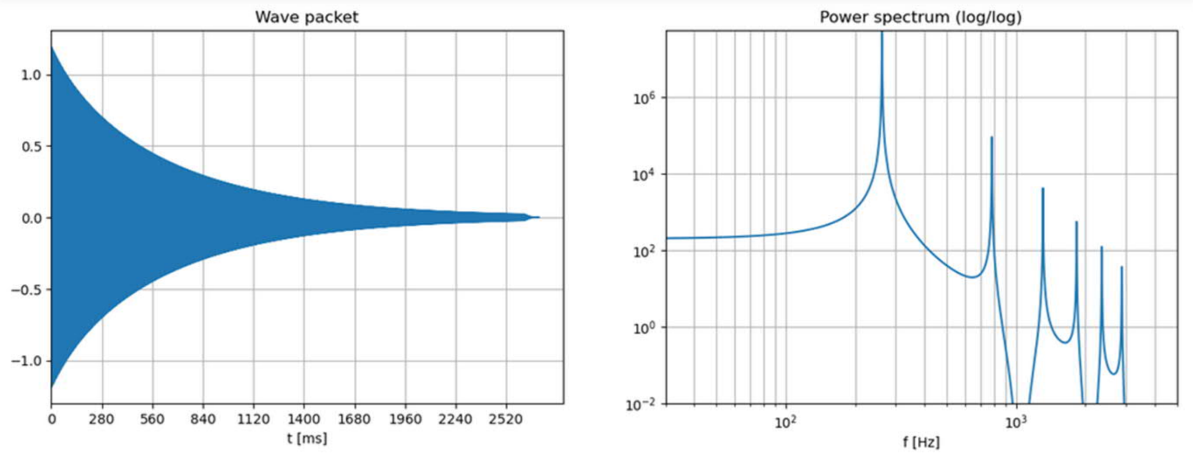
Damped plucked string

OK, let's introduce Γ=1.5 s^{-1}

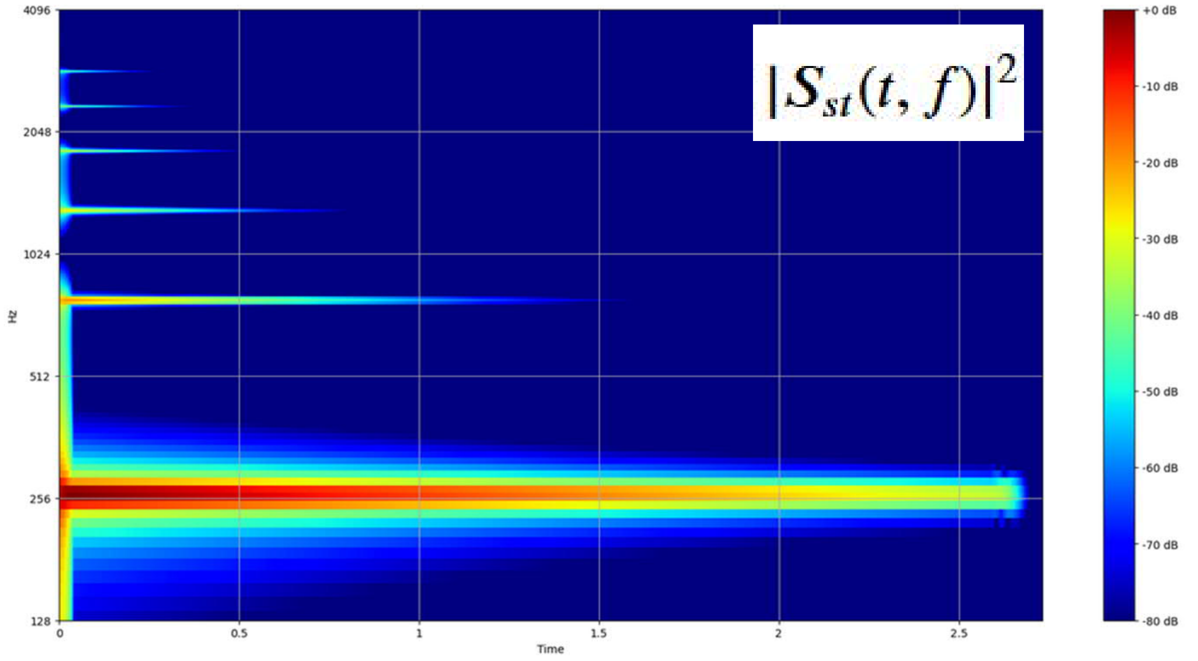
And we take:

K=2 (Harp)

K=20 (Guitar)

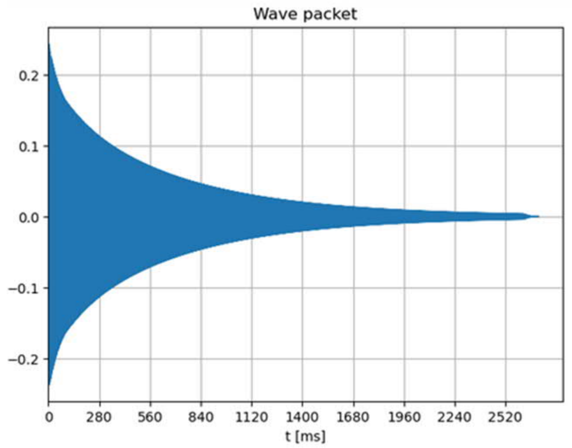


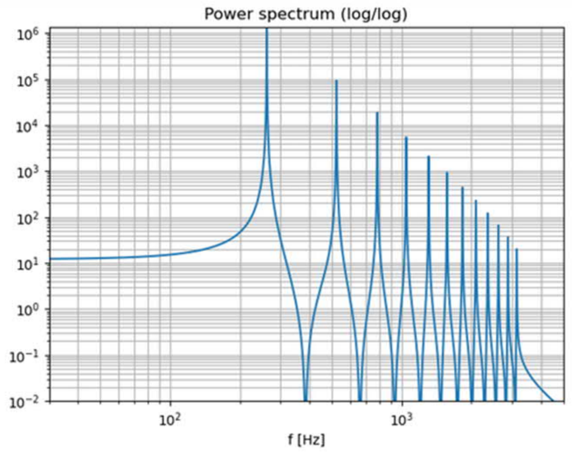

```
for n in range(1, self.harmonics):
  Cn = 1 / n**2 * sin(n * pi / self.k_harp)
  w += Cn * np.sin(2 * np.pi * f * n * t) * \
       np.exp(- t * n * gamma)
```



