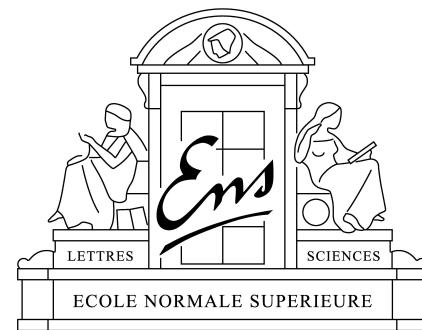


Pitch

Alain de Cheveigné

CNRS / Ecole Normale Supérieure / UCL



Outline

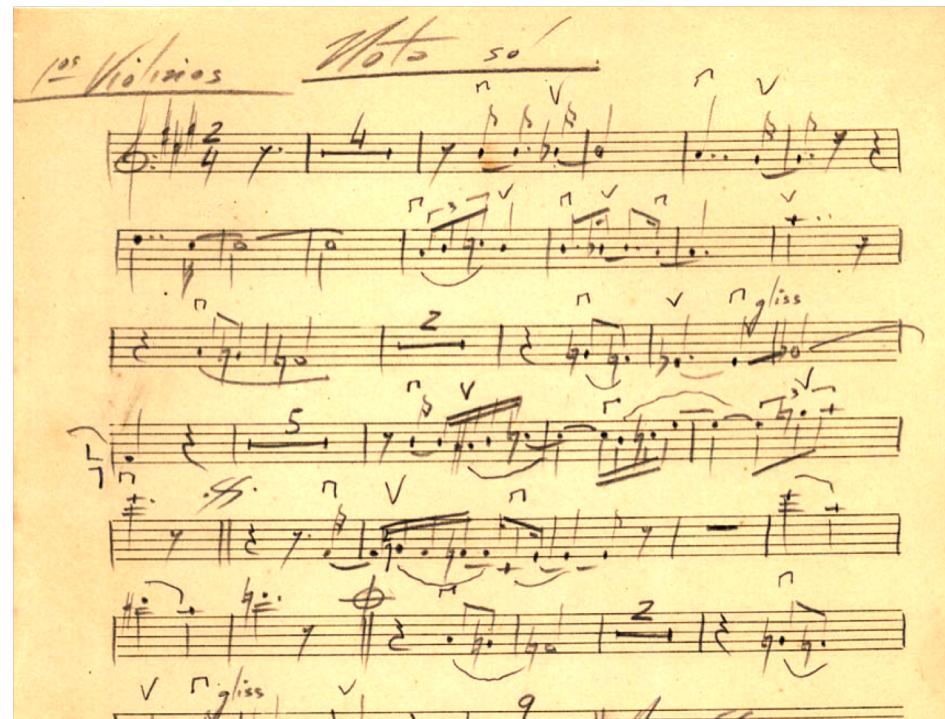
- definition of pitch
- basic facts
- mechanisms, models, and mysteries
- pitch and F0
- F0 estimation YIN

What is pitch?

What is pitch?

everyday definition:

"pitch is the stuff of which music is made"



What is pitch?

psychological definition:

"that attribute of auditory sensation in terms of which **sounds may be ordered** on a scale extending **from low to high**"

ANSI (1973)

What is pitch?

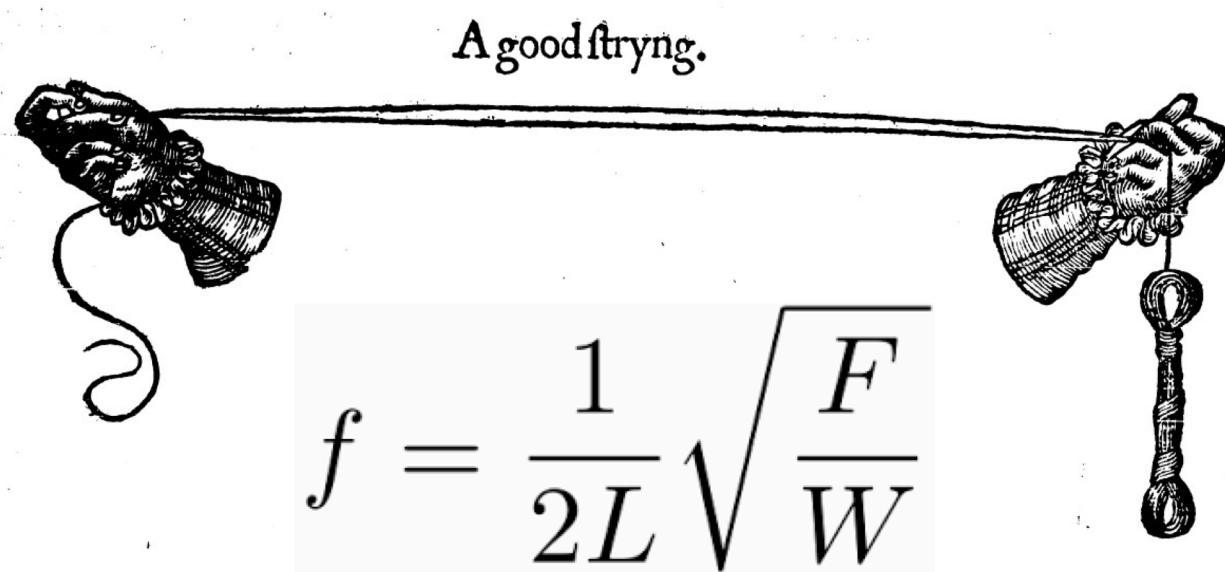
psychophysical definition:

"that attribute of auditory sensation, related to the frequency of a periodic sound, in terms of which sounds may be judged as **dull or sharp** according to whether the **frequency is low or high**"

AFNOR (1977)

The quantitative relation between pitch and frequency was established by Mersenne, (1636)

What is pitch?



$$f = \frac{1}{2L} \sqrt{\frac{F}{W}}$$

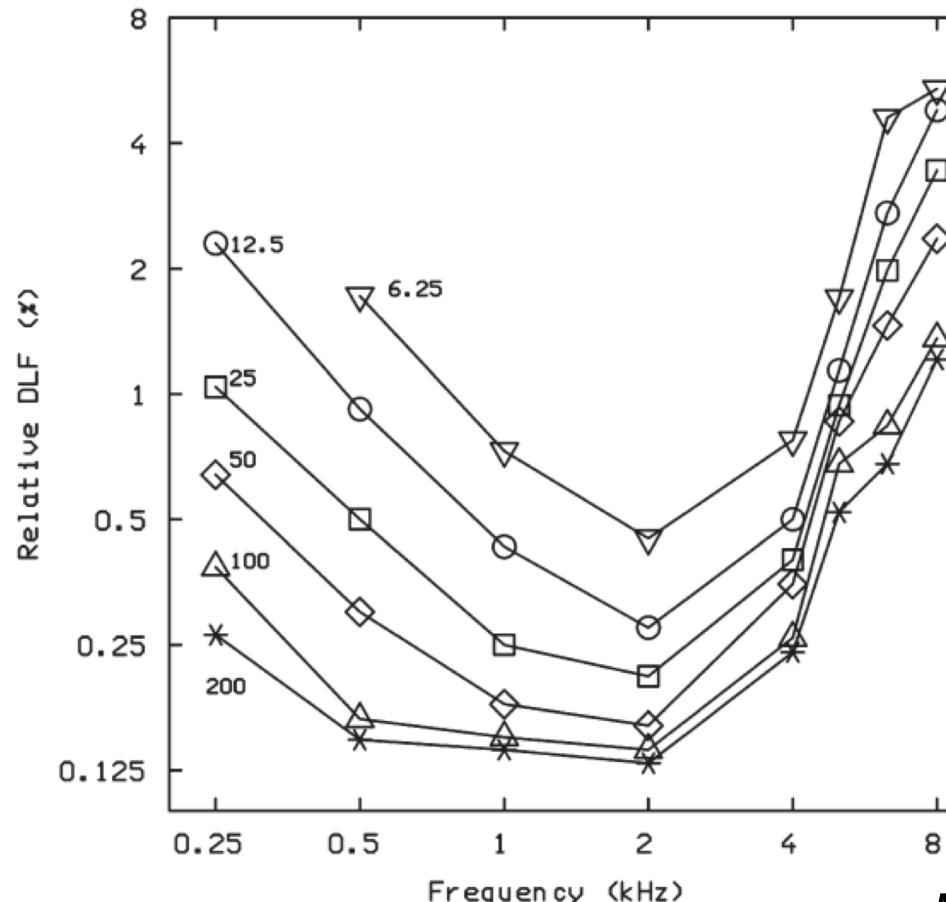


Mersenne, *Traité de l'harmonie universelle* (1636)

Some facts

Some facts

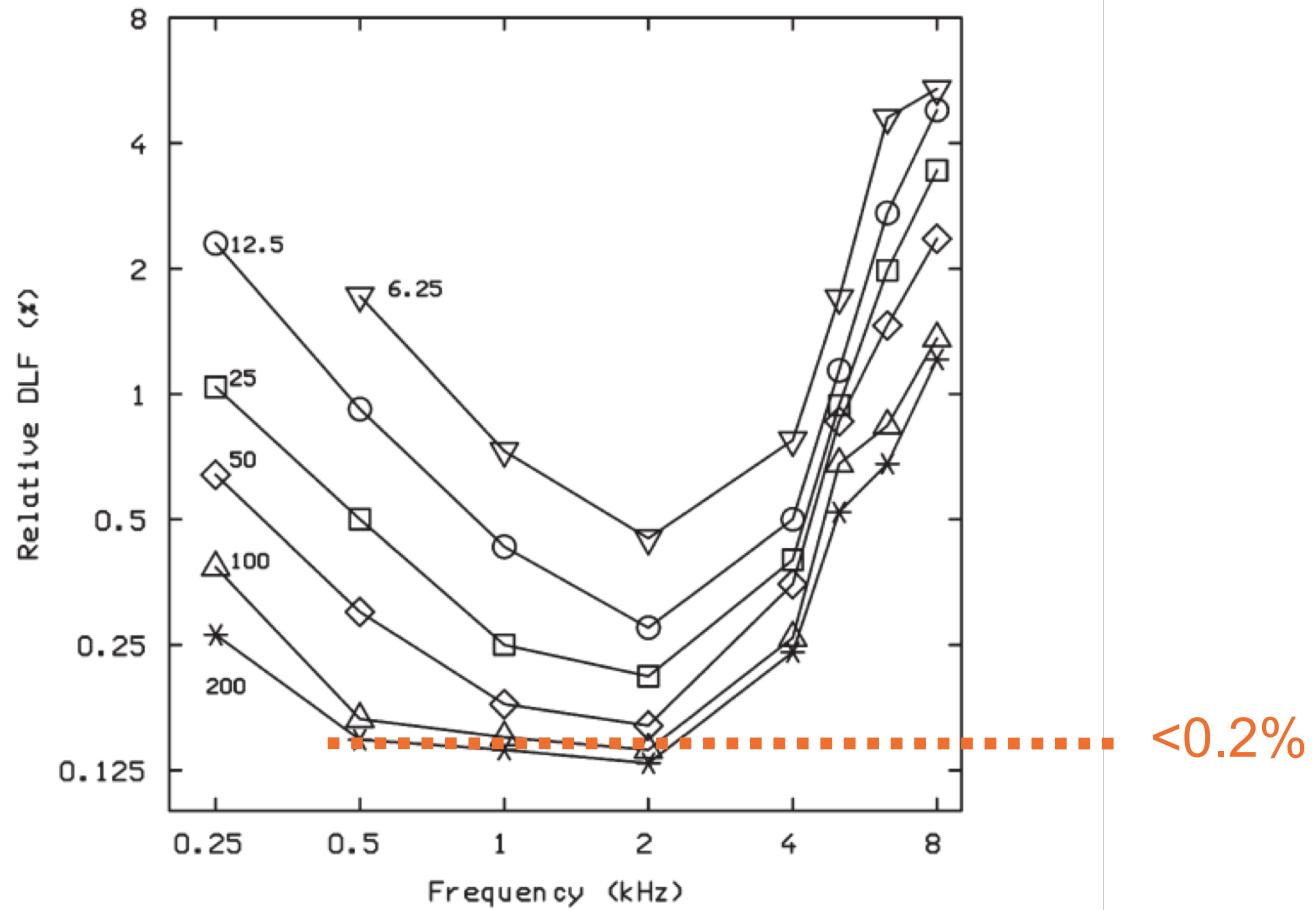
pure tone frequency discrimination limens:



Moore, 1973

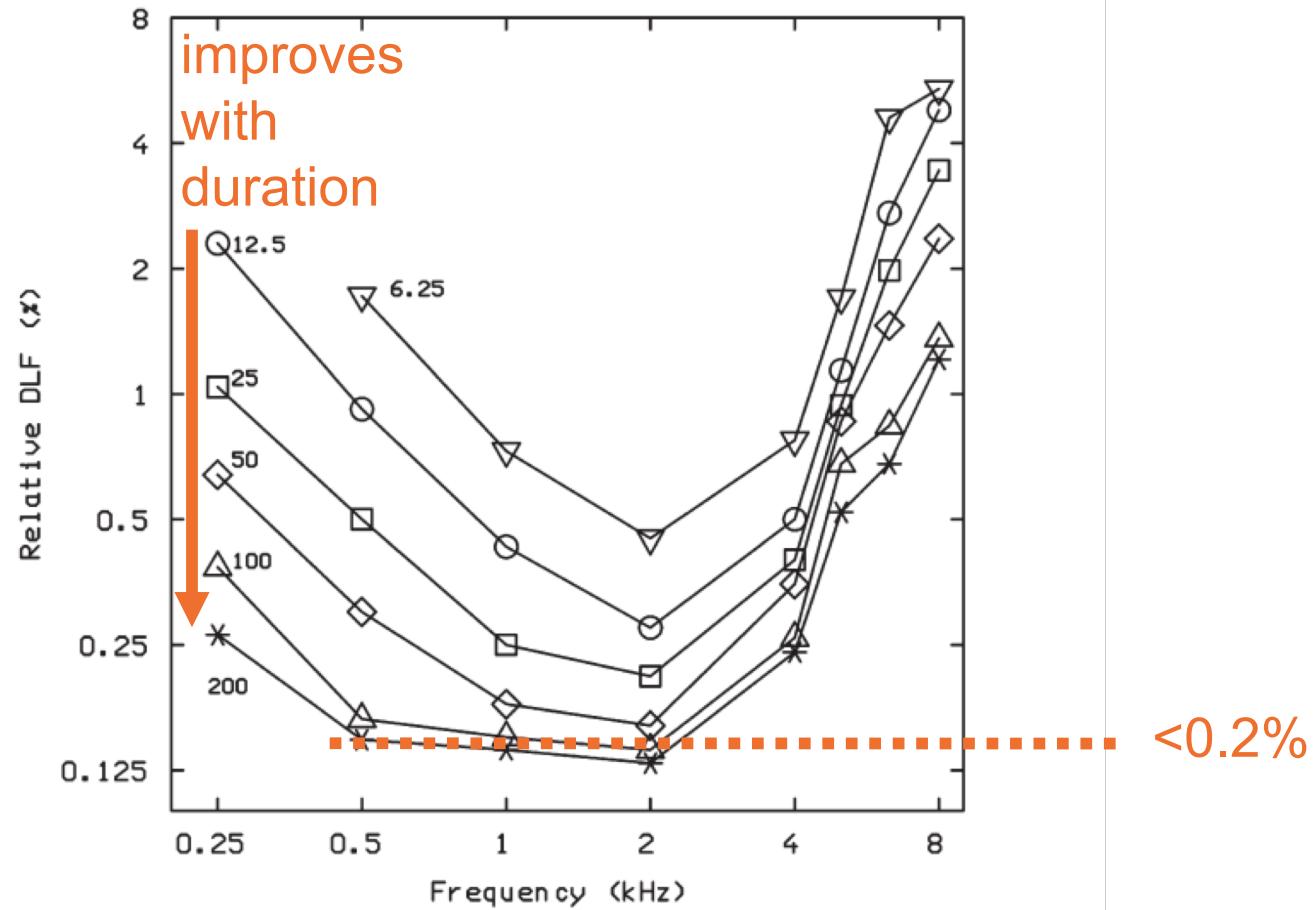
Some facts

pure tone frequency discrimination limens:



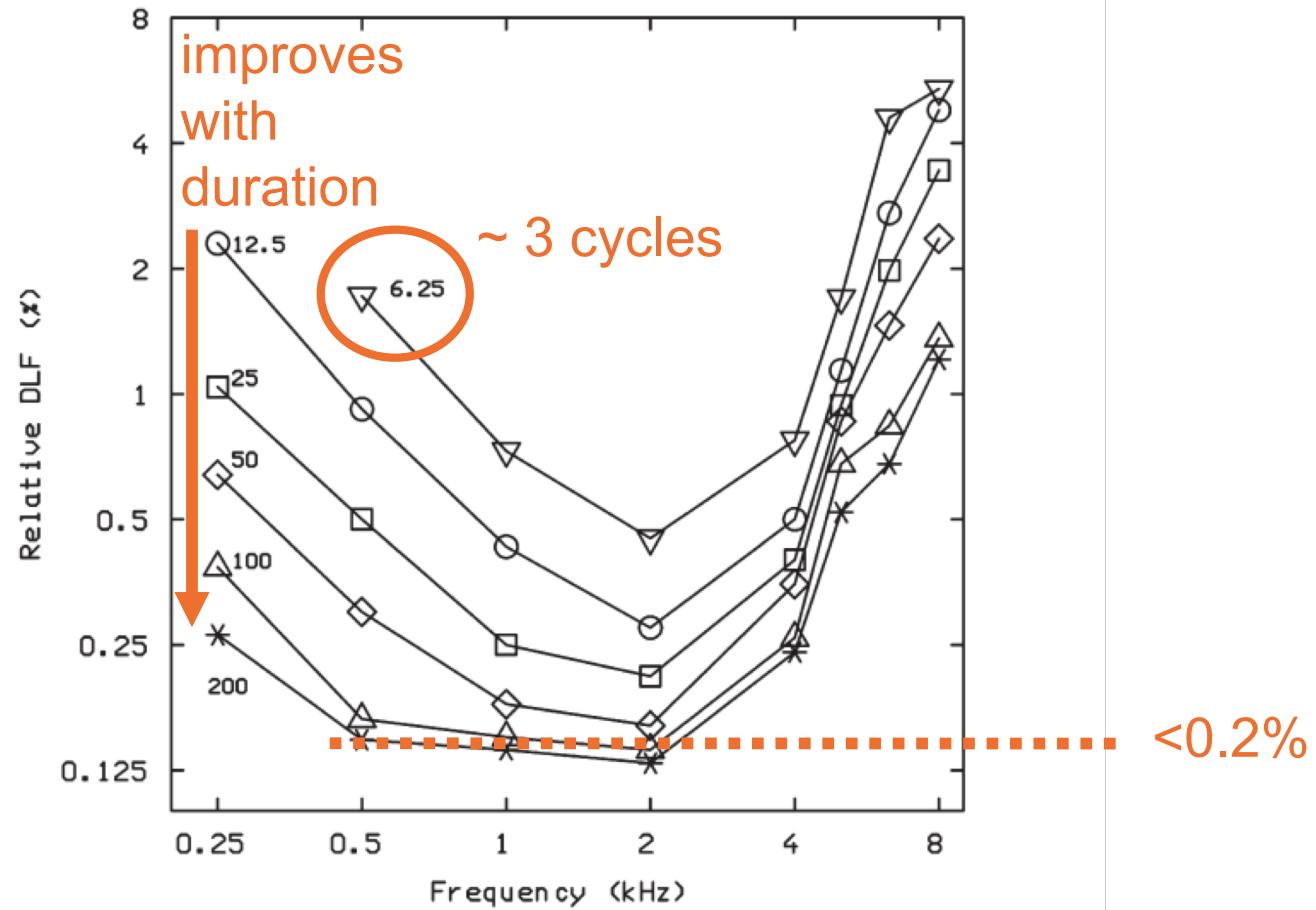
Some facts

pure tone frequency discrimination limens:



