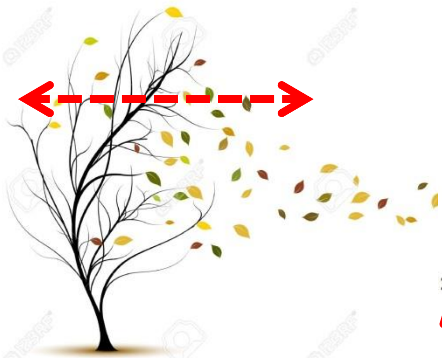
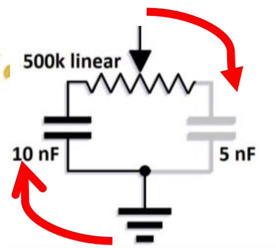
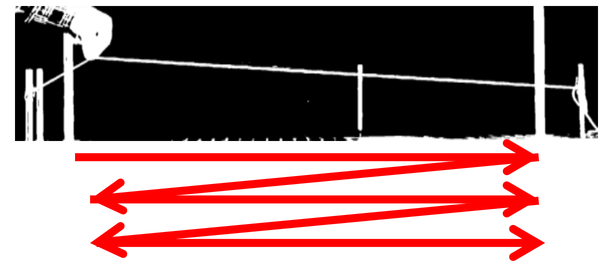
CEPSTRUM
also for
RHYTHM-
detection

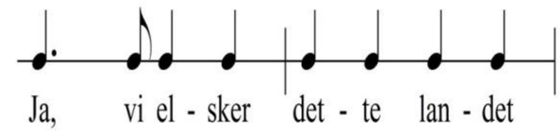
Tor Halmrast
Composer
Acoustician
Head of Acoustics, Statsbygg
Assoc. Prof., Univ. Oslo/Musicology
Norwegian Ac. of Music

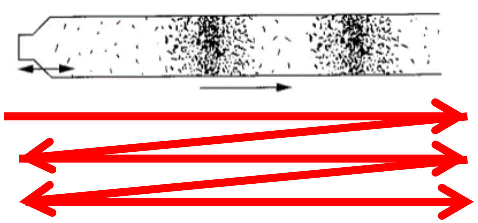
TIMBRE, RYTHM, ECHO = REFLECTIONS (delays)

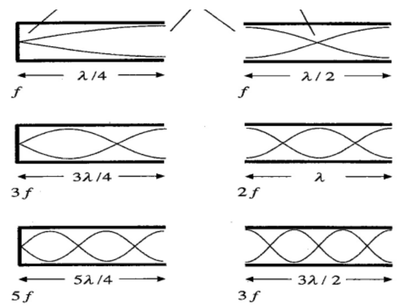


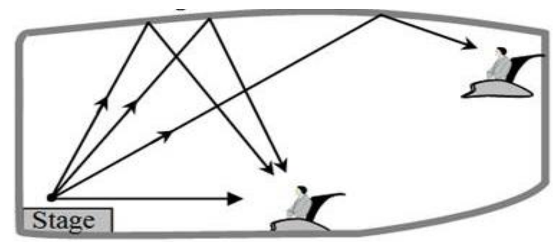


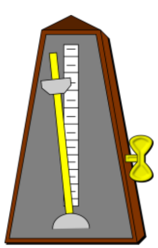












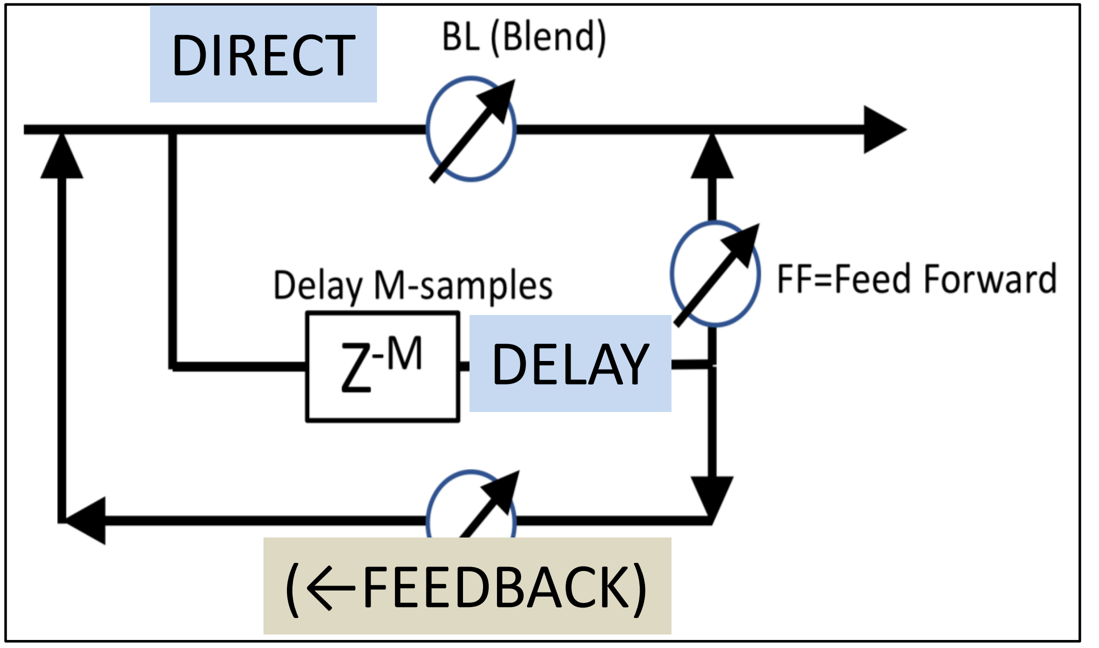
EVERYTHING IS THE SAME!!!
IMPULSE + REFLECTIONS

- Echoes
- Musical instruments
  (flute = reflection+feed back)
- Timbre
  (changes when adding a (short) reflection)
- Flutter Echoes
  (PS! A bit more complicated)
- Rhythm

«EVERYTHING»
YOU CAN SEND A SIGNAL THROUGH
IS A FILTER

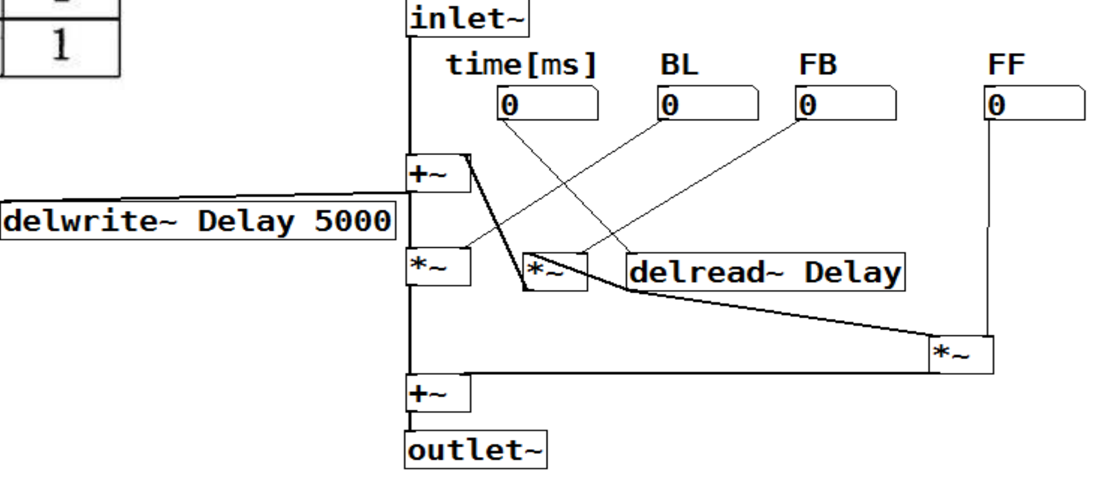
The FILTER that describes it all !!

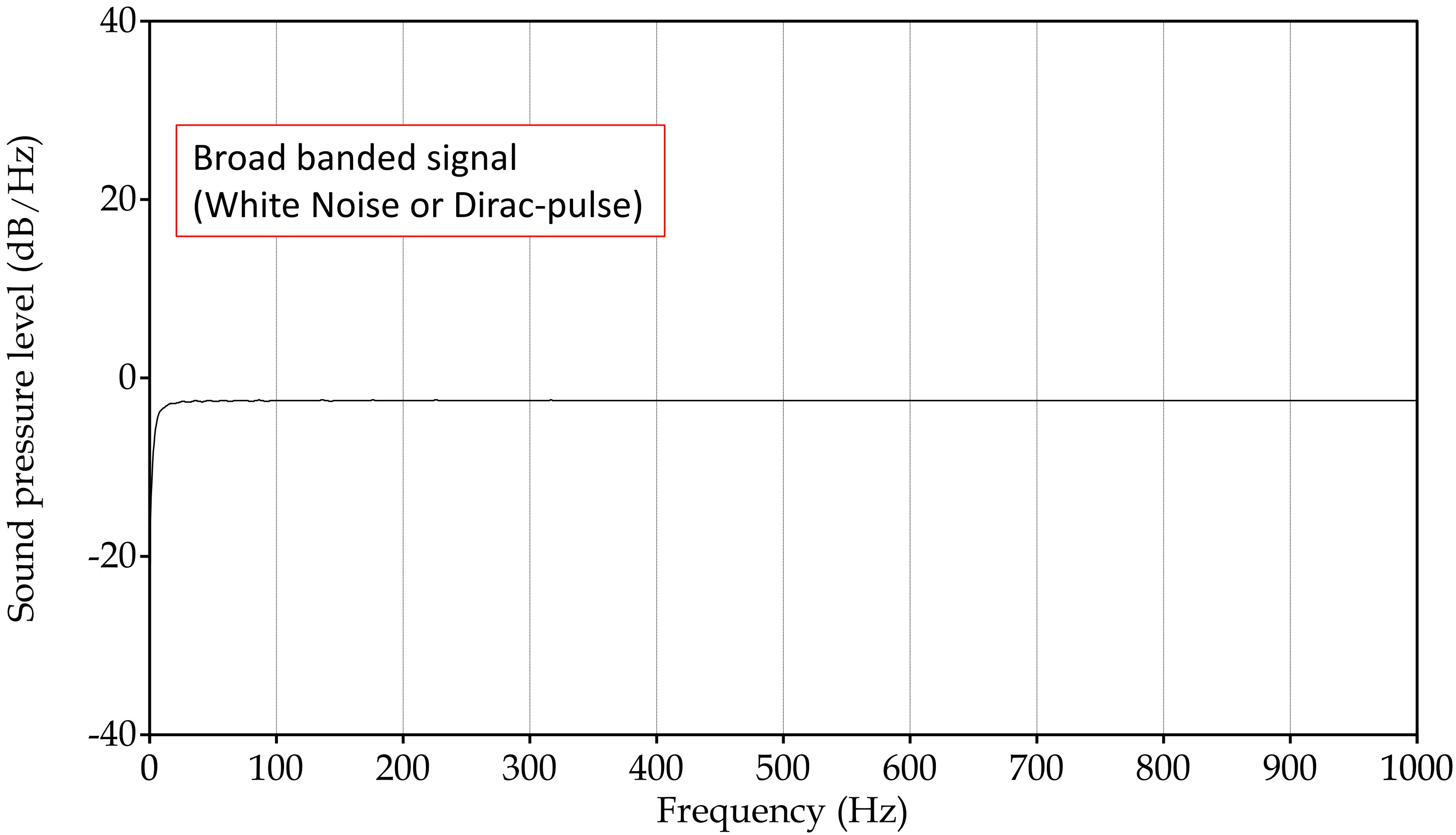
DELAY
(+/- Feed back(

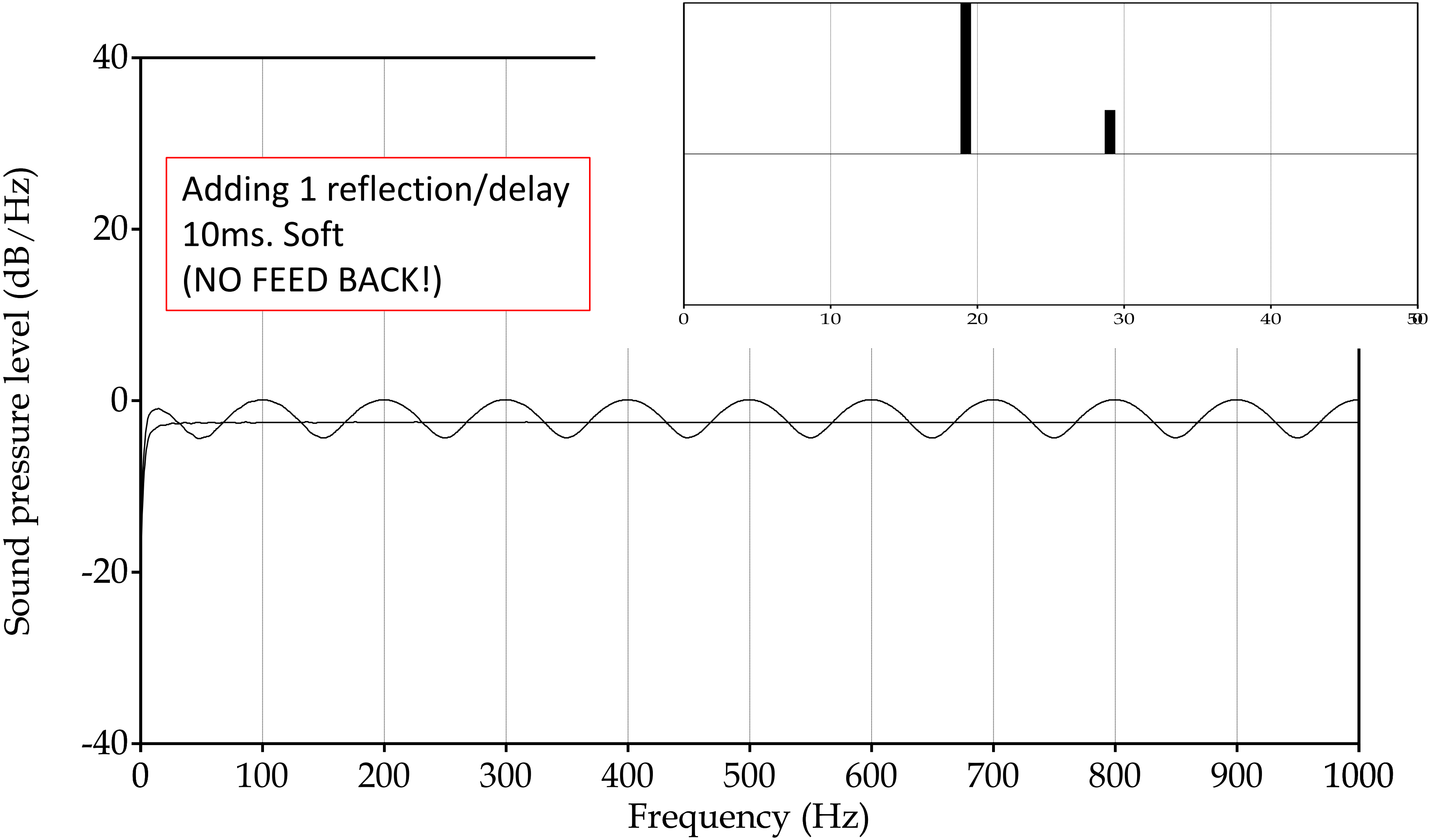


	BL	FB	FF
FIR comb filter	X	0	X
IIR comb filter	1	X	0
allpass	a	$-a$	1
delay	0	0	1

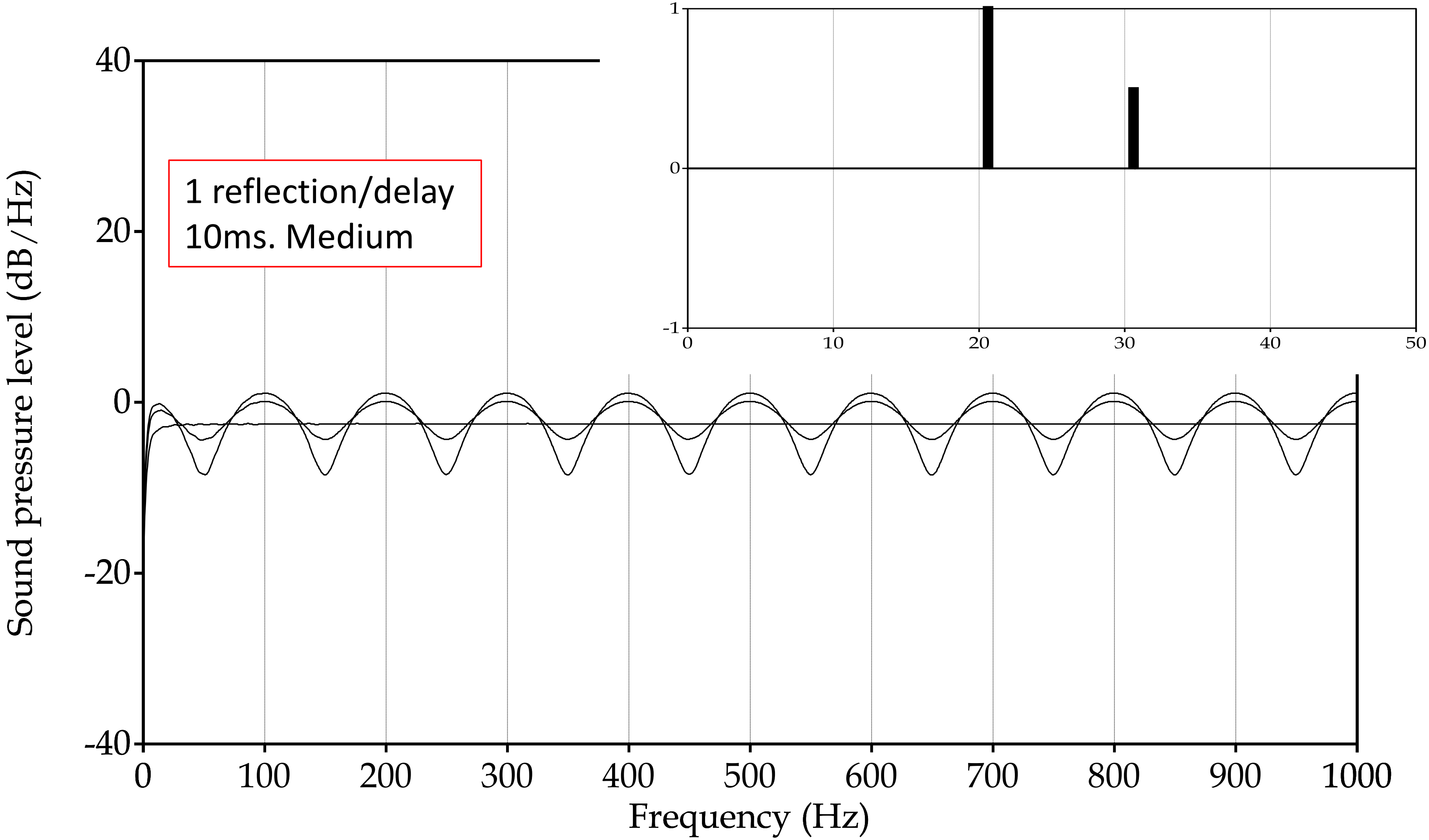
TTT: Timbre Takes Time

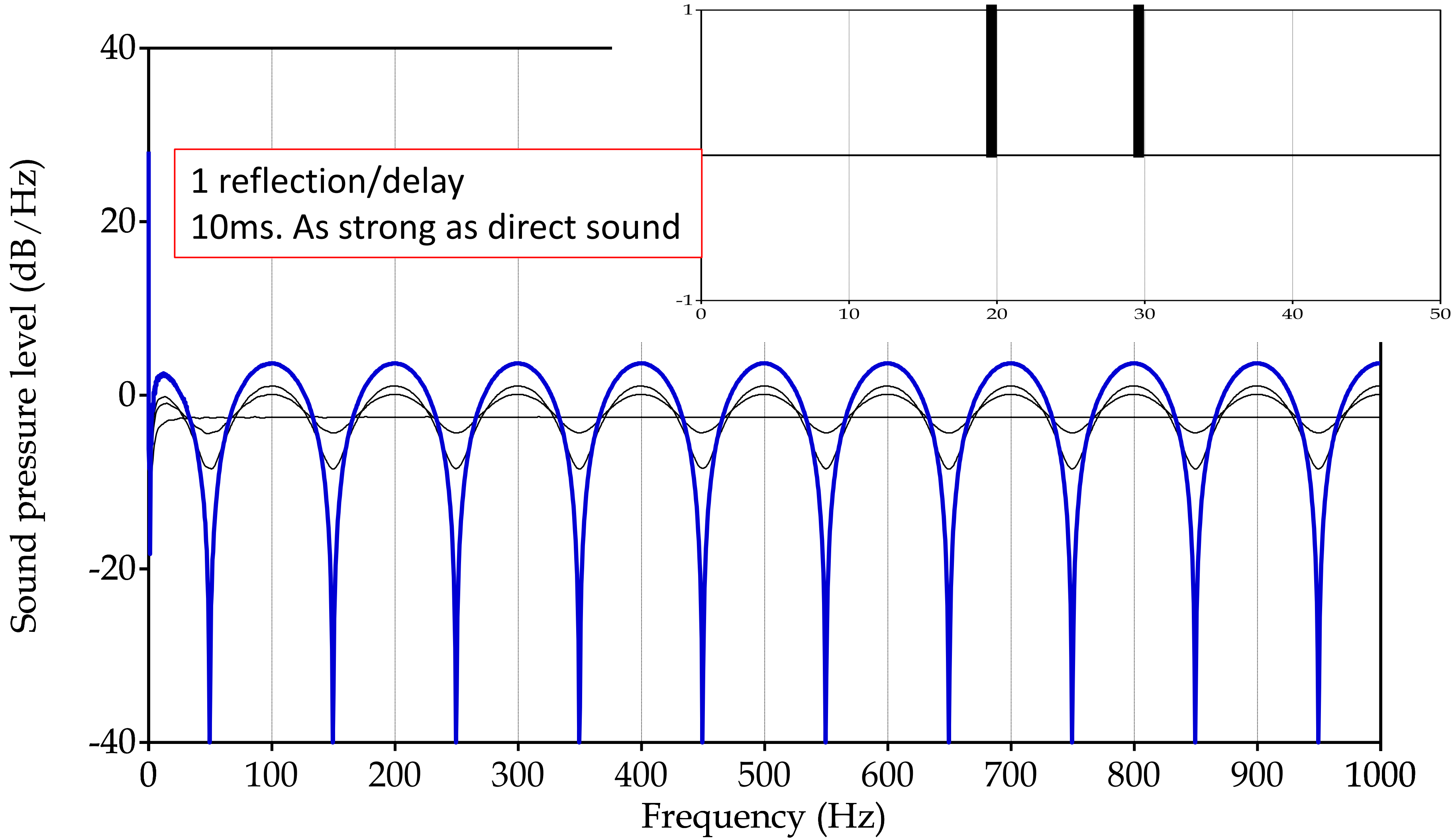



Broad banded signal
(White Noise or Dirac-pulse)
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
100
200
300
400
500
600
700
800
900
1000
Frequency (Hz)

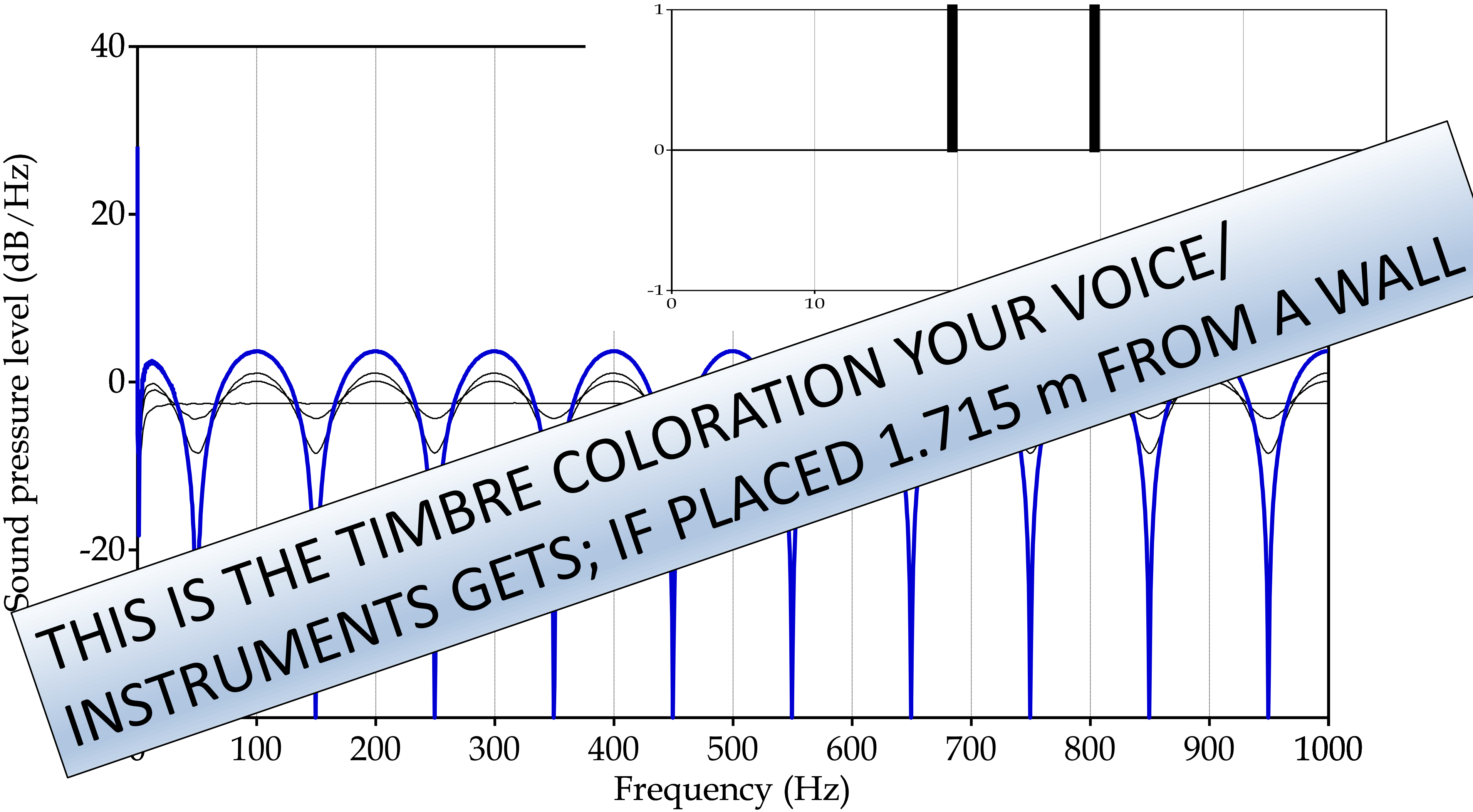


Adding 1 reflection/delay
10ms. Soft
(NO FEED BACK!)
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
100
200
300
400
500
600
700
800
900
1000
Frequency (Hz)
0
10
20
30
40
50