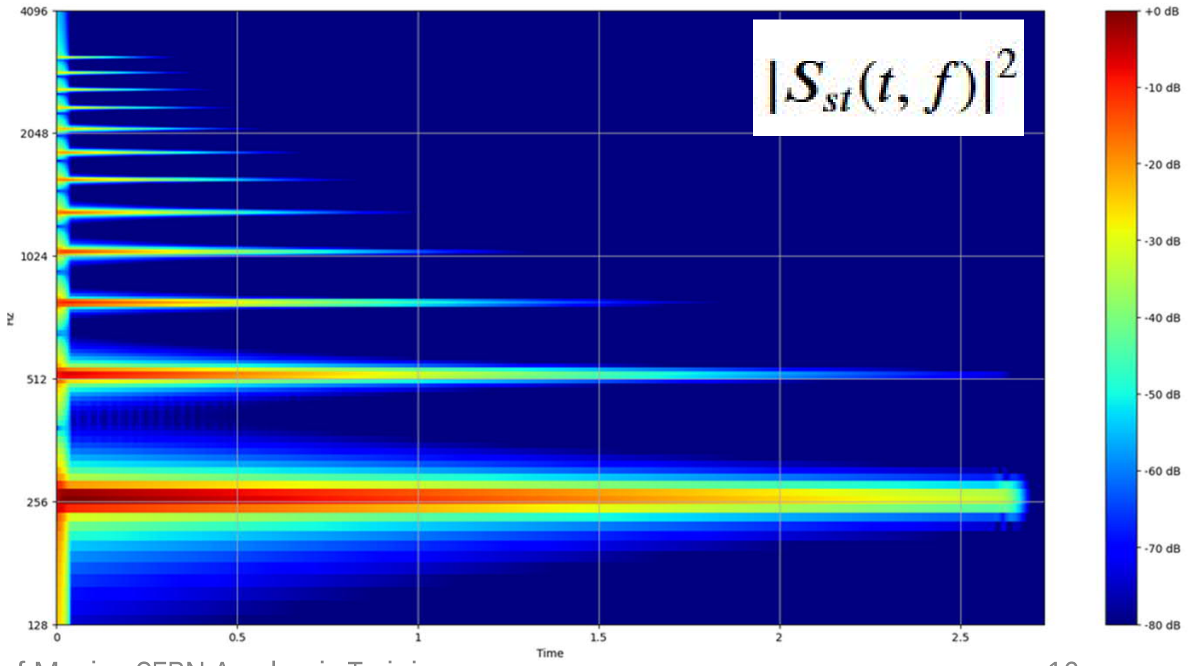
Damped plucked string

OK, let's introduce Γ=1.5 s^{-1}
And we take:
K=2 (Harp)
K=20 (Guitar)





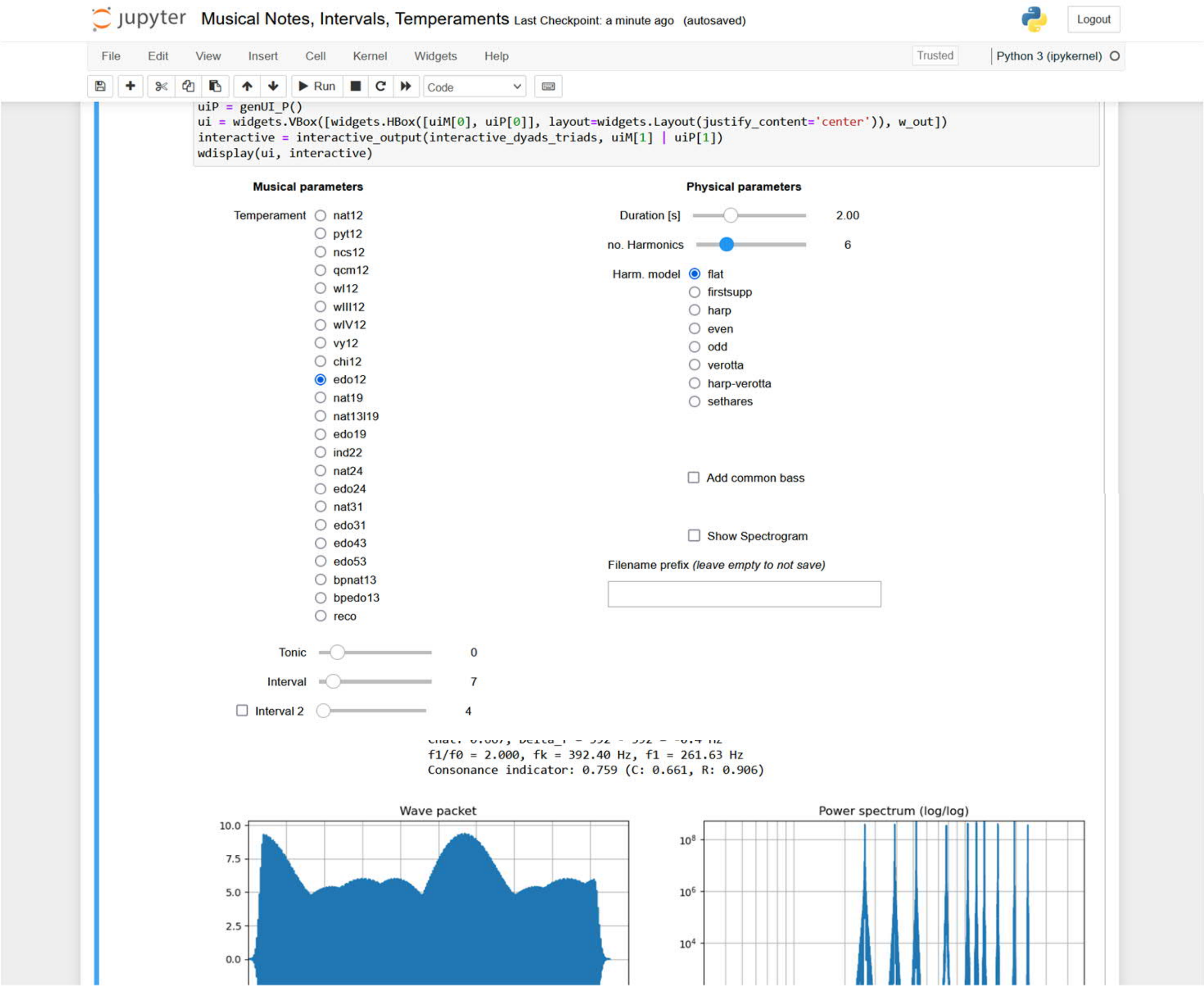

```
for n in range(1, self.harmonics):
  Cn = 1 / n**2 * sin(n * pi / self.k_harp)
  w += Cn * np.sin(2 * np.pi * f * n * t) * \
       np.exp(- t * n * gamma)
```



To be noted that here we generate the stimulus, and deliberately ignore the resonances of a real instrument!

Intermezzo: a toolbox to test and benchmark

- We have developed a set of iPython notebooks to explore sounds and their acoustic effects
  - Based on popular Data Science libraries (`numpy`, `matplotlib`)
  - As shown, we display
    - Time-domain waveform
    - Power spectral density
    - Spectrogram
- We will extensively use this toolbox throughout the series



Sustained Stimulus

What if we have a sustained stimulus as boundary condition?