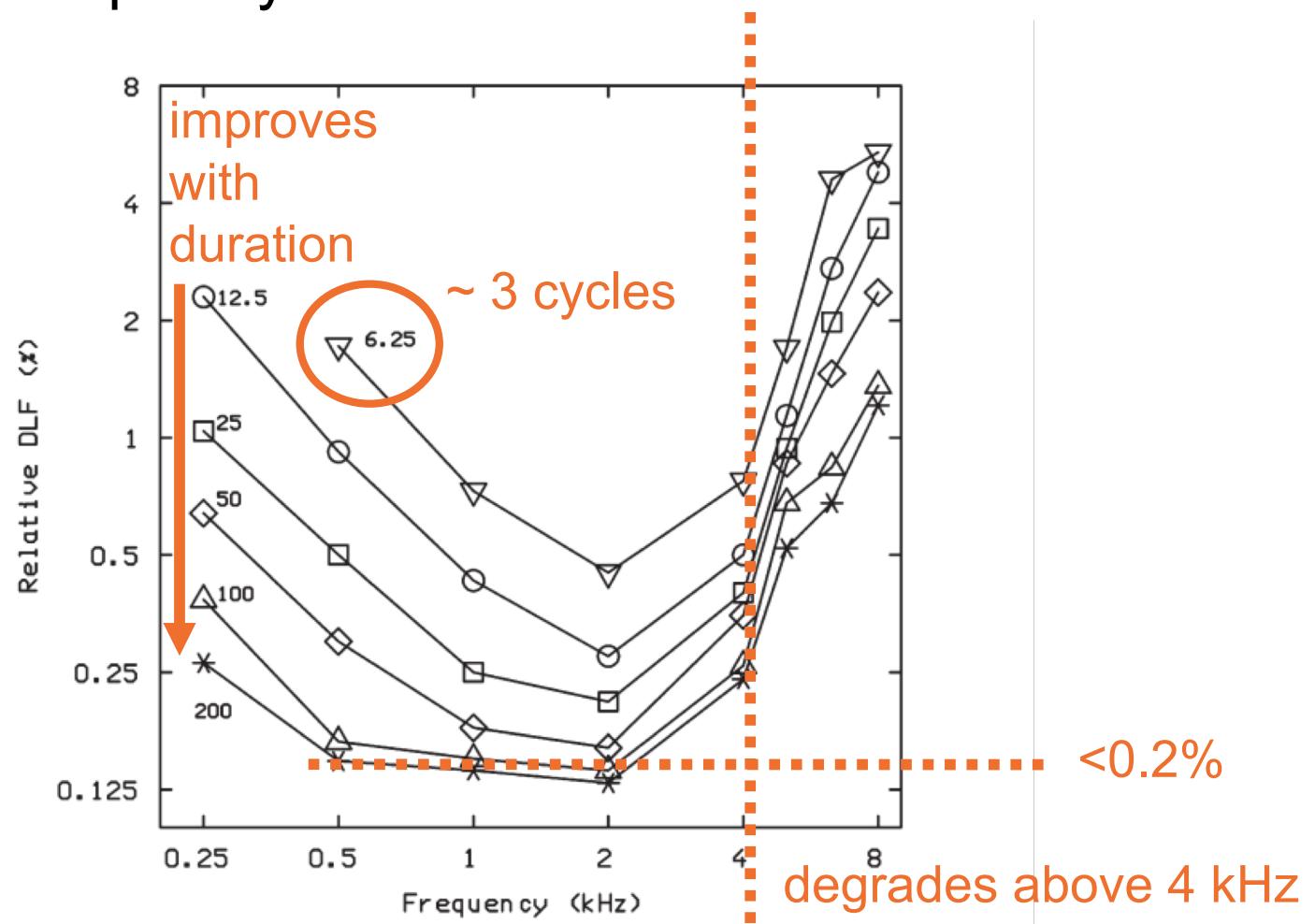
Some facts

pure tone frequency discrimination limens:



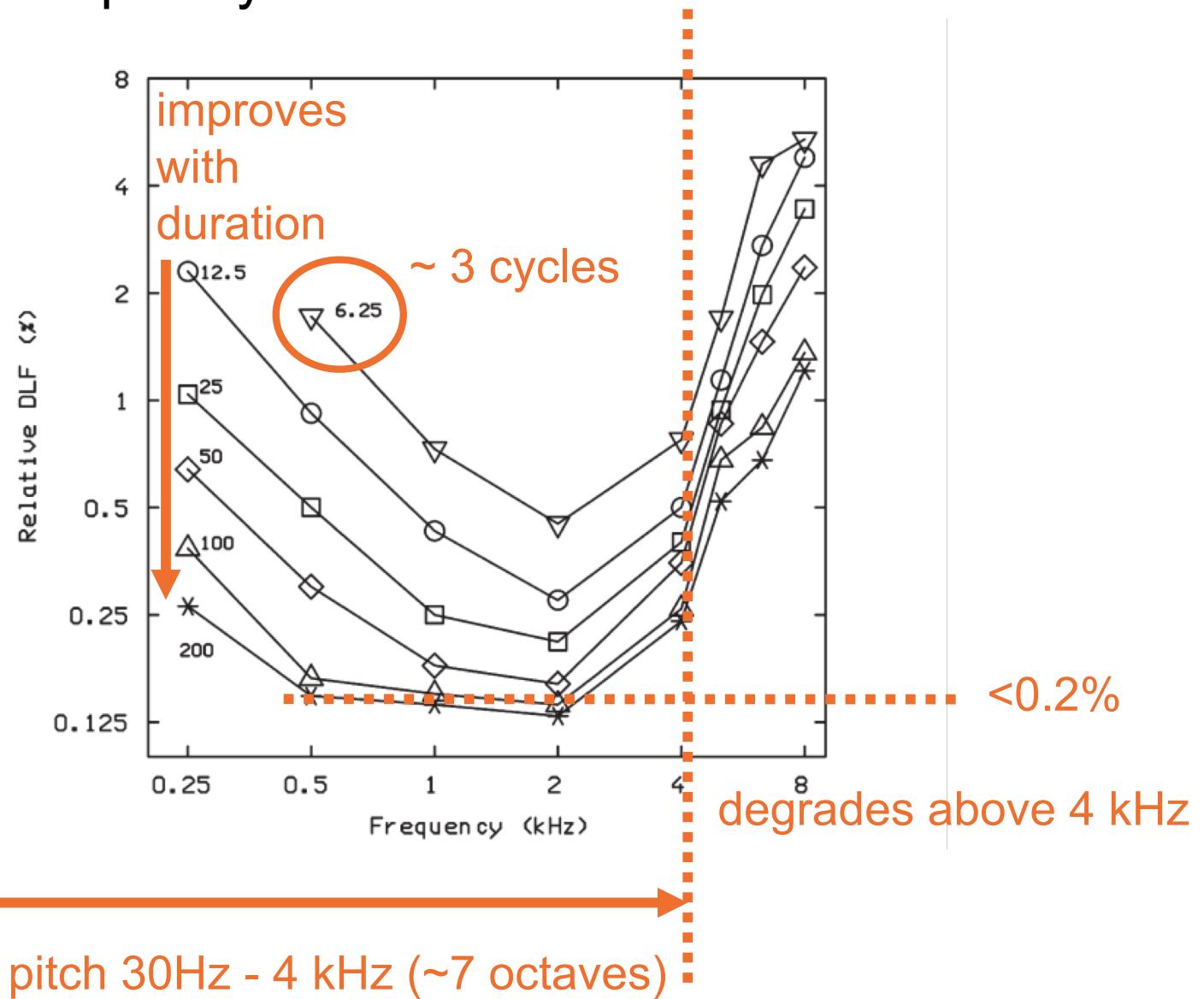
Some facts

pure tone frequency discrimination limens:



Some facts

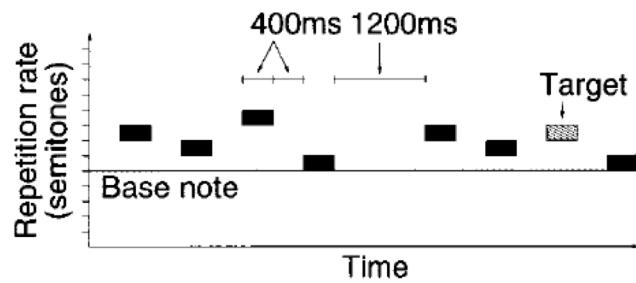
pure tone frequency discrimination limens:



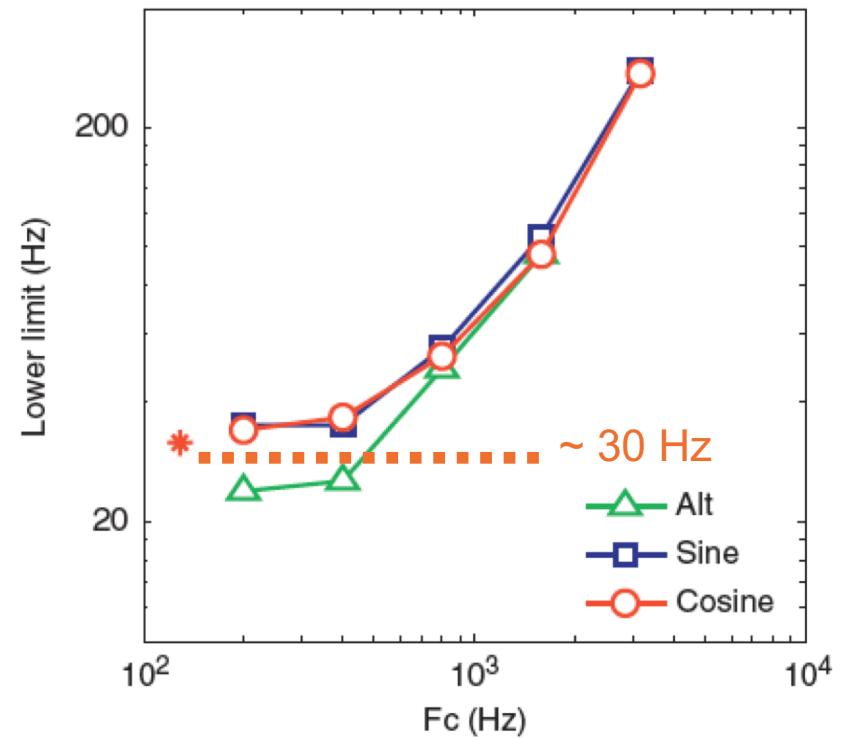
Some facts

lower limit of melodic pitch:

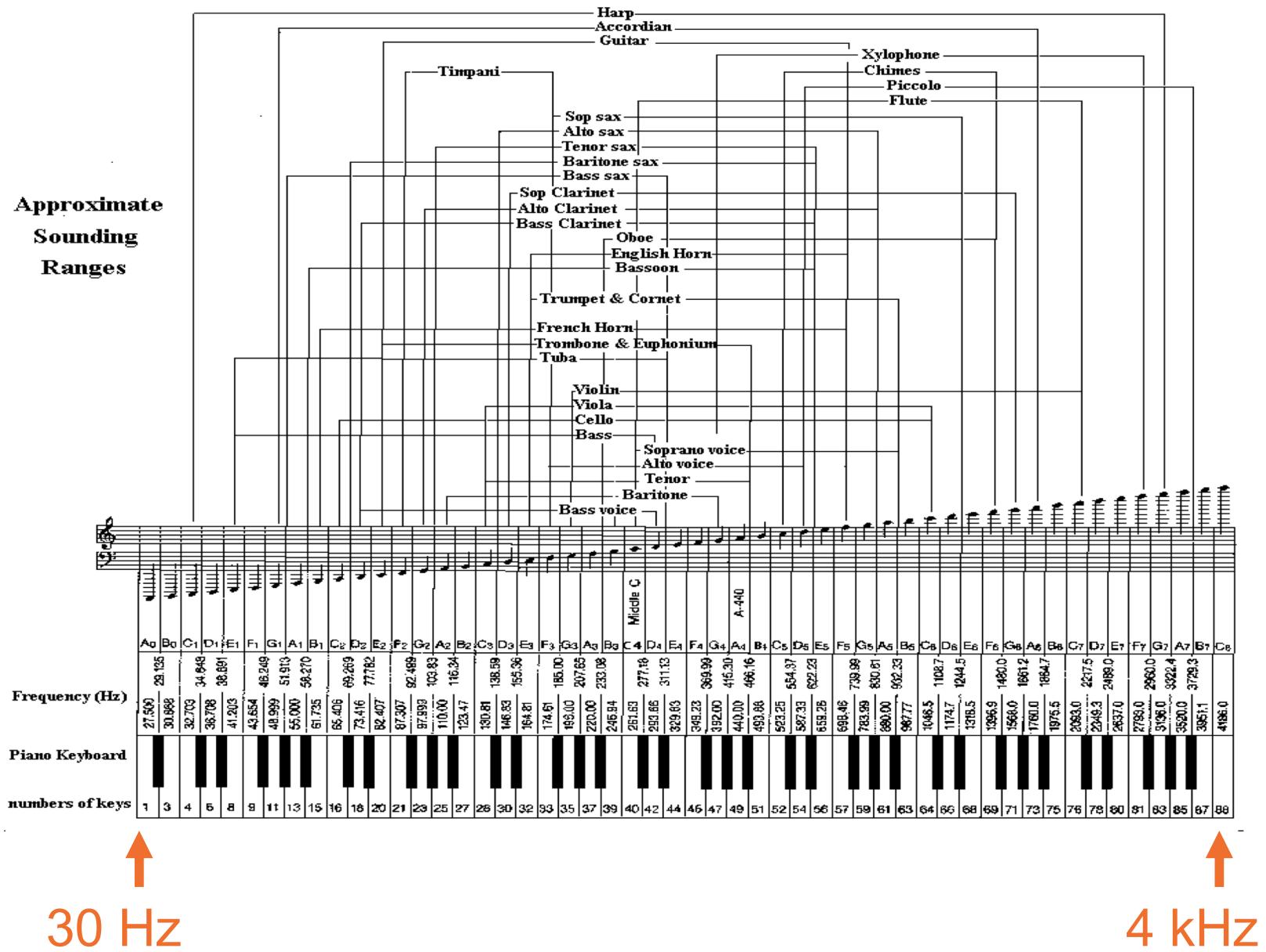
task:



Pressnitzer et al 2001



Some facts



Some facts

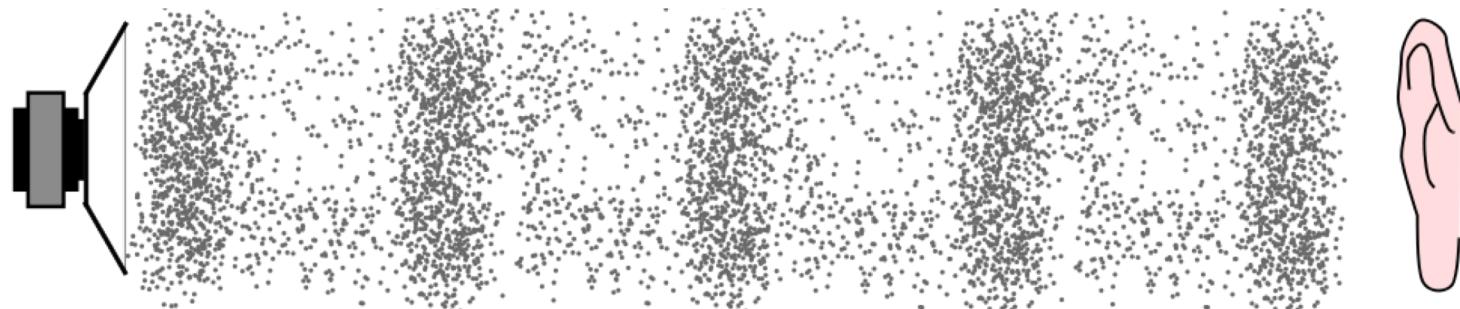
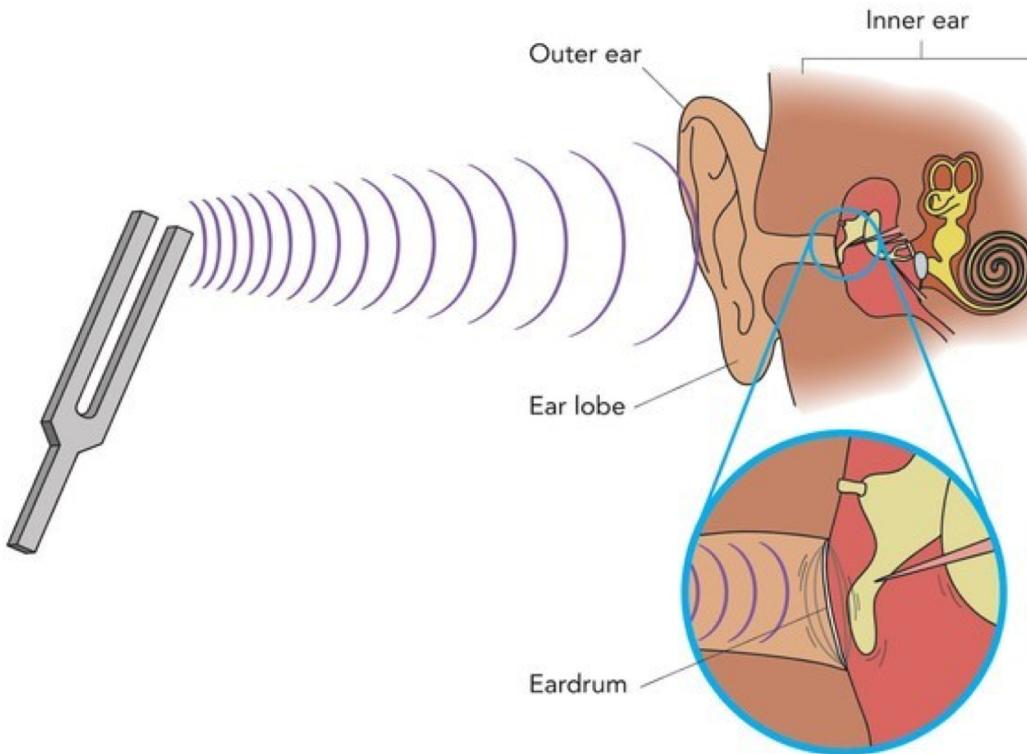
other facts:

- phase insensitive (Ohm's law)
at least for complexes with harmonics of low rank
- stable over wide range of amplitudes
a few % shift for pure tones, less for complex
- stable over wide range of durations
slight shift for very short stimuli

→ pitch is highly invariant to changes over many dimensions

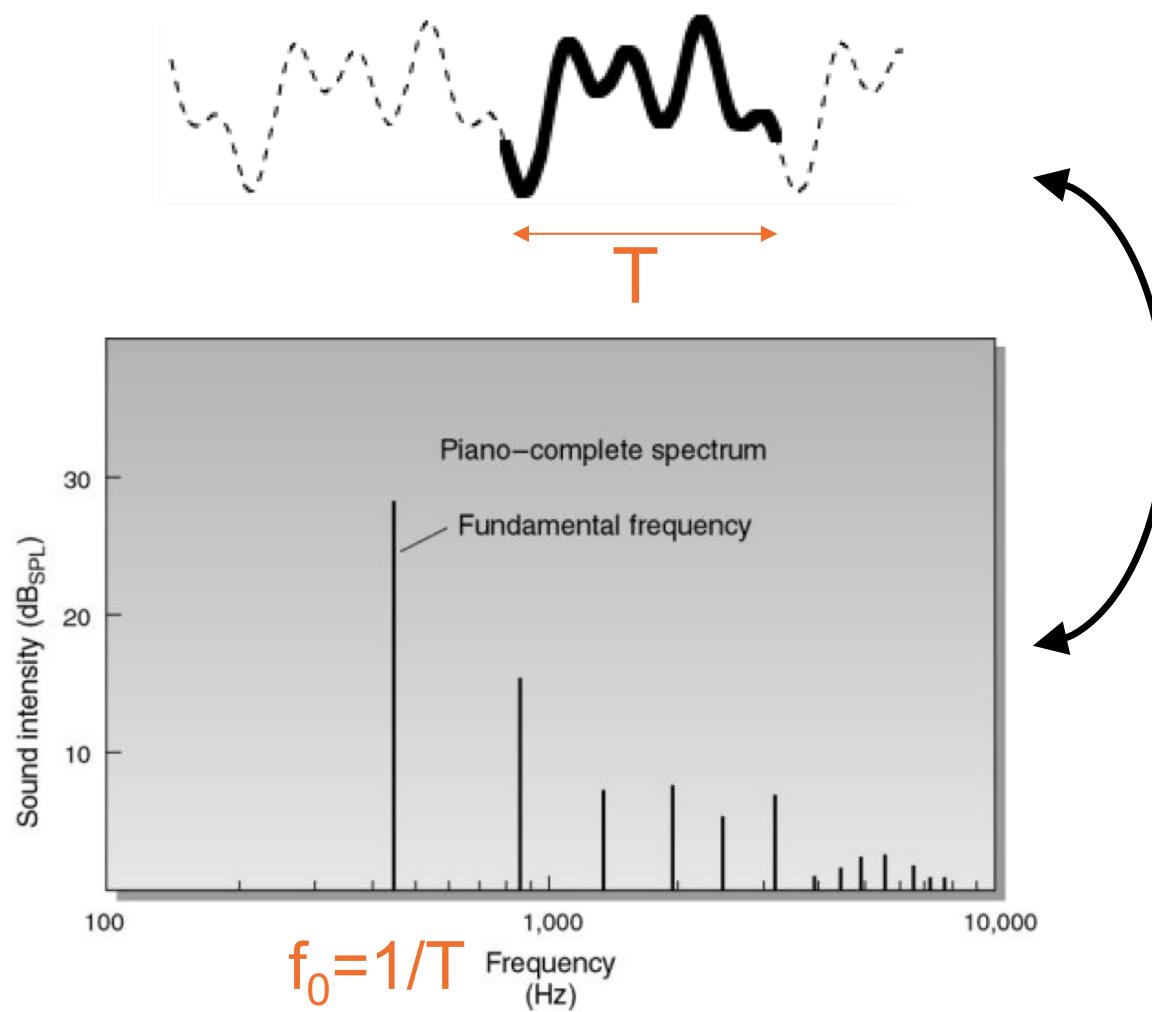
What is pitch (again)?