1 reflection/delay
10ms. Medium
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
100
200
300
400
500
600
700
800
900
1000
Frequency (Hz)
1
0
-1
0
10
20
30
40
50


1 reflection/delay
10ms. As strong as direct sound
Sound pressure level (dB/Hz)
Frequency (Hz)
40
20
0
-20
-40
0
100
200
300
400
500
600
700
800
900
1000
1
-1
0
10
20
30
40
50

COMB-FILTER



THIS IS THE TIMBRE COLORATION YOUR VOICE/
INSTRUMENTS GETS; IF PLACED 1.715 m FROM A WALL

COMB FILTERS

$$CBTB_{=CombBetweenTeethBandwidth} = \frac{1}{\Delta t}$$

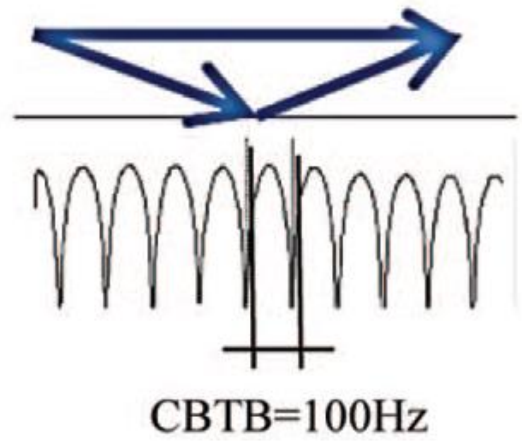
DELAY:

SHORT

MEDIUM
(TIMBRE CHANGES)

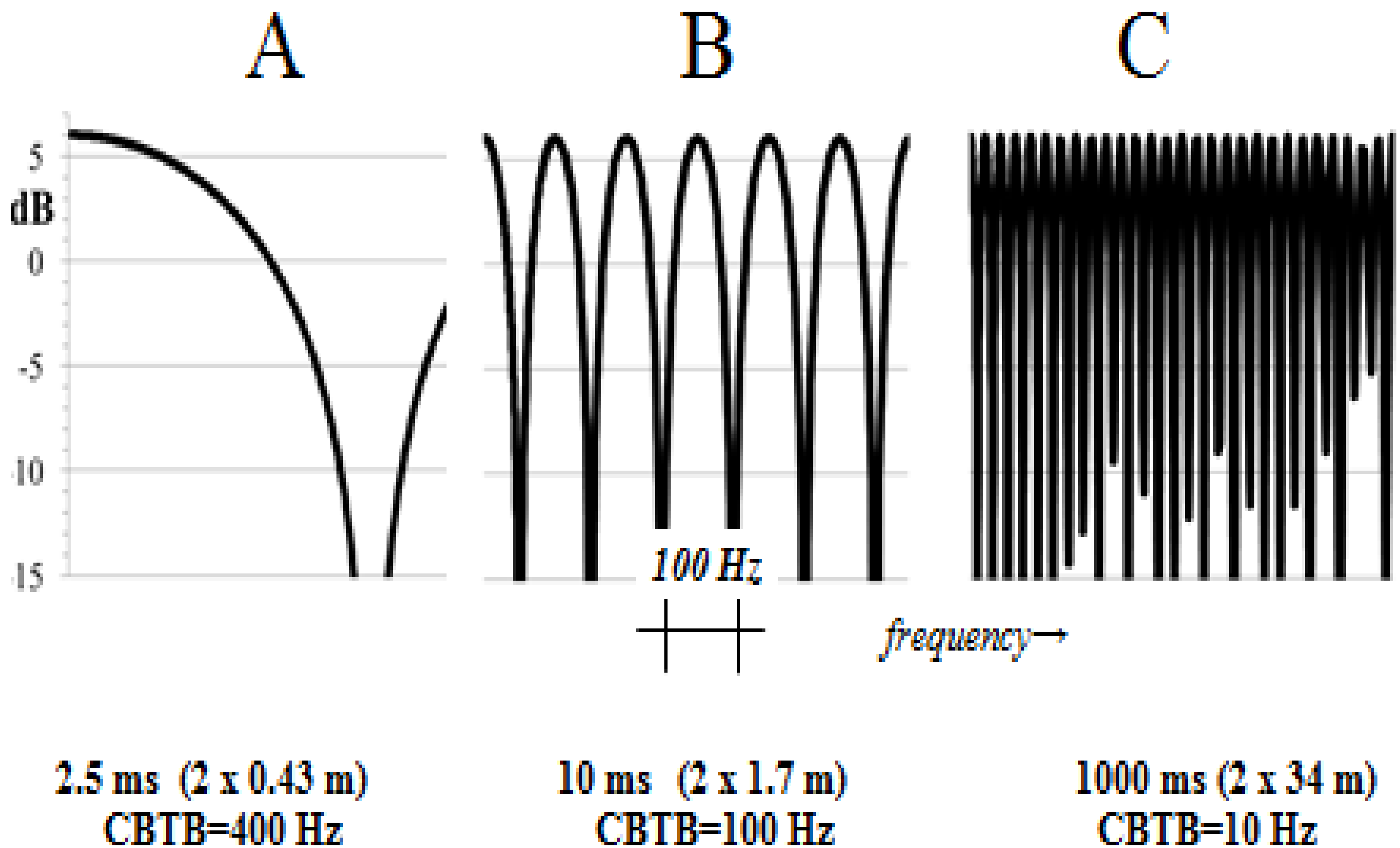
$\Delta t = 10\ ms$

$\Delta x = 2\ x\ 1{,}715 = 3{,}43\ m$



LONG
(ECHO in TIME DOMAIN)

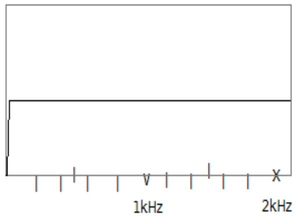
PS! For 50 ms: CBTB=1000/50=20 Hz



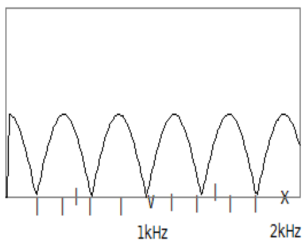
A
B
C
5
dB
0
-5
-10
-15
100 Hz
frequency→
2.5 ms (2 x 0.43 m)
CBTB=400 Hz
10 ms (2 x 1.7 m)
CBTB=100 Hz
1000 ms (2 x 34 m)
CBTB=10 Hz

PERCEIVED «Pitch» of COMB FILTERS

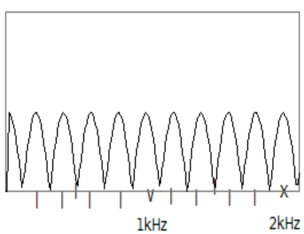
0 ms



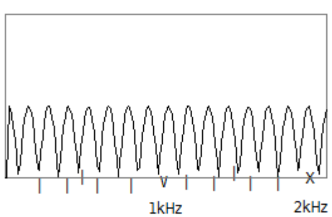
2,5ms CBTB=400Hz *(2x0,43m)*



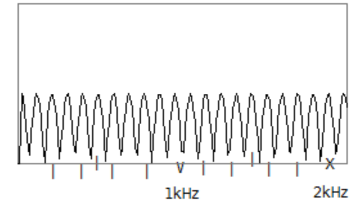
5 ms CBTB=200Hz *(2x0,86m)*



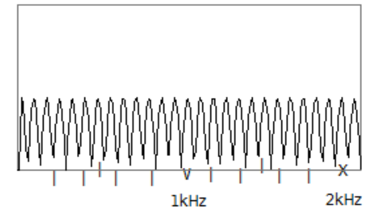
7,5 ms CBTB=137Hz *(2x1,3m)*



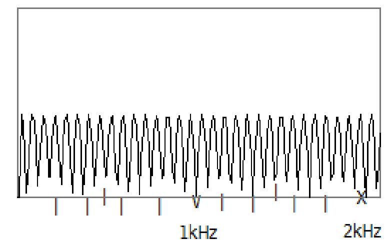
10ms CBTB=100Hz *(2x 1,7m)*



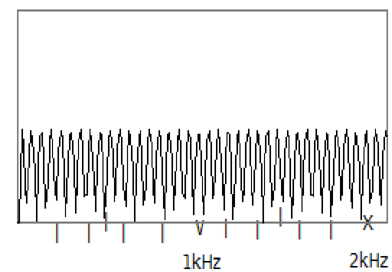
12,5ms CBTB=80Hz *(2x2,14m)*



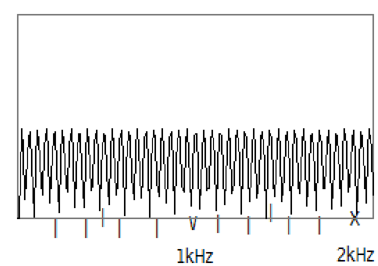
15 ms CBTB=66,7Hz *(2x2,6m)*



17,5ms CBTB=57Hz *(2x3m)*



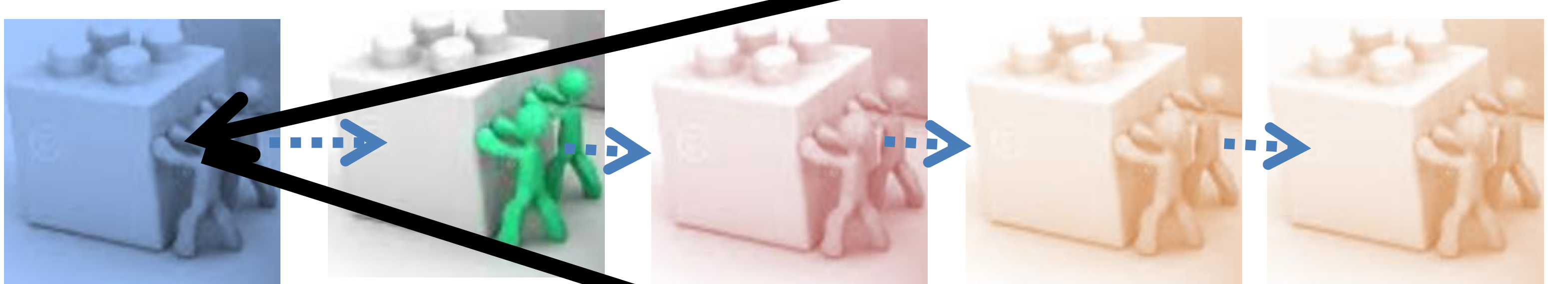
20 ms CBTB=50 Hz *(2x3,4m)*



50 ms CBTB=20 Hz *(2x8,6m)*

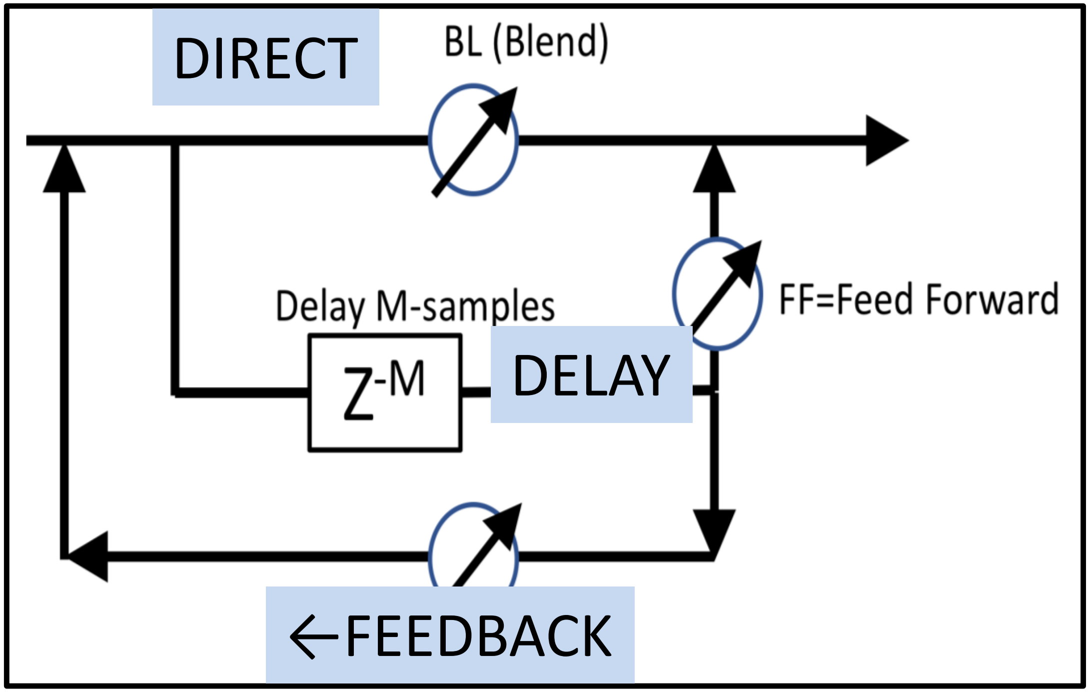
Constant white noise
approximation: Water fountain, Walking on gravel etc.

Composition made by abruptly moving a reflecting wall



The **R**ational Anthem of Norway

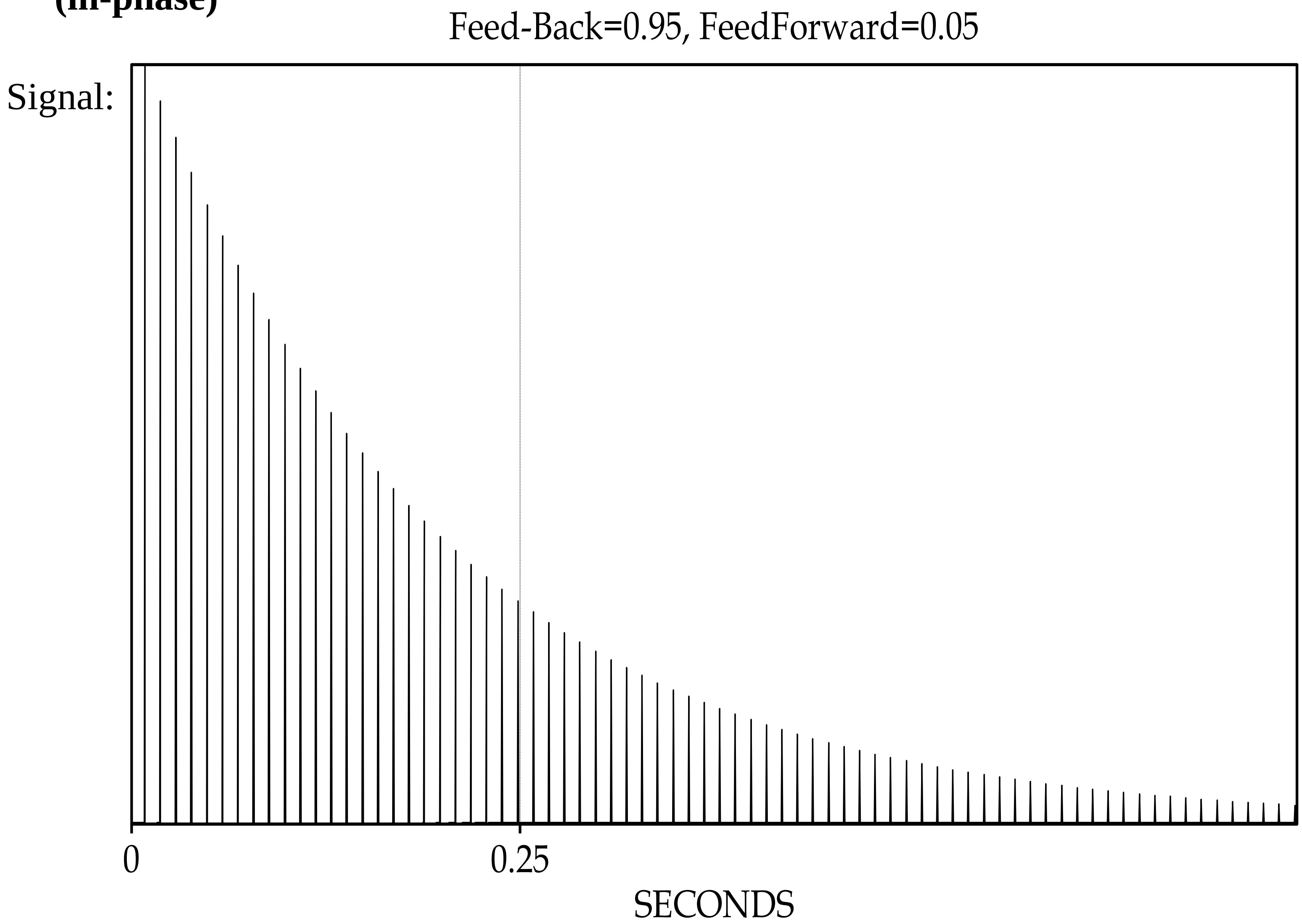
DELAY +/- Feed back

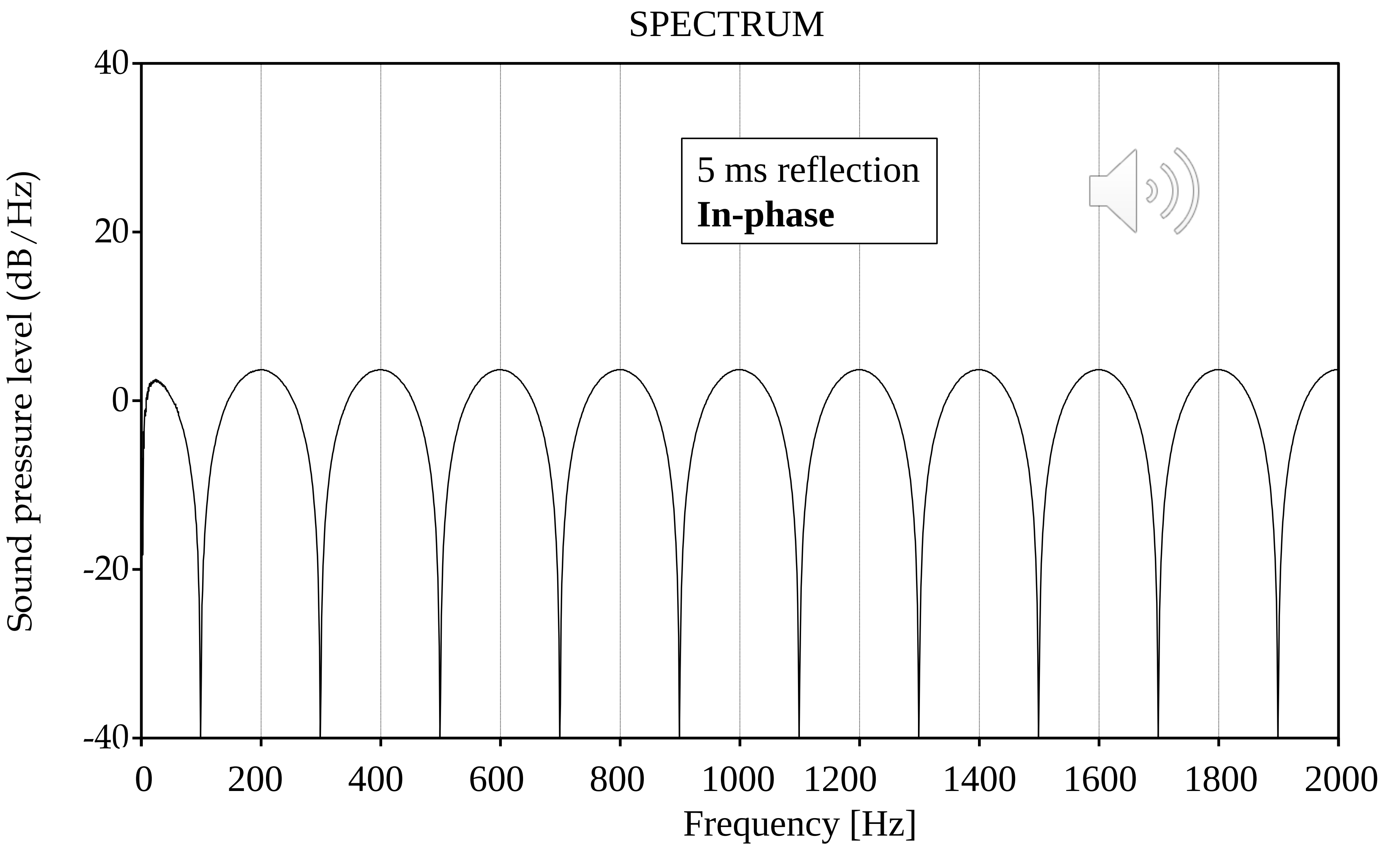


	BL	FB	FF
FIR comb filter	X	0	X
IIR comb filter	1	X	0
allpass	a	$-a$	1
delay	0	0	1