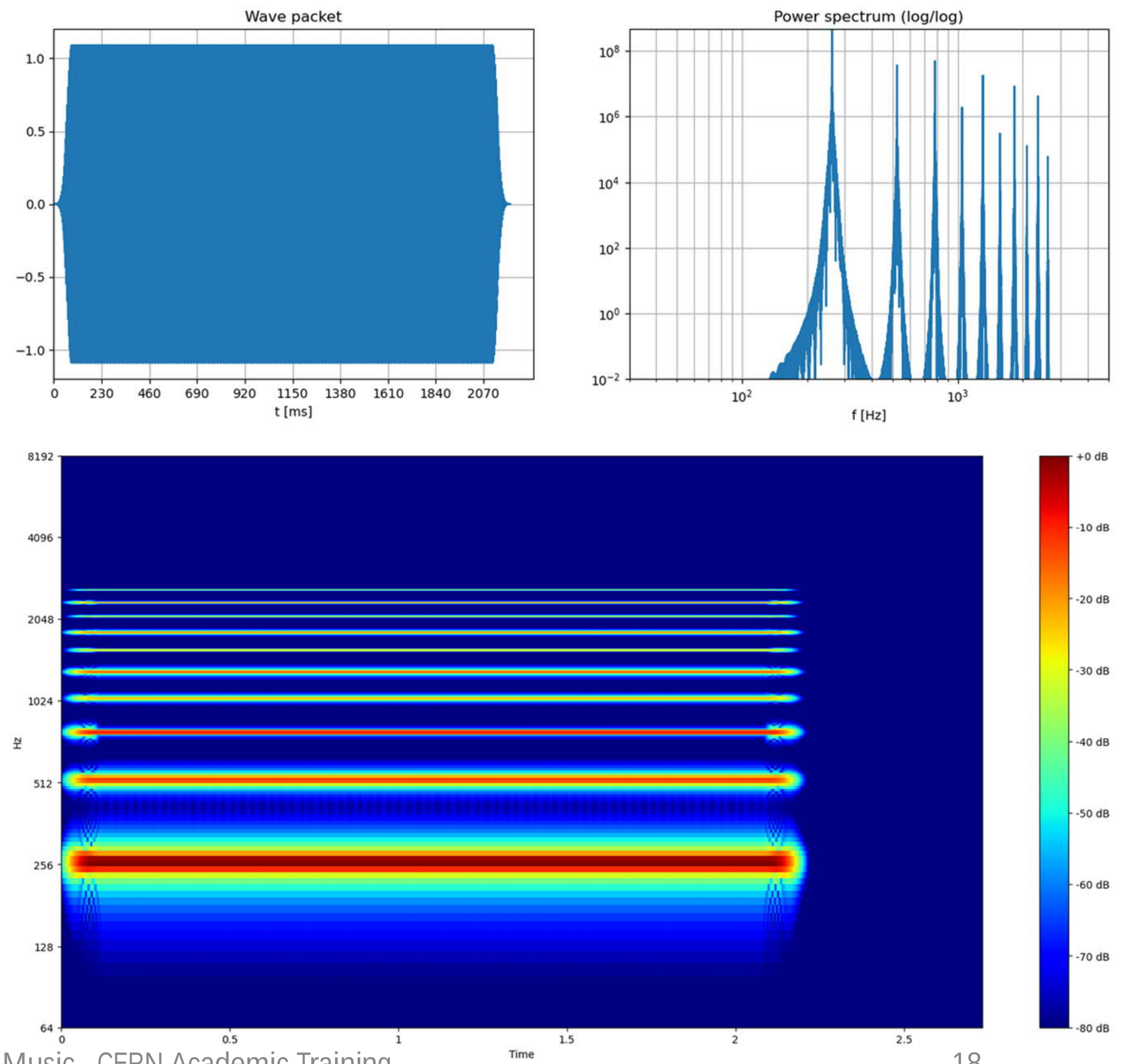
- Same general solution
- Typically, smooth start and end

Common examples:

- The voice!
- Strings
- Woodwinds, pipes

An example: harmonics with relative amplitudes going as

```
Cn = 1 / n**(2 if n % 2 == 0 else 1.1)
```



COMBINING FREQUENCIES

- Now let's combine two sounds with frequency f_1 and f_2, and explore the auditory effects when hearing them together:
  - When the frequencies are "very close": *primary beatings*
  - "Close" frequencies: significant *roughness* in the composed sound
  - Far away frequencies: combined sound relatively less rough
  - "Islands" of *consonance* when $f_2 = k f_1$, k = "simple" ratio
    - Related to a "simpler" waveform?
  - *Secondary beatings* close to such ratios

- How "close", how "far away"? Let's introduce
  - The Critical Bandwith
  - The Discrimination Limen

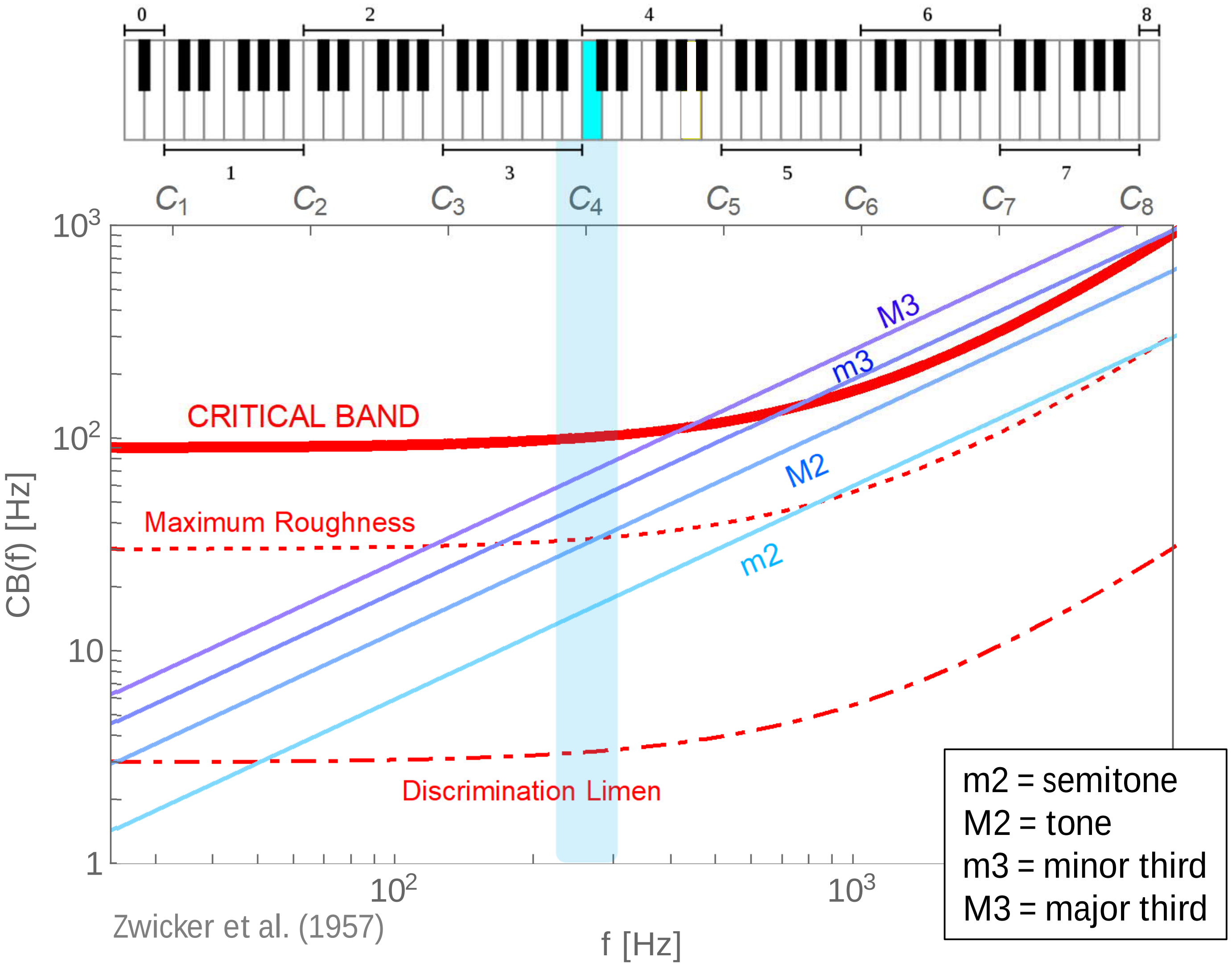
Critical Bandwidth and Discrimination Limen

CB: the frequency band within which a second sound *interferes* in the perception of the first

DL: the frequency difference which makes two sounds perceived as *different pitches* when heard separately

Also known as *Just Noticeable Difference*

For frequences around C_4 (262 Hz), CB is about 100 Hz.
Beats are heard until about 12Hz, roughness between 12Hz and 100Hz, with a maximum at about 33 Hz.