What is pitch (again)?



What is pitch (again)?

periodic sound signal: $s(t) = s(t+T), \forall t$

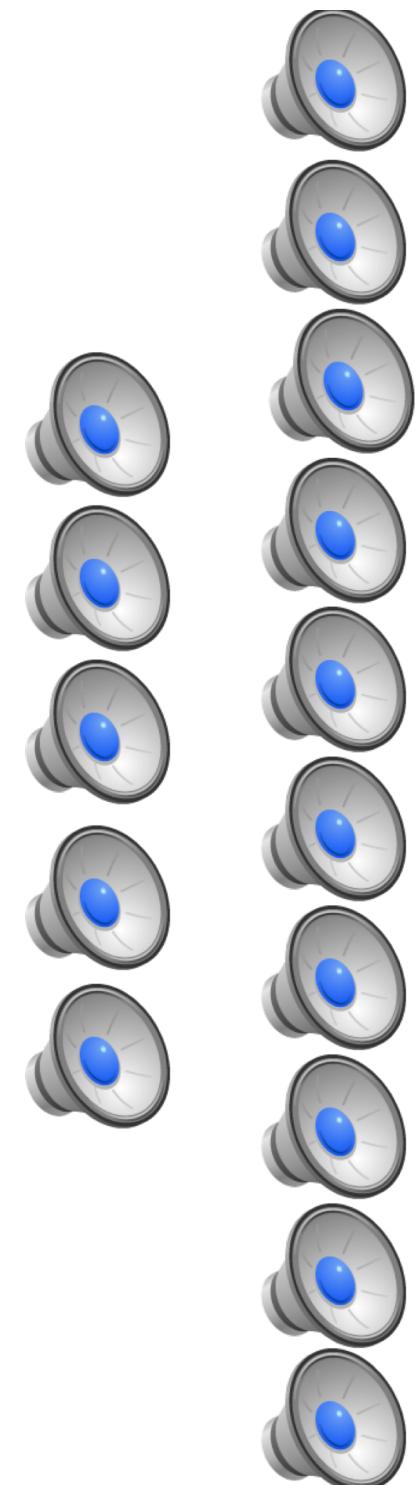
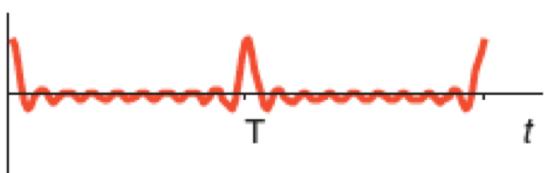
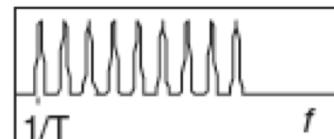
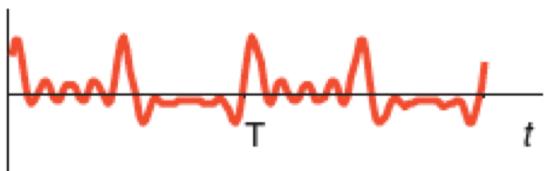
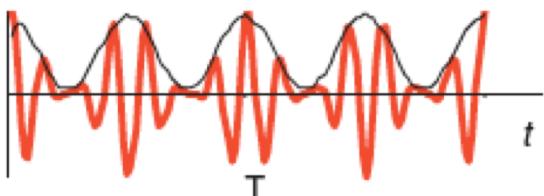
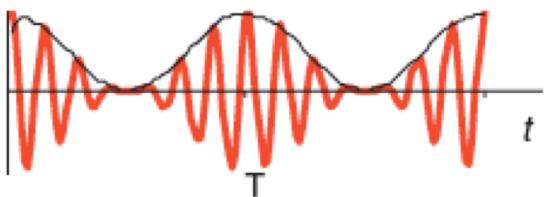
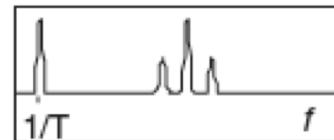
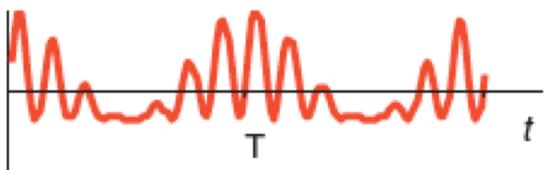
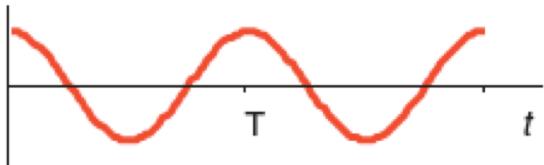


Jean-Baptiste-Joseph Fourier
(1768-1830)

Fourier transform

What is pitch (again)?

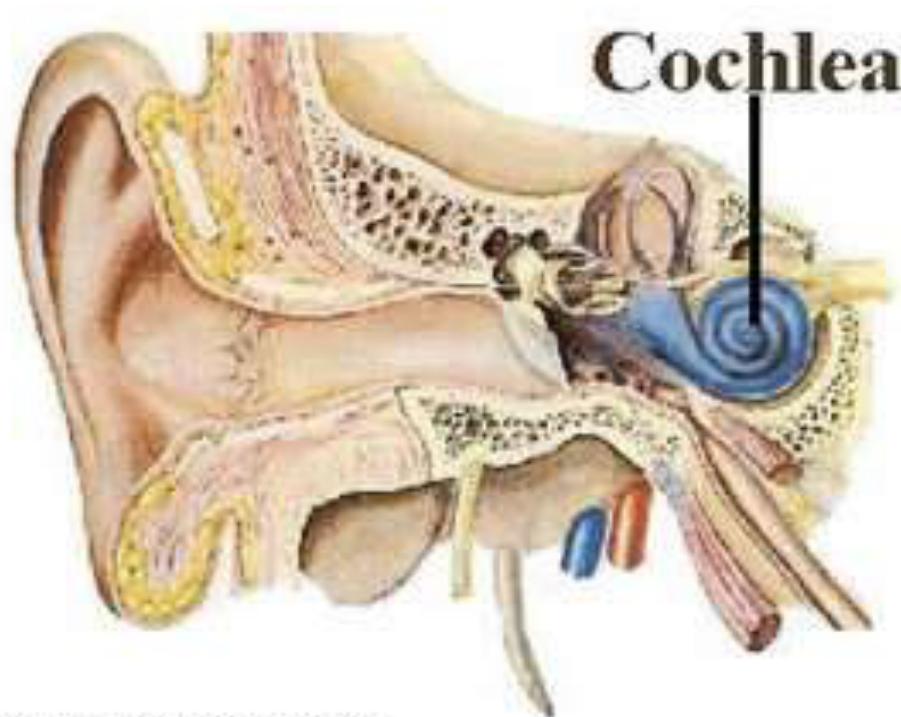
- periodic sounds evoke pitch
- pitch value (tone height) depends on $f_0=1/T$
- does not depend on shape



Pitch is an abstraction, a many-to-one mapping

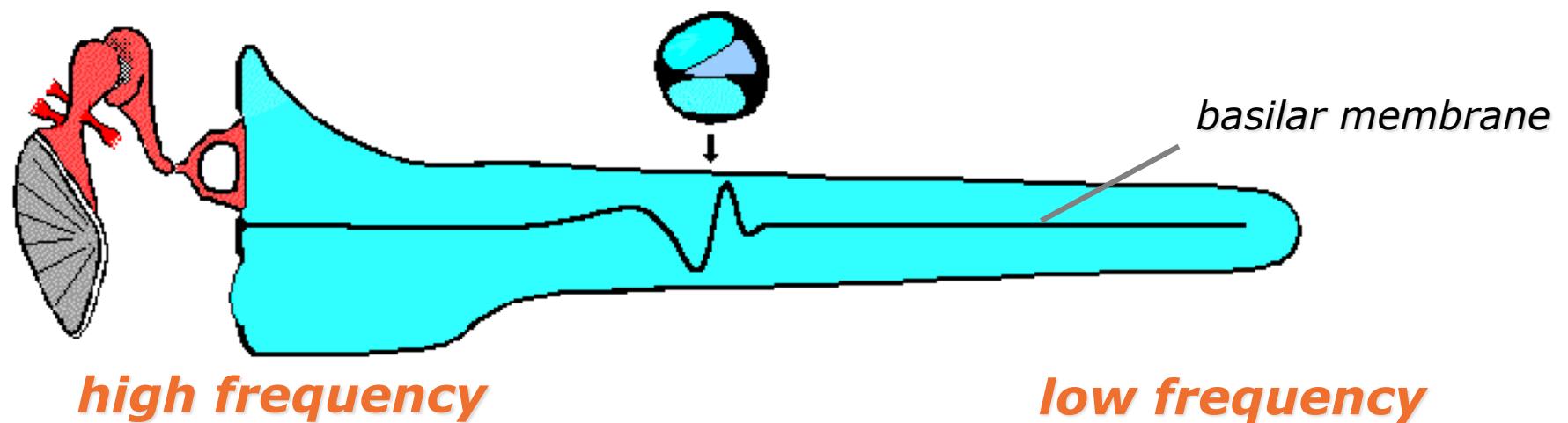
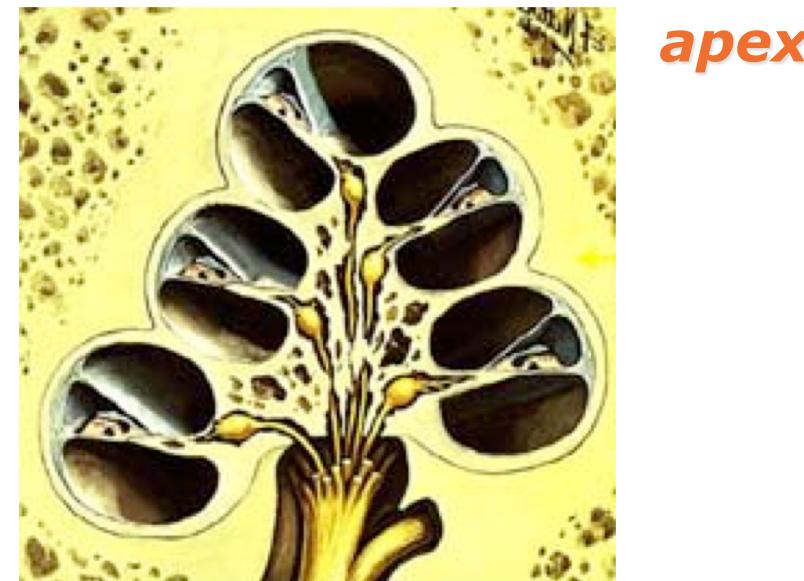
Physiological basis

Physiological basis

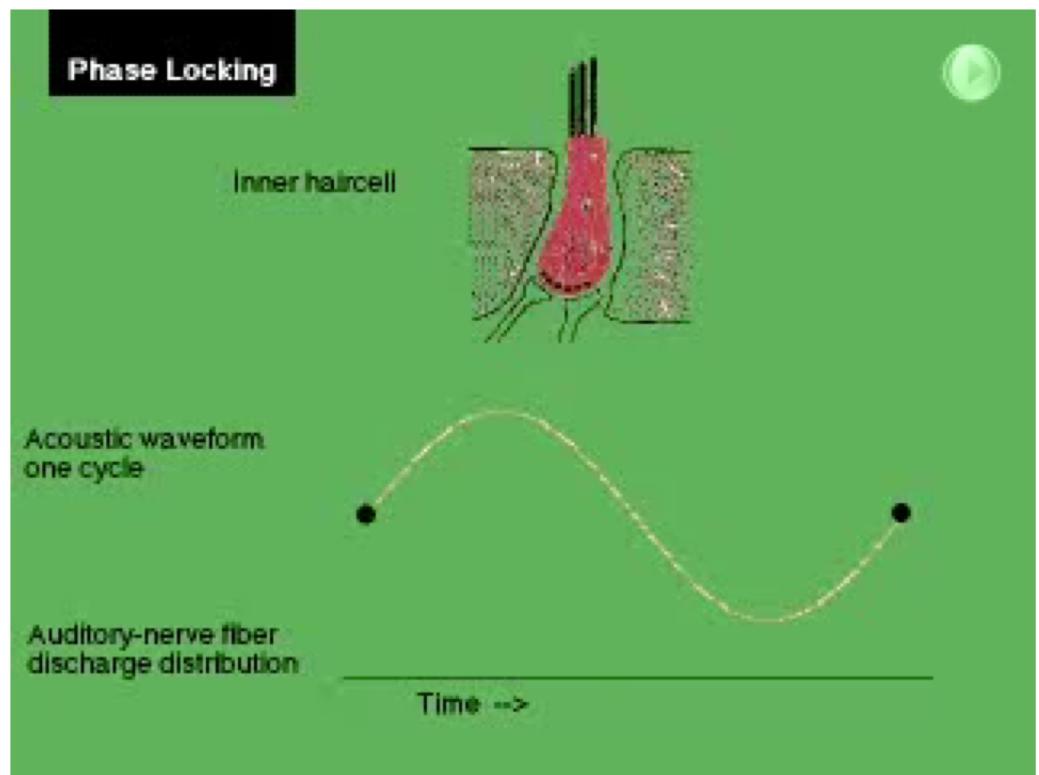
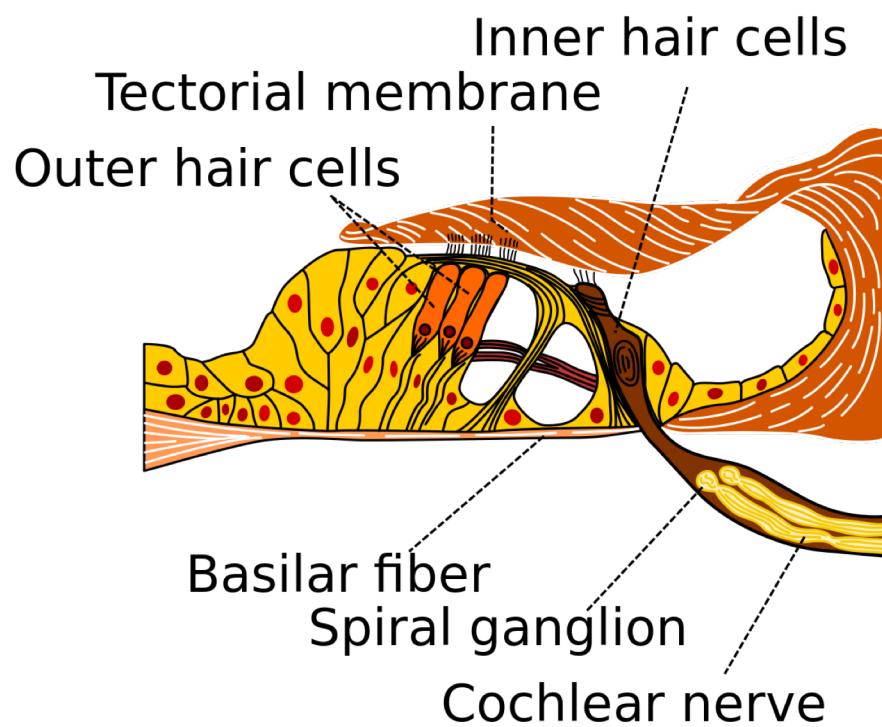


Alec H. Salt, Washington University

Physiological basis

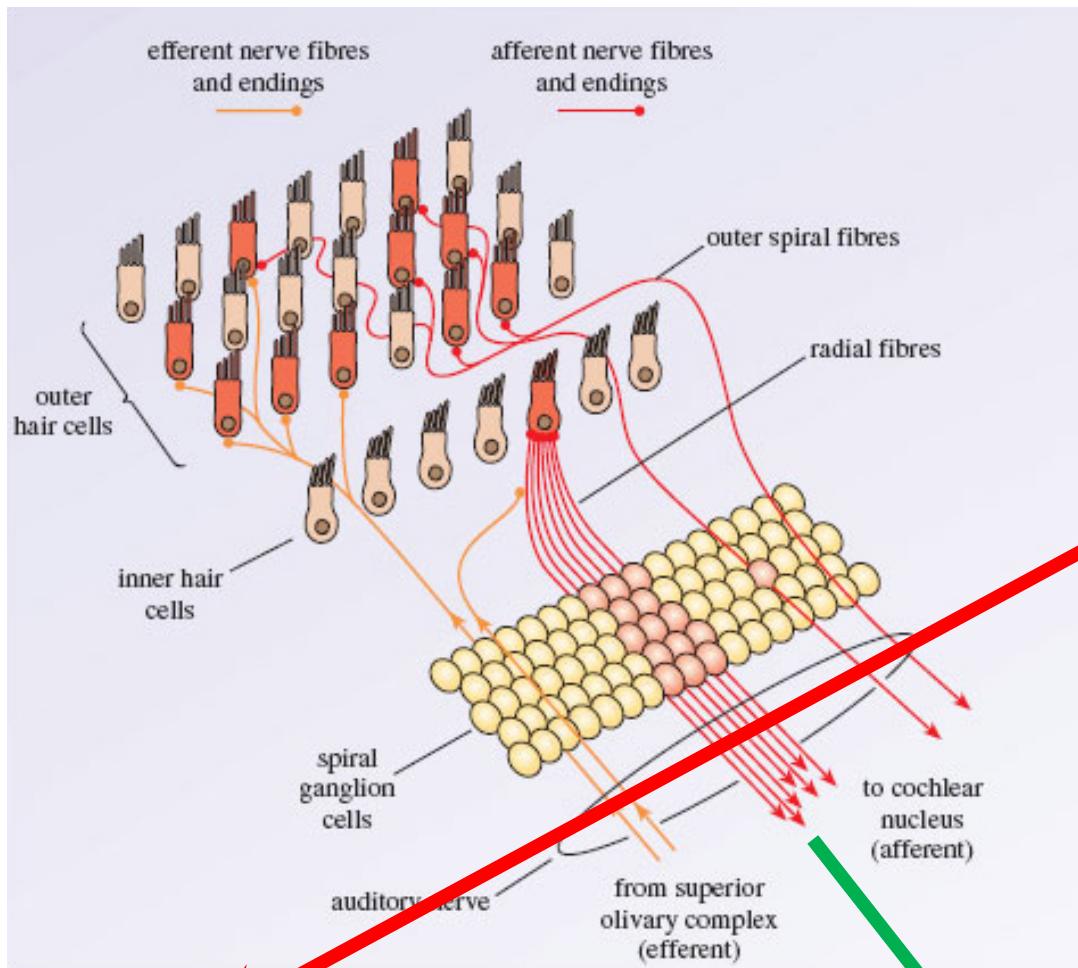


Physiological basis



→ to brain

Physiological basis



two coding principles:

- tonotopy (place)
- time

f

t

tonotopy:

TRAITE
DE
L'ORGANE
DE L'OUIE,

CONTENANT LA STRUCTURE,
les Usages & les Maladies de toutes
les parties de l'Oreille.