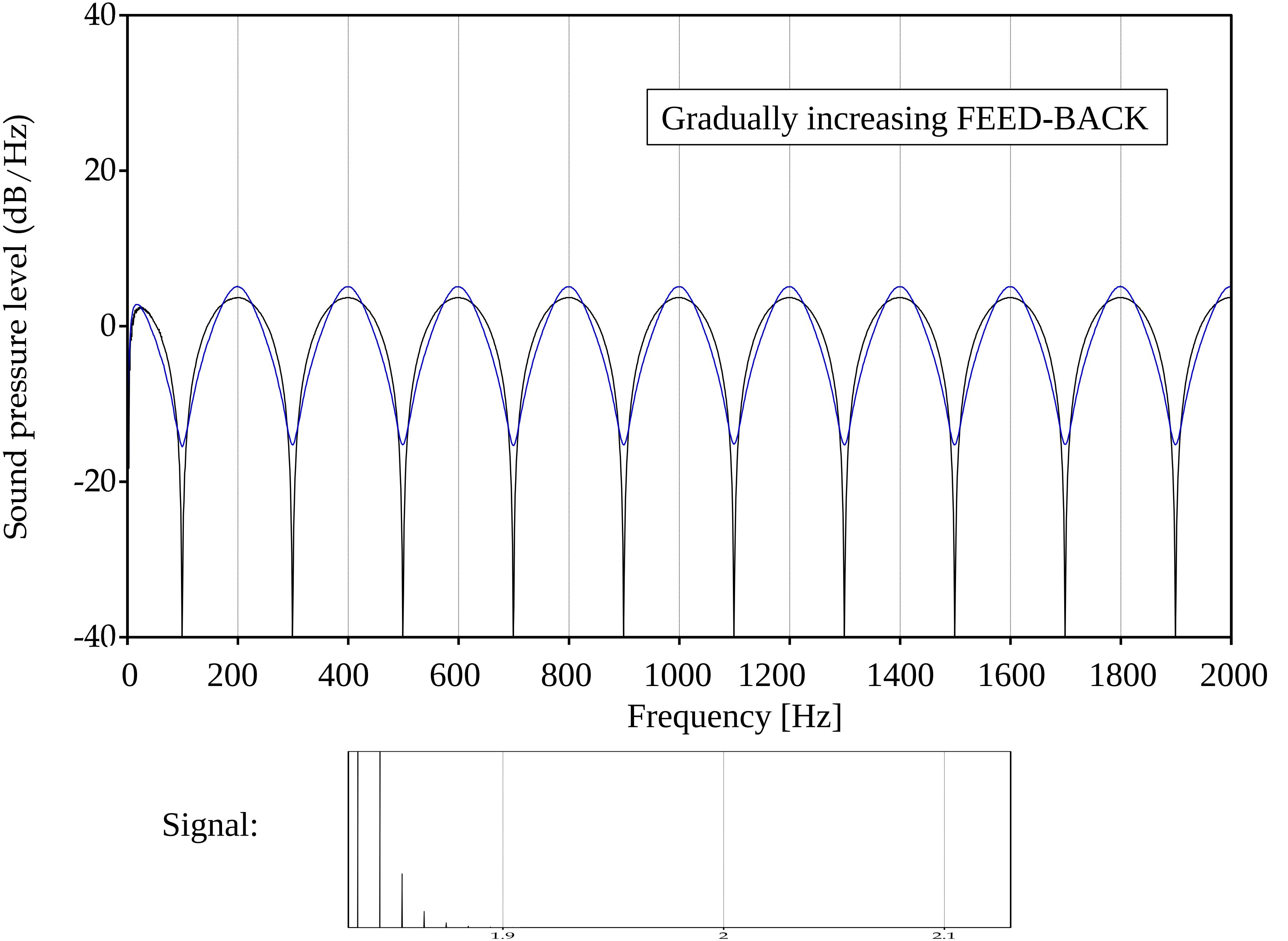
TTT: Timbre Takes Time

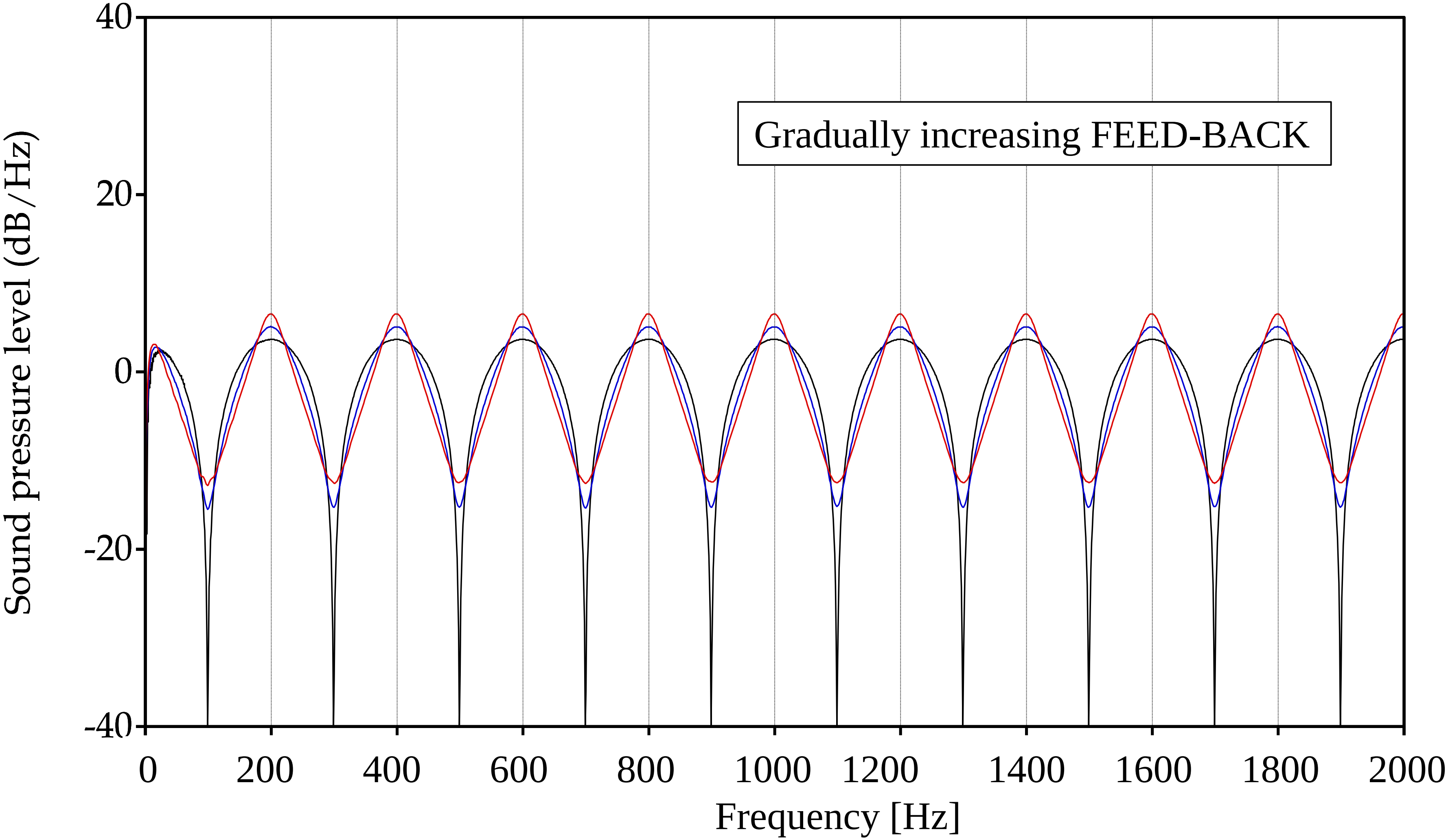
WITH FEED BACK
(in-phase)



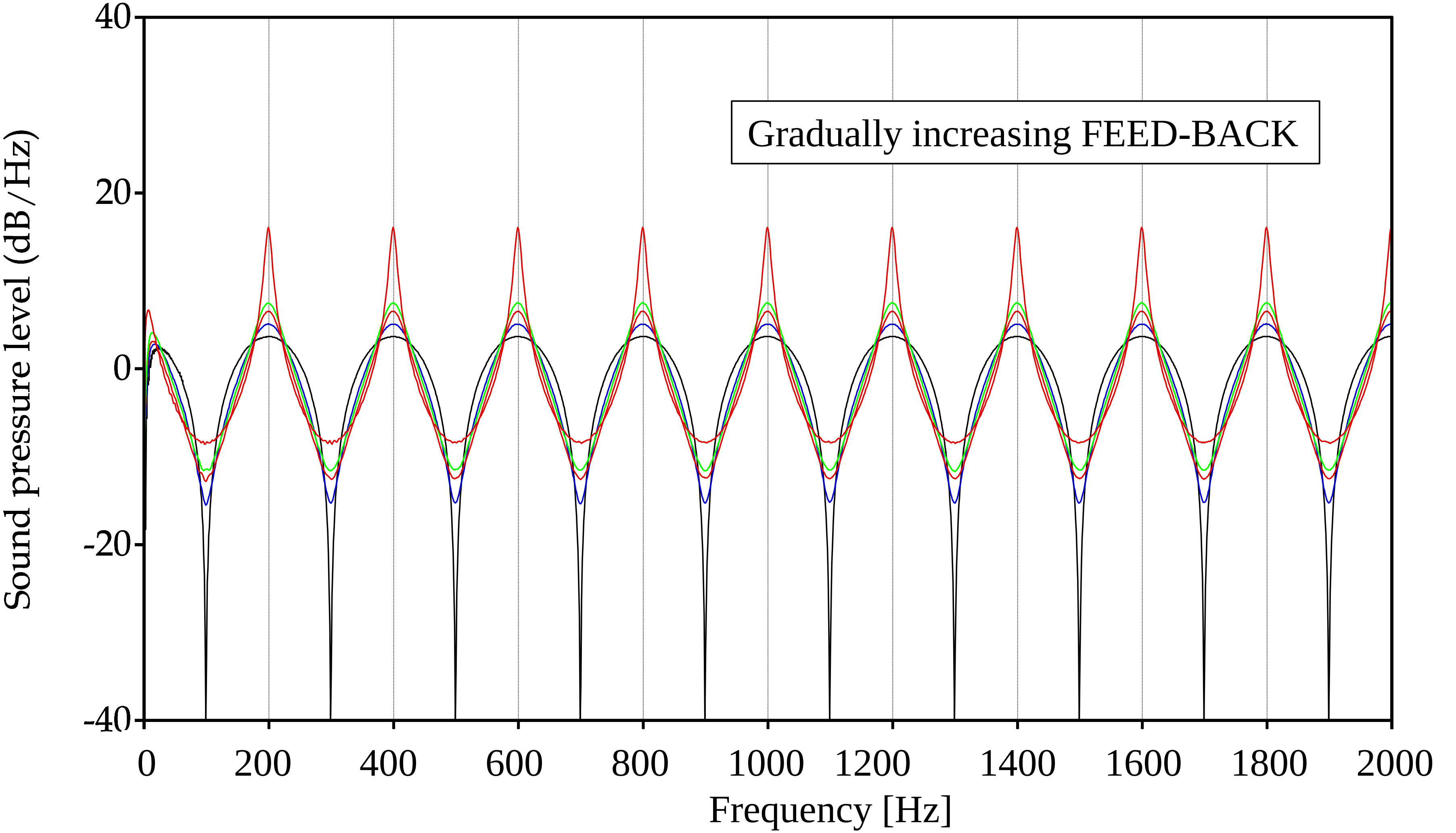


SPECTRUM
5 ms reflection
In-phase
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Frequency [Hz]

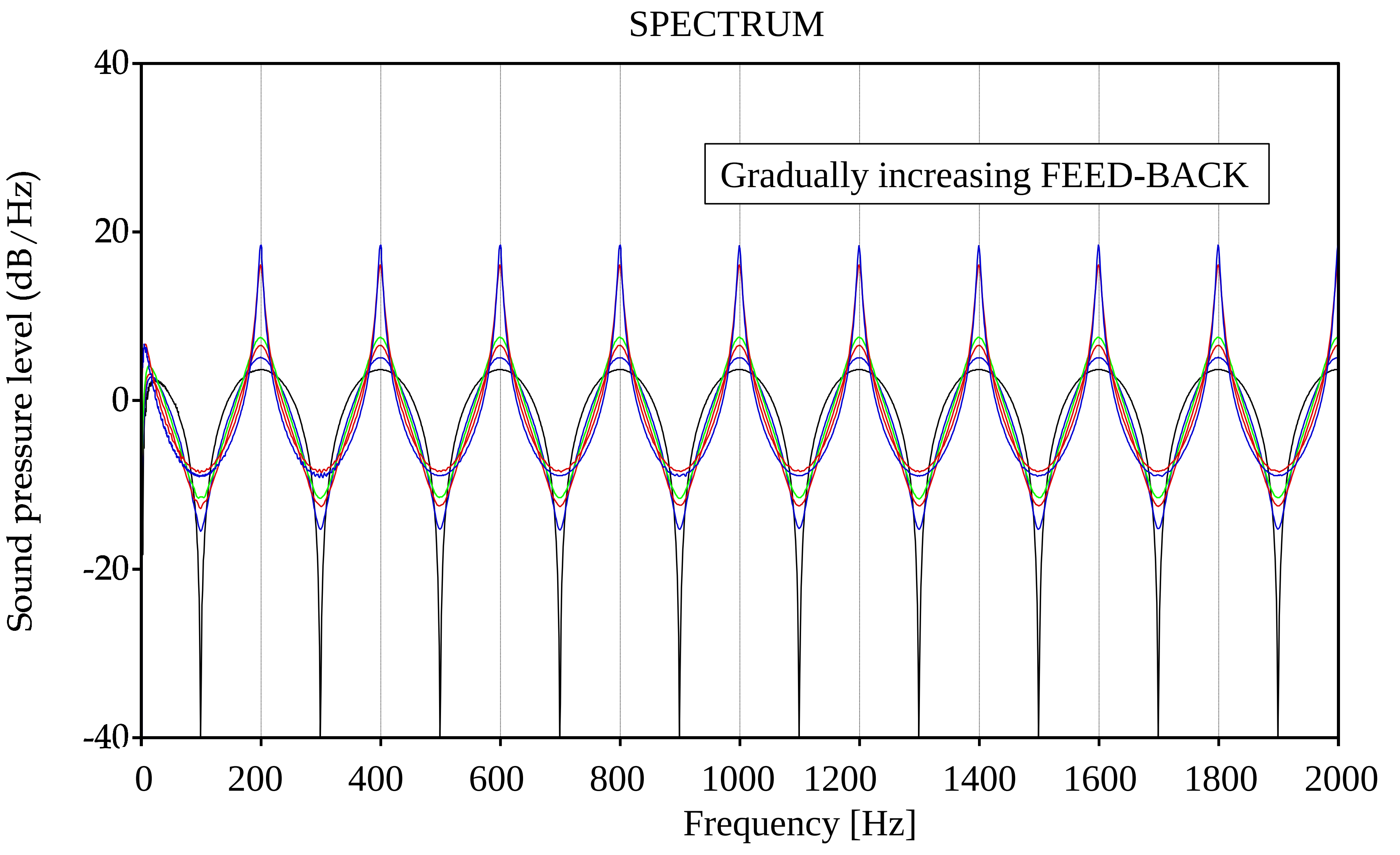

Gradually increasing FEED-BACK
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Frequency [Hz]
Signal:
1.9
2
2.1



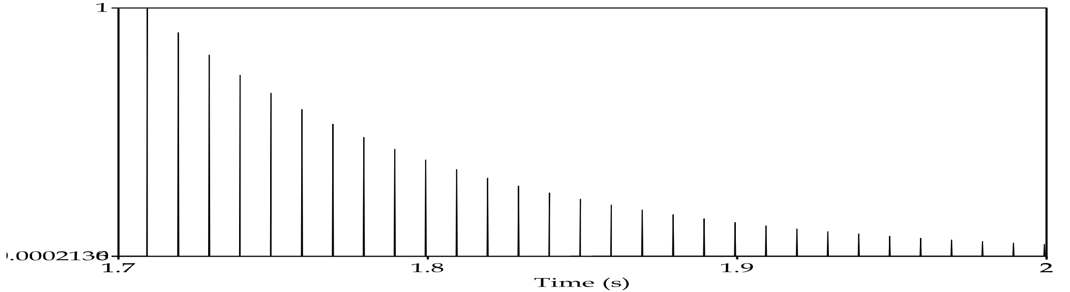
Gradually increasing FEED-BACK
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Frequency [Hz]

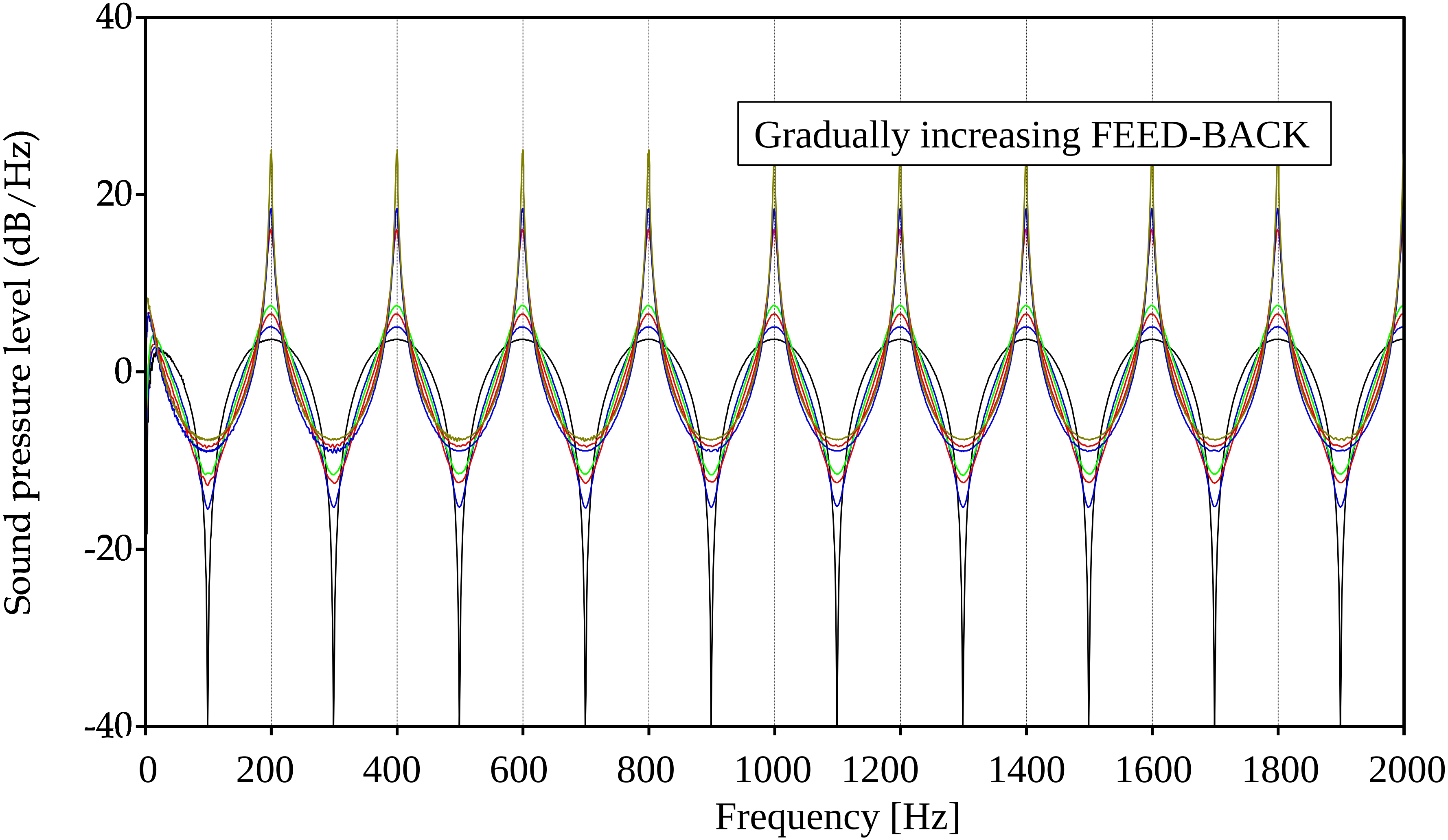


Gradually increasing FEED-BACK
Sound pressure level (dB/Hz)
Frequency [Hz]
40
20
0
-20
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000


SPECTRUM
Gradually increasing FEED-BACK
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Frequency [Hz]

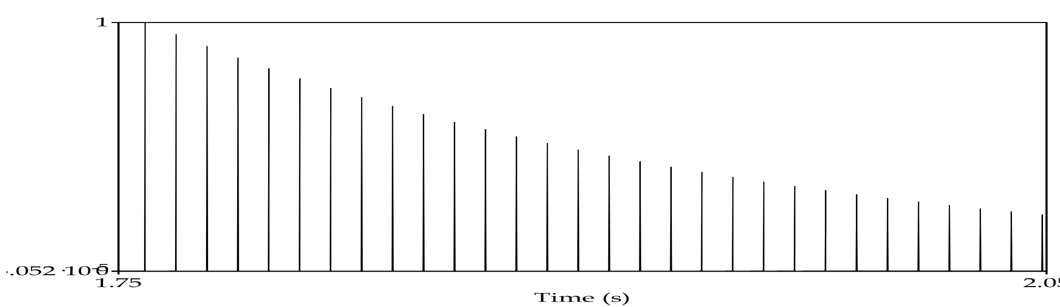
Signal:


1
.0002136
1.7
1.8
1.9
2
Time (s)

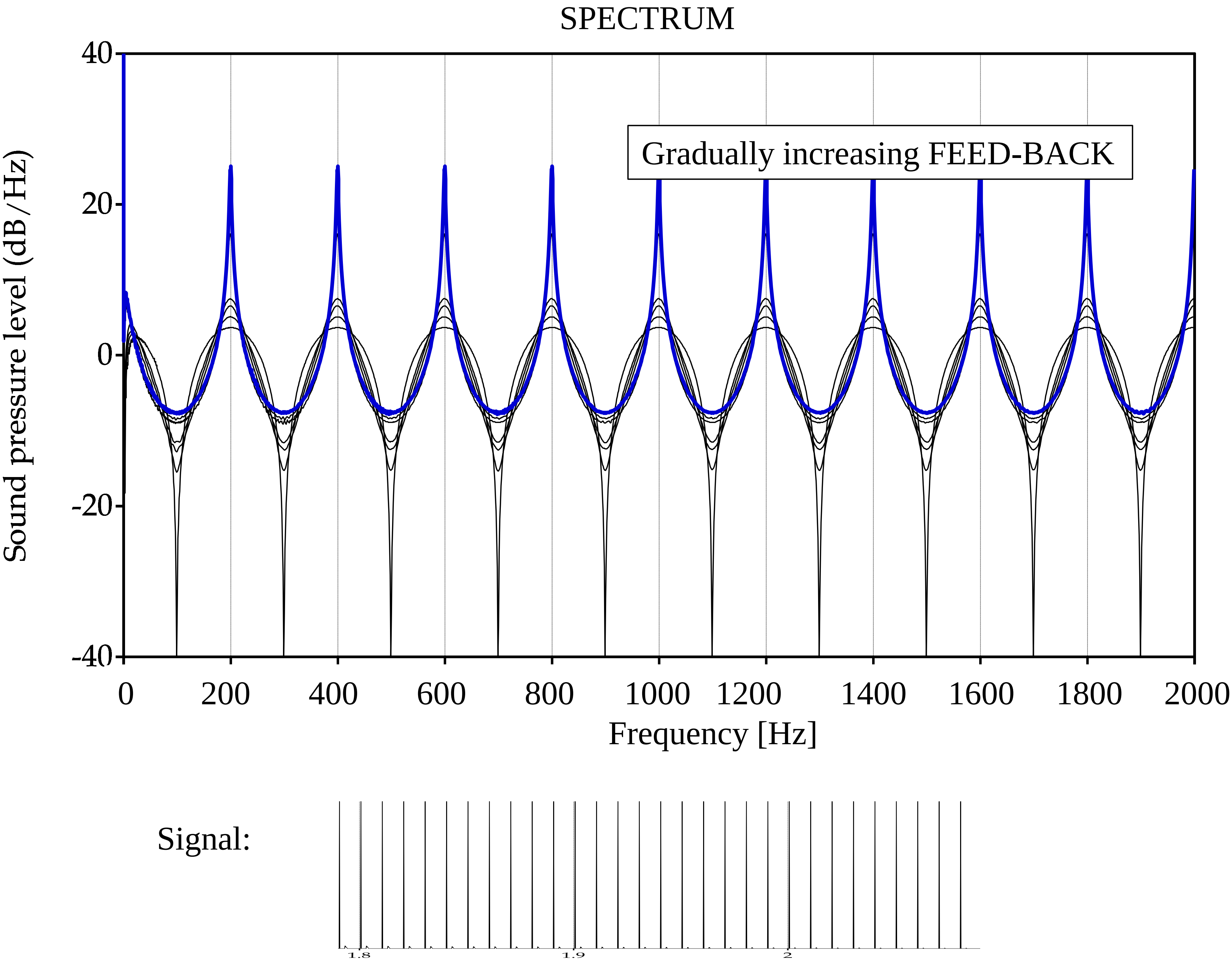


Gradually increasing FEED-BACK
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Frequency [Hz]

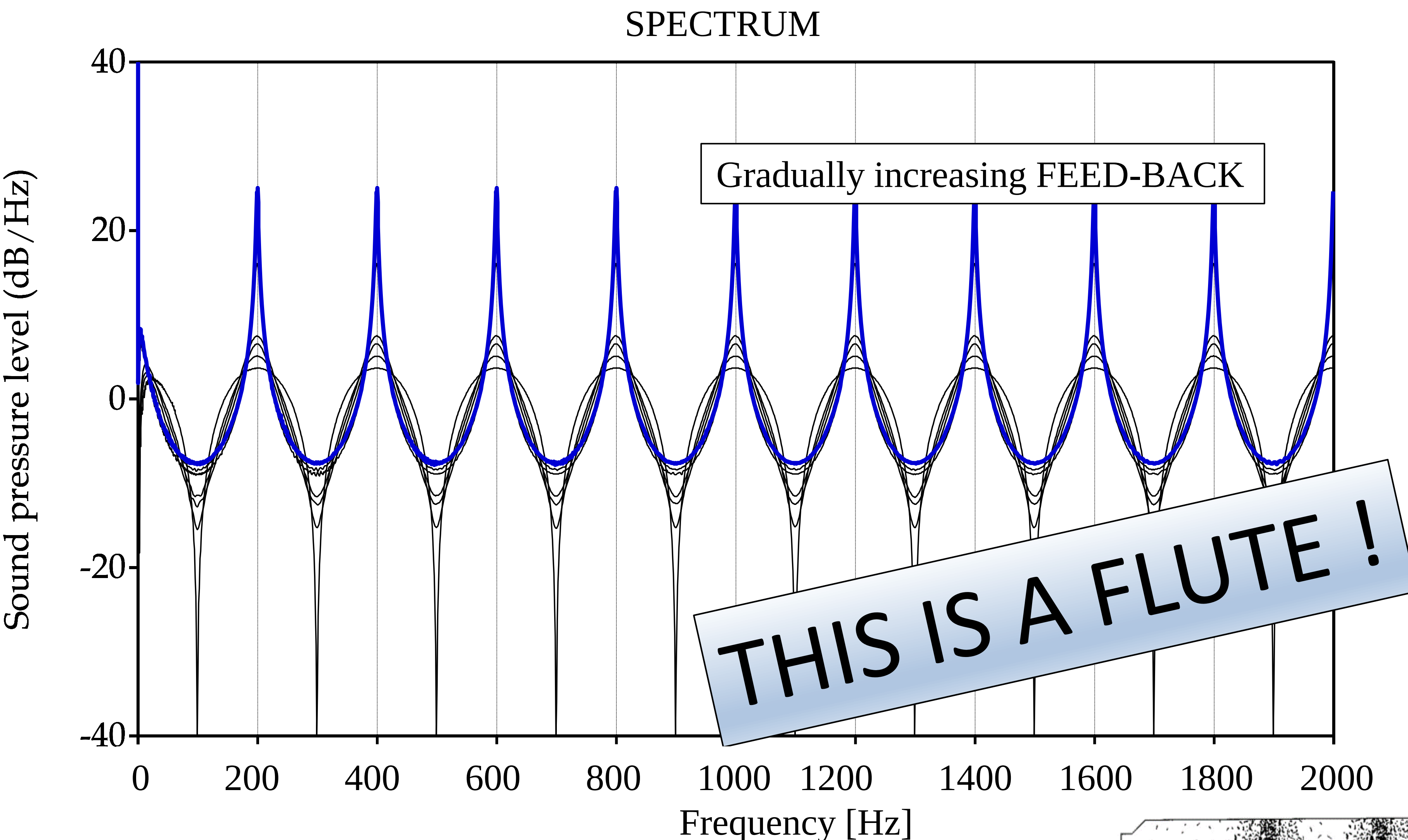
Signal:



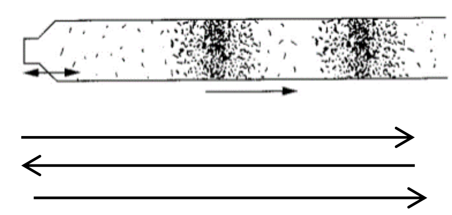
1
1.75
2.05
Time (s)



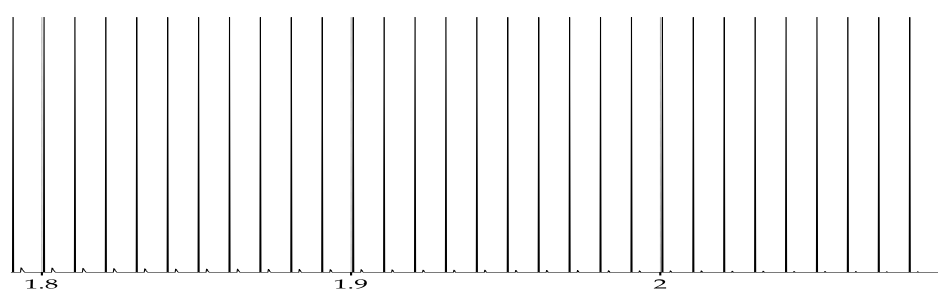
SPECTRUM
Gradually increasing FEED-BACK
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Frequency [Hz]
Signal:
1.8
1.9
2



SPECTRUM
Gradually increasing FEED-BACK
THIS IS A FLUTE !
Sound pressure level (dB/Hz)
40
20
0
-20
-40
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Frequency [Hz]



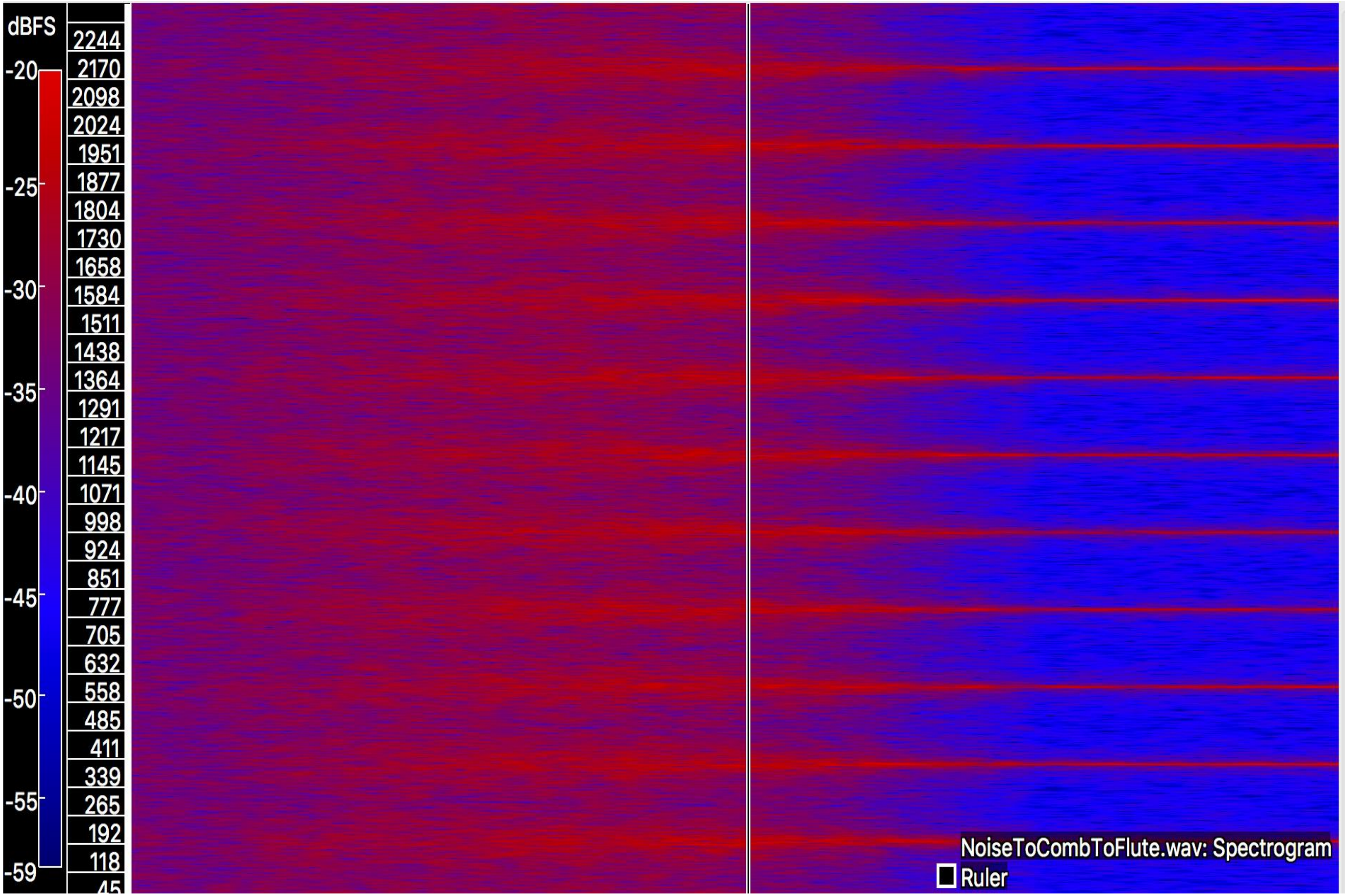
Signal:



1.8
1.9
2

Freq [Hz]
linear

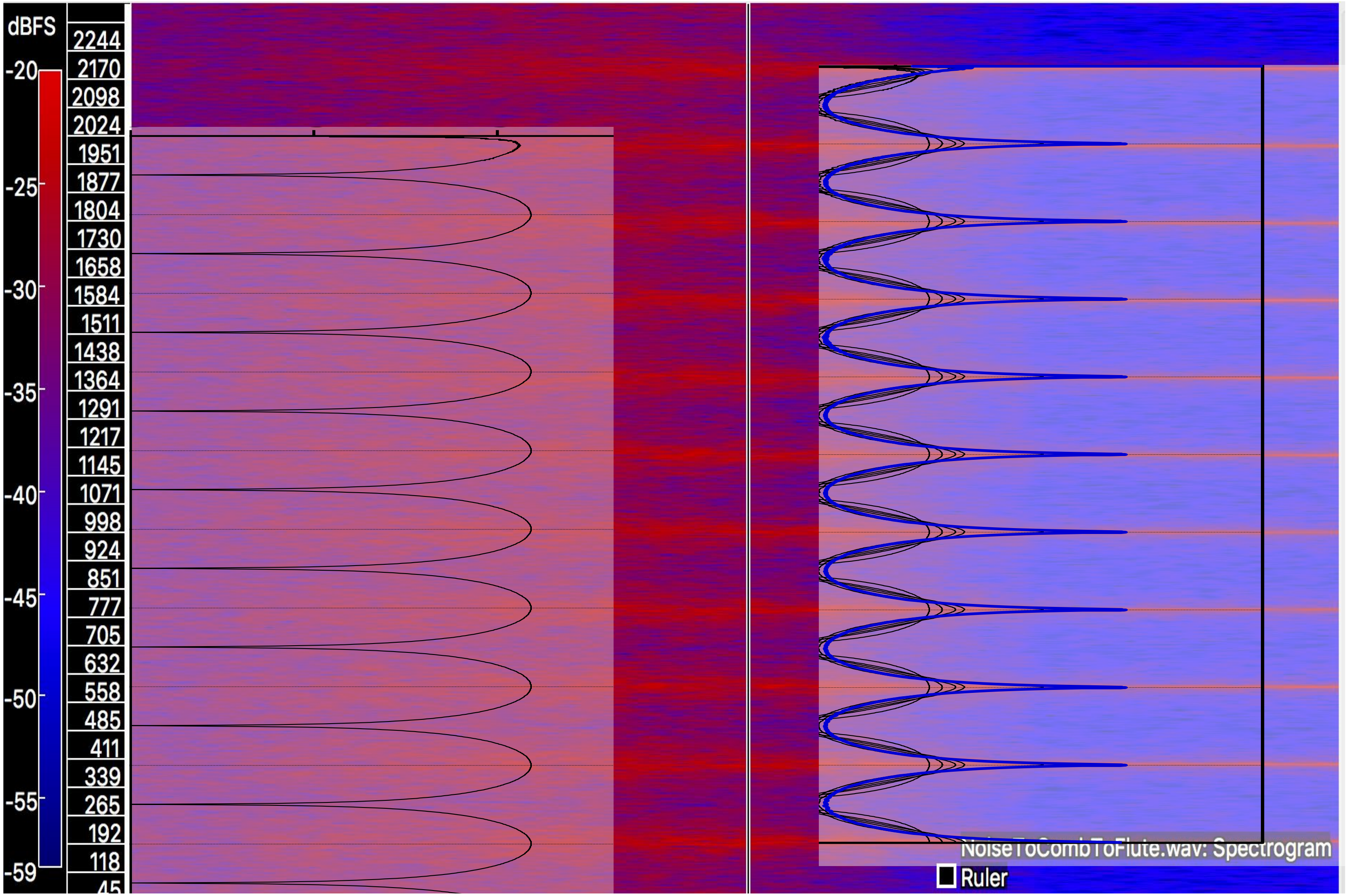
SPECTROGRAM



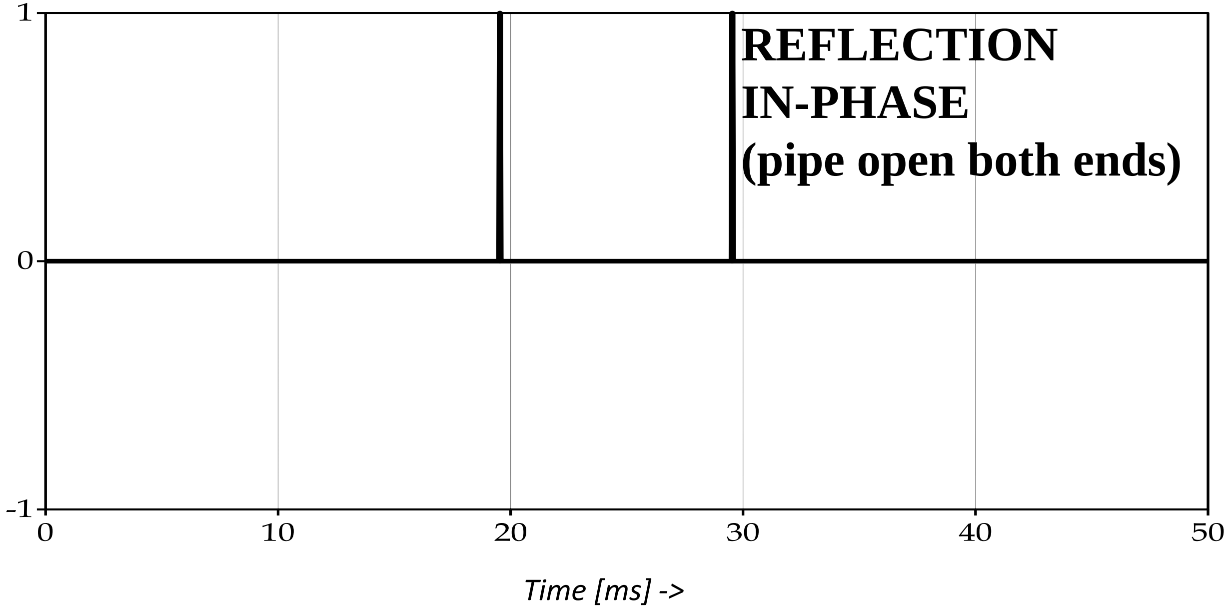
Time s

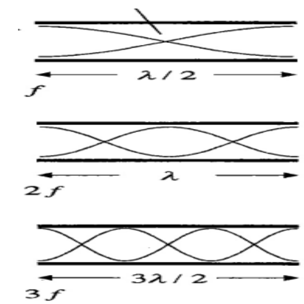
Freq [Hz]
linear

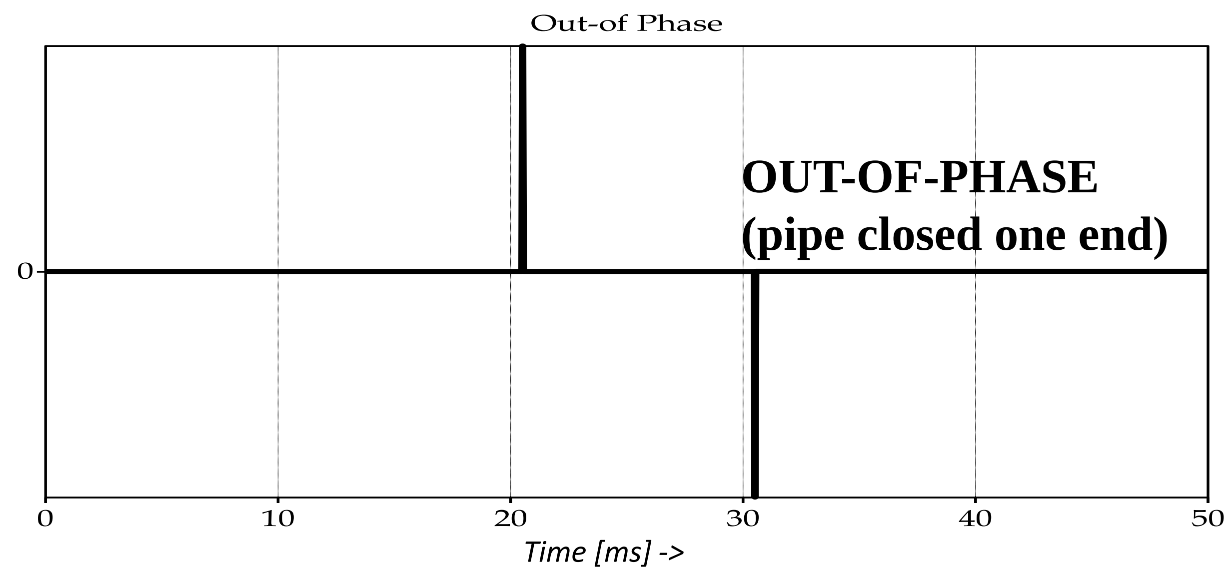
SPECTROGRAM

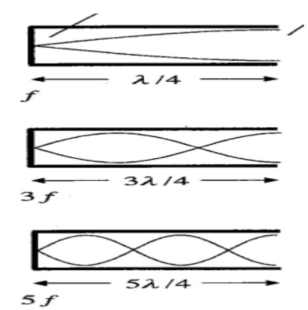


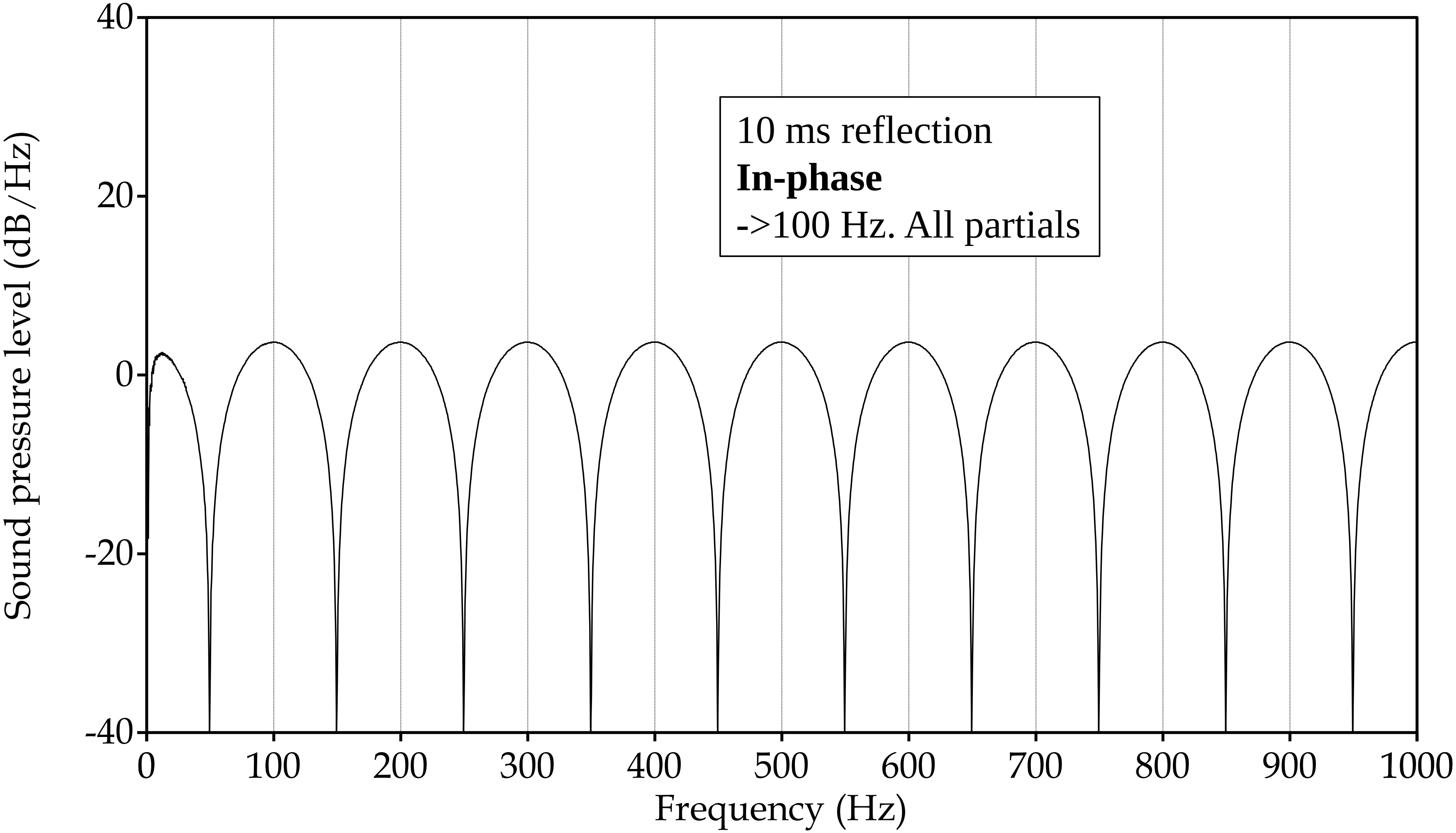
Time s


REFLECTION
IN-PHASE
(pipe open both ends)
1
0
-1
0
10
20
30
40
50
Time [ms] ->

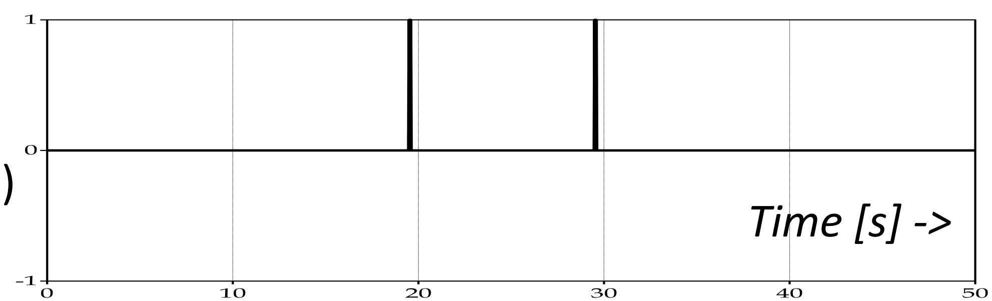

λ / 2
f
λ
2 f
3λ / 2
3 f

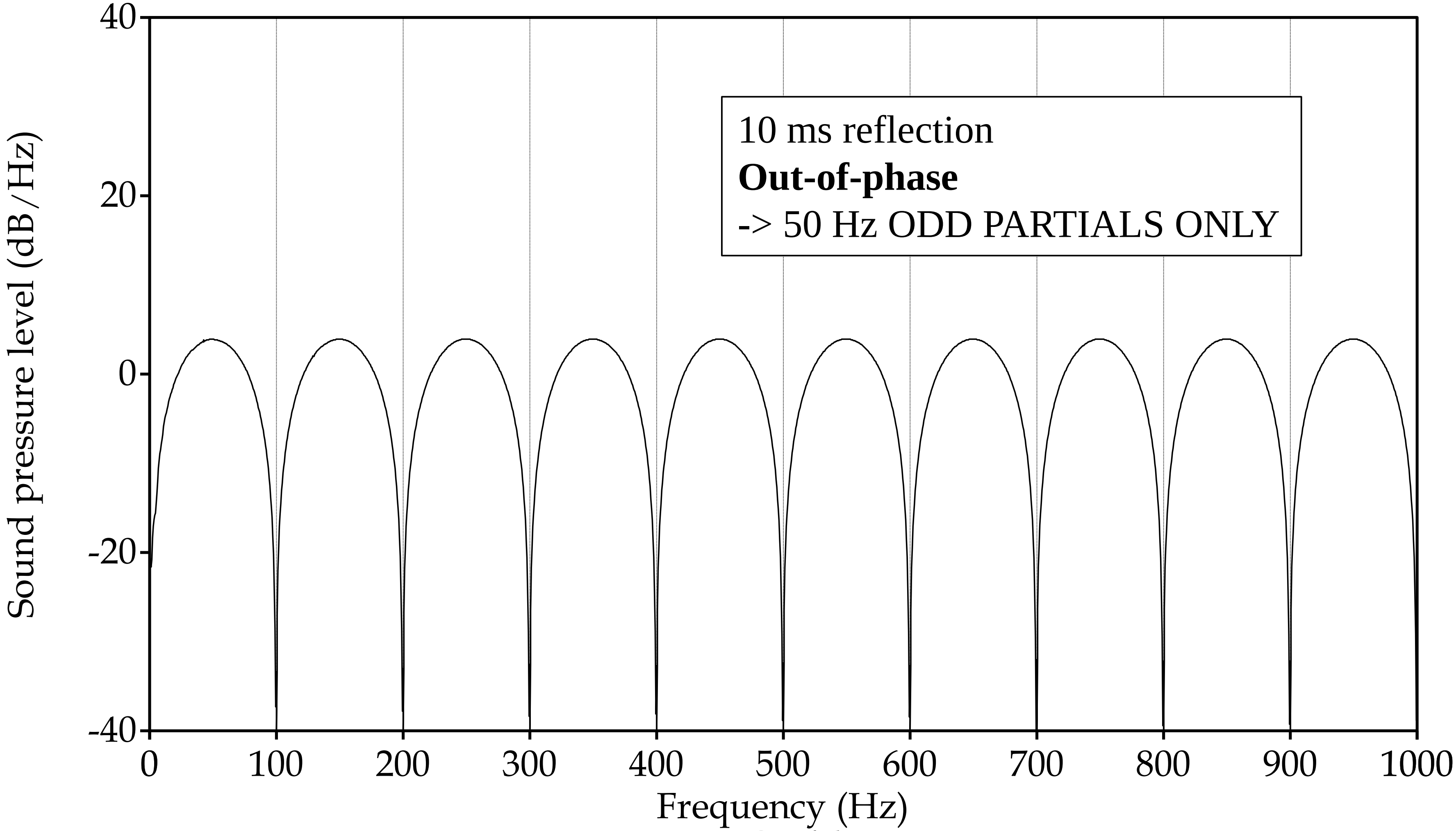

Out-of Phase
OUT-OF-PHASE
(pipe closed one end)
0
0
10
20
30
40
50
Time [ms] ->


λ /4
f
3λ /4
3 f
5λ /4
5 f

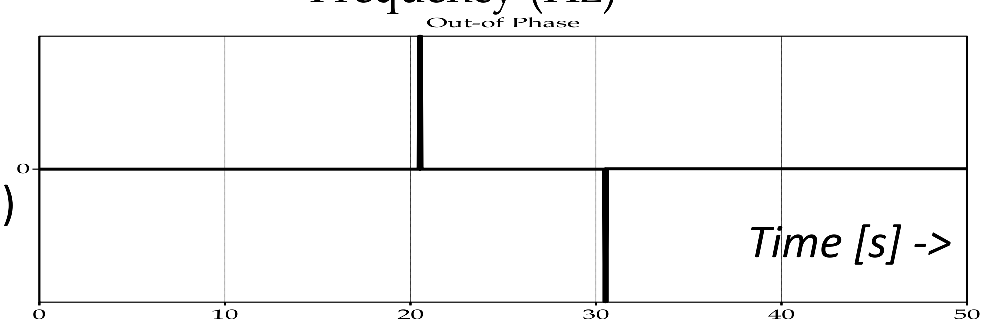


REFLECTION
(IMPULS-RESPONSE)





REFLECTION
(IMPULS-RESPONSE)

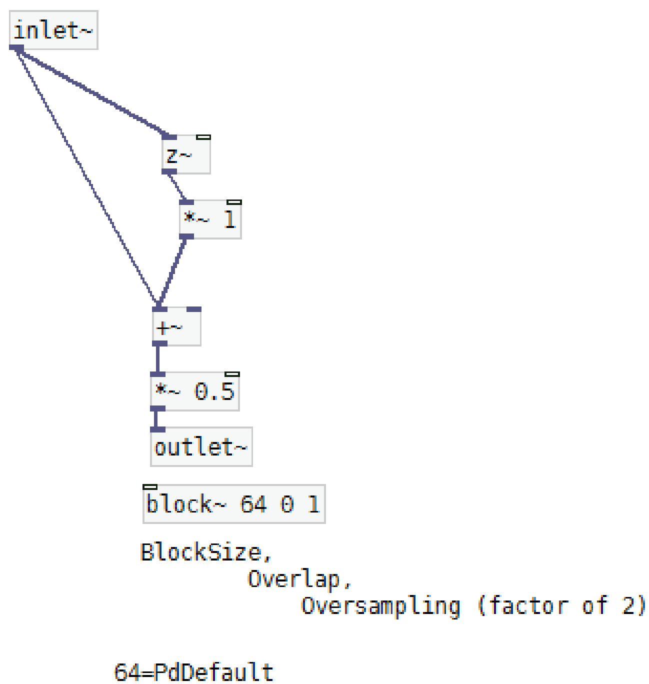


SIDESTEP

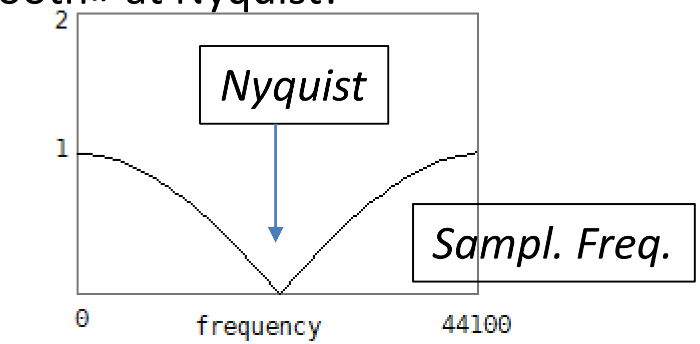
Mean of two and two samples is often considered to give a **Low Pass filter**.