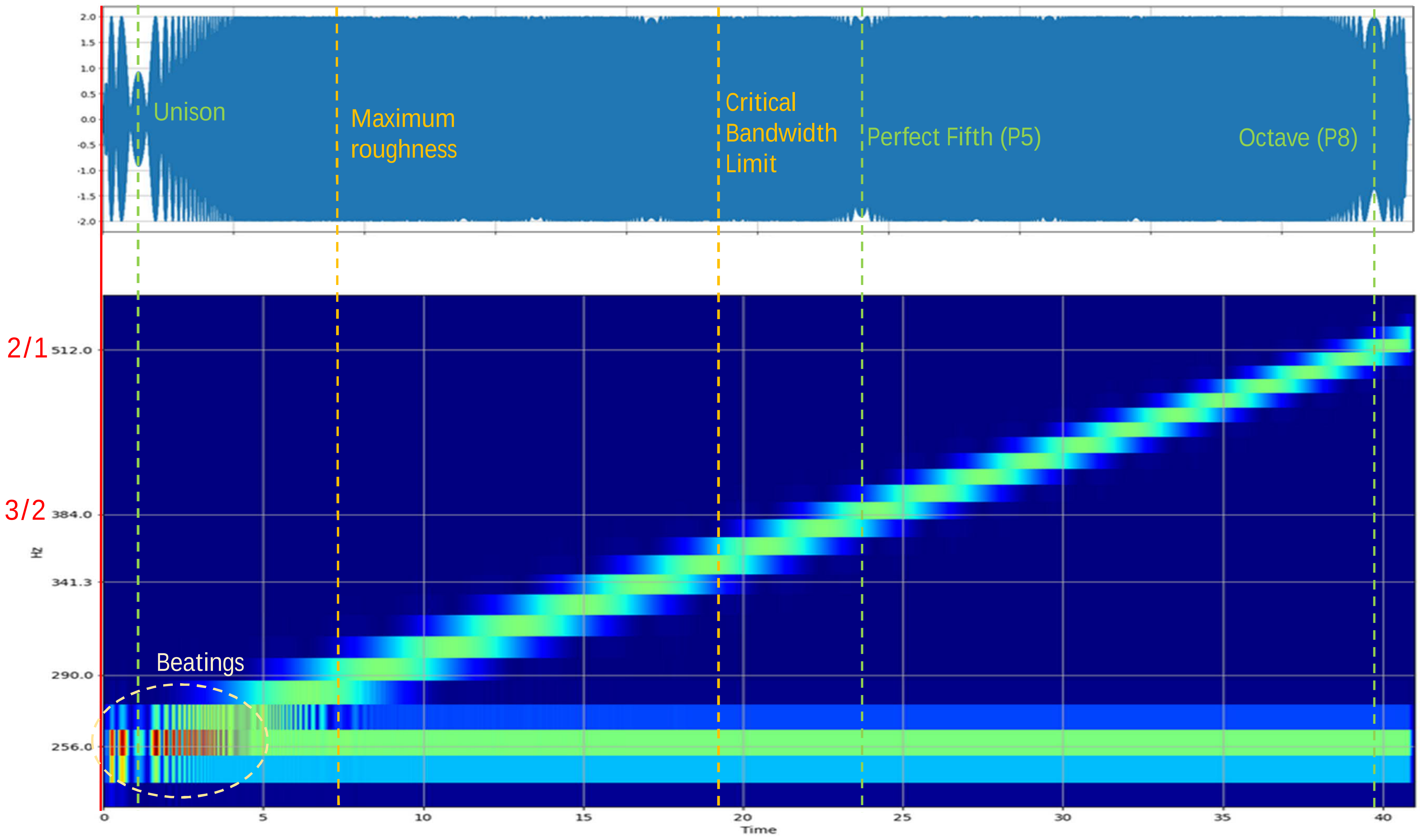
Combining Frequencies: tests

- Let's try these effects: we sum two sounds with frequencies f_1 and f_2
  - f_1 = 256 Hz (A bit flatter than C_4)
  - f_2 sweeps from 250 Hz to 520 Hz = from slightly below f_1 to slightly above $2\ f_1$
  - To account for the logarithmic perception of the hear, we let

$$f_2(t) = 2^{t/T} f_1 \quad \text{taking e.g. T = 40s}$$

  - *Numerically, we just use* `scipy.chirp(method='log') !`

- What do we expect?
  - We're looking for beatings and roughness until $f_2 - f_1 \approx 100$ Hz