Par M. Du VERNEY, de l'Academie
Royale des Sciences, Conseiller, Medecin
Ordinaire du Roy, & Professeur en
Anatomie & en Chirurgie au Jardin
Royal des Plantes.

Fig. IV.



1683

tonotopy:

Du Verney (1683):

“[the spiral lamina] ... is not only capable of receiving the quivers of air, but its structure necessarily suggests that it can respond to all of their different characters

--> *concept of sympathetic vibration*

tonotopy:

Du Verney (1683):

“[the spiral lamina] ... is not only capable of receiving the quivers of air, but its structure necessarily suggests that it can respond to all of their different characters

... being wider at the beginning of the first turn than at the end of the last, one can say that the wider parts can be caused to vibrate while the others do not...

--> *concept of selectivity*

tonotopy:

Du Verney (1683):

“[the spiral lamina] ... is not only capable of receiving the quivers of air, but its structure necessarily suggests that it can respond to all of their different characters

... being wider at the beginning of the first turn than at the end of the last, one can say that the wider parts can be caused to vibrate while the others do not...

and are capable of **slower** vibrations and consequently respond to **deeper** tones, whereas if the narrower parts are hit their vibrations are **faster** and consequently respond to **sharper** tones...

--> *relation between frequency and pitch*
(Mersenne, Galileo ~ 1638)

tonotopy:

Du Verney (1683):

“[the spiral lamina] ... is not only capable of receiving the quivers of air, but its structure necessarily suggests that it can respond to all of their different characters

... being wider at the beginning of the first turn than at the end of the last, one can say that the **wider** parts can be caused to vibrate while the others do not...

and are capable of **slower** vibrations and consequently respond to deeper tones, whereas if the **narrower** parts are hit their vibrations are **faster** and consequently respond to sharper tones...

--> *concept of tonotopy*

tonotopy:

Du Verney (1683):

“[the spiral lamina] ... is not only capable of receiving the quivers of air, but its structure necessarily suggests that it can respond to all of their different characters

... being wider at the beginning of the first turn than at the end of the last, one can say that the wider parts can be caused to vibrate while the others do not...

and are capable of slower vibrations and consequently respond to deeper tones, whereas if the narrower parts are hit their vibrations are faster and consequently respond to sharper tones...

... such that according to the various motions of the **spiral lamina**, the spirits of the **nerve**... receive different impressions that represent in the **brain** the various aspects of tones.”

--> *concept of tonotopic projection*

tonotopy:

Du Verney (1683):

“[the spiral lamina] ... is not only capable of receiving the quivers of air, but its structure necessarily suggests that it can respond to all of their different characters

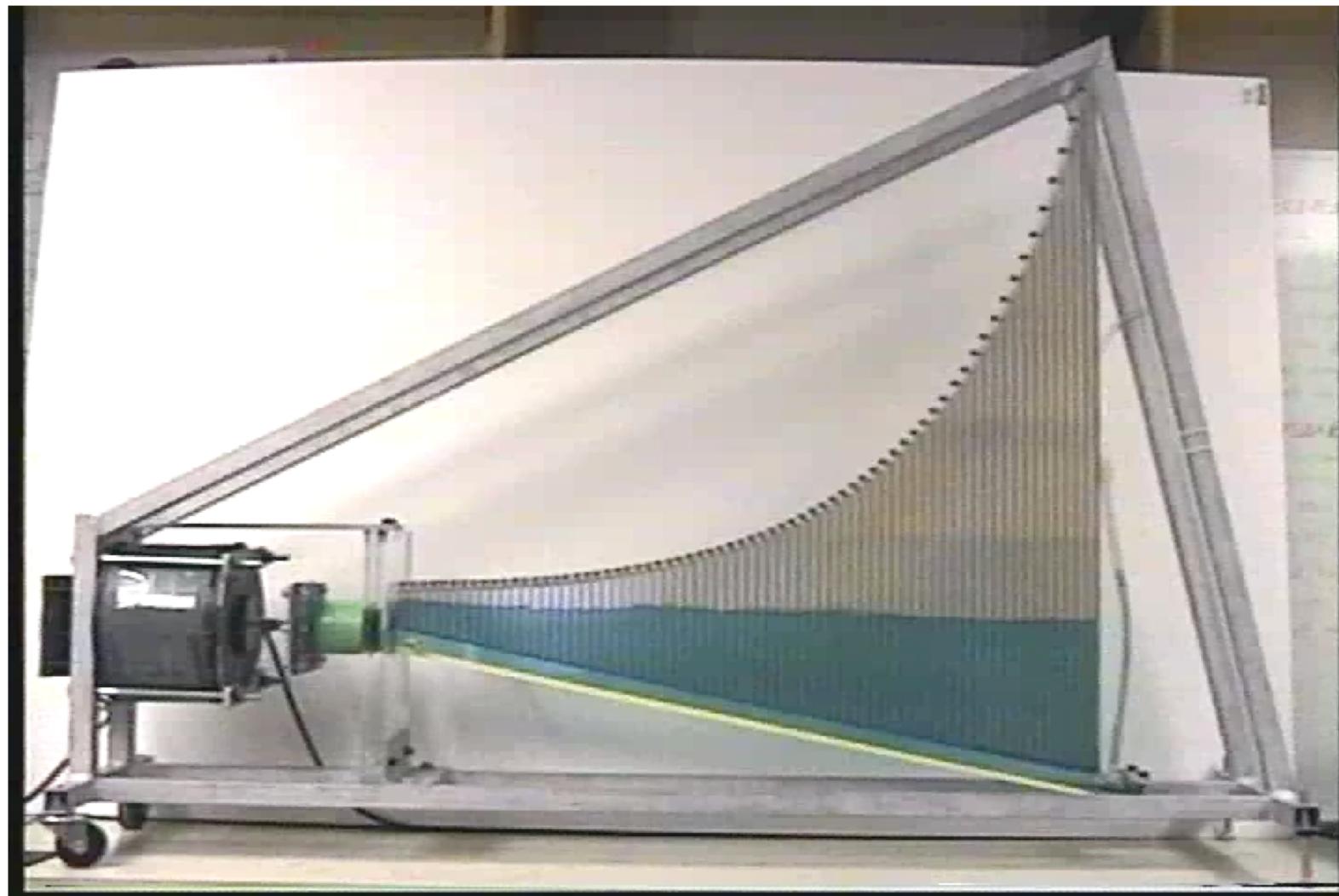
... being wider at the beginning of the first turn than at the end of the last, one can say that the wider parts can be caused to vibrate while the others do not...

and are capable of slower vibrations and consequently respond to deeper tones, whereas if the narrower parts are hit their vibrations are faster and consequently respond to sharper tones...

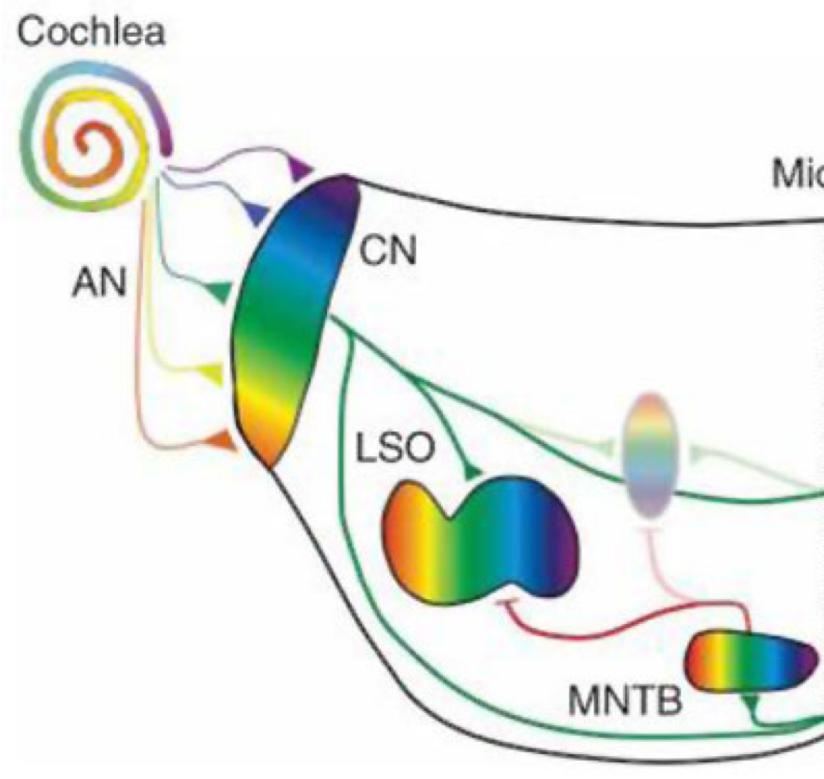
... such that according to the various motions of the **spiral lamina**, the spirits of the **nerve**... receive different impressions that represent in the **brain** the various aspects of tones.”

--> *concept of tonotopic projection*

tonotopy:



tonotopy:



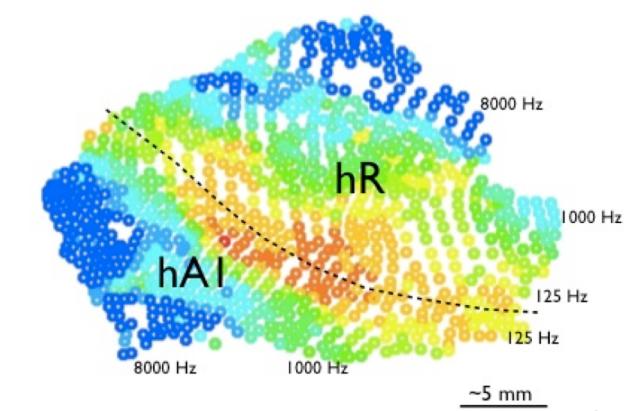
cochlea



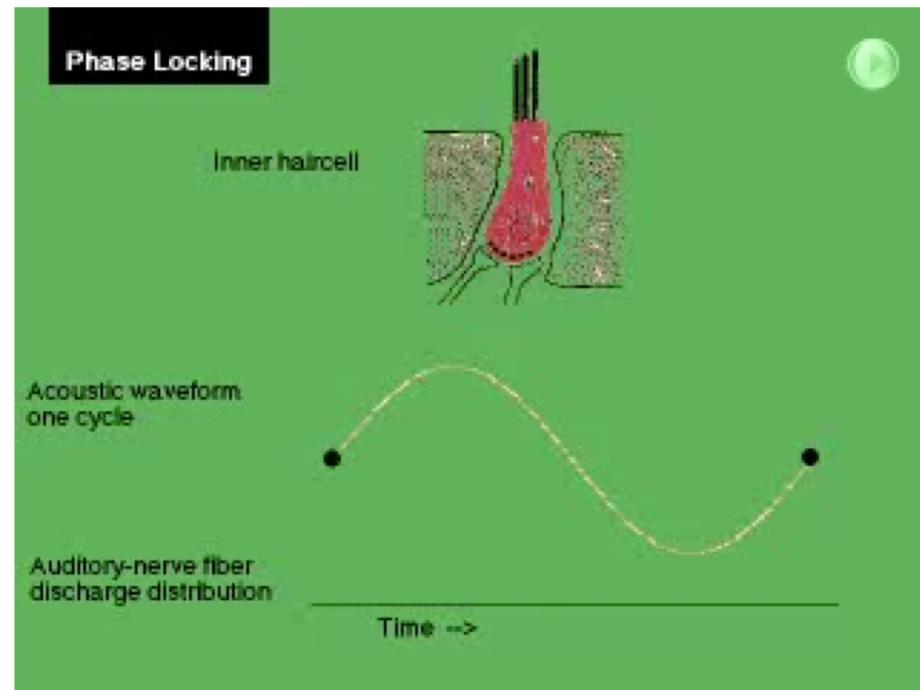
brainstem



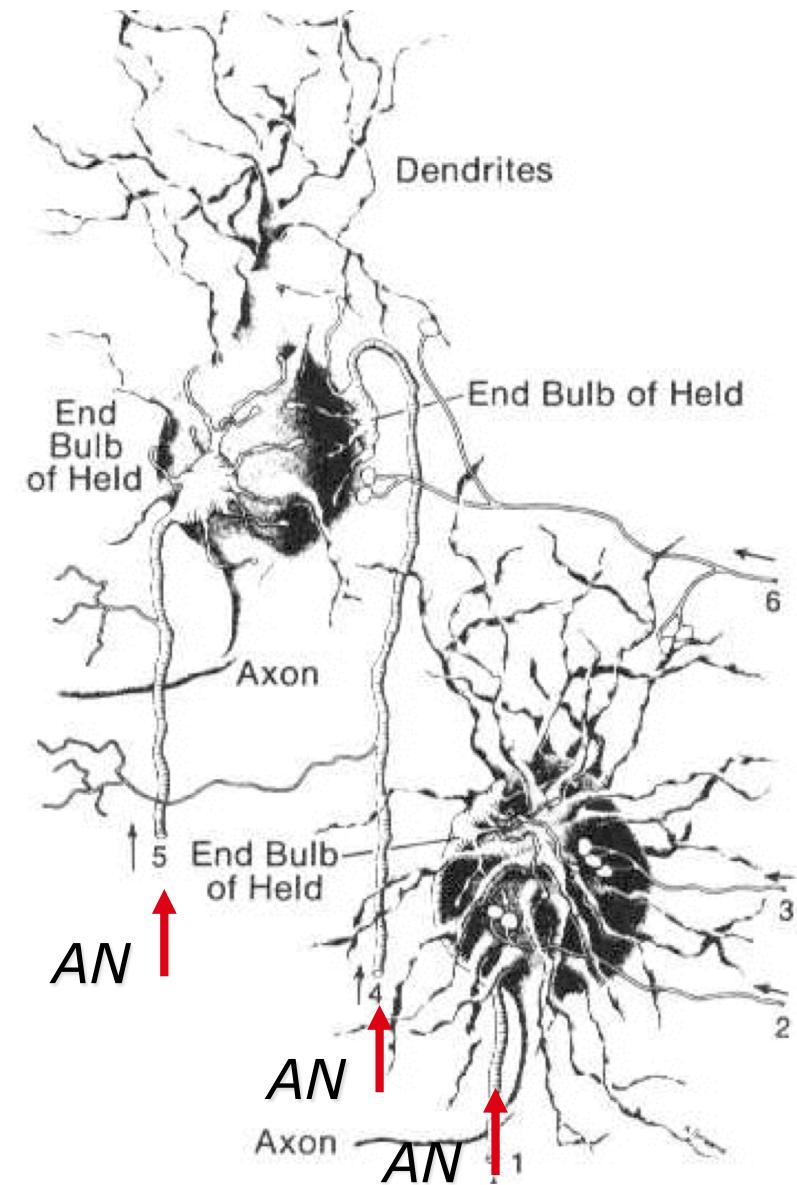
cortex



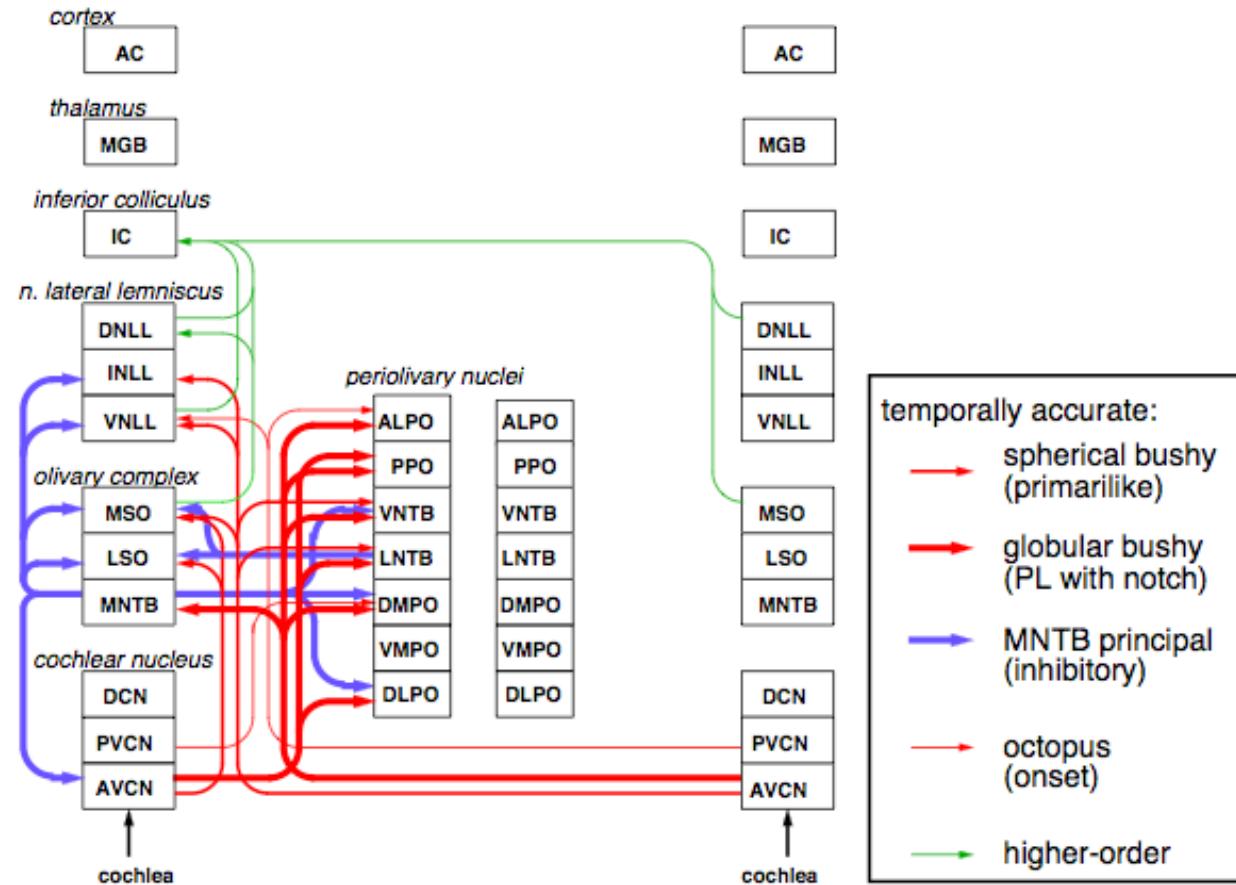
time:



"temporal fine structure"



temporal information:



... transmitted to many stages within the auditory system

How do we explain pitch?