Ever wondered if that is the complete answer?

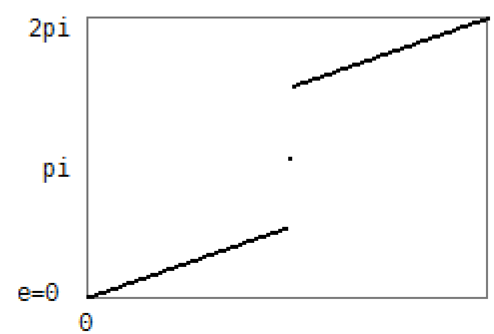
It is **actually a COMB FILTER**, with the first dip/ «tooth» at Nyquist!



Frequency Response



Phase Response

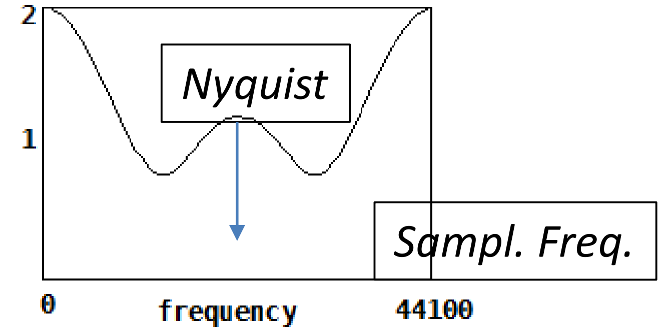


With a Group Delay (derivate of phase) of 0.5, constant for all frequencies.

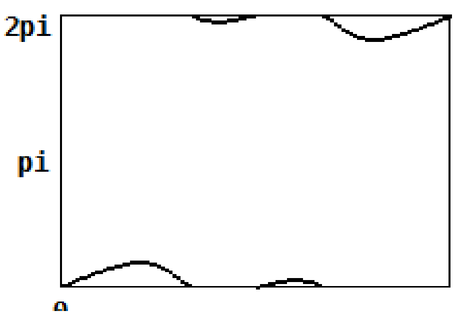
And the mean of 3 samples gives a Comb Filter with first tooth at Nyquist/3, and thus a LowPass for low freq, but HiPass for very high freq, towards Nyquist

The mean of 3 samples gives a Comb Filter
with first dip /«tooth» at Sampl.Rate/3,

Frequency
Response



Phase
Response

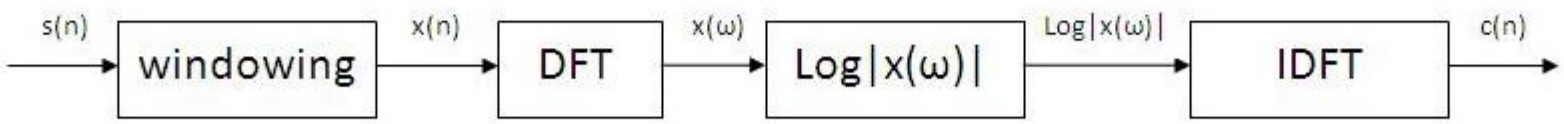


a LowPass for low freq, but
HiPass for very high freq, towards Nyquist

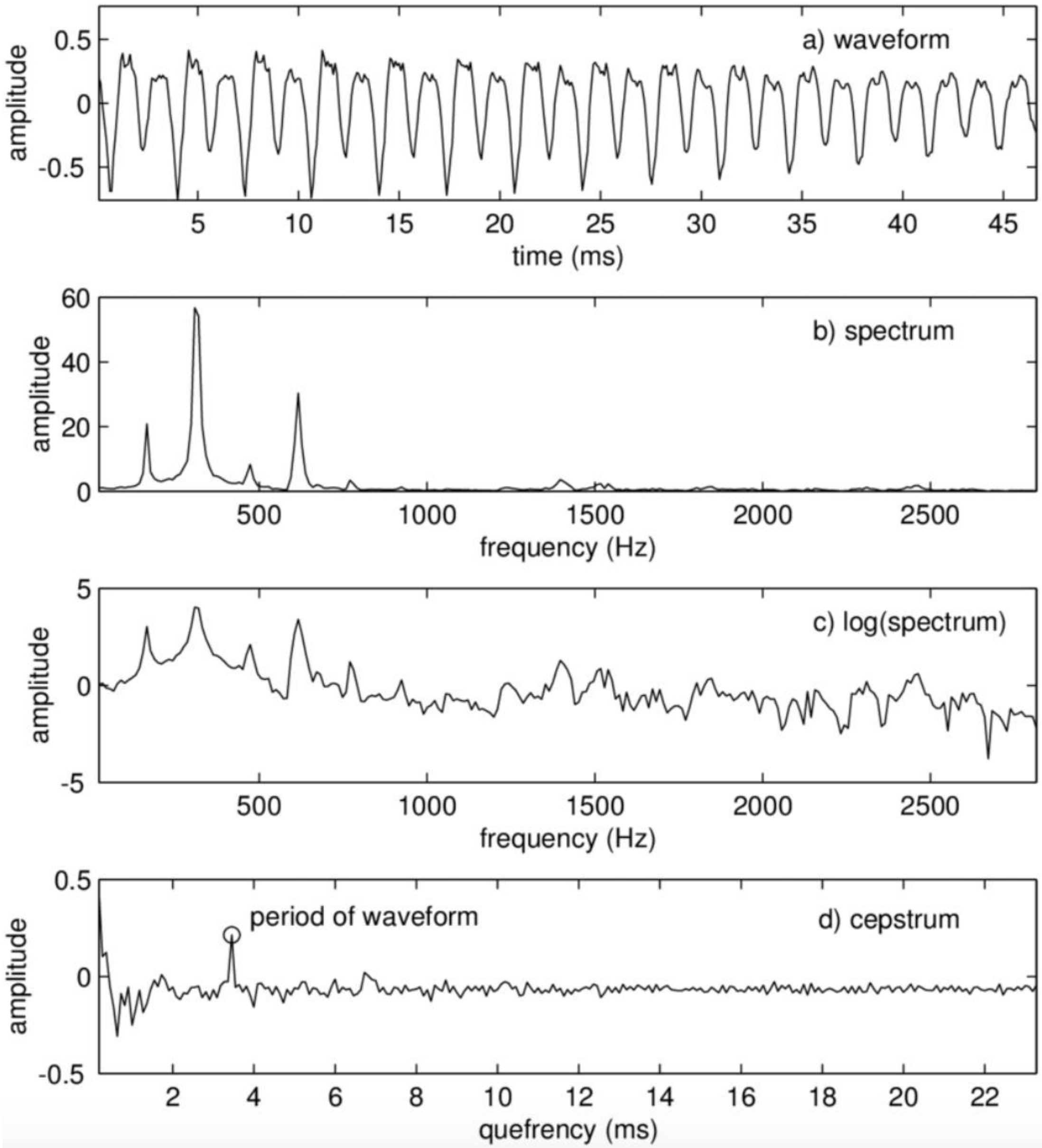
"CEPSTRUM" =reversing the first four letters of "SPECTRUM".

CEPSTRUM

the inverse Fourier transform of the log magnitude spectrum.



complex cepstrum,
real cepstrum,
power cepstrum,
phase cepstrum



SIGNAL

Fourier->

SPECTRUM contains harmonics at evenly spaced intervals, whose **magnitude decreases quite quickly as frequency increases**

LOG SPECTRUM, **compresses the dynamic range and reduces amplitude differences in the harmonics**

Fourier ->
CEPSTRUM

x-axis: QUEFRENCY

FREQUENCY

Quefrency,
“”TIME”” [s] or [samples]

CEPSTRUM

Analysis of
«Rythmic Behavior»
in the frequency domain.

CEPSTRUM:
de-convolution

Separate **Source** and **Filter TIMBRE**

Pitch-Detector

Echo-Detector

Rhythm-Detector(?)

convolution of two signals = addition of their complex cepstra

$$x_1 * x_2 \rightarrow x_1' + x_2'$$

CEPSTRUM:
de-convolution

vocal excitation (pitch) and **vocal tract** (formants) are **additive** in the logarithm of the power spectrum and thus clearly separate.

The **cepstrum** projects all the slowly varying components in log magnitude spectrum to the low frequency region and fast varying components to the high frequency regions.

In the log magnitude spectrum, the slowly varying components represent the envolope corresponds to the vocal tract and the fast varying components to the excitation source.

As a result the vocal tract and excitation source components get represented naturally in the spectrum of speech.

The initial few values in the cepstrum typically 13-15 cepstral values represent the vocal tract information

The **power cepstrum** =squared magnitude of the inverse Fourier transform of the logarithm of the squared magnitude of the Fourier transform of a signal

$$= \left| \mathcal{F}^{-1} \left\{ \log(|\mathcal{F}\{f(t)\}|^2) \right\} \right|^2$$

The *complex cepstrum* is defined as the Inverse Fourier transform of the logarithm (with unwrapped phase) of the Fourier transform of the signal. This is sometimes called the spectrum of a spectrum.

complex cepstrum of signal = IFT(log(FT(the signal))+$j2\pi m$)

(where m is the integer required to properly unwrap the angle or imaginary part of the complex log function)

Many texts define the process as FT → abs() → log → IFT,
i.e., that the cepstrum is the "inverse Fourier transform of the log-magnitude Fourier spectrum".
(the difference between squaring or taking the absolute value amounts to an overall factor of 2)

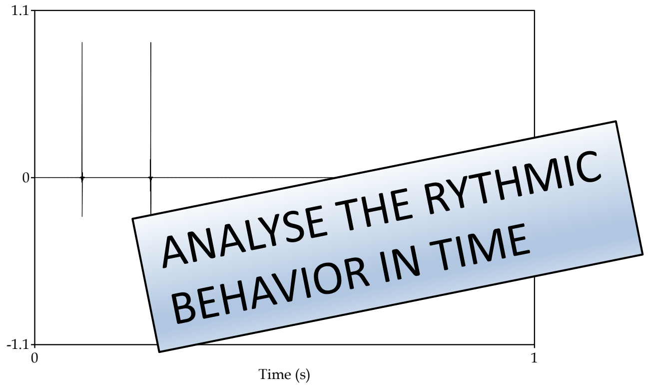
MFCC Mel Filter Cepstral Coefficients. (Speech recognition)
Combines the advantages of the cepstrum with a frequency scale based on critical bands

LPC Linear Prediction Cepstral Coefficients

SIFT *Simplified Inverse Filter Tracking algorithm*
Encompasses the desirable properties of both autocorrelation and cepstral pitch analysis techniques.

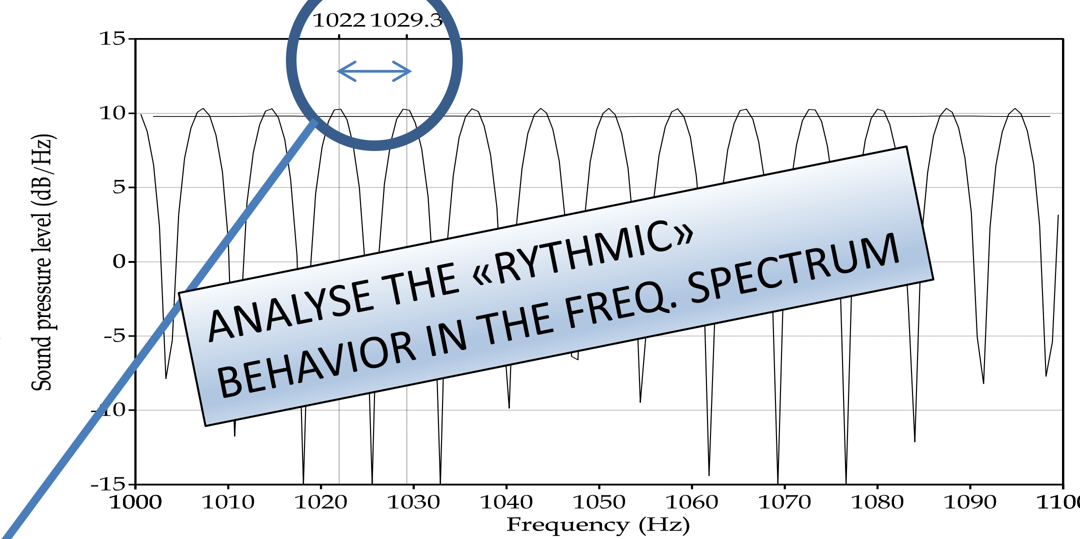
Dirac Pulse + Reflection

Δt =137 ms

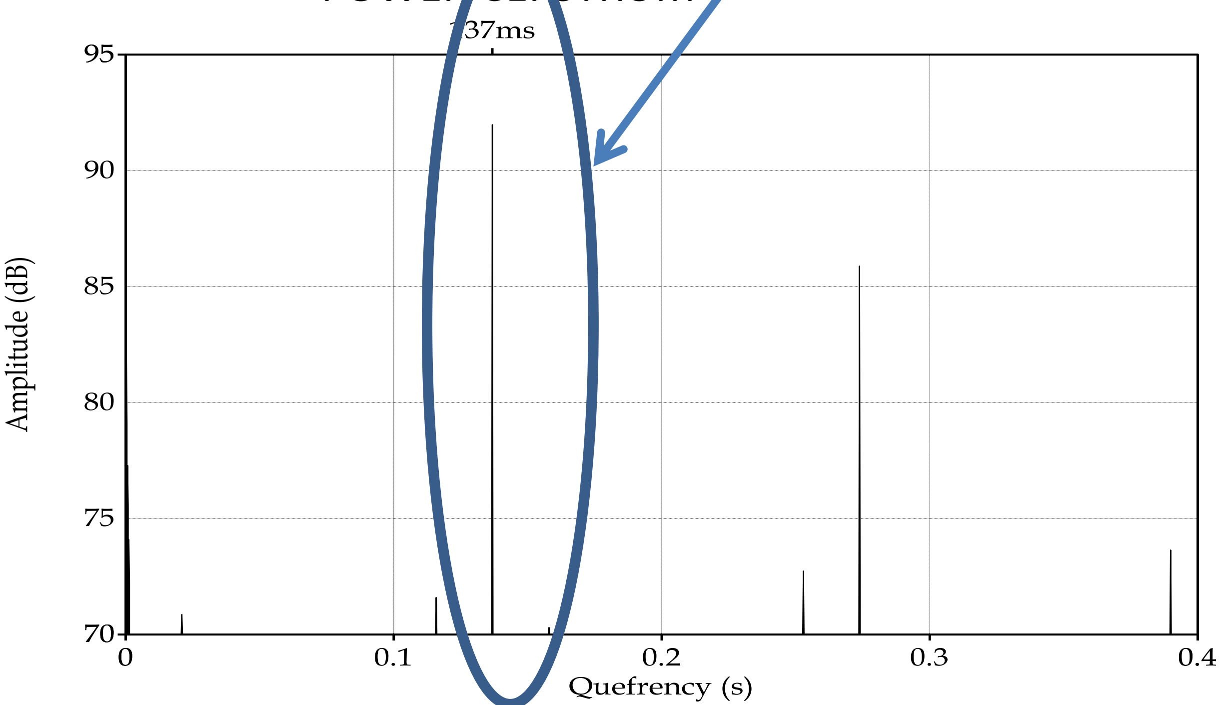


SPECTRUM CBTB= 1/Δt

CBTB=1000/137=7,3 Hz (1029.3-1022)



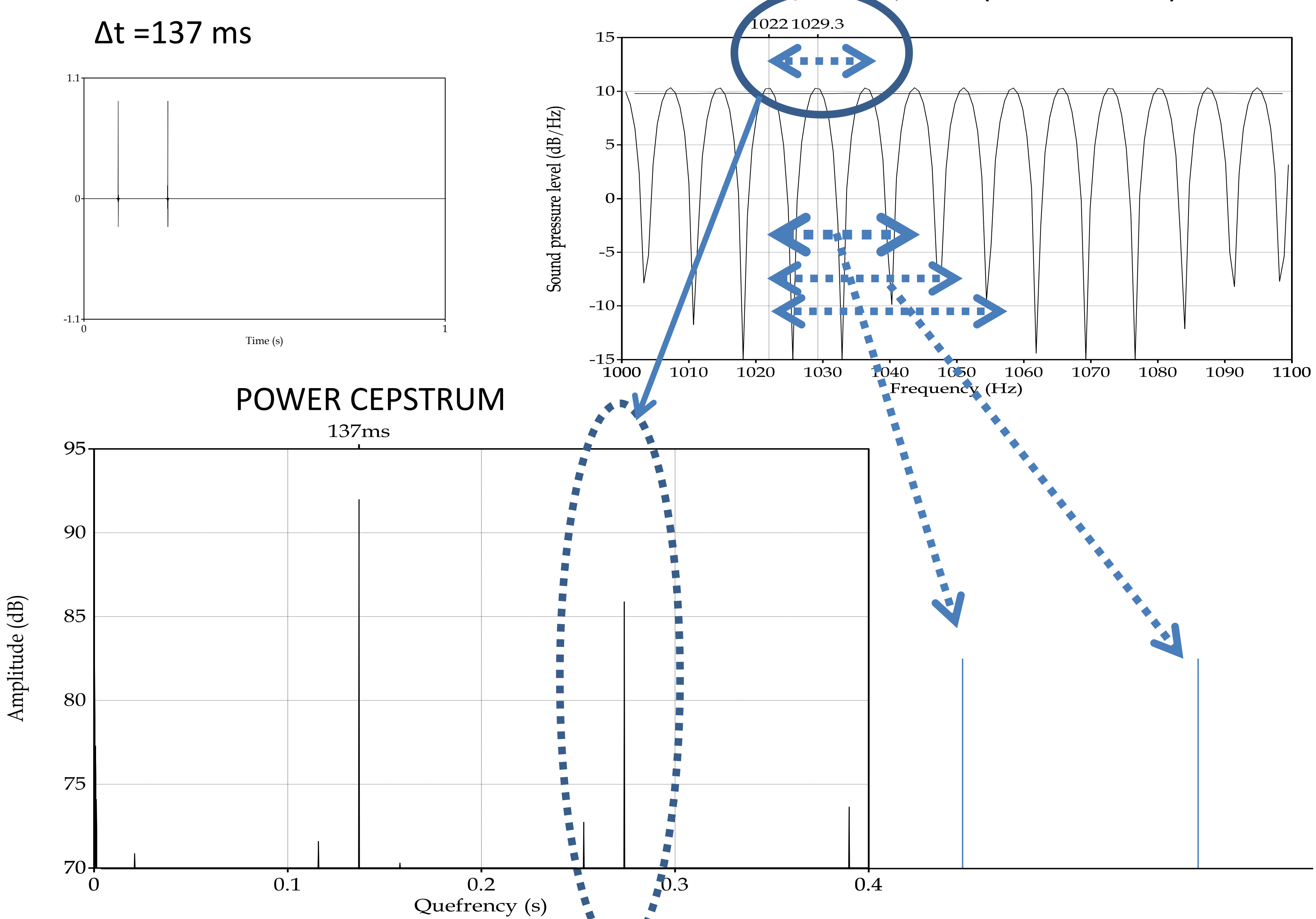
POWER CEPSTRUM



Dirac + reflection

SPECTRUM CBTB= 1/Δt
CBTB=1000/137=7,3 Hz (1029.3-1022)

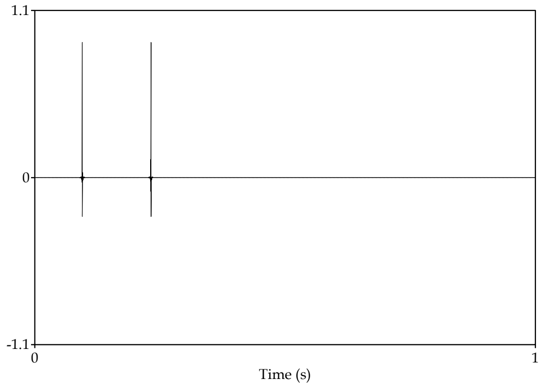
Δt =137 ms



POWER CEPSTRUM

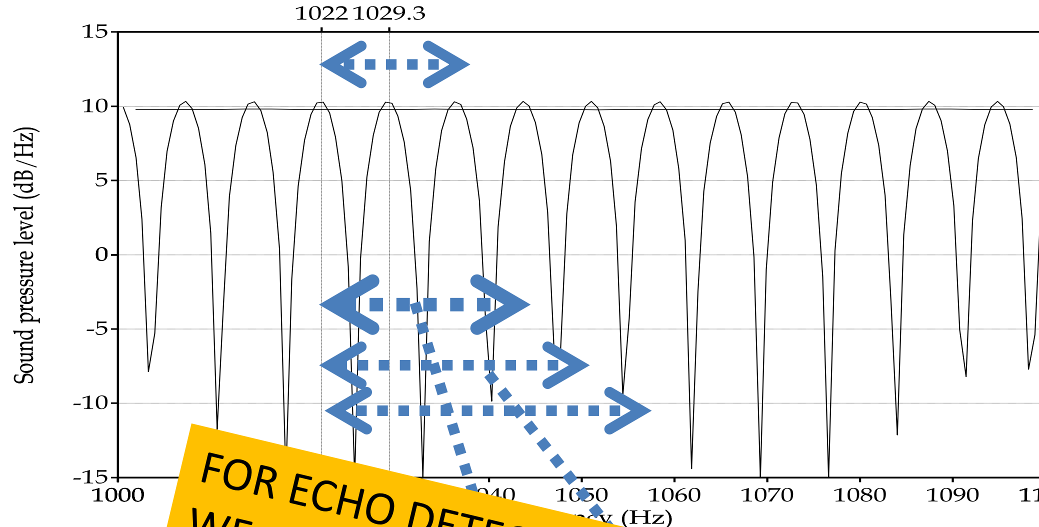
Dirac + reflection

Δt =137 ms

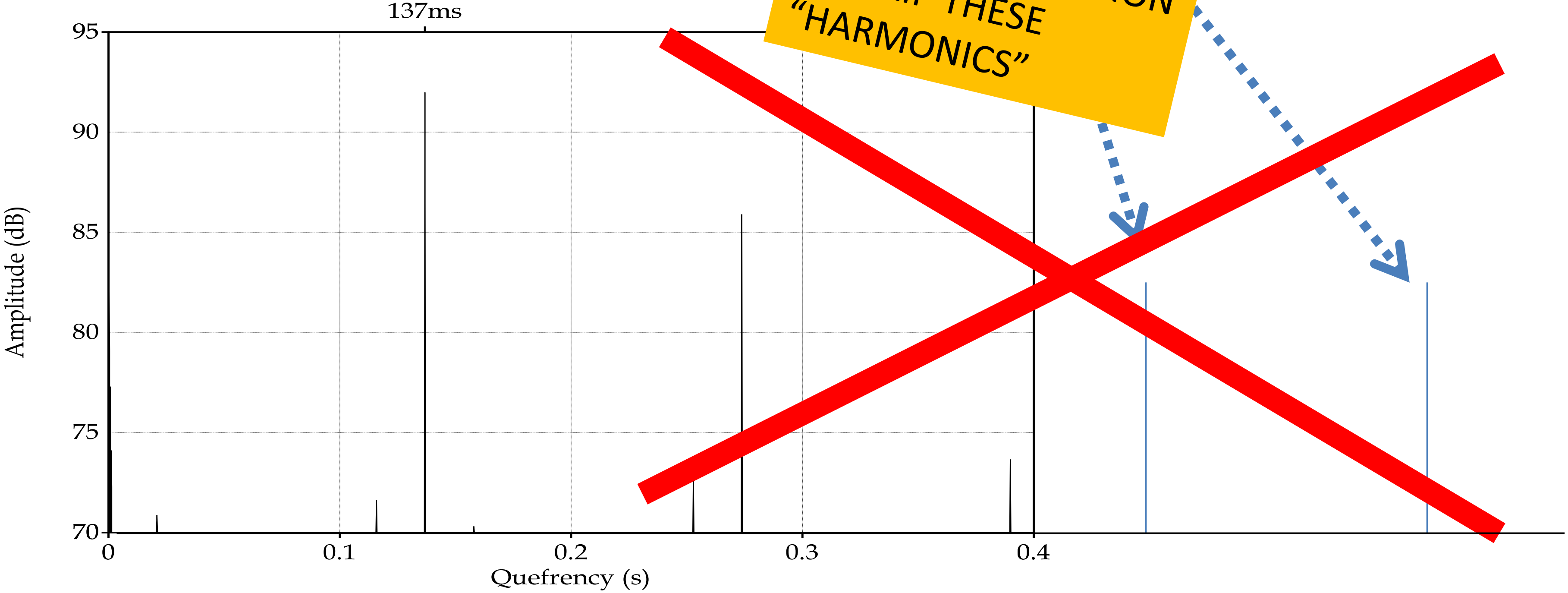


SPECTRUM CBTB= 1/Δt

CBTB=1000/137=7,3 Hz (1029.3-1022)