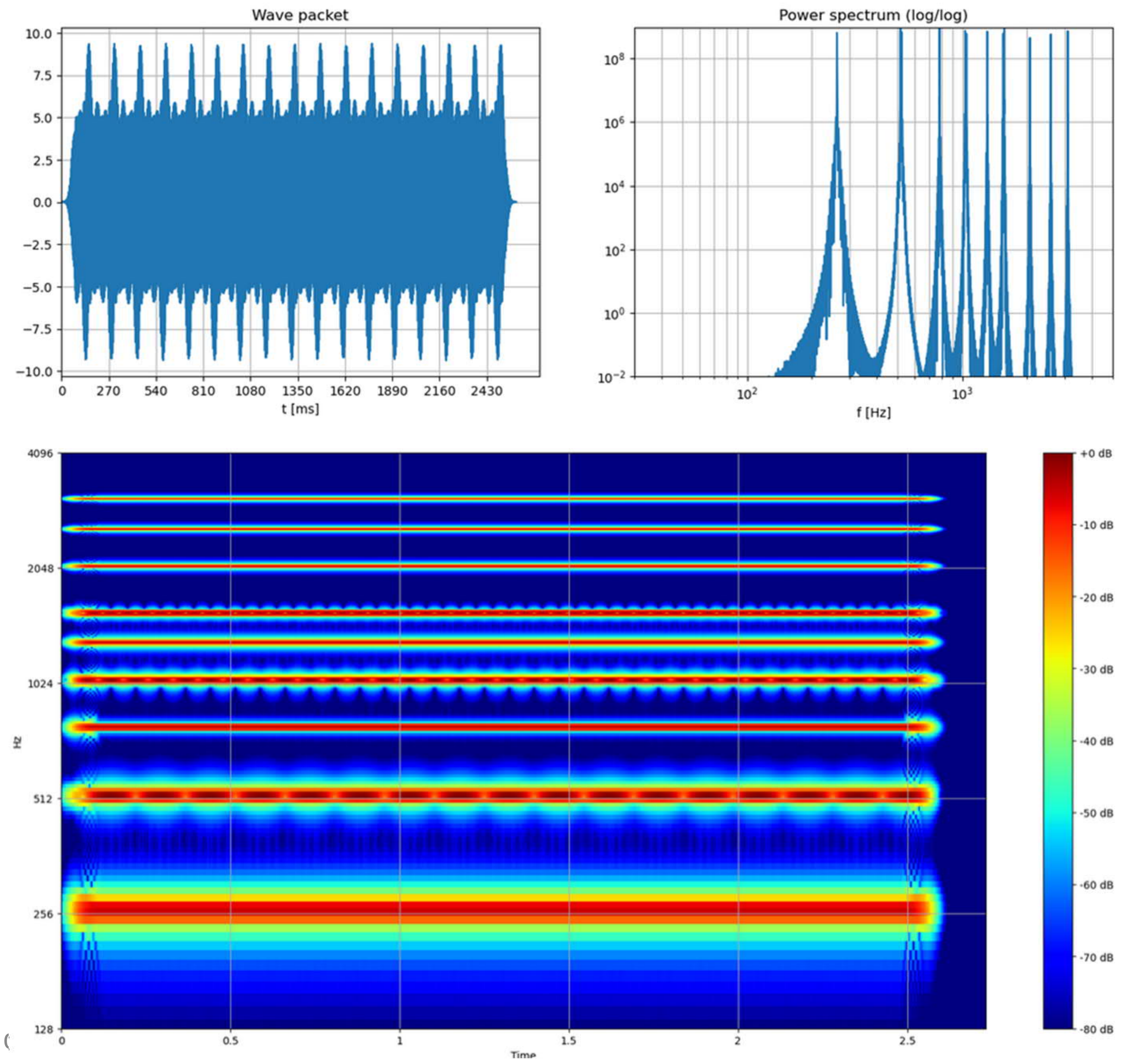


Unison
Maximum roughness
Critical Bandwidth Limit
Perfect Fifth (P5)
Octave (P8)
2/1
3/2
512.0
384.0
341.3
290.0
256.0
Hz
Beatings
0
5
10
15
20
25
30
35
40
Time
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0

Combining Frequencies: P8

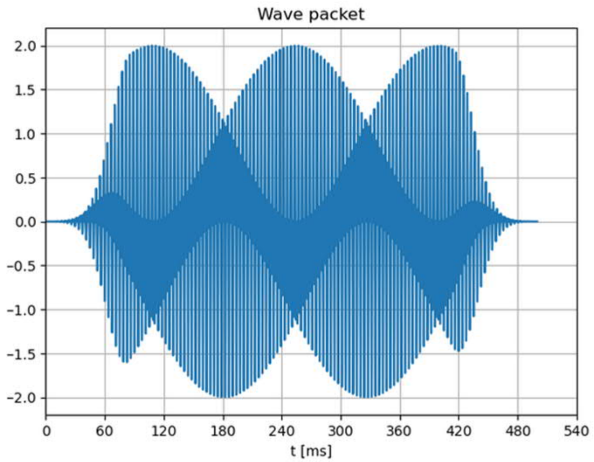
- Secondary beatings:
  $f_1 = C_4$, $f_2 \approx 2\, f_1$

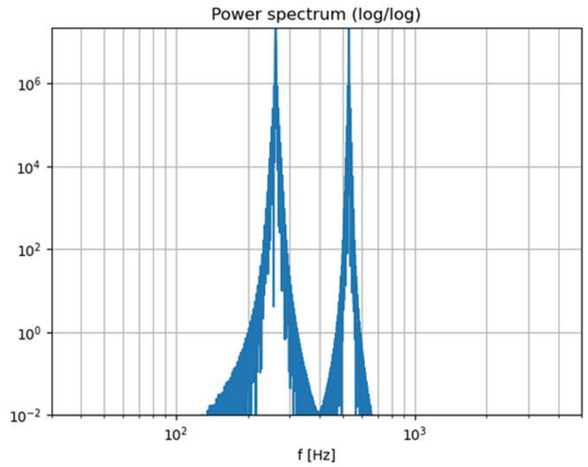
- We call this Mistuned Octave
  - Note the beatings pattern in the Spectrogram for the higher harmonics



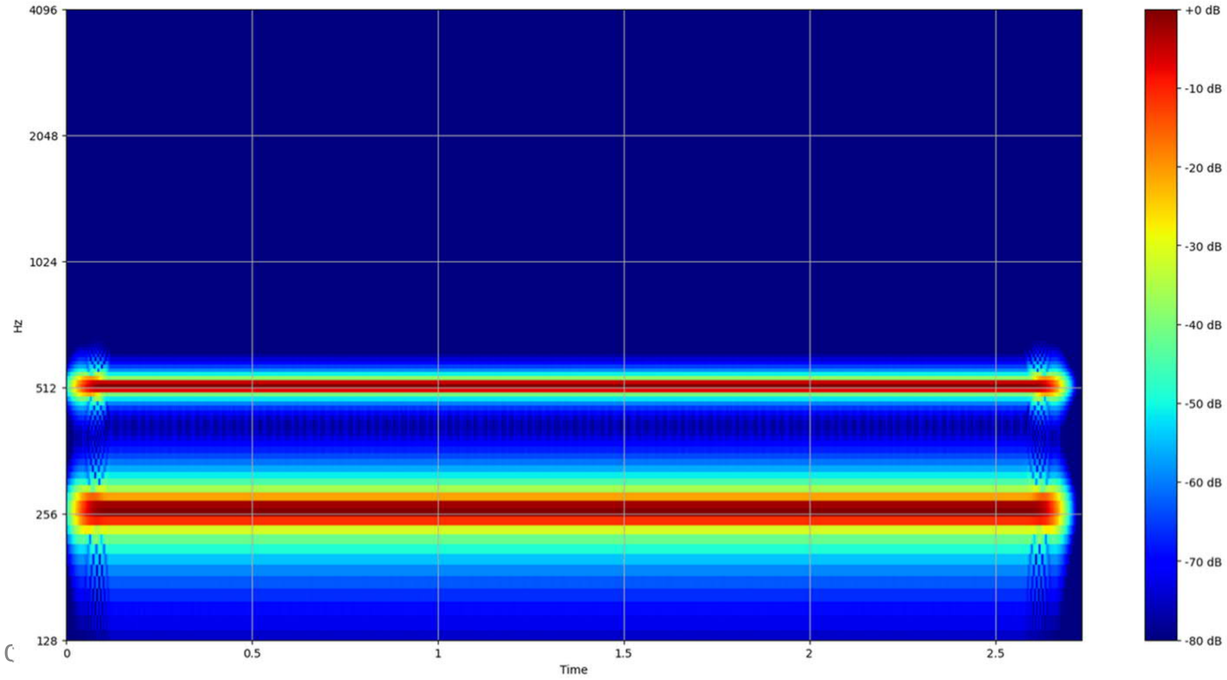
Combining Frequencies: P8

- Secondary beatings:
  $f_1 = C_4$, $f_2 \approx 2\, f_1$



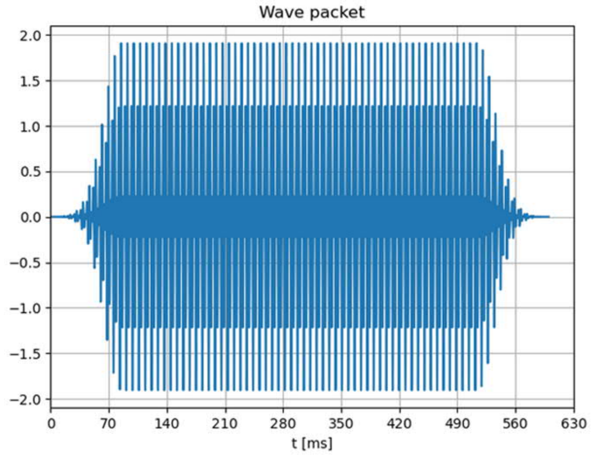


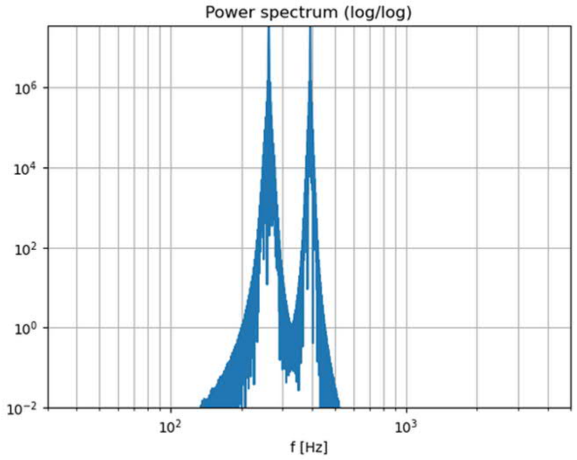
- We call this Mistuned Octave
  - Note the beatings pattern in the Spectrogram for the higher harmonics
  - The beatings do not vanish if we keep only the fundamentals!