How do we explain pitch?

4 main theories:

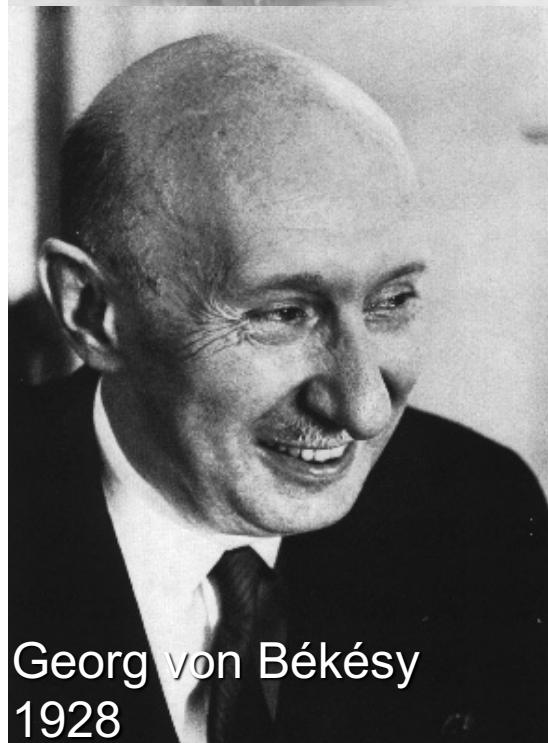
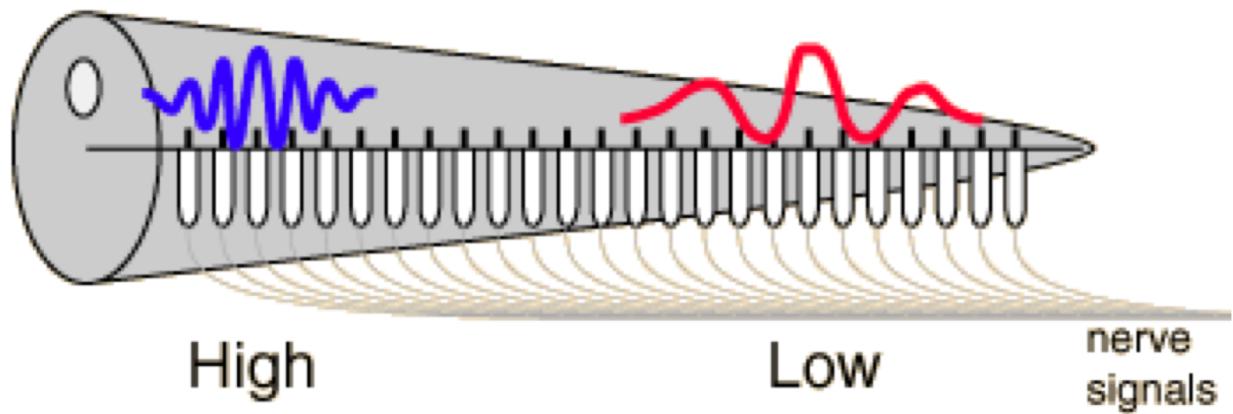
- place
- time
- pattern-matching
- autocorrelation

2 things to explain:

- exquisite sensitivity
- invariance



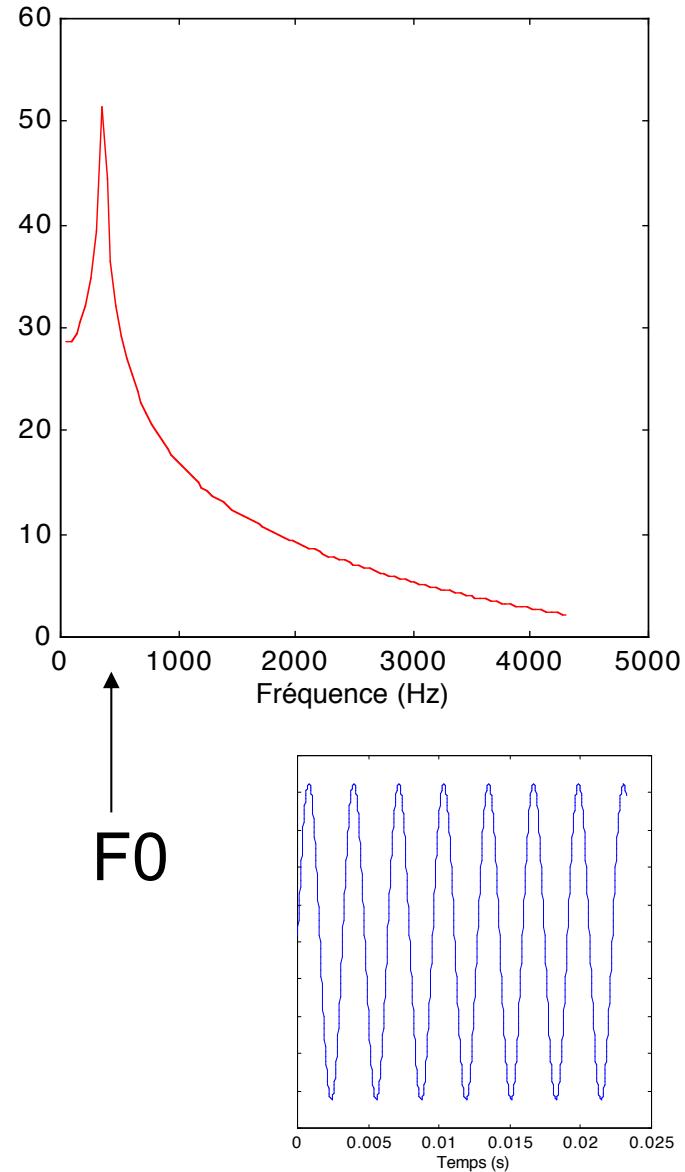
1. Place



position of maximum stimulation--> pitch

early roots: Duverney, etc.

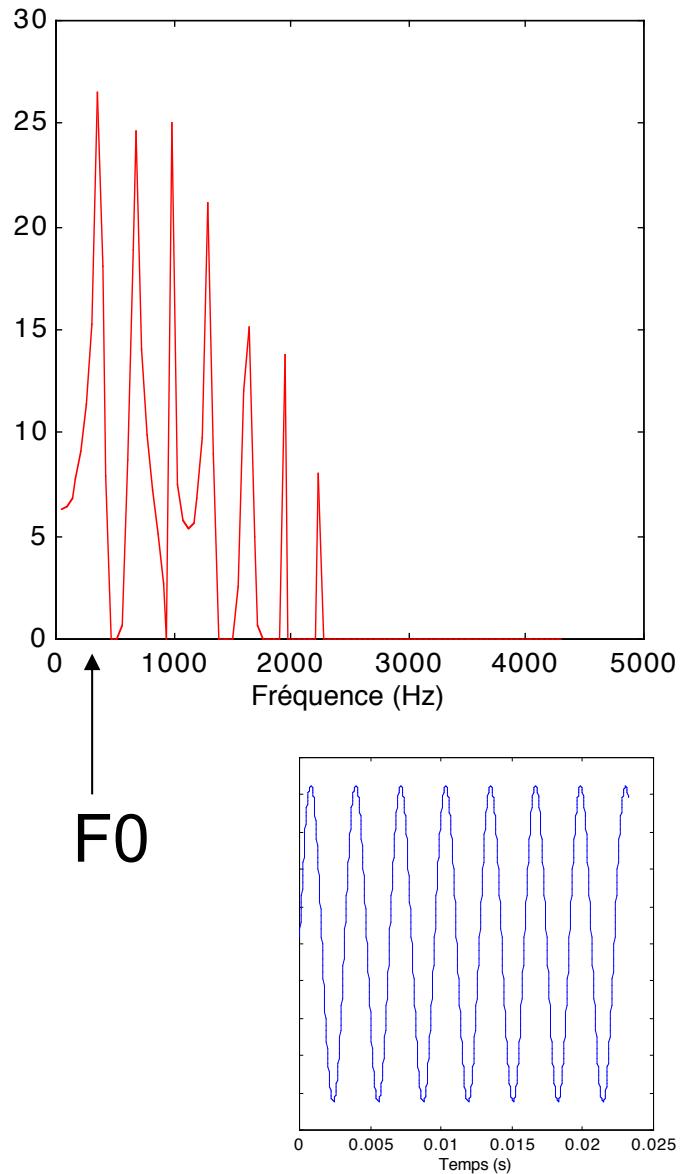
problems with place...



peak of spectrum?

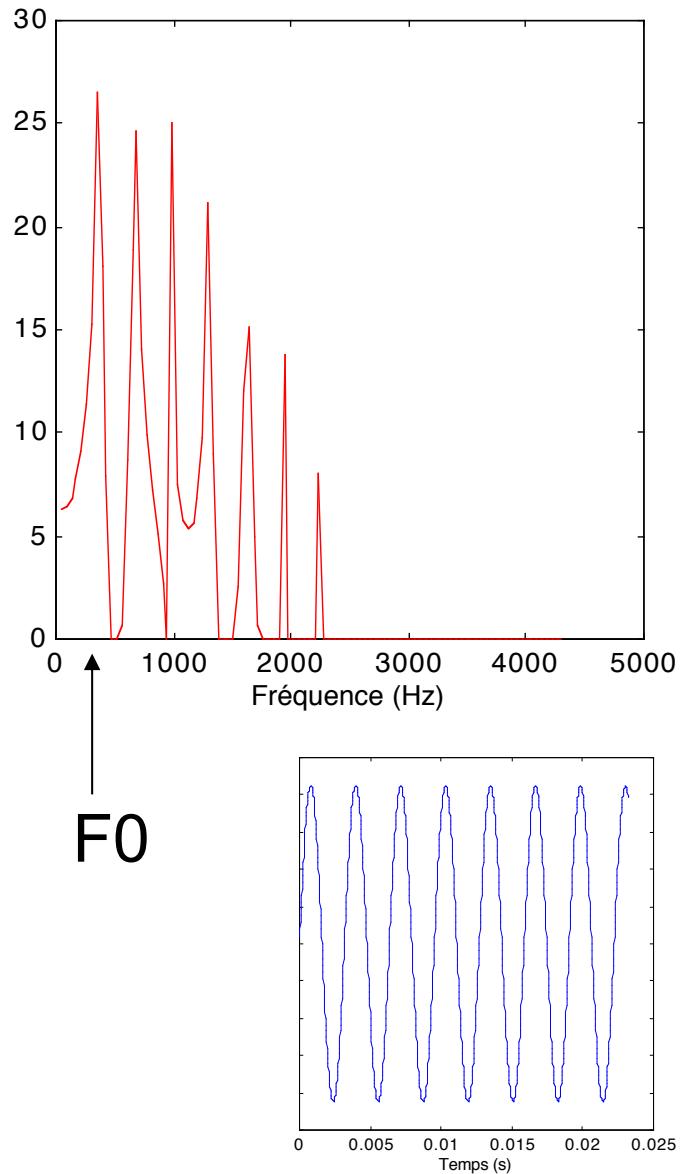
F0

problems with place...



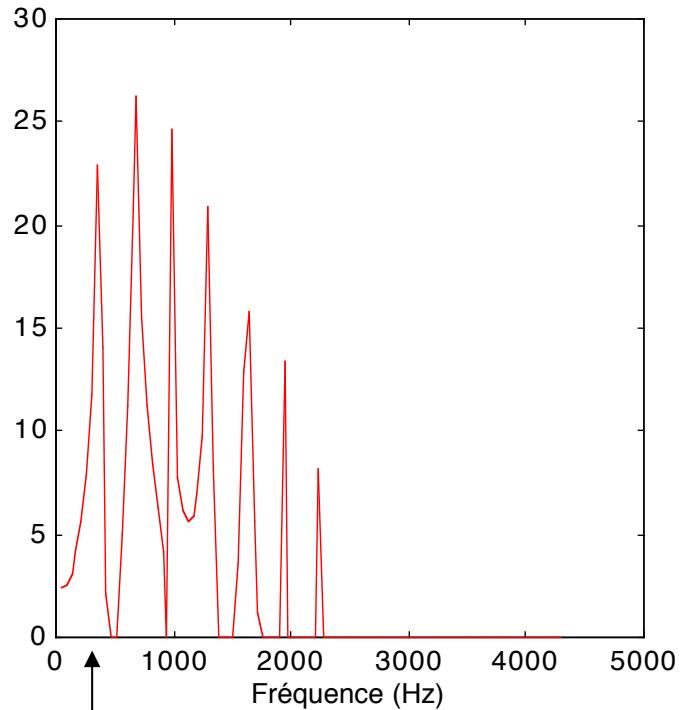
peak of spectrum?
→ fails if multiple peaks

problems with place...



peak of spectrum?
→ fails if multiple peaks
highest peak?

problems with place...



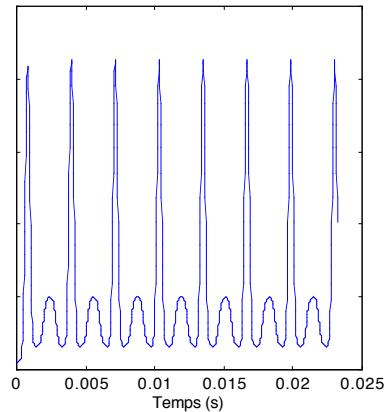
peak of spectrum?

→ fails if multiple peaks

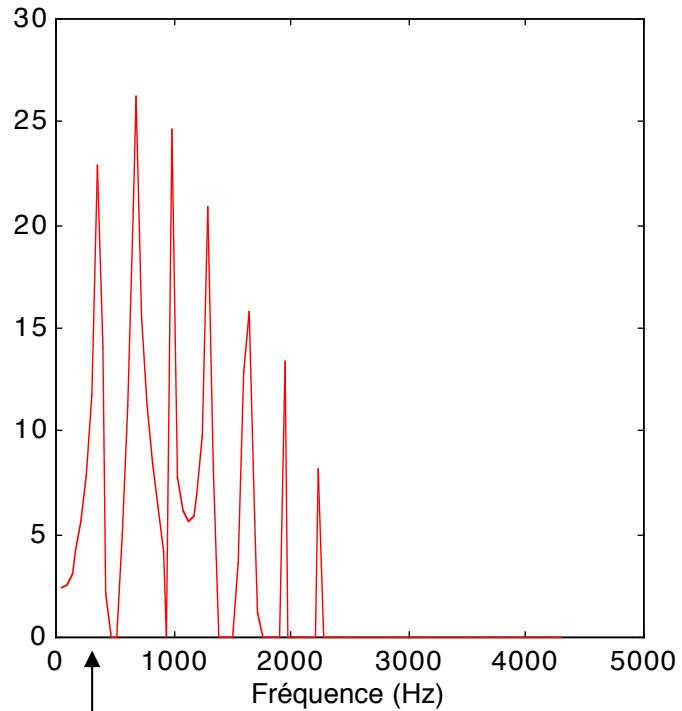
highest peak?

→ fails if harmonic stronger

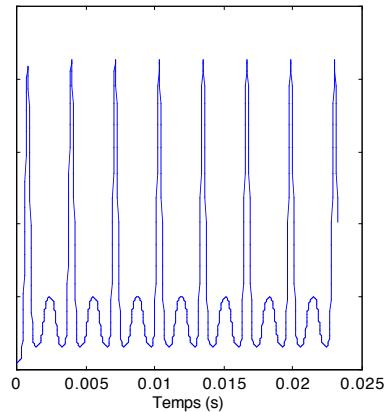
F0



problems with place...



F0



peak of spectrum?

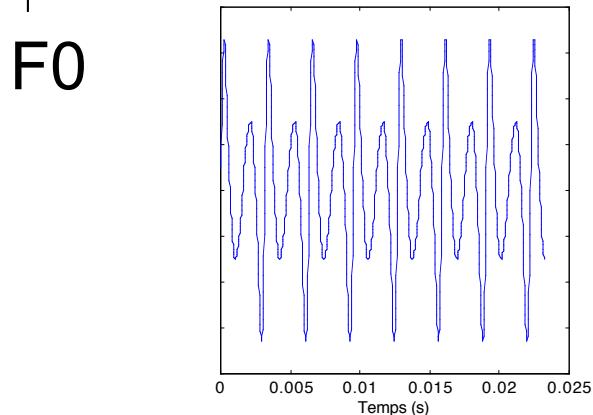
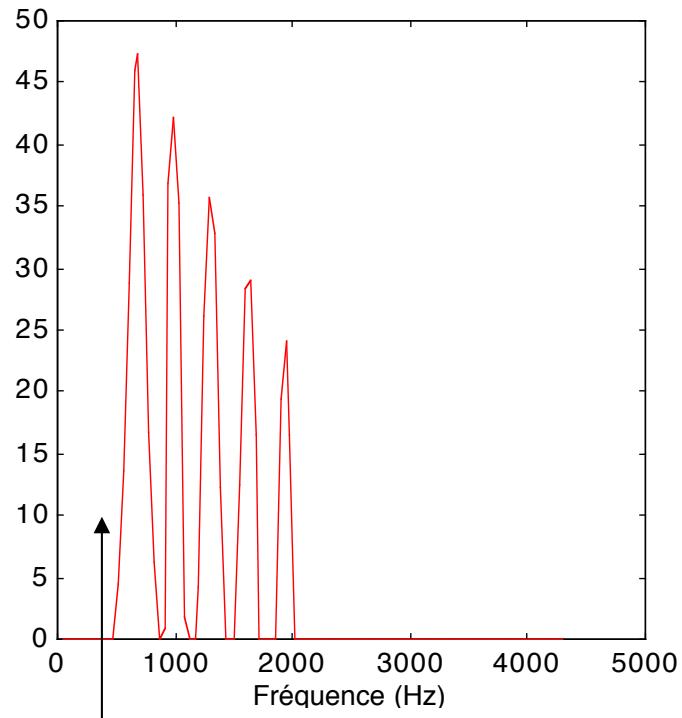
→ fails if multiple peaks

highest peak?

→ fails if harmonic stronger

first peak?

problems with place...



peak of spectrum?