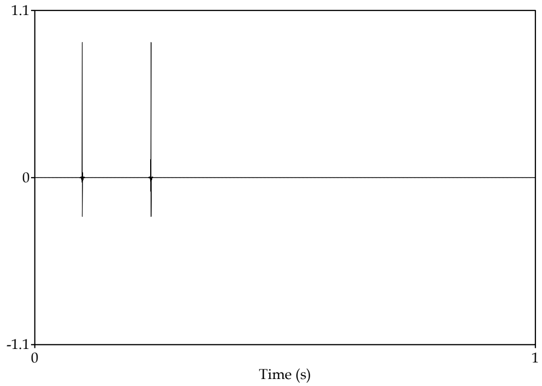


POWER CEPSTRUM



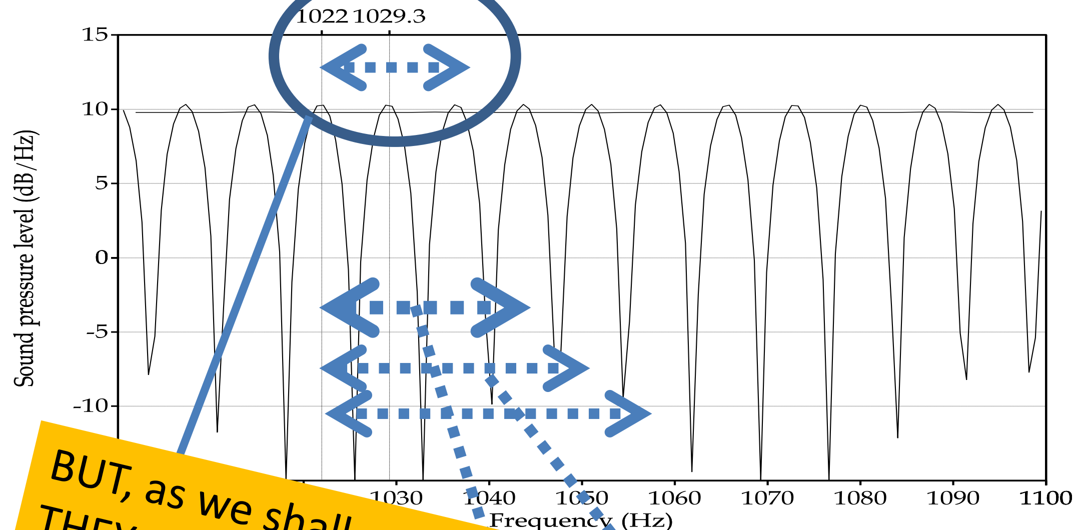
Dirac + reflection

Δt =137 ms



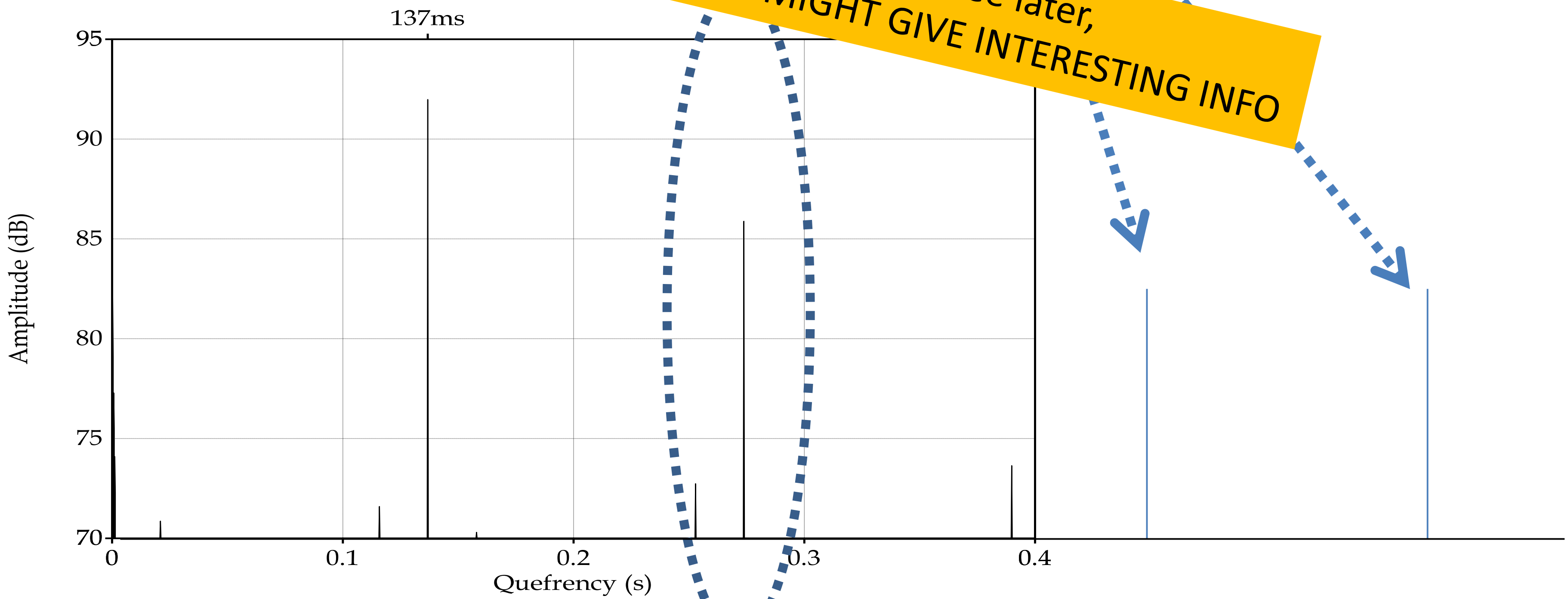
SPECTRUM CBTB= 1/Δt

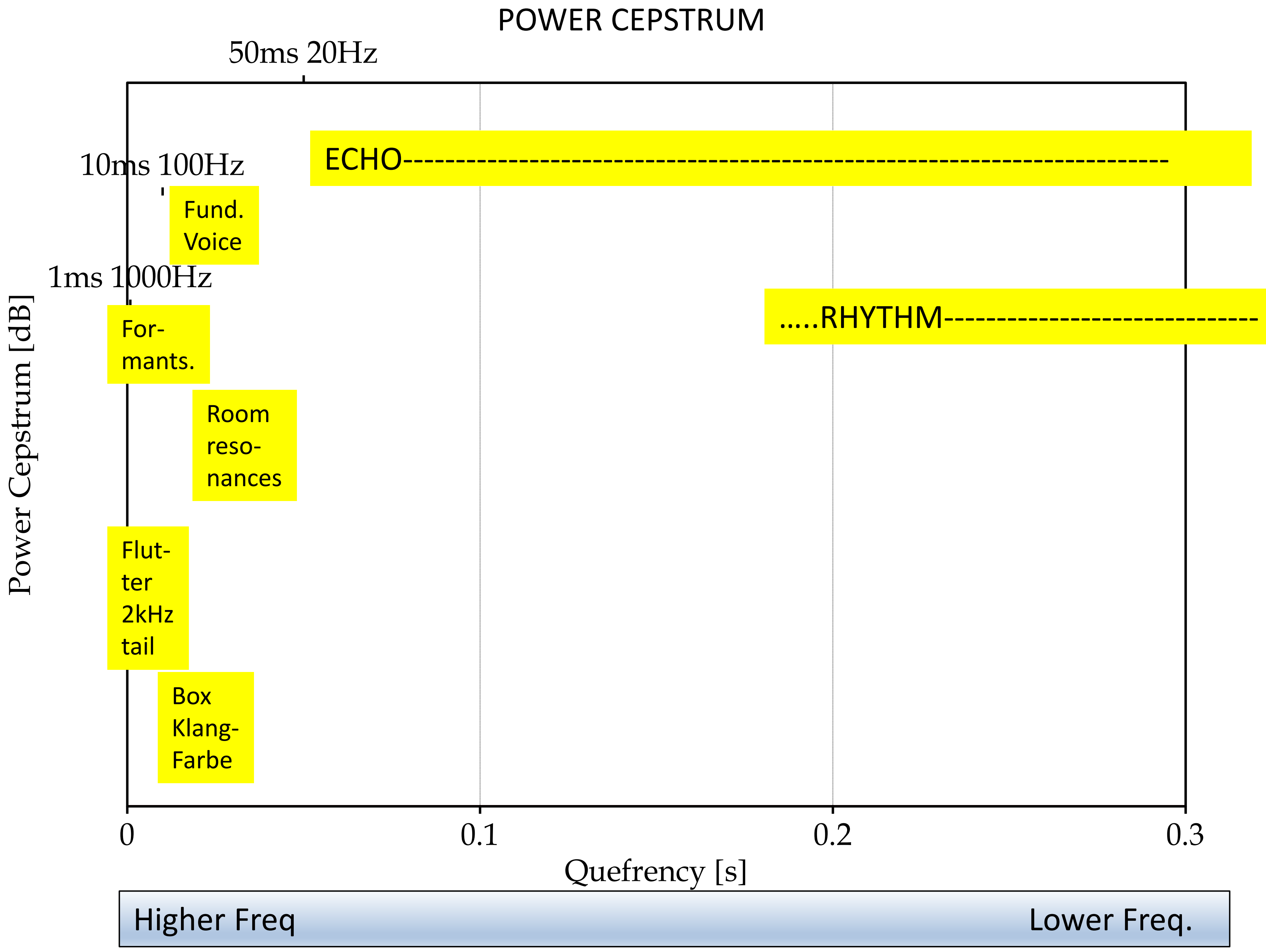
CBTB=1000/137=7,3 Hz (1029.3-1022)



BUT, as we shall see later,
THEY MIGHT GIVE INTERESTING INFO

POWER CEPSTRUM




POWER CEPSTRUM
50ms 20Hz
10ms 100Hz
1ms 1000Hz
ECHO--
Fund. Voice
.....RHYTHM----------------------------
For-mants.
Room reso-nances
Flut-ter 2kHz tail
Box Klang-Farbe
Power Cepstrum [dB]
0
0.1
0.2
0.3
Quefrency [s]
Higher Freq
Lower Freq.

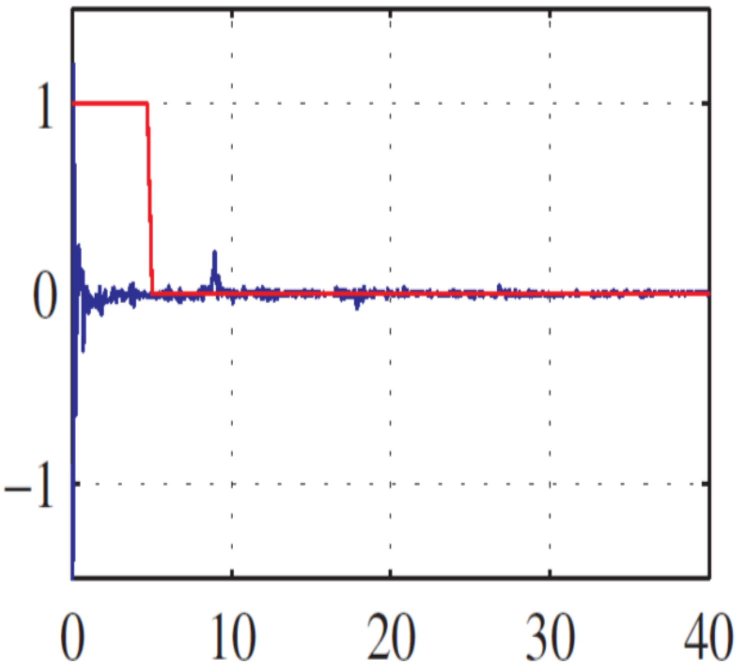
LIFTERING =

FILTERING in the Quefrency domain

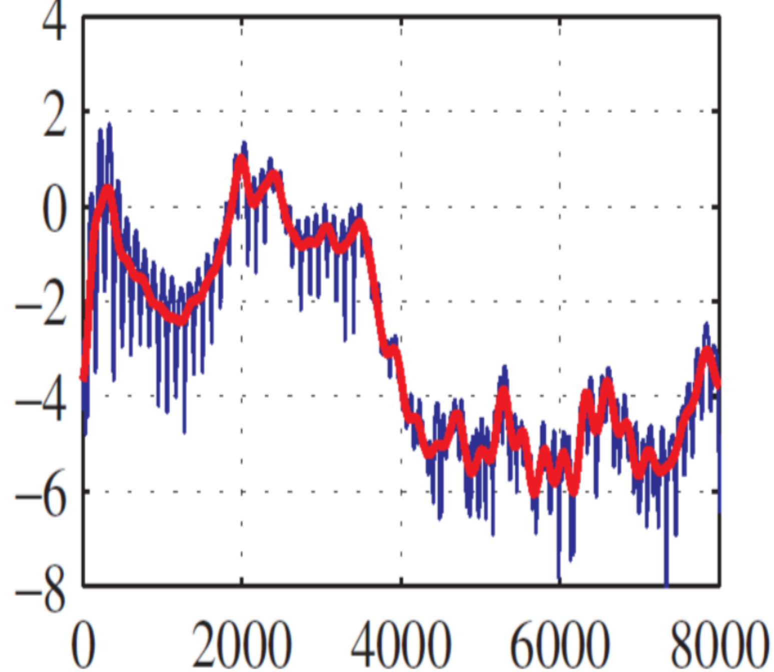
Looking at just the region of quefrencies we want

Liftering in the cepstral domain

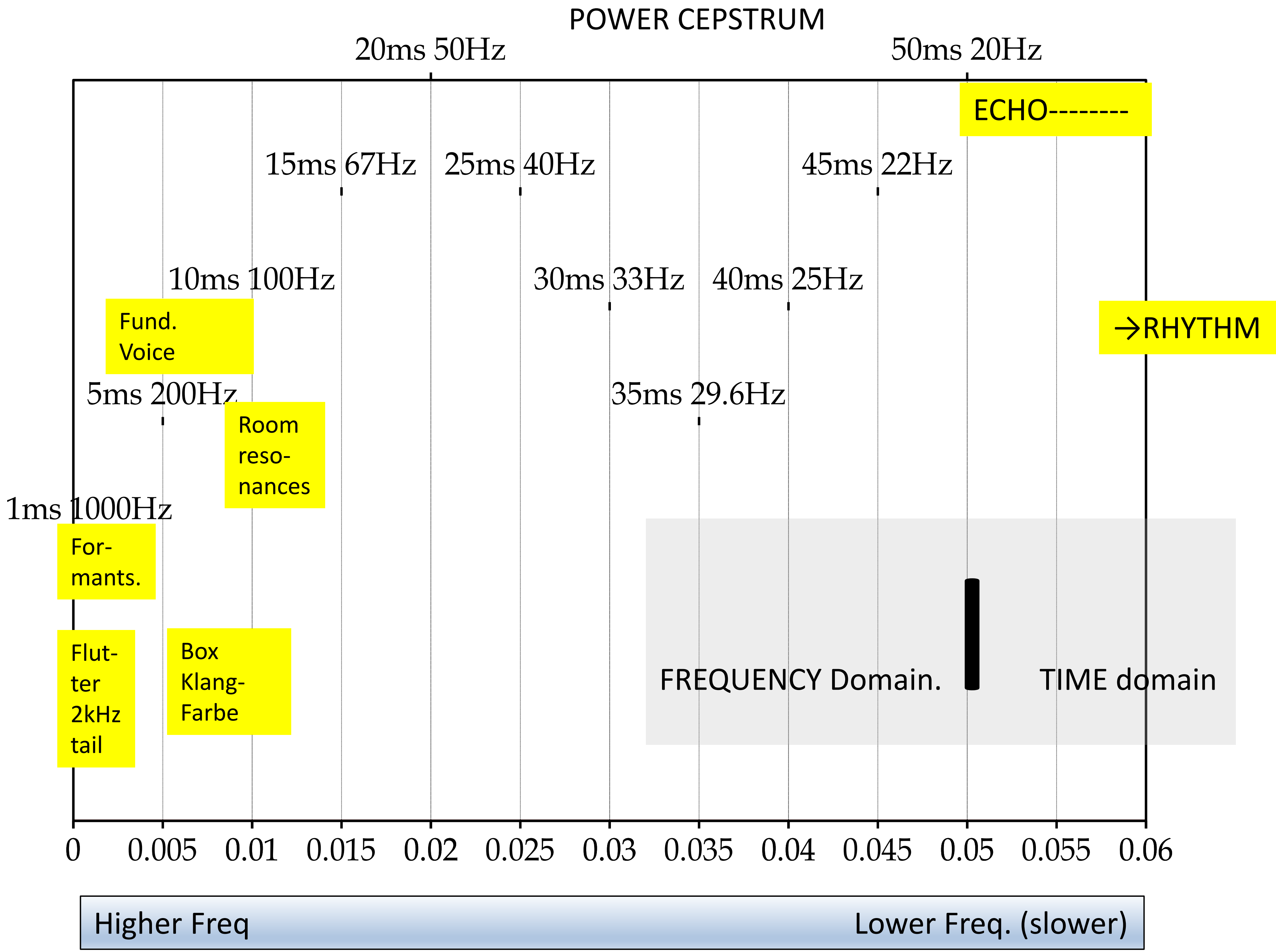
The red “Lifter” removes the things happening fast (low in [s]), means smoothening of the spectrum



Time (or Samples) ->



Frequency [Hz]->


POWER CEPSTRUM
20ms 50Hz
50ms 20Hz
ECHO--------
15ms 67Hz
25ms 40Hz
45ms 22Hz
10ms 100Hz
30ms 33Hz
40ms 25Hz
Fund. Voice
→RHYTHM
5ms 200Hz
35ms 29.6Hz
Room reso-nances
1ms 1000Hz
For-mants.
Flut-ter 2kHz tail
Box Klang-Farbe
FREQUENCY Domain.
TIME domain
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
0.045
0.05
0.055
0.06
Higher Freq
Lower Freq. (slower)

CEPSTRUM and AUTOCORRELATION
to detect reflections (echoes)

Original, waveform

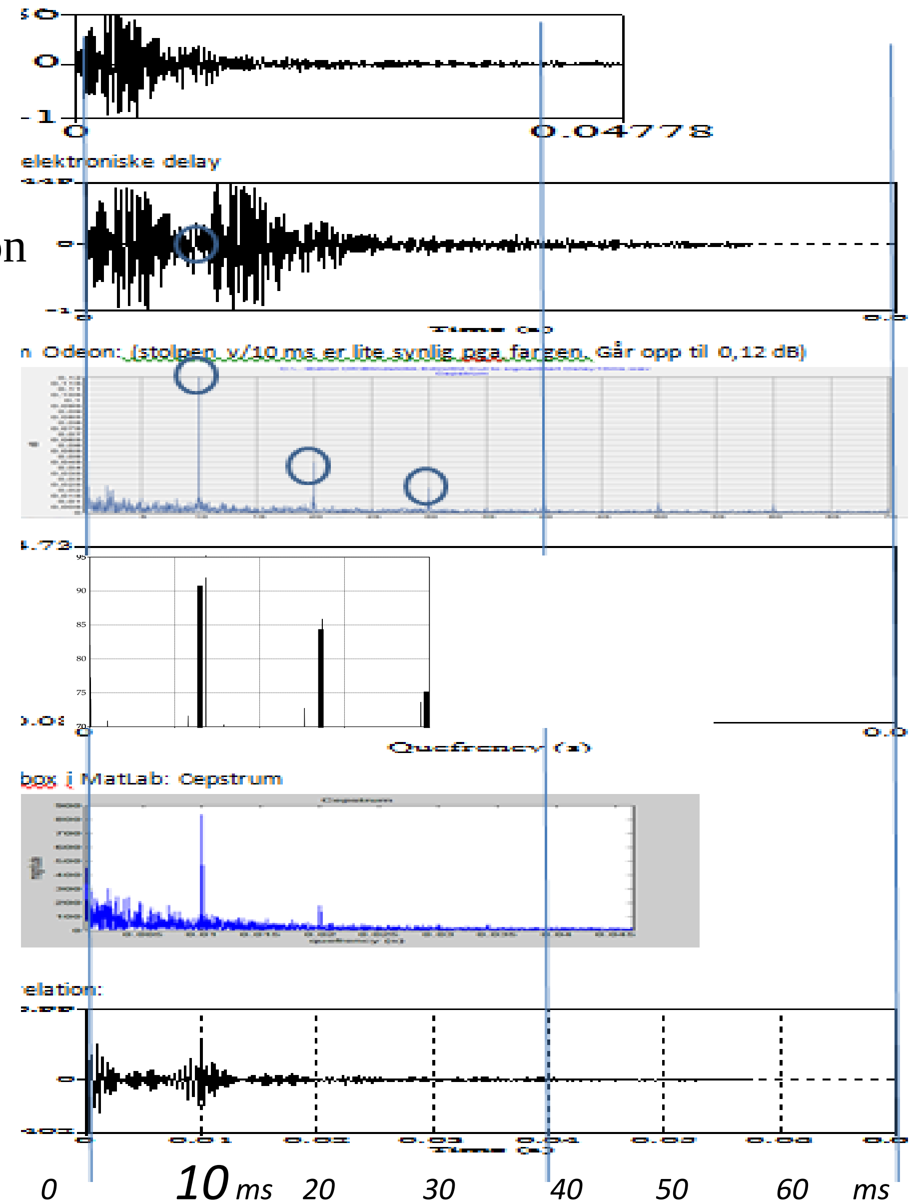
Original + 10 ms delayed reflection

Cepstrum (Odeon)

Cepstrum (Praat)

Cepstrum MIR-toolbox (MatLab)

Autocorrelation (Praat)



UiO : Department of Musicology
University of Oslo

SAM PHILLIPS´ SLAP BACK ECHO; *LUCKILY IN MONO*

Tor Halmrast
Head of Acoustics, Statsbygg
Assoc. Prof., Univ. Oslo/Musicology
Norwegian Ac. Of Music

Art of Record Production, St.holm 1-3 dec 2017

Sam Phillips
The Memphis Recording Service.
SUN Studios

"…150 to 200ms with just one repeat at almost the same amplitude as the original signal".

"..60 -120 milliseconds?"

"At 60 Milliseconds, you don't hear much doubling,
but everything sounds a little thicker and fatter"
"At 120 Milliseconds you are REALLY starting to hear
the distinct doubling of the sound".