Combining Frequencies: P5

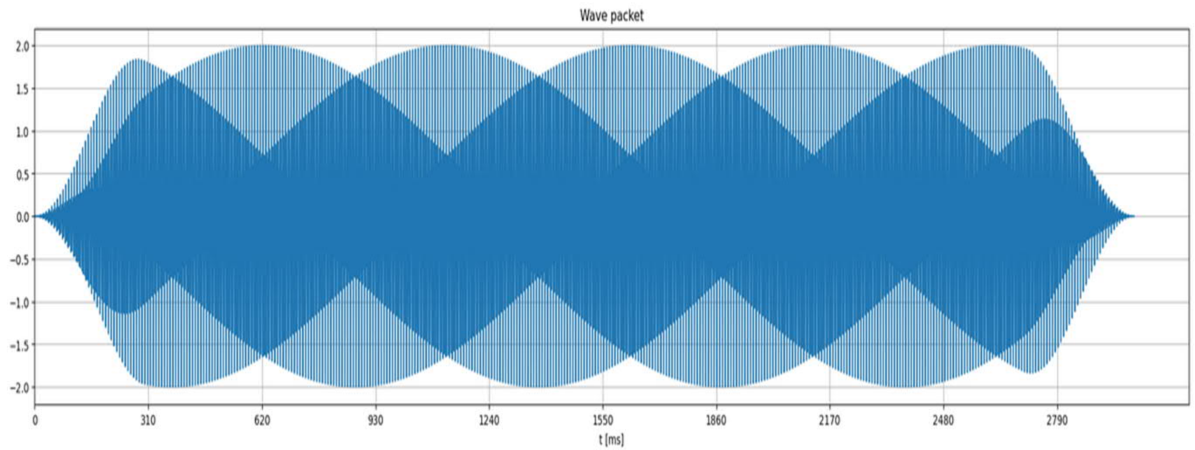
- Let's now focus on
  $f_2 = 3/2\ f_1$
- So called a Perfect Fifth or P5
  (We will see tomorrow why "Fifth")





Combining Frequencies: P5

- Let's now focus on
  $f_2 = 3/2\ f_1$
- So called a Perfect Fifth or P5
  (We will see tomorrow why "Fifth")
- Secondary beatings
  if $f_2 = 3/2\ f_1 - 1\ \text{Hz} \approx 3/2\ f_1$



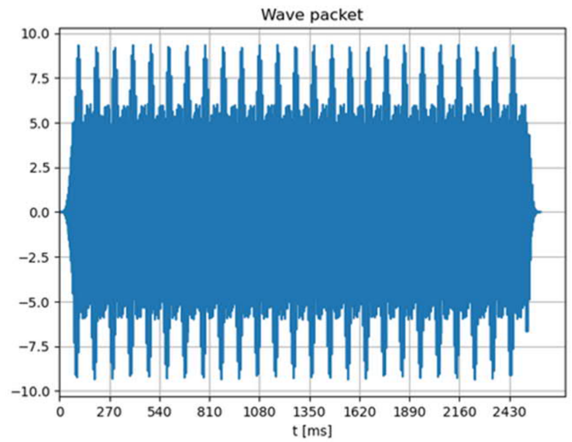
The hear is not sensitive to a difference of 1 Hz. However, the hear is extremely sensitive to beatings: this is why we "tune" musical instruments based on beatings!

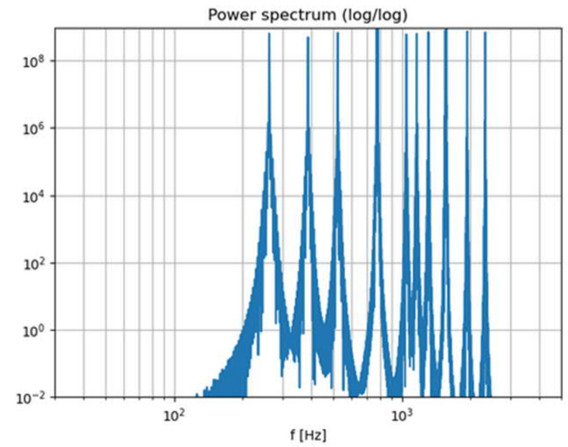
- We call this Mistuned Fifth
  - Similar considerations as for the Mistuned Octave

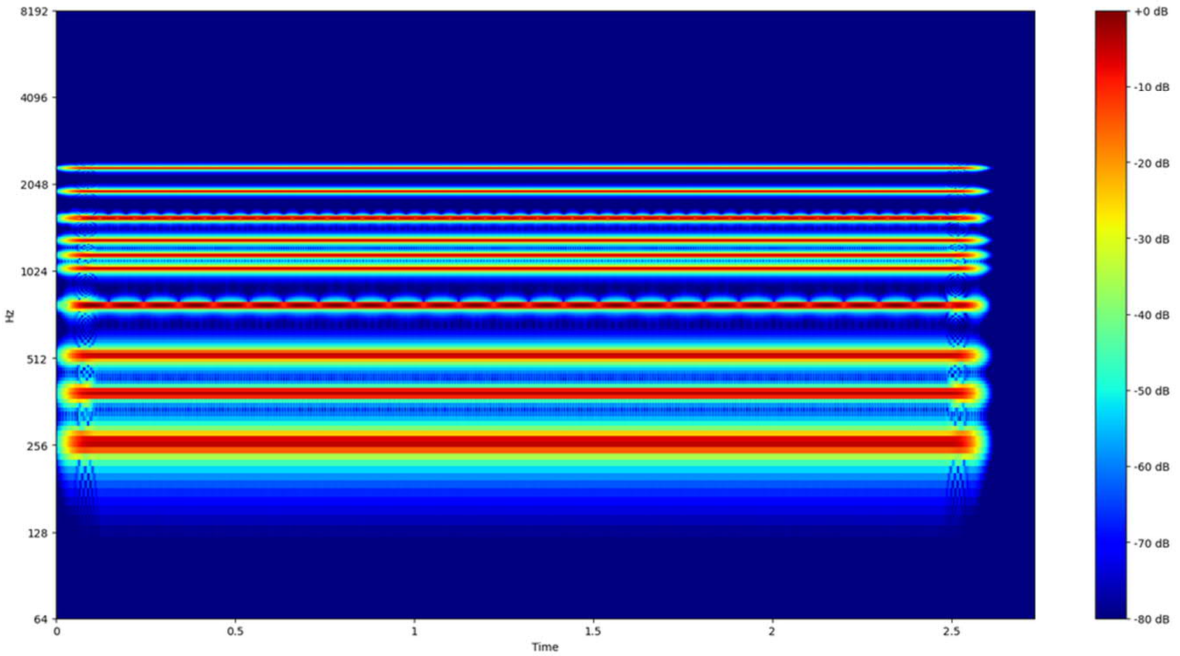
Combining Frequencies: P5

- Let's now focus on
  $f_2 = 3/2\ f_1$
- So called a Perfect Fifth or P5
  (We will see tomorrow why "Fifth")
- Secondary beatings
  if $f_2 = 3/2\ f_1 - 1\ \text{Hz} \approx 3/2\ f_1$