→ fails if multiple peaks

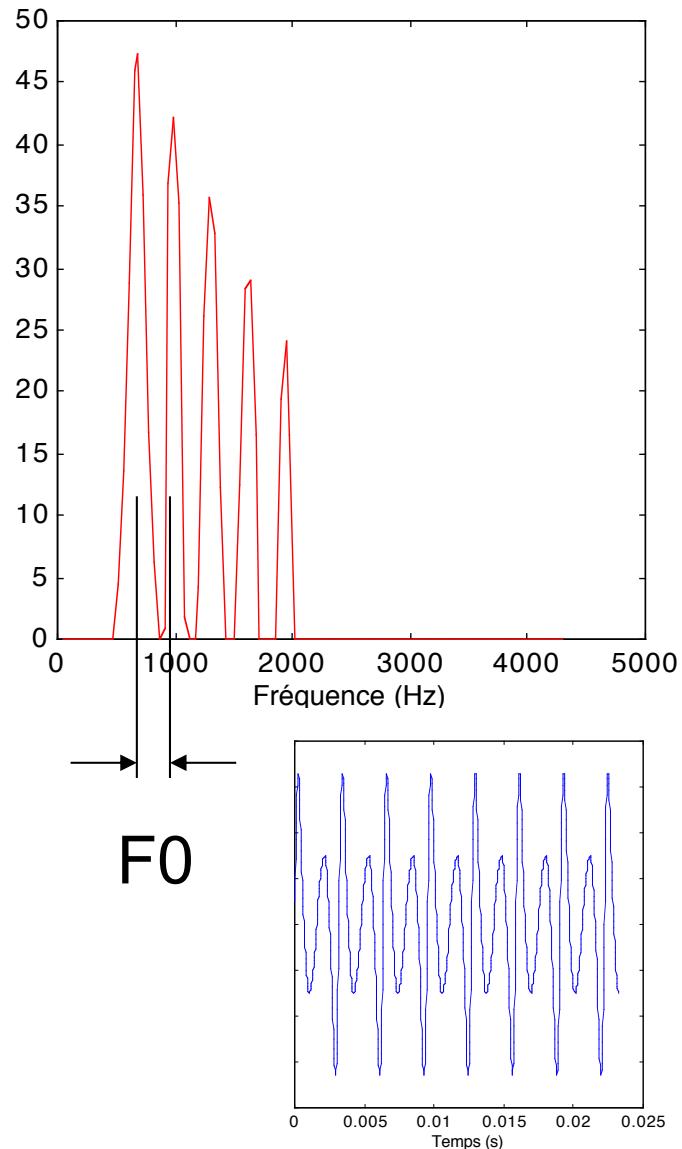
highest peak?

→ fails if harmonic stronger

first peak?

→ fails if missing fundamental

problems with place...



peak of spectrum?

→ fails if multiple peaks

highest peak?

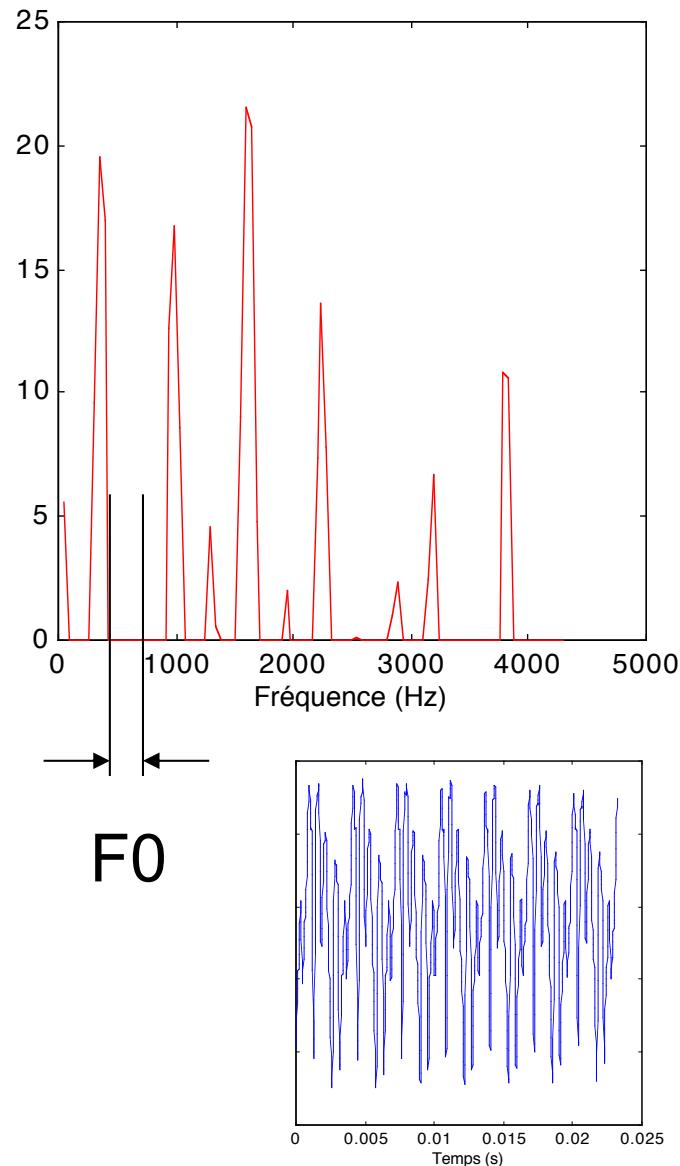
→ fails if harmonic stronger

first peak?

→ fails if missing fundamental

spacing between partials?

problems with place...



peak of spectrum?

→ fails if multiple peaks

highest peak?

→ fails if harmonic stronger

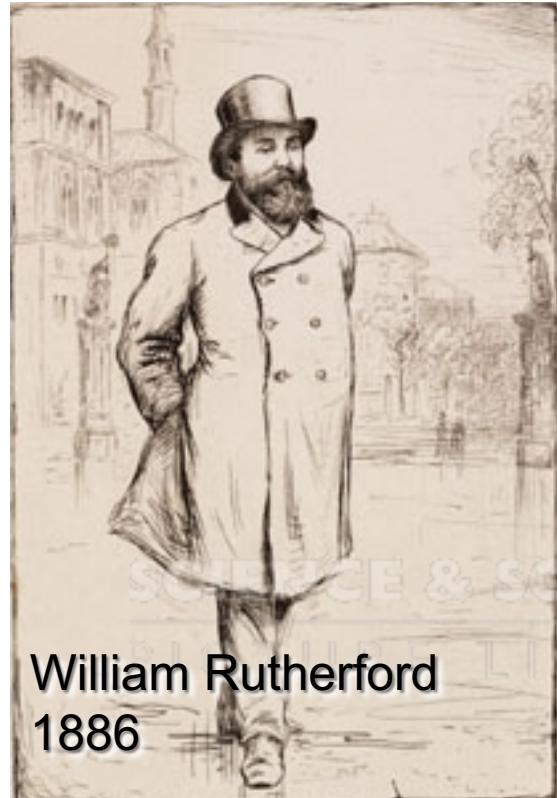
first peak?

→ fails if missing fundamental

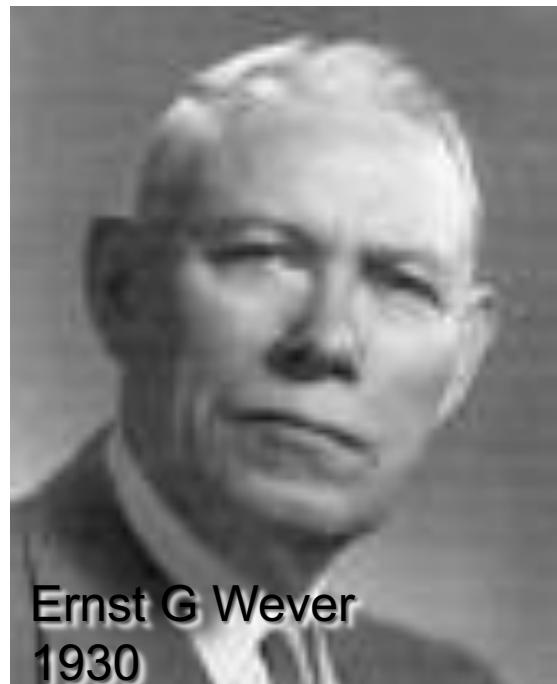
spacing between partials?

→ fails if irregular

... not easy to fix

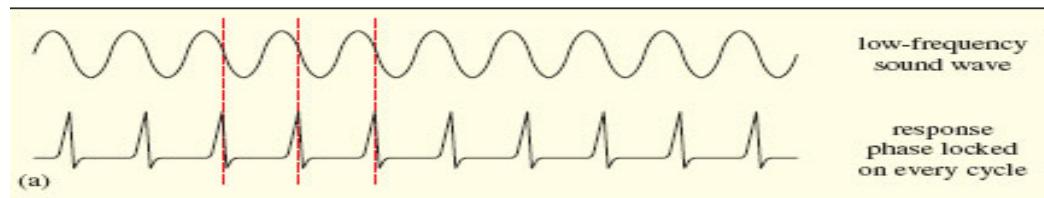


William Rutherford
1886



Ernst G Wever
1930

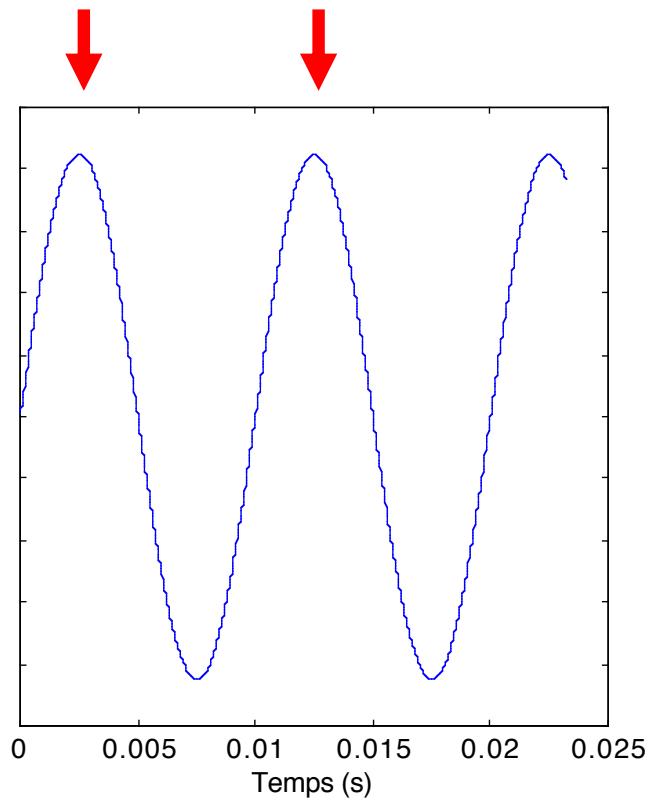
2. Time



interval between pulses → pitch

earlier roots: Nichomachus (2nd century AD)

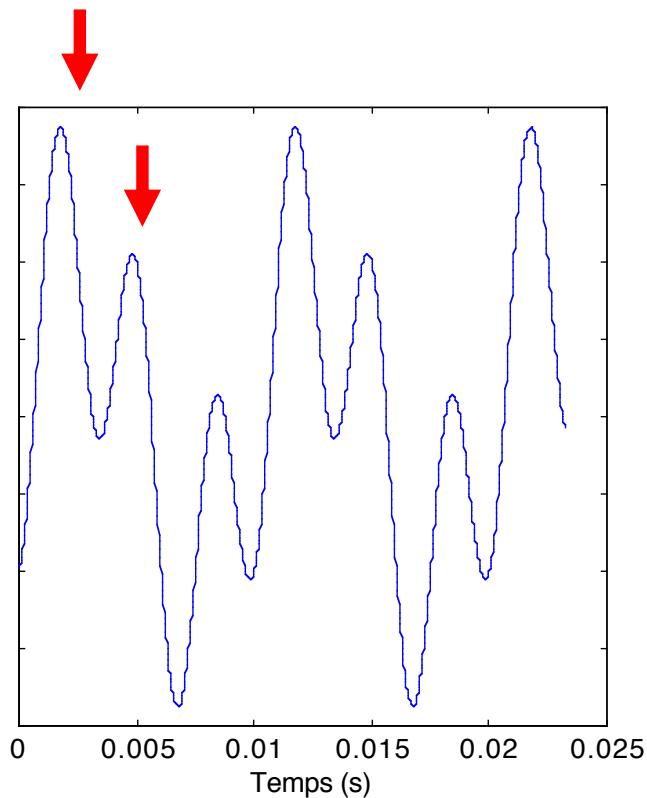
problems with time...



markers on peaks?

inter-marker interval
→ period → pitch

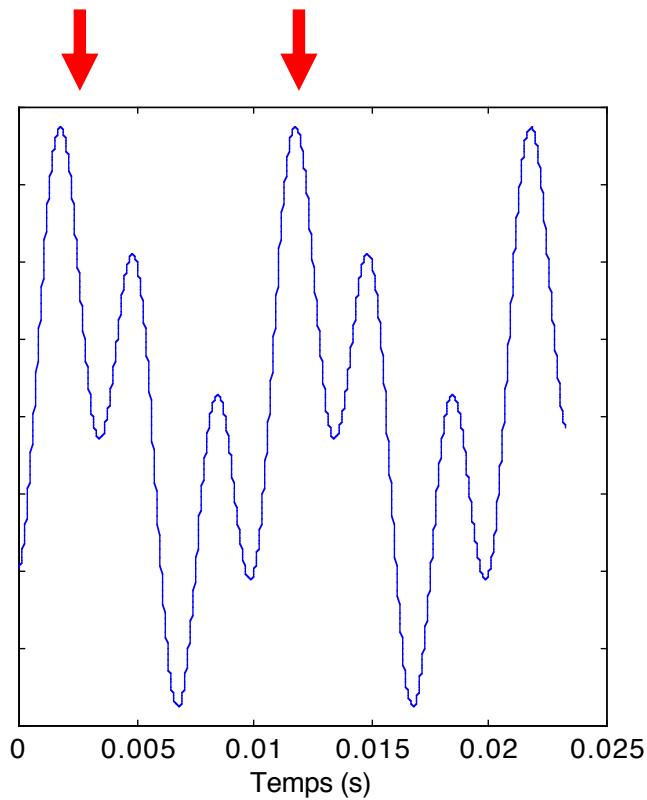
problems with time...



markers on peaks?
→ fails if multiple peaks

inter-marker interval
→ period → pitch

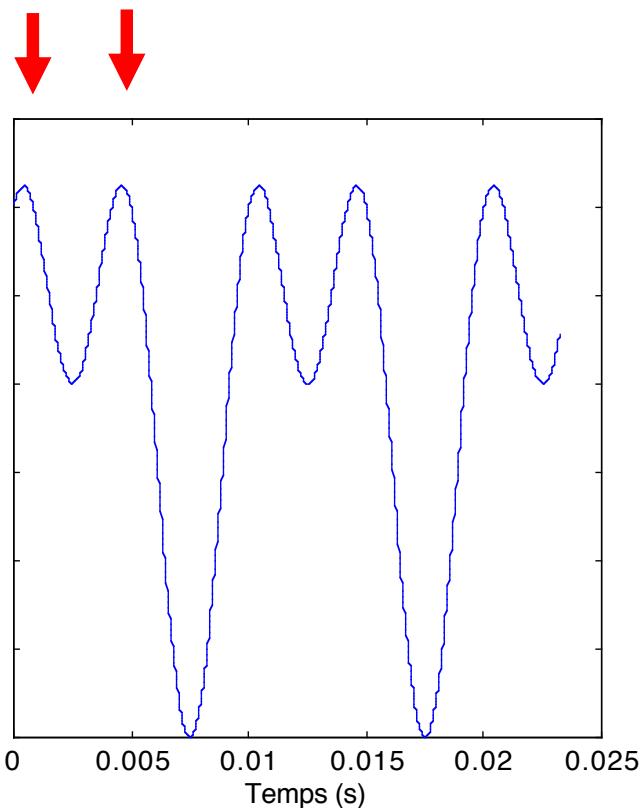
problems with time...



markers on peaks?
→ fails if multiple peaks
markers on biggest peaks?

inter-marker interval
→ period → pitch

problems with time...

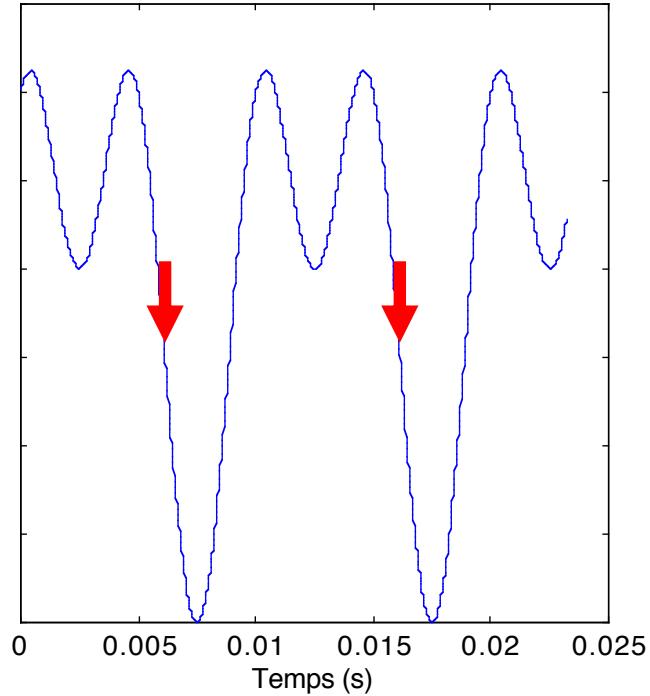


markers on peaks?
→ fails if multiple peaks

markers on biggest peaks?
→ fails if multiple "biggest"

inter-marker interval
→ period → pitch

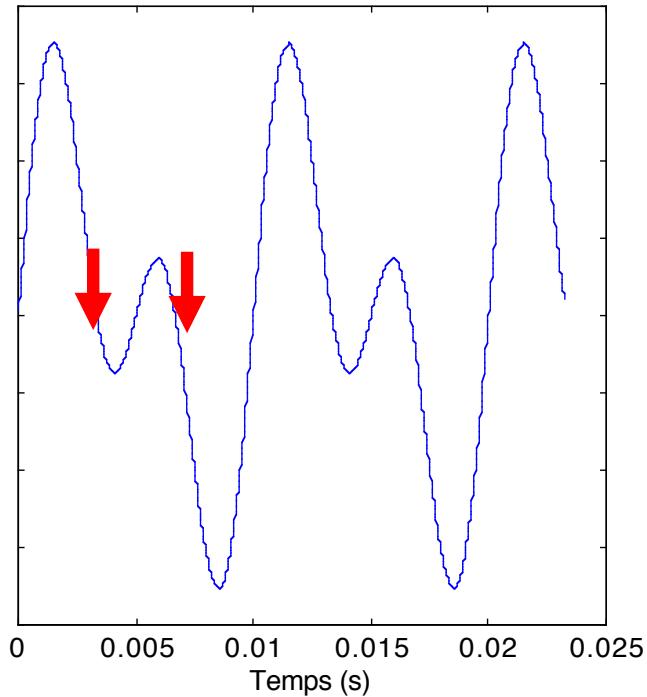
problems with time...



- markers on peaks?
 - fails if multiple peaks
- markers on biggest peaks?
 - fails if multiple "biggest"
- markers on zero crossings?

inter-marker interval
→ period → pitch

problems with time...



markers on peaks?

→ fails if multiple peaks

markers on biggest peaks?

→ fails if multiple "biggest"

markers on zero crossings?