«..75 to 250 milliseconds), with little or no feedback»

Sam Philips: **Single TAPE-ECHO**
(later: RCA-simulation: used a hallway as Rev.Chamber)

"…2 Ampex 350 recorders"

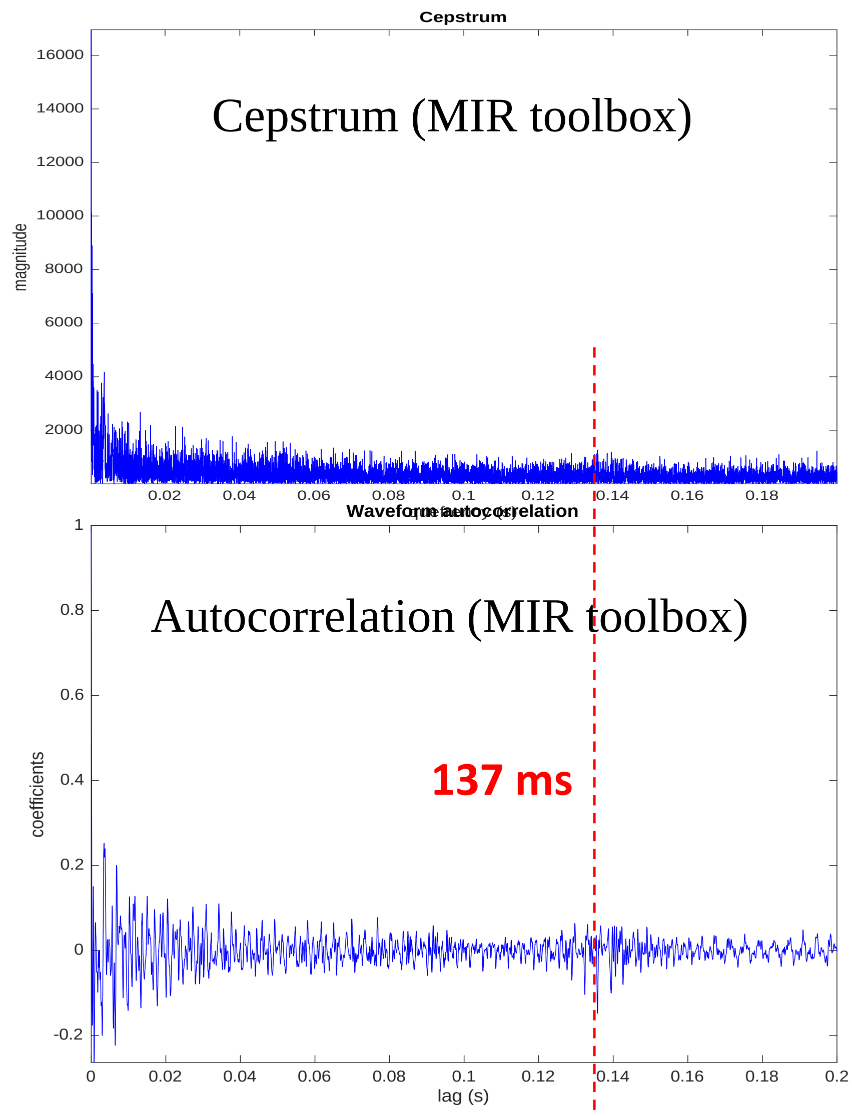
«The physical space between heads, the speed of the tape, and the chosen volume being the main controlling factors»

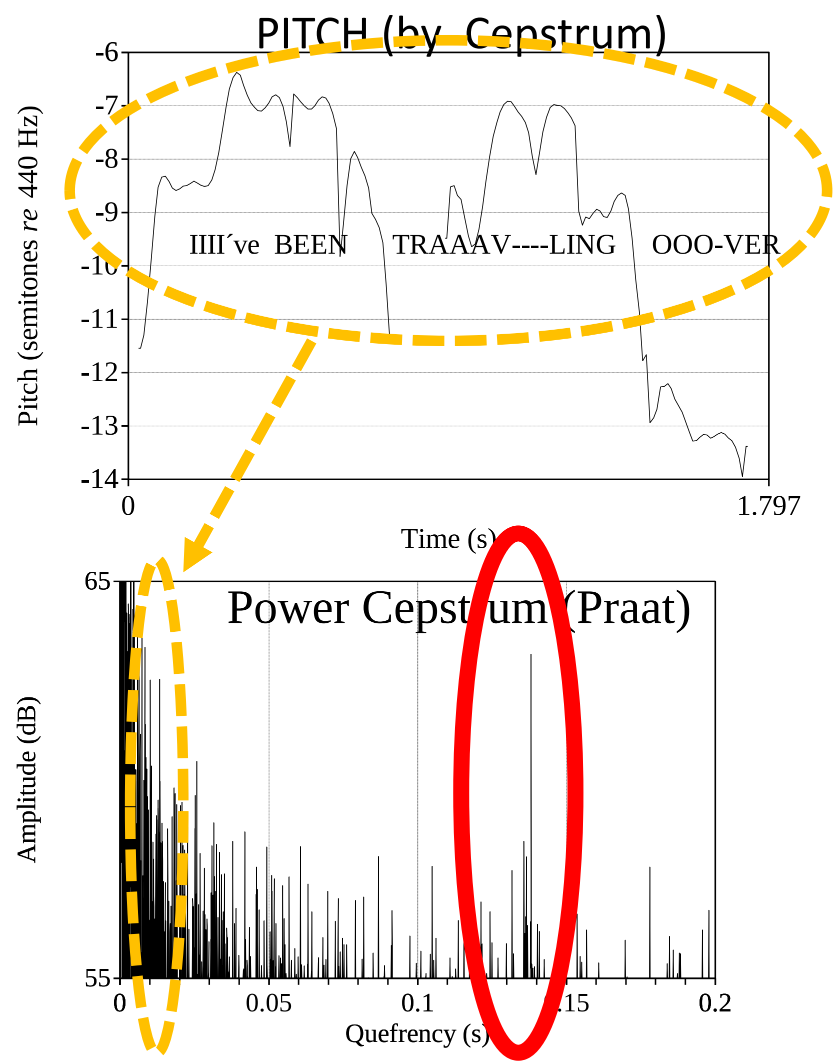
"530ms will be too long, **163 ms** seemed to sound about right.
…a little either side of a 1/16 note might be what it really comes down to".

"..Scotty Moore's guitar-work on "That's All Right"

Tryin´to get to you

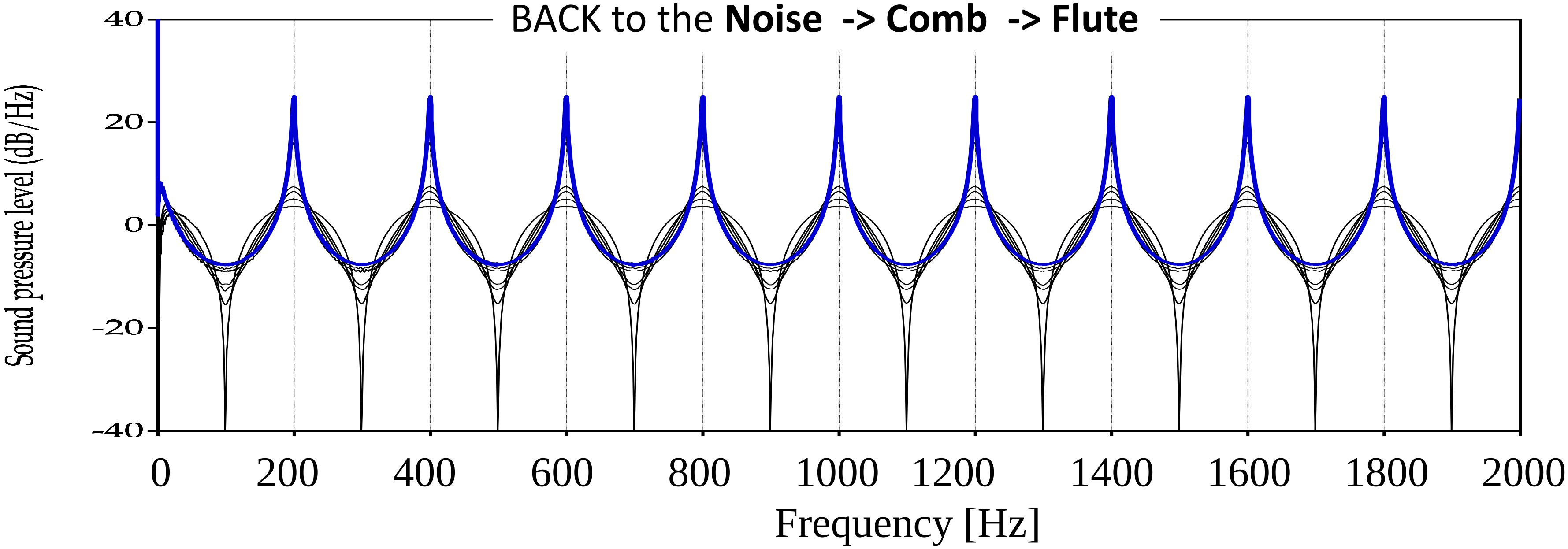
Sun Studios

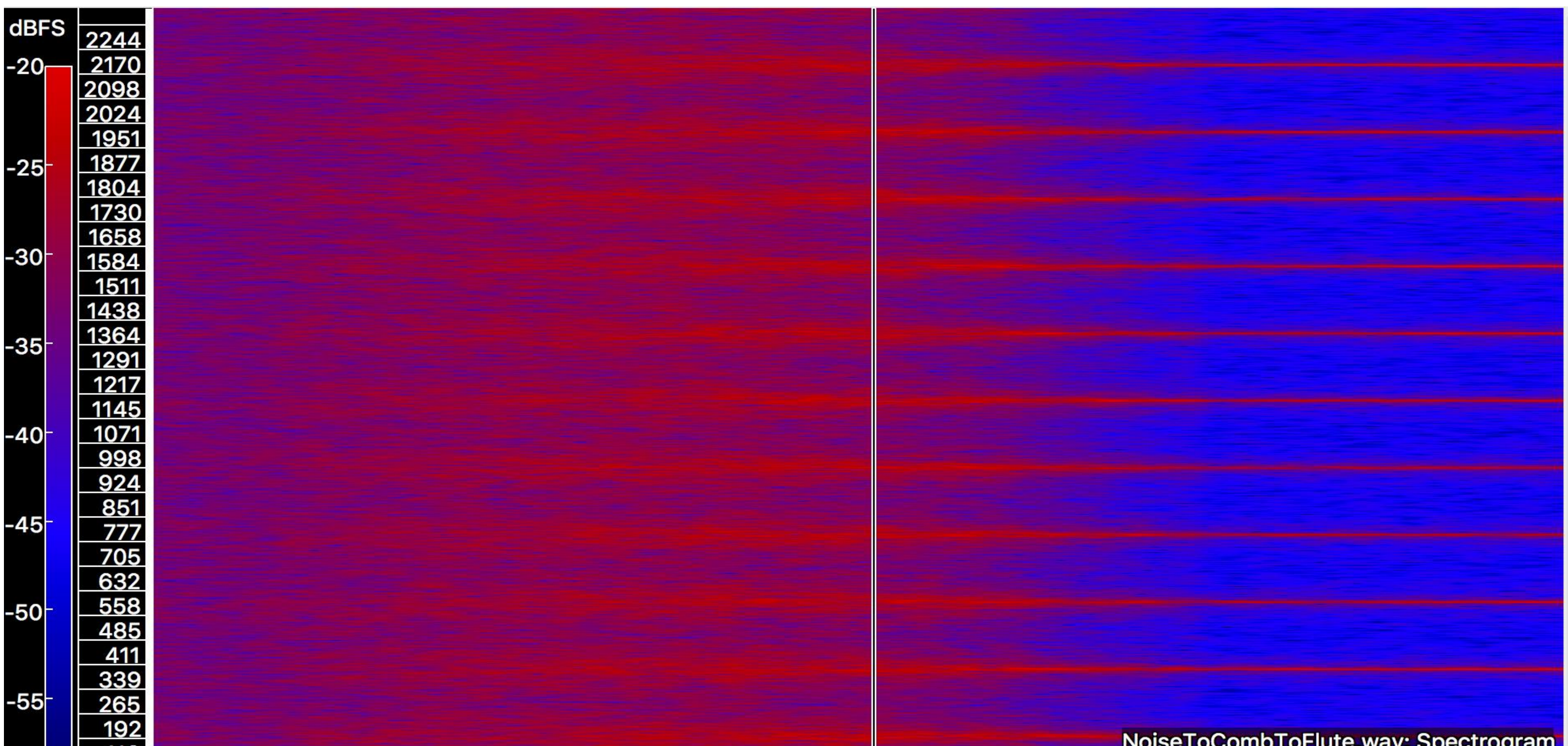




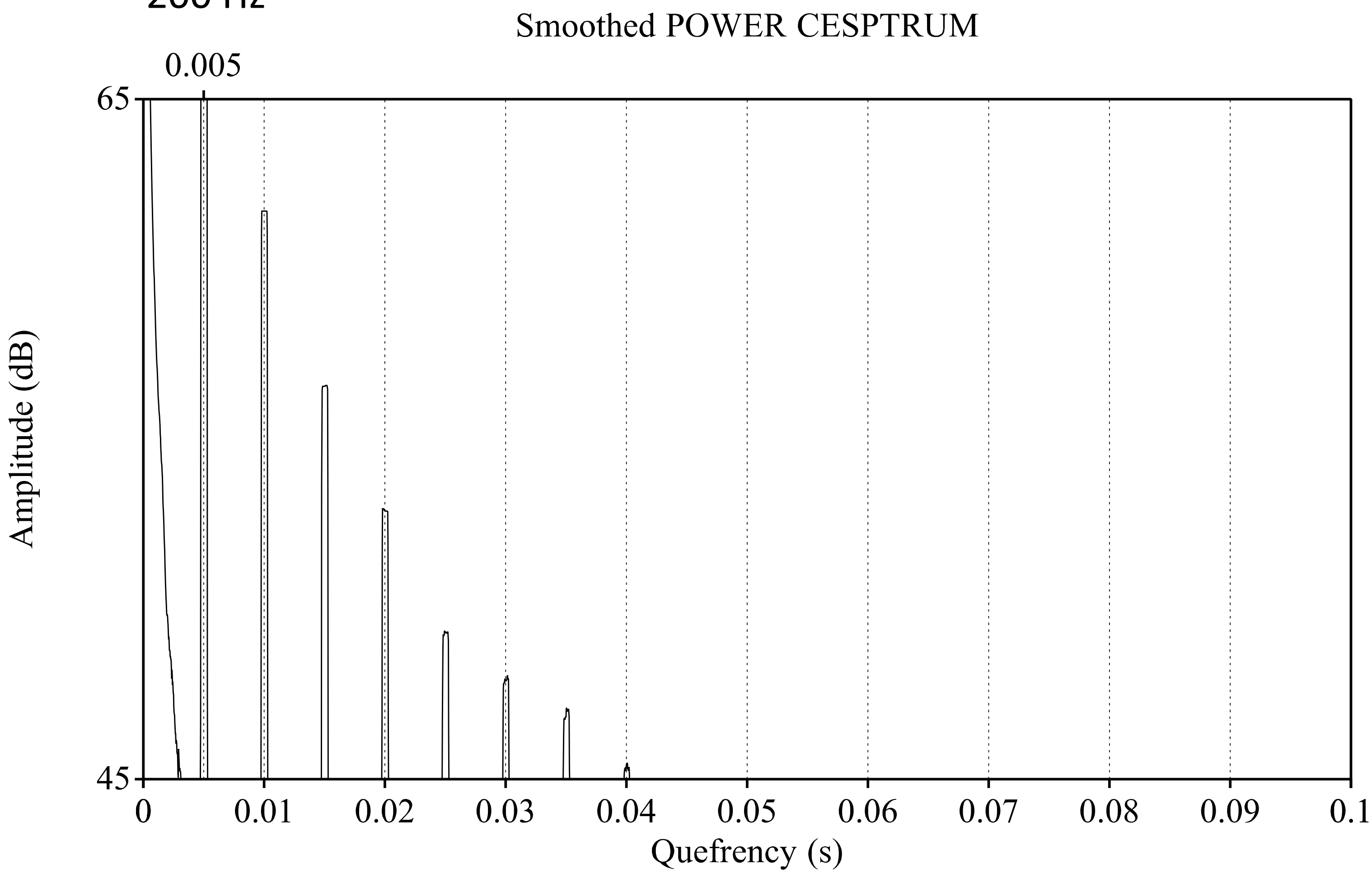
137 ms

BACK to the **Noise -> Comb -> Flute**

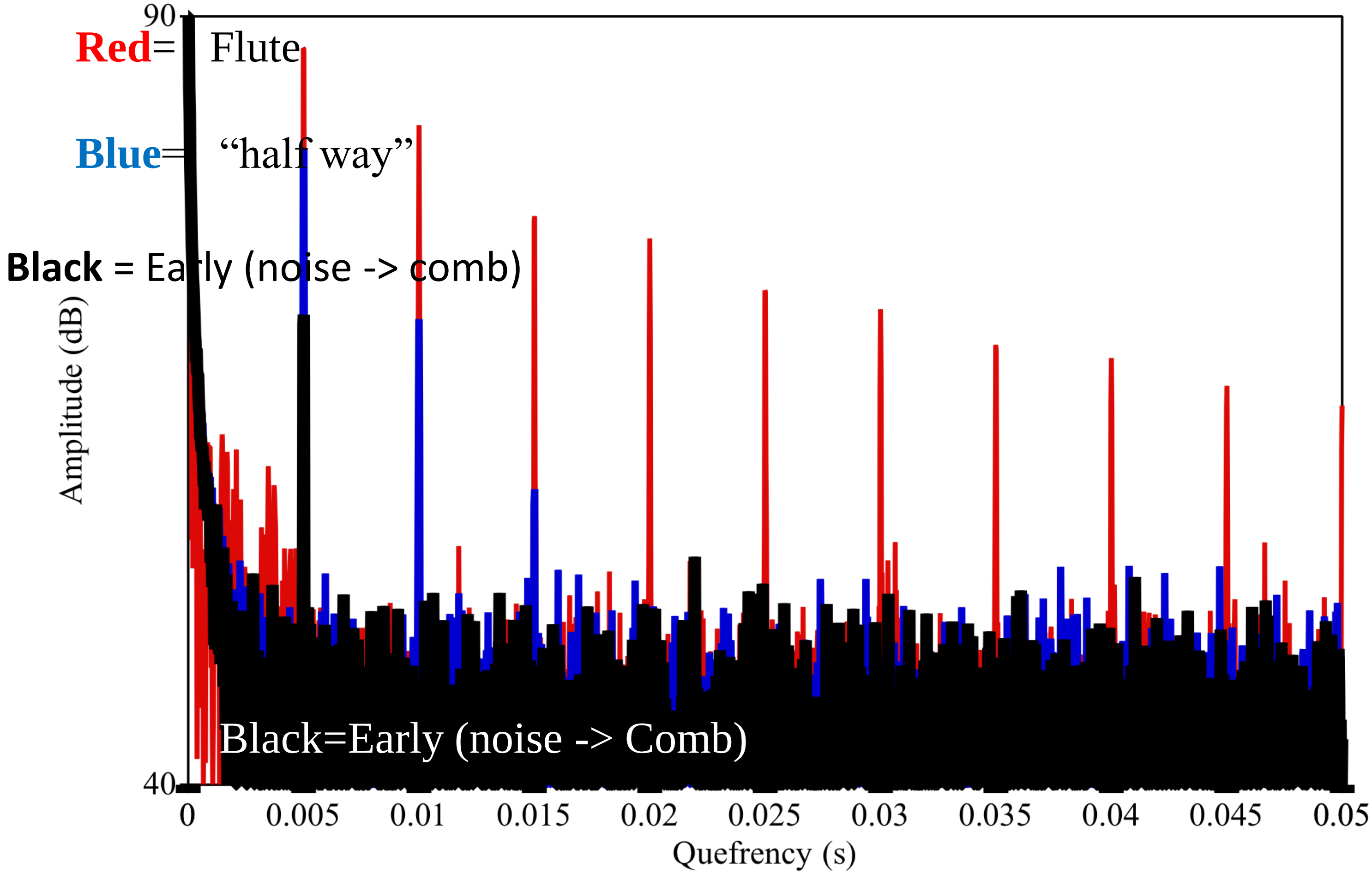




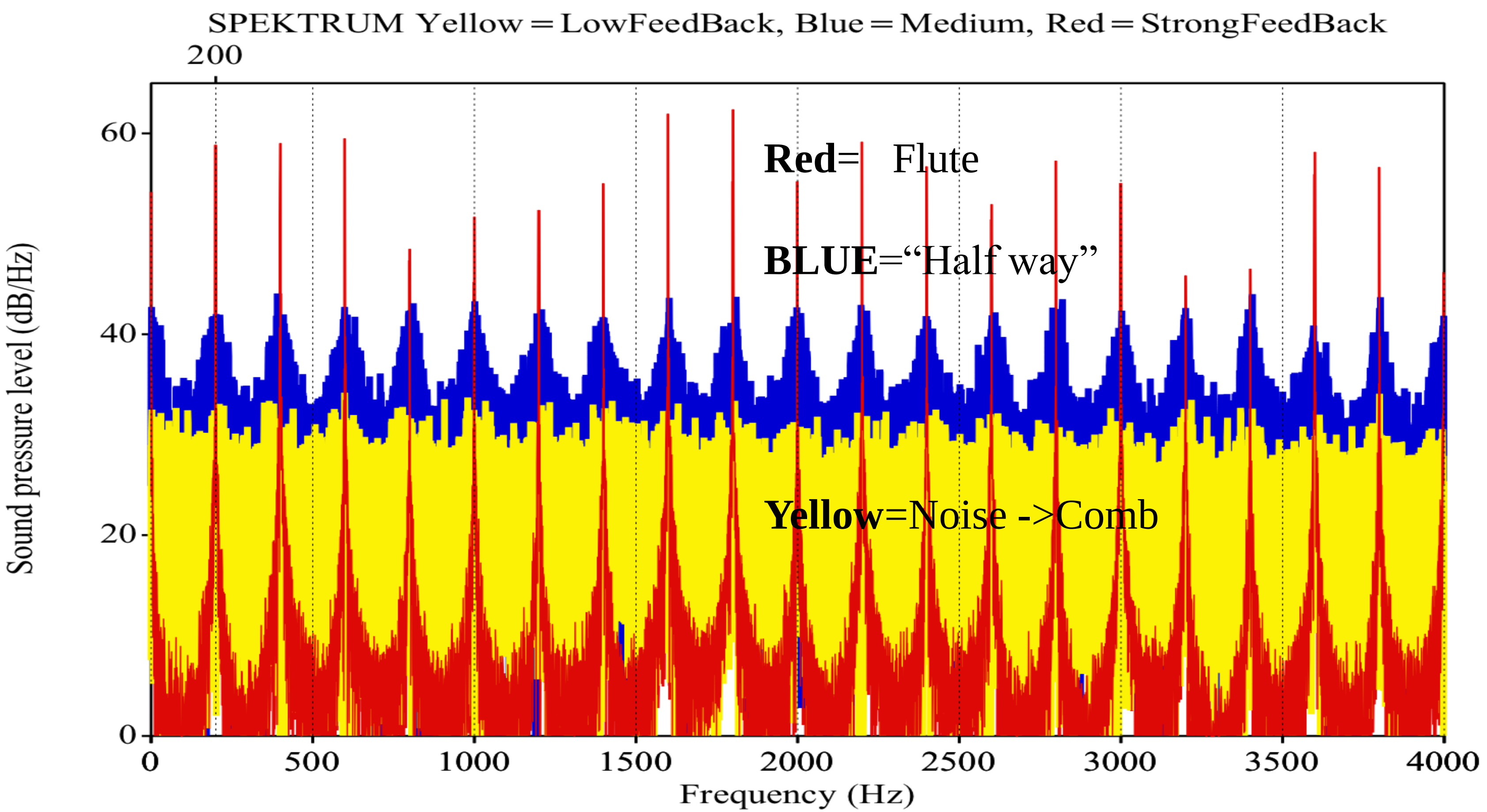
1/005s =
200 Hz


Smoothed POWER CESPTRUM
0.005
65
45
Amplitude (dB)
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
Quefrency (s)

Gradual change of
Power **Cepstrum**
over time for the
Noise->Comb->Flute sound
(increasing feed-back)