- We call this Mistuned Fifth
  - Similar considerations as for the Mistuned Octave
  - Higher harmonics reinforce beatings







Combining Frequencies: the Fundamental Bass

- Let's now assume:

$$\frac{f_2}{f_1} = \frac{m}{n} \quad m, n \in \mathbb{N} \text{ and coprime}$$

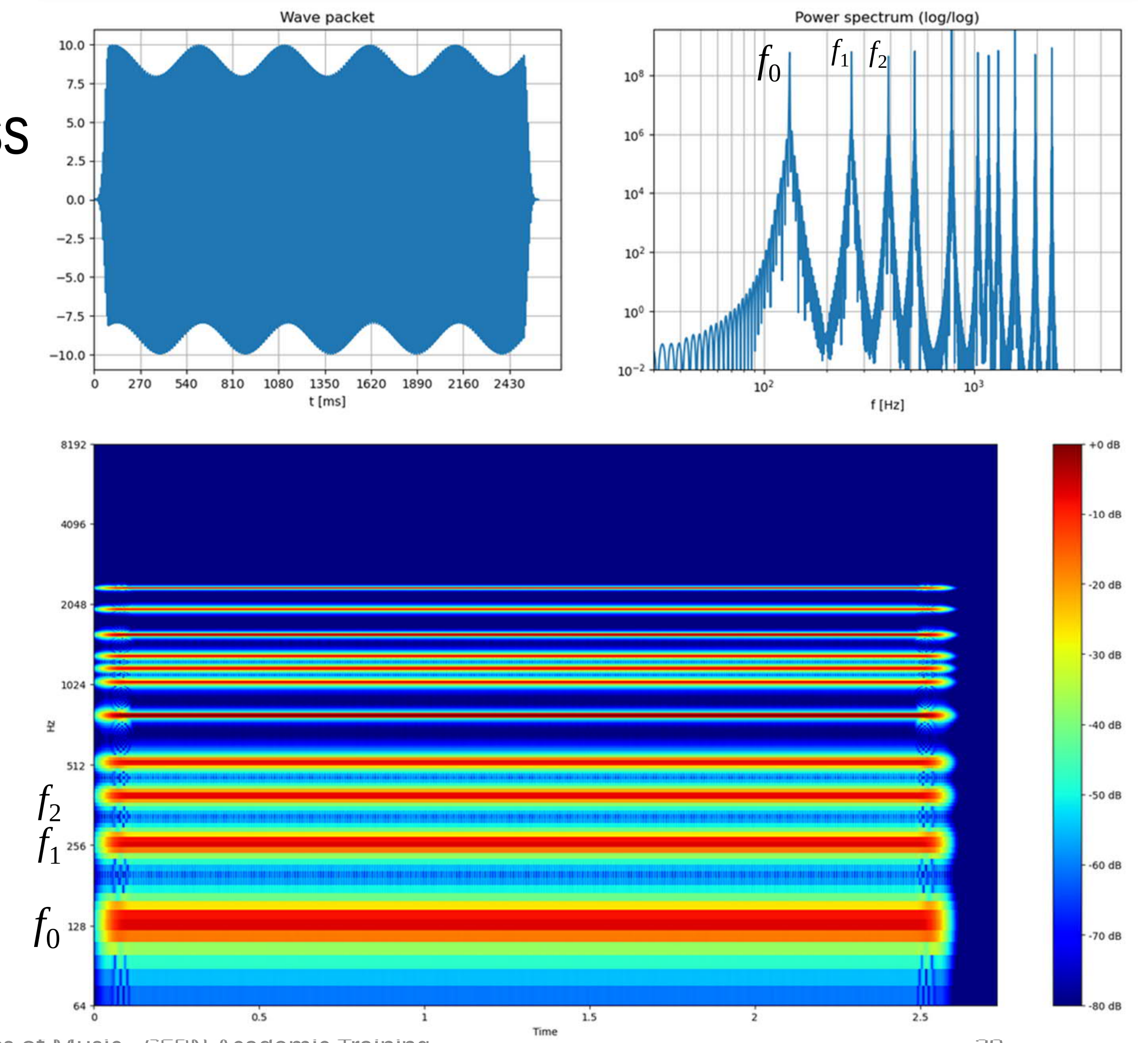
- We can define a new frequency f_0 , which we denote the Fundamental Bass Frequency, as:

$$f_0 = \frac{f_2}{m} = \frac{f_1}{n}$$

- This is also known as the Common Bass or Missing Fundamental, and its inverse represents the period of the combined waveform

Combining Frequencies: P5 + Fundamental Bass

- Can we detect the Fundamental Bass when playing e.g. a P5?
  - What about exploiting our sensitivity to beatings?
- Let's play a P5 and add a single frequency $f = f_0 + 2$ Hz
  - Incidentally, the most prominent *Combination Tone*, so-called *Tartini's sound*, has the same frequency as the FB for a P5



Combining Frequencies: Recap

- We have introduced the Psychoacoustic concepts of Roughness, Critical Bandwidth, Discrimination Limen, Fundamental Bass
- We have observed different forms of beatings and their waveforms
  - The ear is sensitive to “well tuned” vs “mistuned” simple ratios

- We observe a varying degree of “consonance” or “dissonance” when changing the f_2 / f_1 ratio
  - This appears to be correlated to the physical properties of the resulting wave
  - “Cultural”, “subjective” component also present to some degree