→ fails if more than one

inter-marker interval

→ period → pitch

... not easy to fix



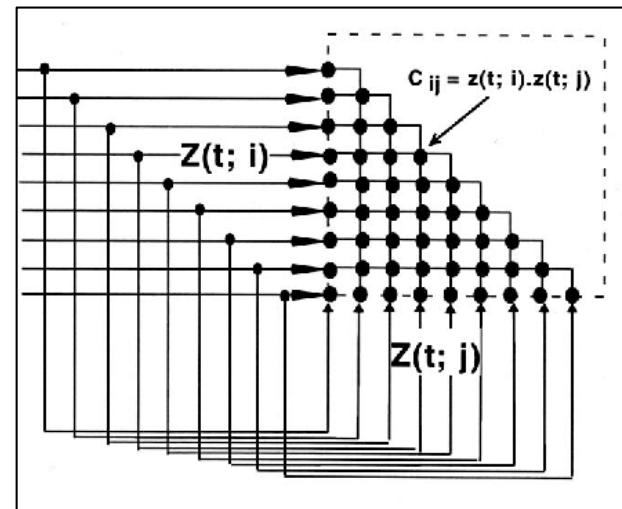
Egbert de Boer
1956



Ernst Terhardt
1974

3. Pattern matching

spectrum →

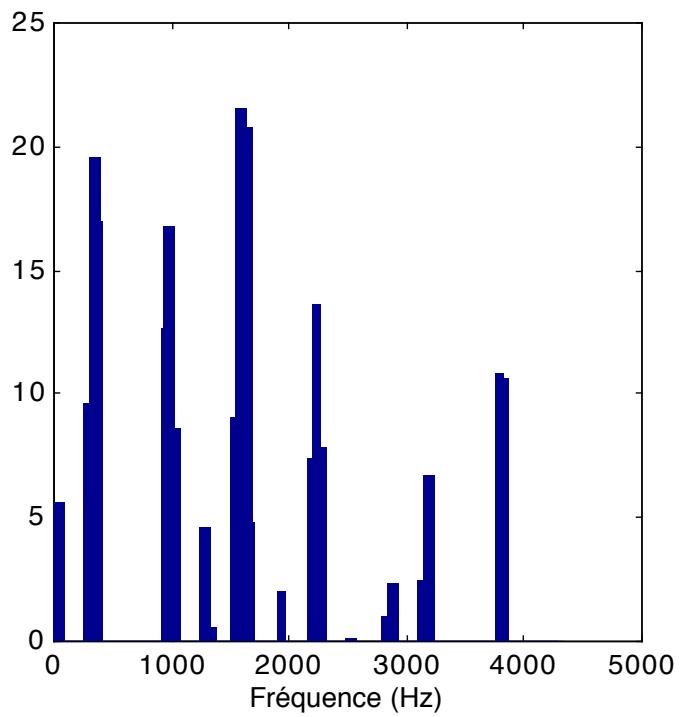


→ pitch

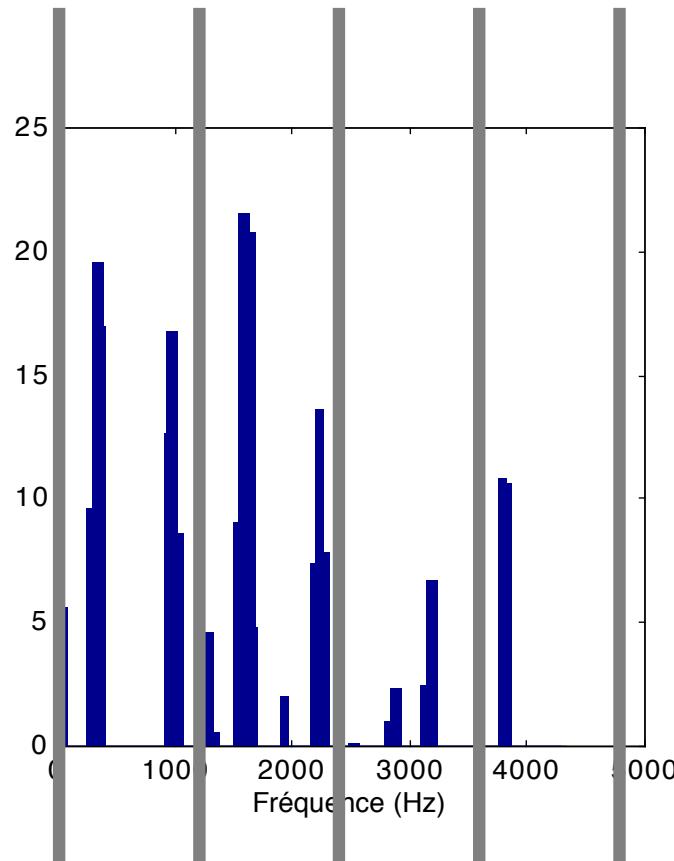
harmonic pattern matcher

earlier roots: Helmholtz, Alhazen (11 th century)

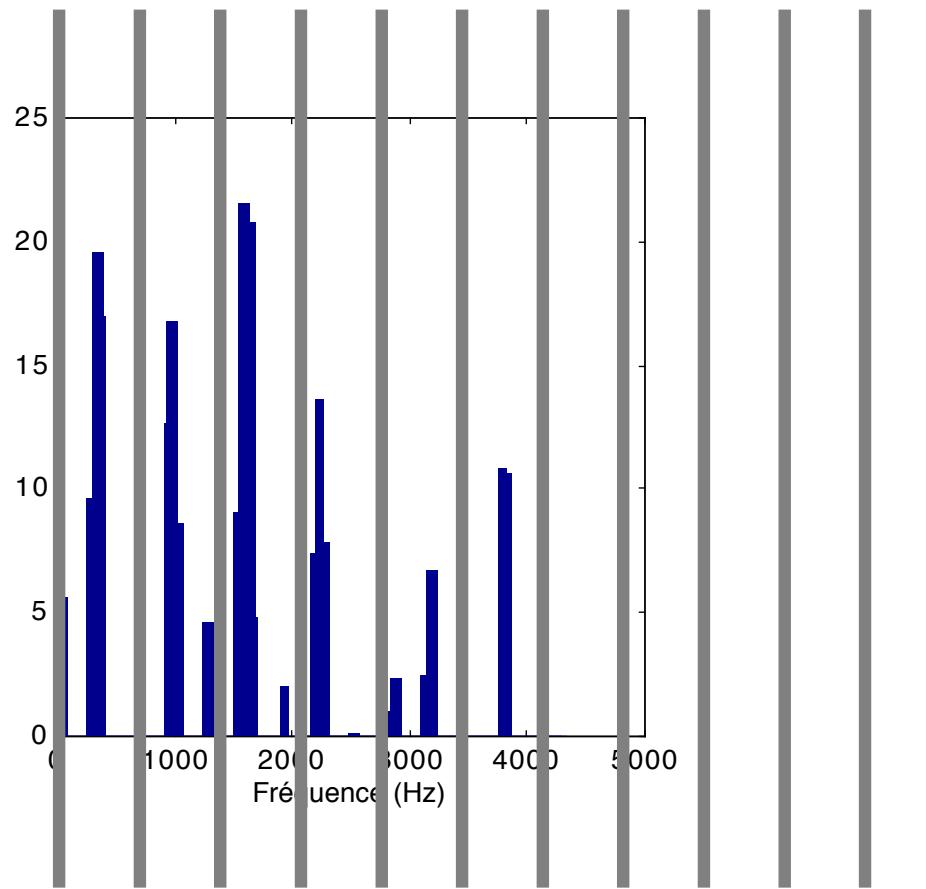
Pattern matching



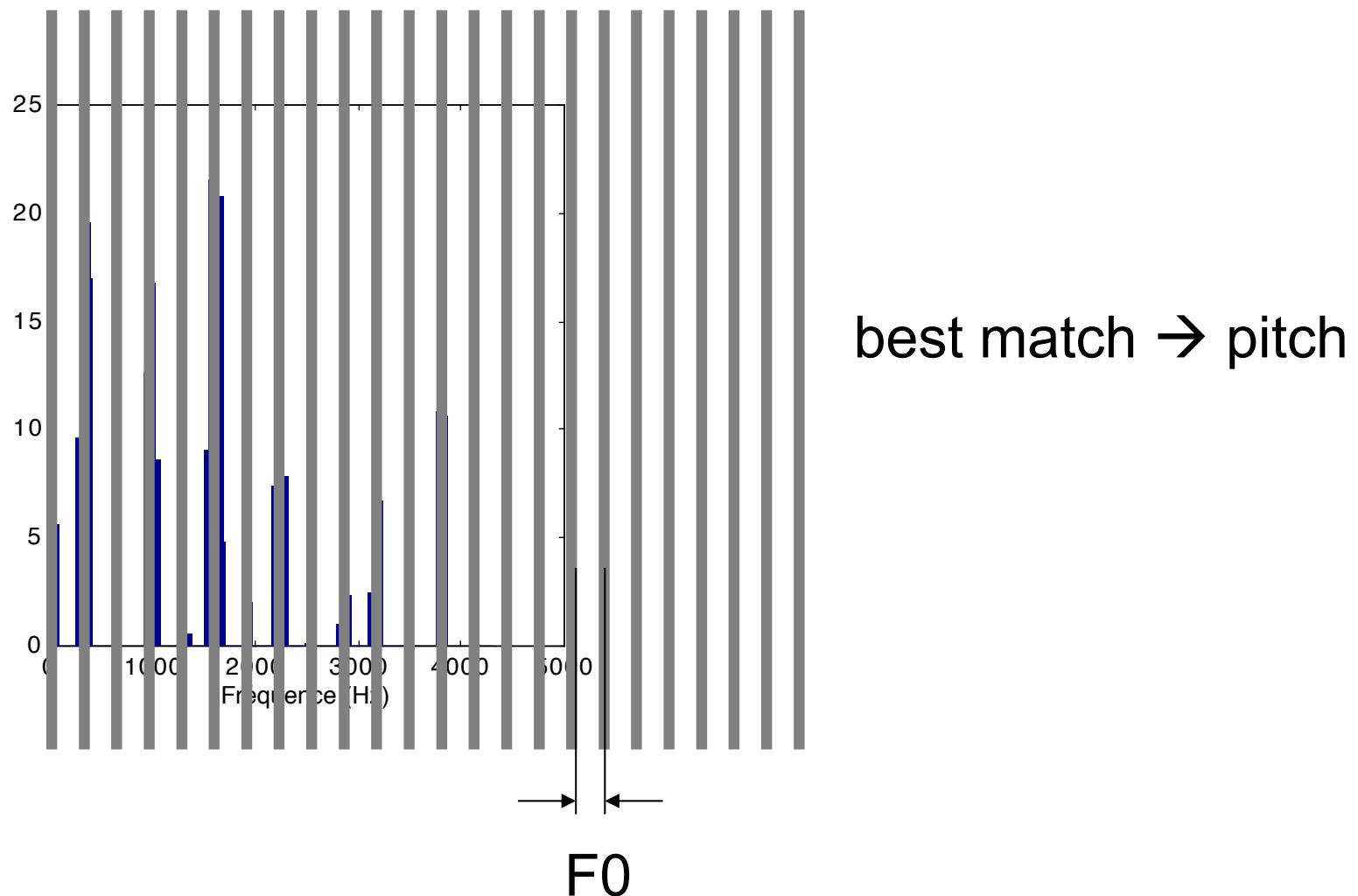
Pattern matching



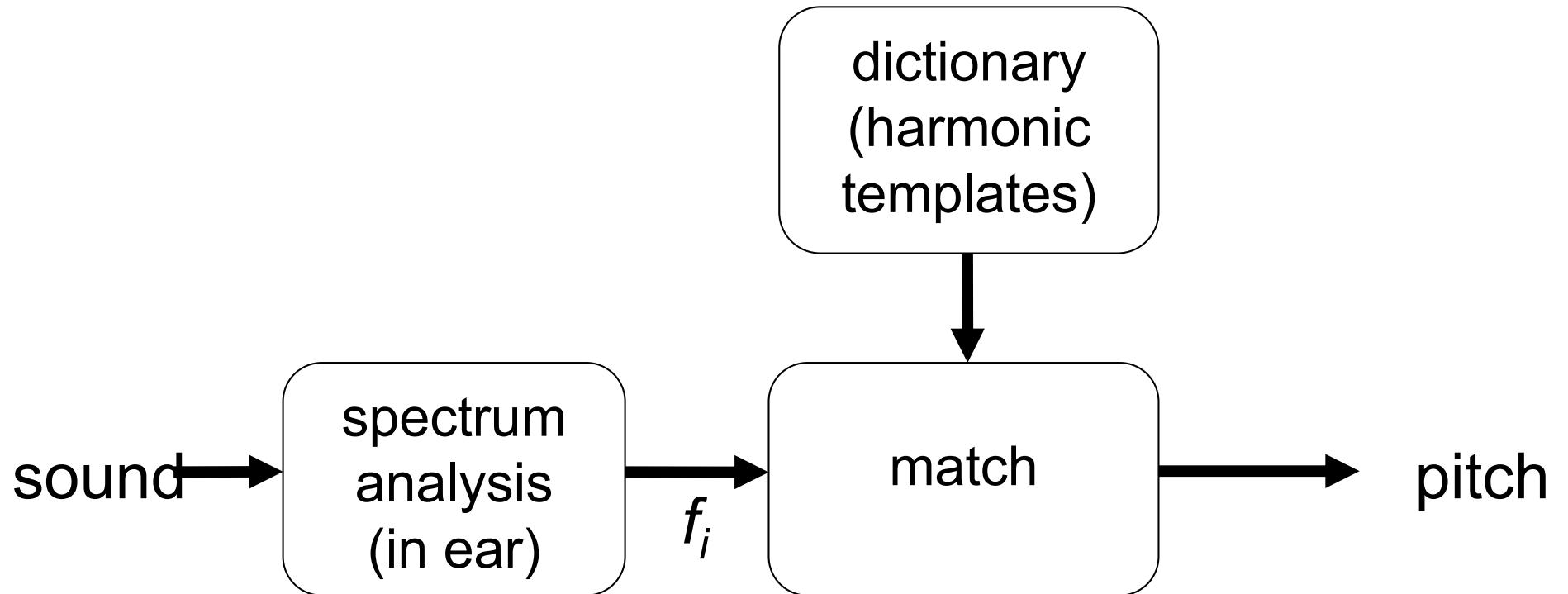
Pattern matching

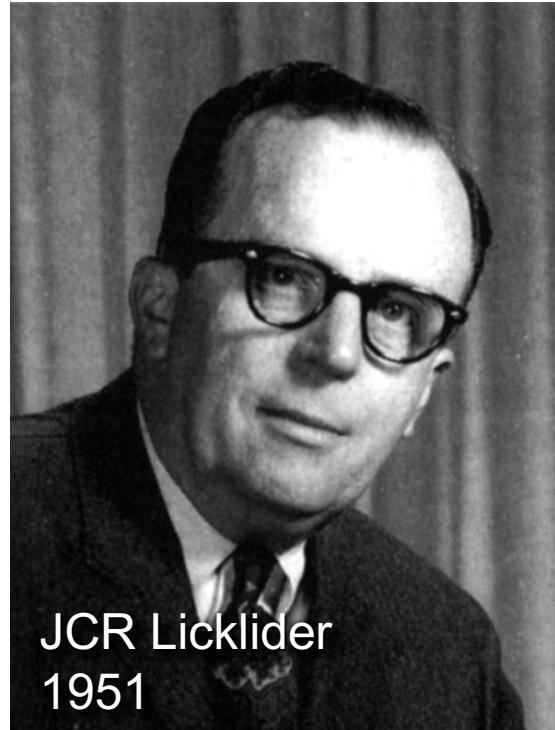


Pattern matching

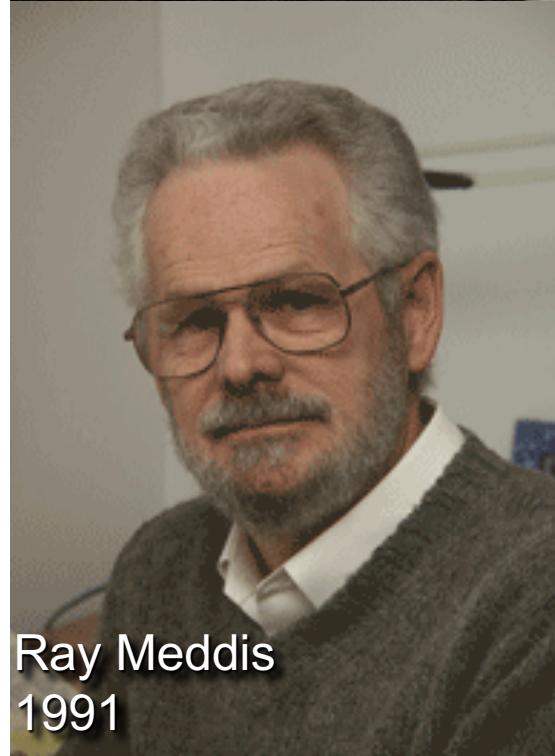


Pattern matching





JCR Licklider
1951



Ray Meddis
1991

4. Autocorrelation

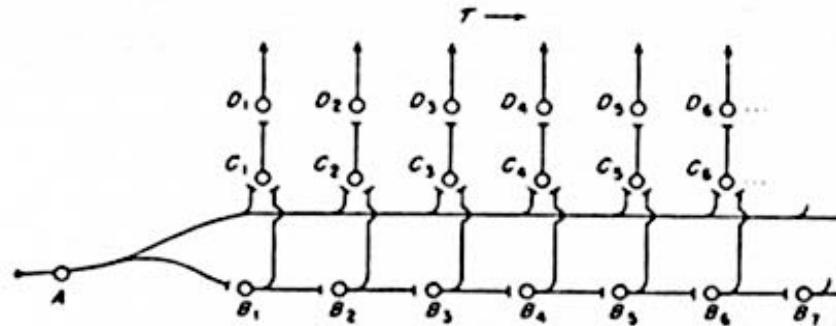
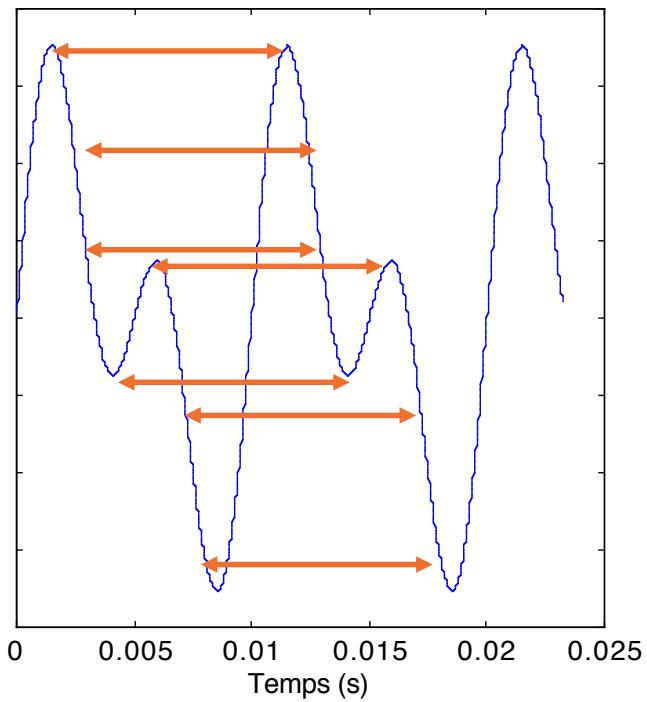


Fig. 1. – Basic schema of neuronal autocorrelator. A is the input

peak in autocorrelation function → pitch

Autocorrelation



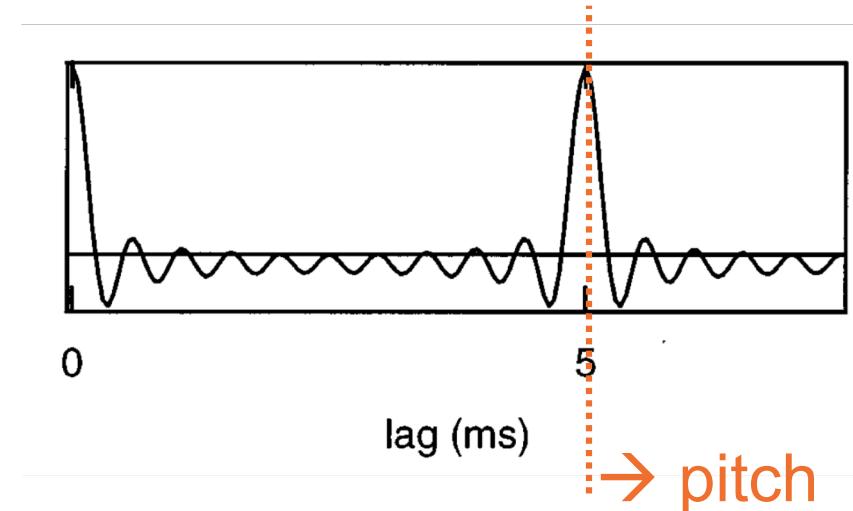
"shift and compare"

$$A_t(\tau) = \sum_W x(t)x(t - \tau)$$

product

time lag

integration window



Pros & cons:

place:

- time honored (Duverney, Helmholtz, von Békésy), gives role to cochlea
- only works for pure tones

time:

- time honored (Nicomachus, Rutherford, Wever)
- phase-sensitive, "brittle", no role for cochlea

pattern matching:

- works for all stimuli (if partials can be resolved)
- requires extra "pattern matching" stage, fails if partials can't be resolved

autocorrelation:

- works for all stimuli
- no neural correlate found, can't explain all aspects of psychophysics

Summary:

- pitch \leftrightarrow F0 (mainly)
- exquisite discrimination (~0.2 %)
- invariant to other dimensions
- still not sure how it is perceived

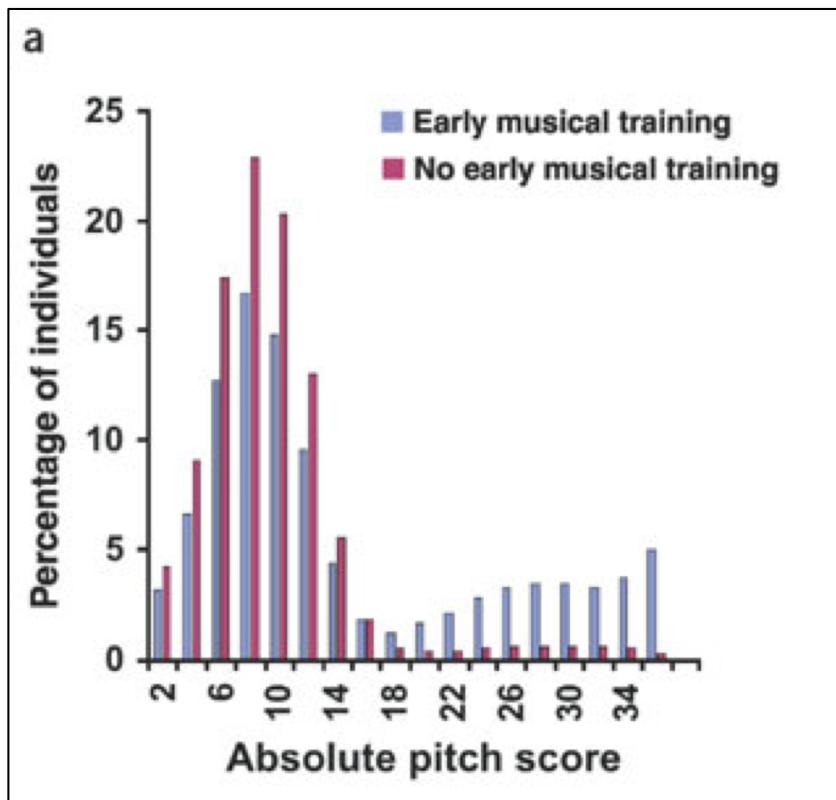
Some mysteries

relative vs absolute pitch

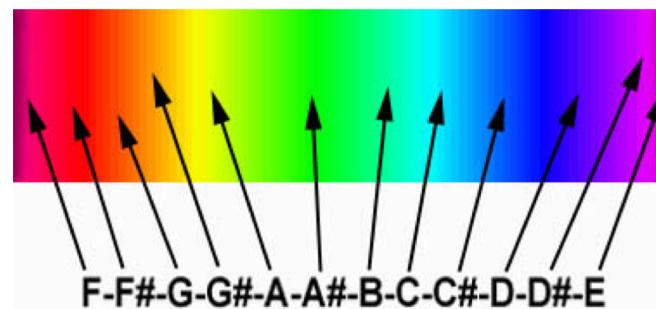
- most are sensitive to pitch *interval* (*F0 ratio*)
- few are sensitive to *absolute* pitch (*F0*)

all of our models predict absolute pitch...

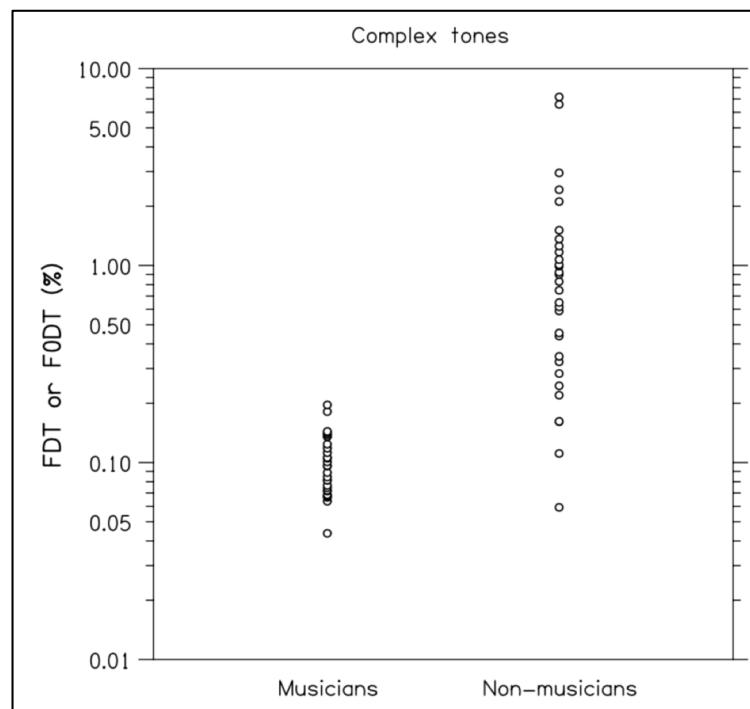
absolute pitch:



often associated with early music training



sometime associated with synesthesia



one aspect of inter-individual differences

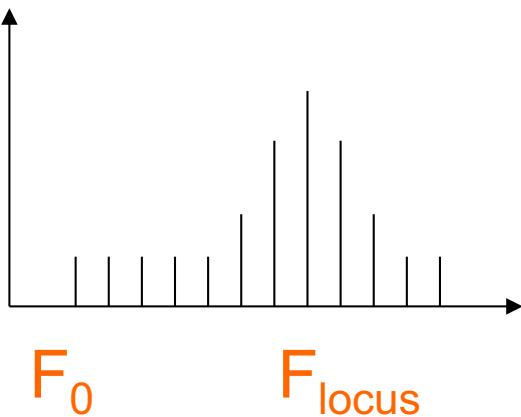
multiple pitches

- a stimulus sometimes evokes *multiple* pitches:
 - *ambiguous* pitch
 - *concurrent* pitches

our models assume a single pitch...

multiple pitches

throat singing:



courtesy Tran Van Quai

polyphony:



Bach's musical offering
(orchestrated by Webern)

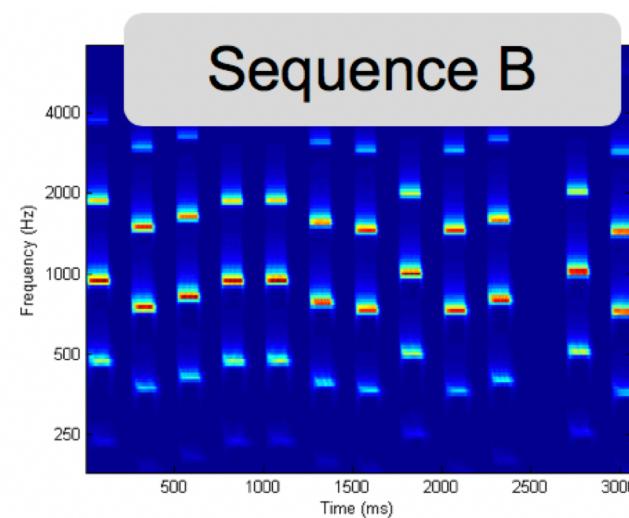
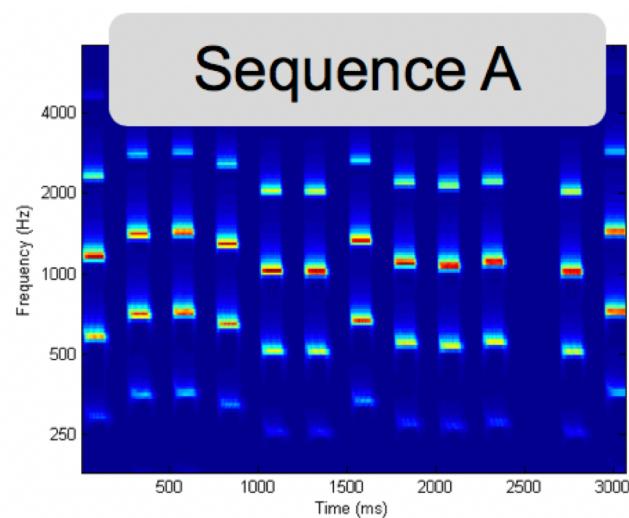
effect of context

- pitch, interval depend on context
 - *harmonic*
 - *melodic*
- tonality (related to pitch?)

our models are mostly context-blind...

effect of context

You will hear a sequence of tones, a short pause, and then two final tones.
Does the pitch go up or down between the two final tones?



courtesy
Daniel Pressnitzer
(Chambers et al. 2014)



last two tones are physically identical!

Pitch estimation, YIN

~~Pitch estimation, YIN~~
~~F0~~

YIN

YIN, a fundamental frequency estimator for speech and music^{a)}

Alain de Cheveigné^{b)}

Ircam-CNRS, 1 place Igor Stravinsky, 75004 Paris, France

Hideki Kawahara

Wakayama University

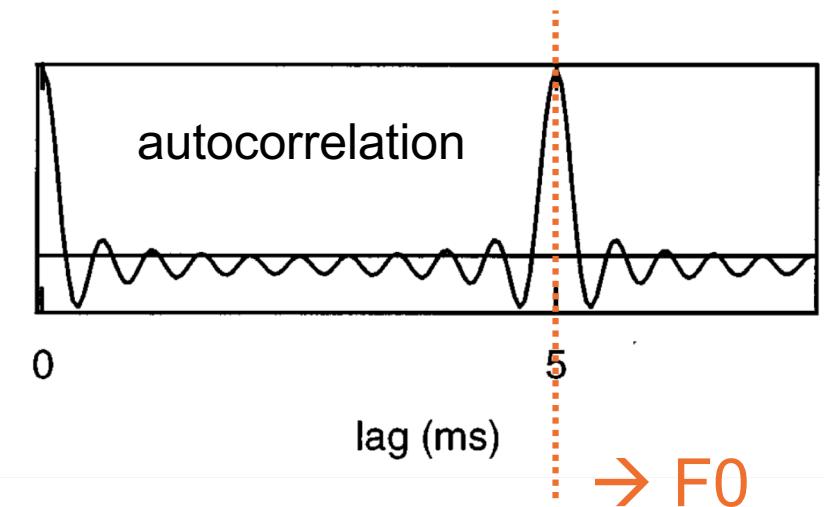
J. Acoust. Soc. Am. 2002

~1900 citations, Google scholar

YIN

$$A_t(\tau) = \sum_W x(t)x(t - \tau)$$

"shift and compare"



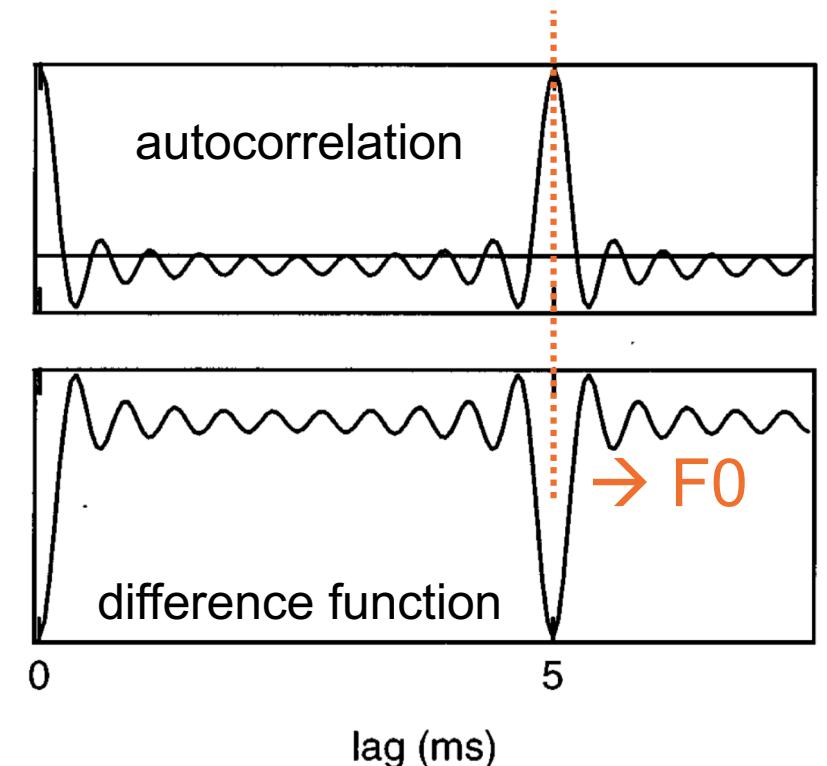
YIN

$$A_t(\tau) = \sum_W x(t)x(t - \tau)$$

"shift and compare"

$$D_t(\tau) = \sum_W [x(t) - x(t - \tau)]^2$$

"shift and compare"

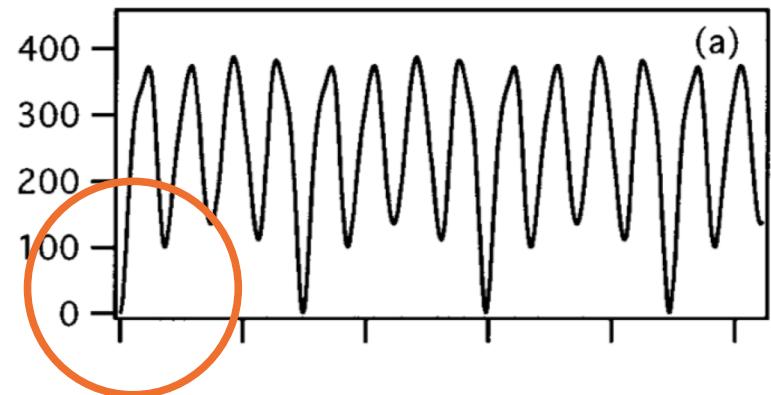


$$D_t(\tau) = A_t(0) + A_{t-\tau}(0) - A_t(\tau)$$

equivalent to autocorrelation, should be no better...

YIN

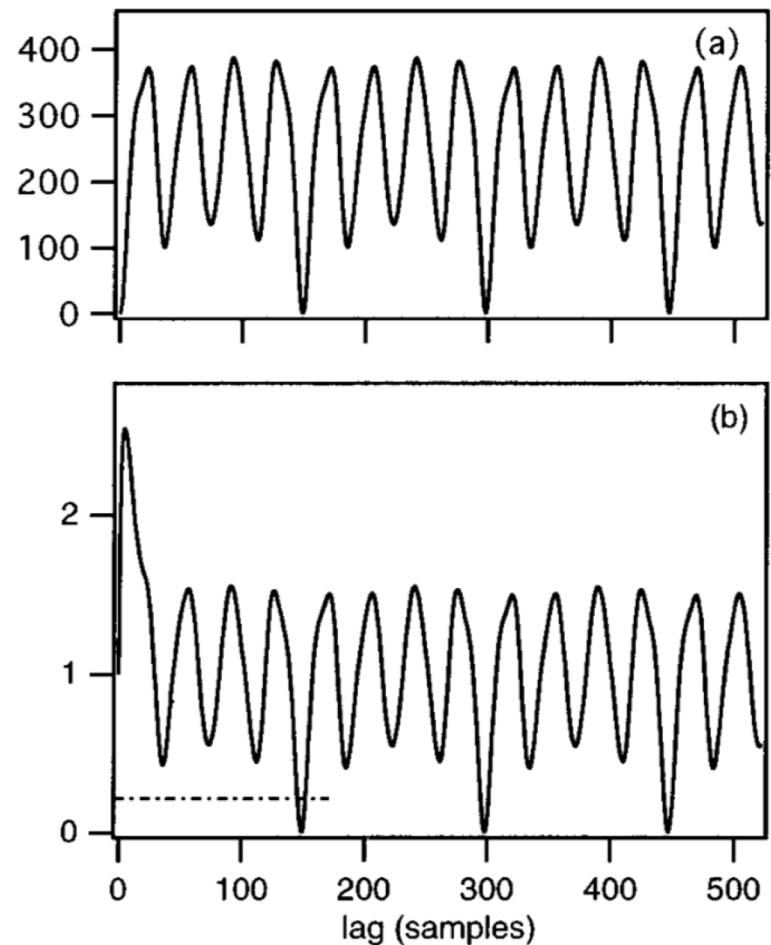
the YIN trick:



how to avoid locking on to these ? (similar problem for ACF)

YIN

the YIN trick:



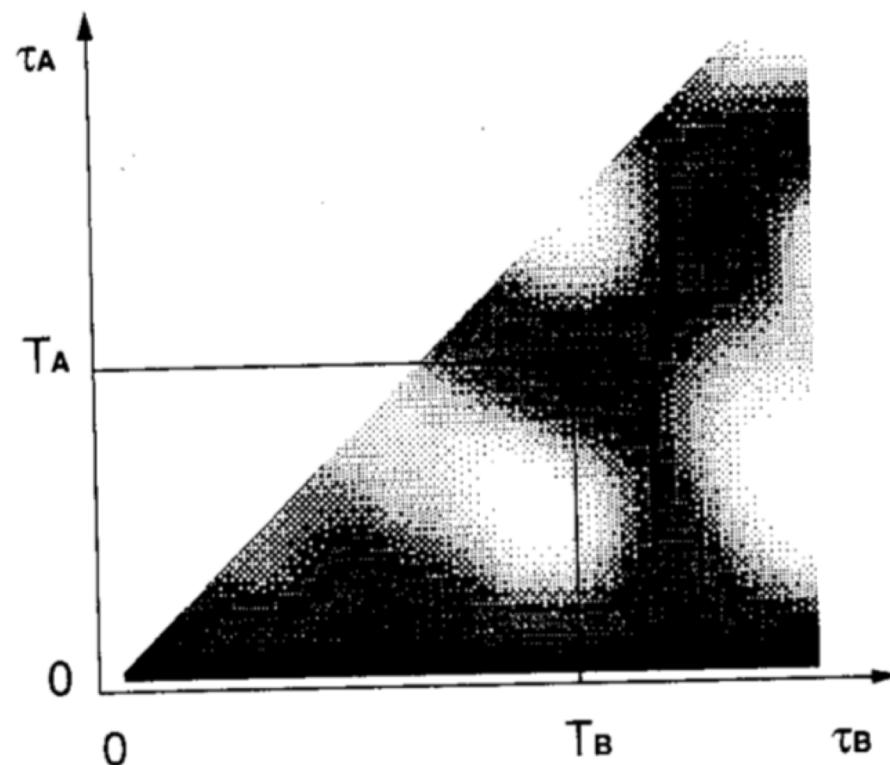
"mean-normalized" DF:

$$D'_t(\tau) = \frac{D'_t(\tau)}{1/\nu \sum_{\nu} D'_t(\nu)}$$

YIN

can be extended to multiple F0s:

$$D_t(\tau_A, \tau_B) = \sum_W [x(t) - x(t - \tau_A) - x(t - \tau_B) + x(t - \tau_A - \tau_B)]^2$$



F0 estimation

More in Rachel and Johanna's talks !

Summary

- pitch is a crucial perceptual quantity, complex & not fully understood
- the main signal correlate is F0
- pitch \neq F0
- tons of methods to estimate F0, tons of applications!