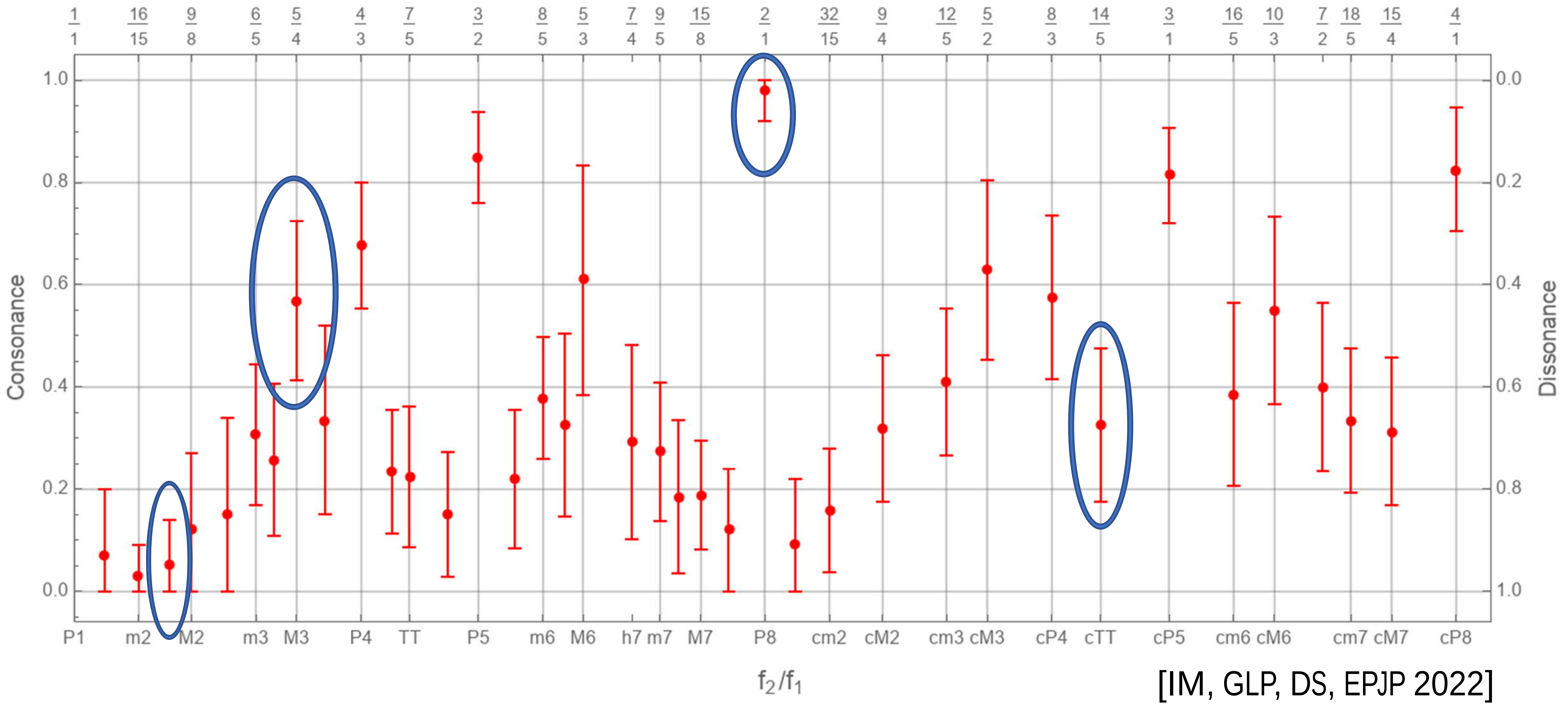
A CONSONANCE TEST

- Can we try and formalize this concept?
  Can we produce a psychoacoustic test?
  - Yes! Several scientists and musical theorists did it in the past, limited to one octave and using a specific musical instrument
    - From Foderà (early 1800) to Bowling et al. (2018)

  - We tried it ourselves, too, with some variants: we built a harmonic timbre that does not resemble too much a familiar instrument
    - slow exponential decay + partials with 1/n amplitude
  - We chose 38 ratios, with $f_1 = C_4$ and f_2 running up to $4 f_1$ = 2 octaves
  - We let volunteers hear several dyads and provide a degree of perceived consonance within a scale 1 to 5:

1 = very dissonant, 3 = neither diss. nor cons., 5 = very consonant

Test Results



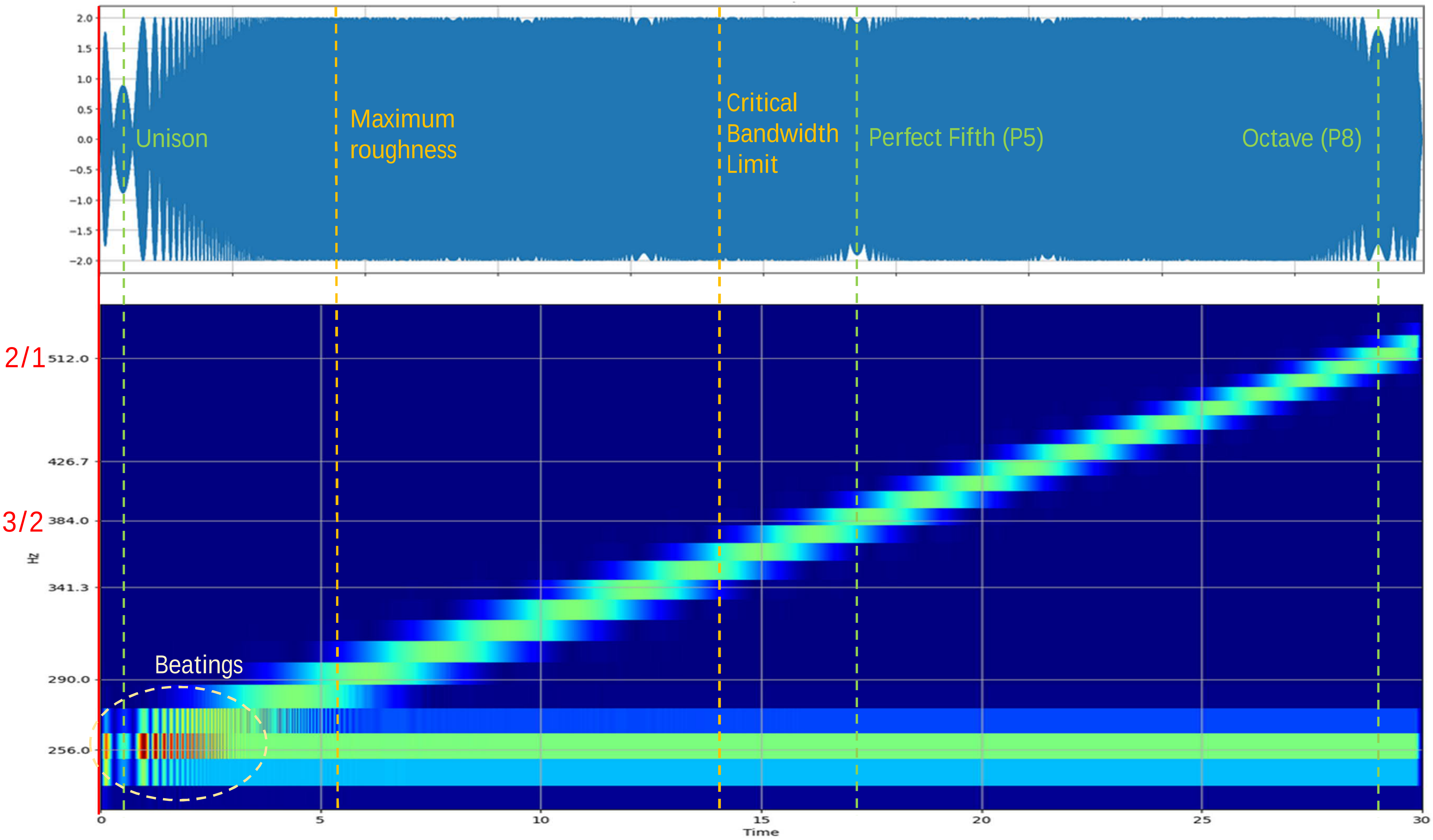
[IM, GLP, DS, EPJP 2022]

Results from 20 individuals. Error bars represent one standard deviation.

Test Results

- How can we interpret those results?
- Can we describe a model of C&D that fits them, in terms of the “toolbox” we have analyzed?

→ Tomorrow: Theoretical Physics to the rescue!



2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
Unison
Maximum roughness
Critical Bandwidth Limit
Perfect Fifth (P5)
Octave (P8)
2/1
512.0
426.7
3/2
384.0
Hz
341.3
290.0
Beatings
256.0
0
5
10
15
20
25
30
Time