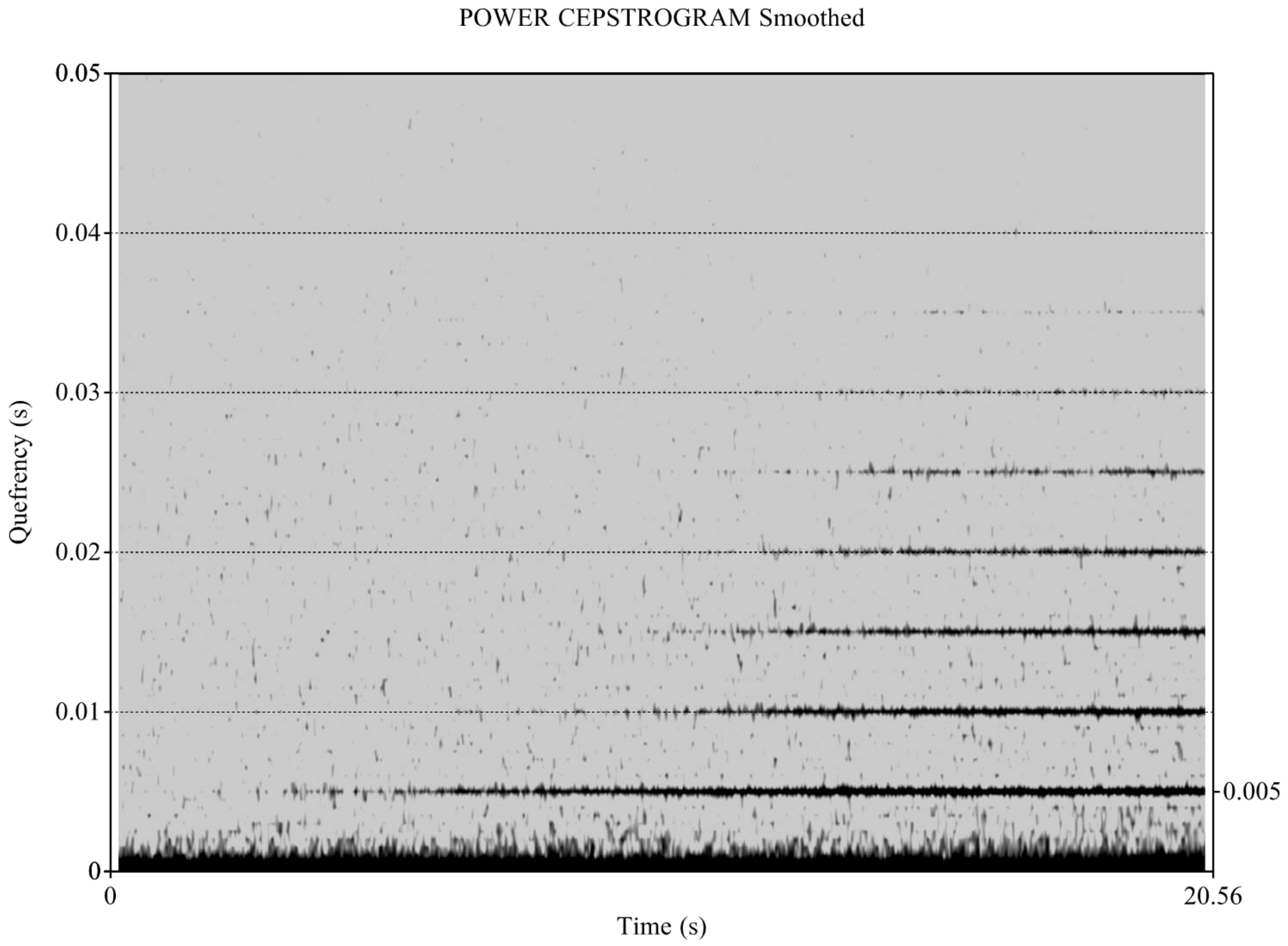


Gradual change of
SPECTRUM
over time for the
Noise->Comb->Flute sound



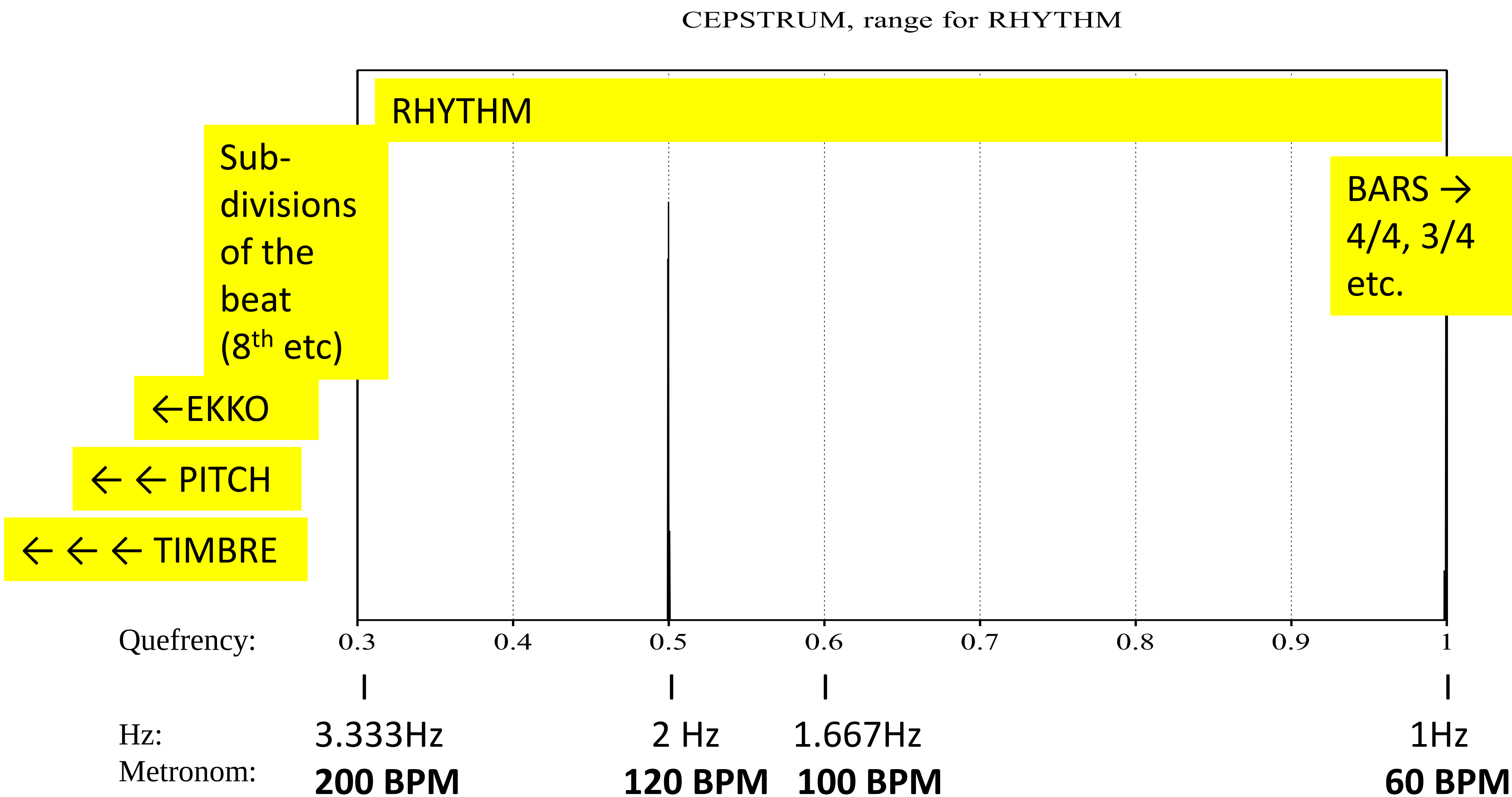
CEPSTROGRAM

NoiseToCombToFlute.wav

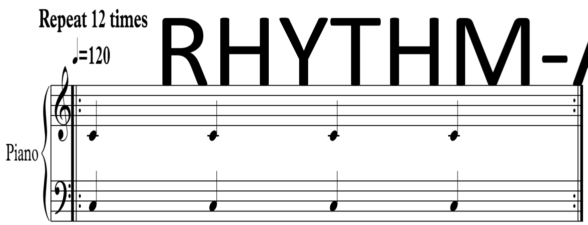


RHYTHM-ANALYSIS using CEPSTRUM?

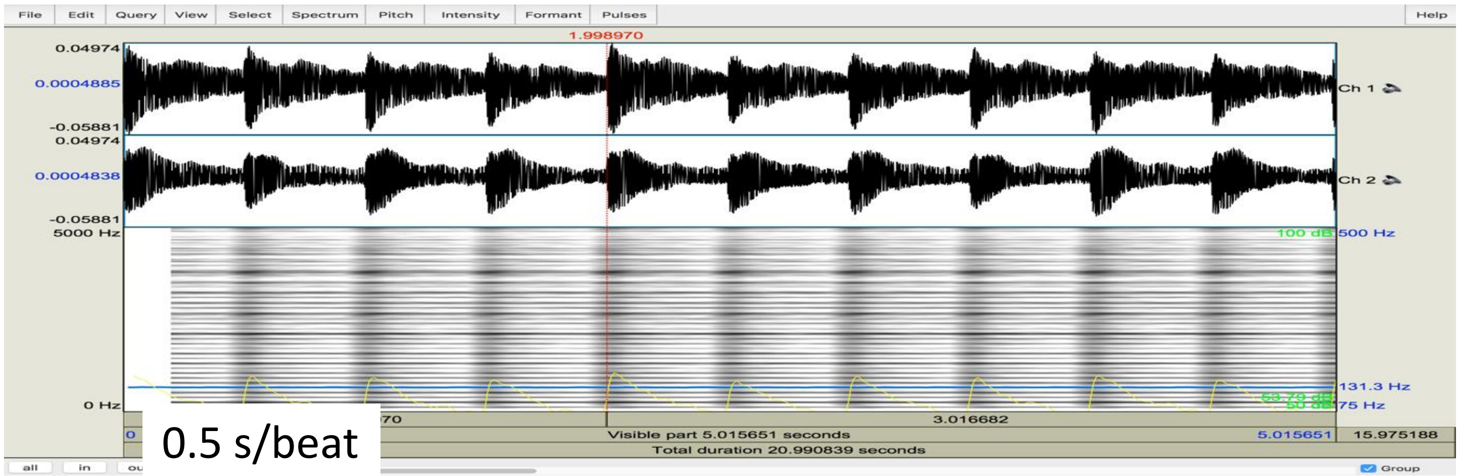
Zoom In (Liftering)



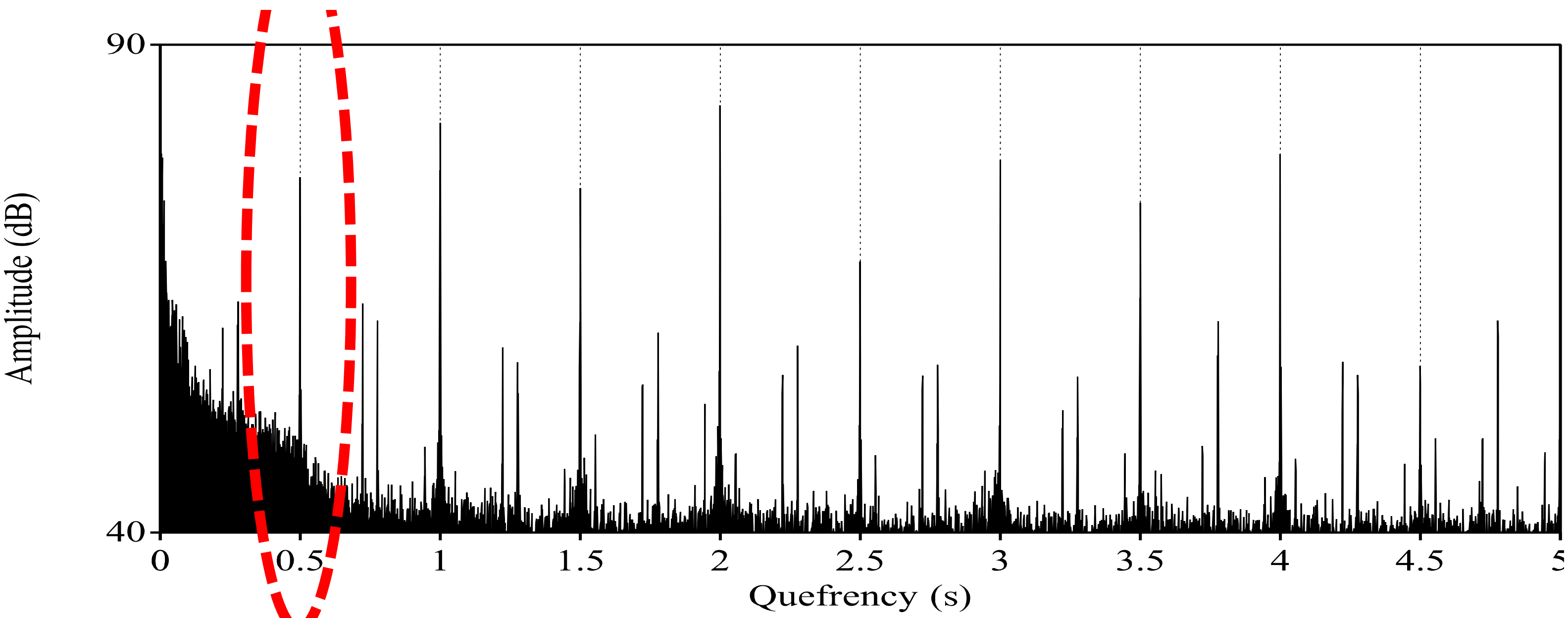
RHYTHM-ANALYSIS using CEPSTRUM?



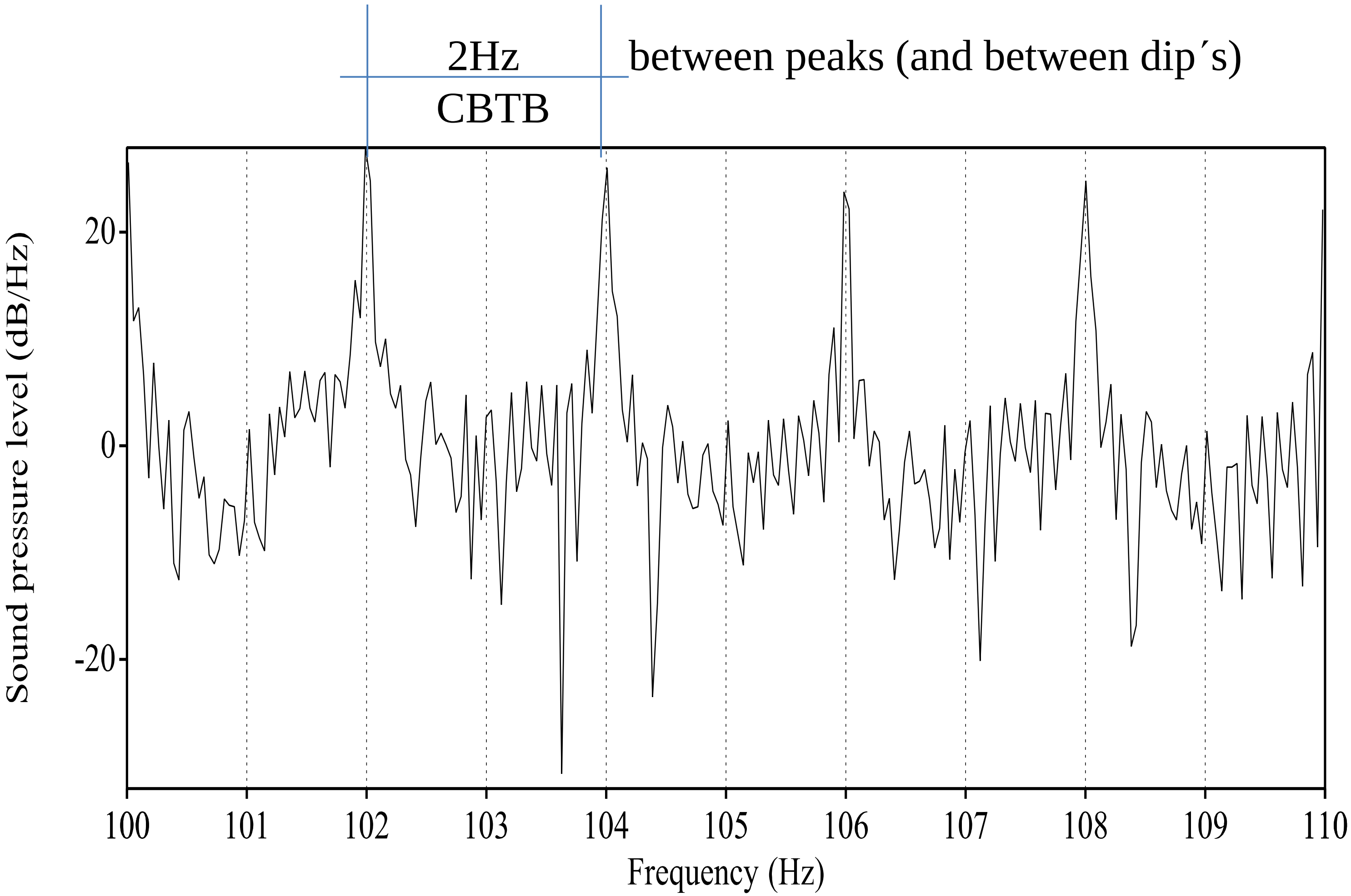
PIANO
(Sibelius)



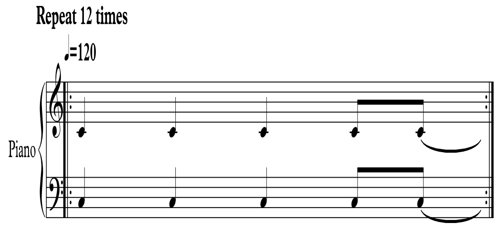
0.5 s/beat
2 Hz
120 BMP



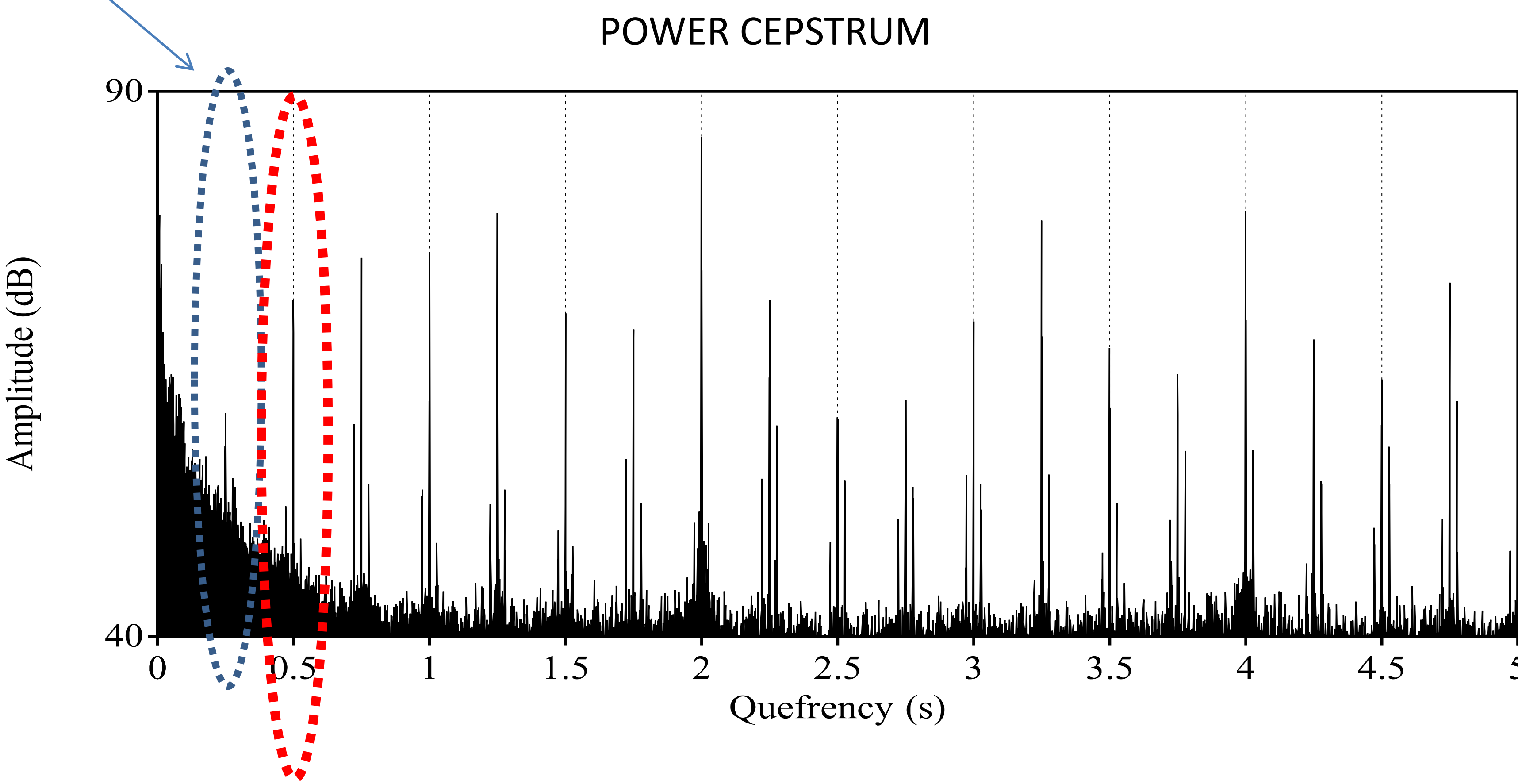
Actually, this rhythm is shown also in a ZOOM-IN of the SPECTRUM:

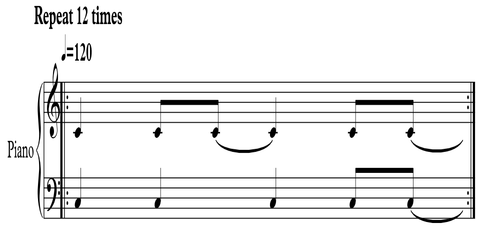


But these changes in spectrum is not perceived (<<<Critical Bandwidth)

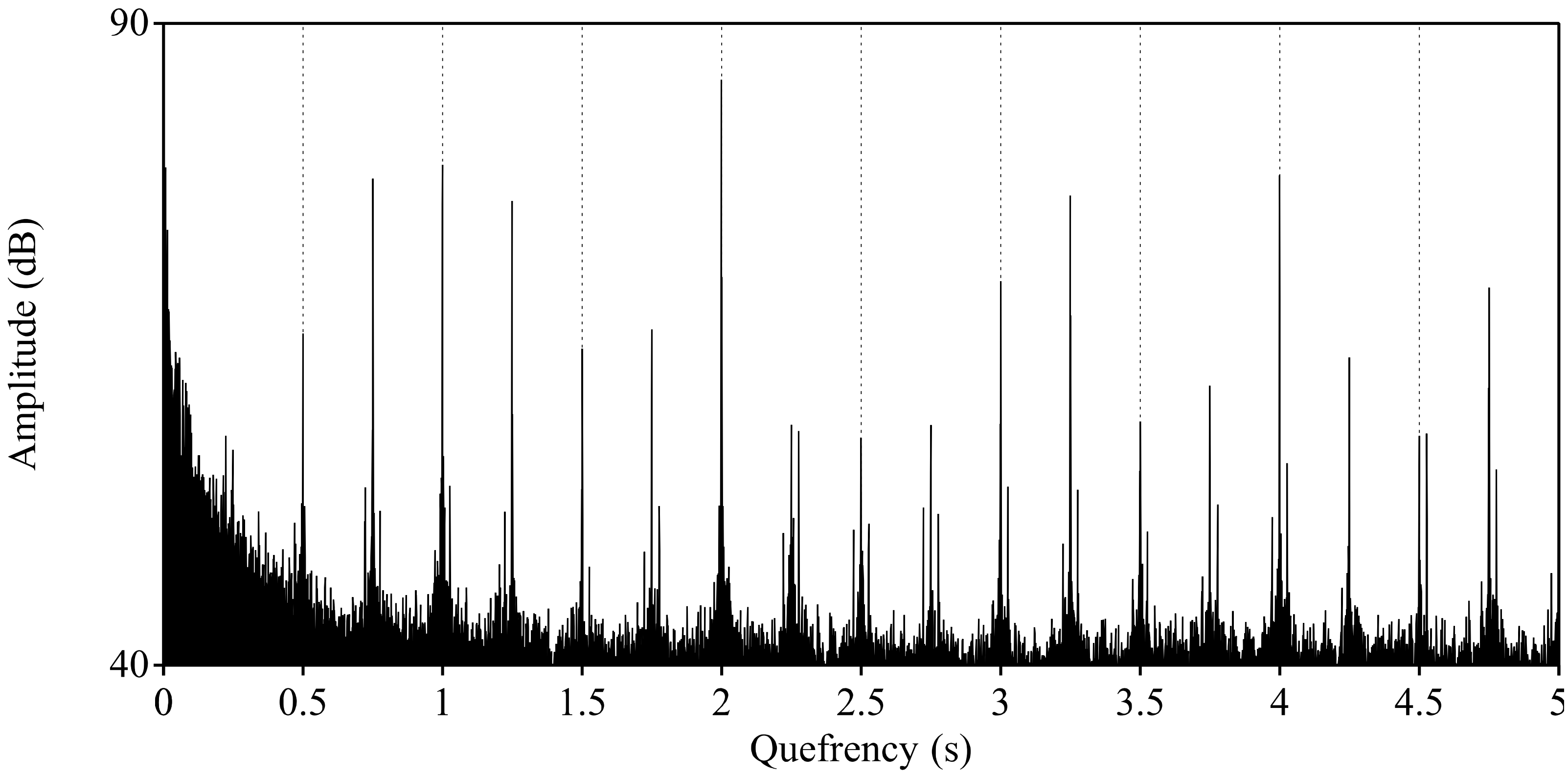


The syncopation adds a quefrency of half the beat

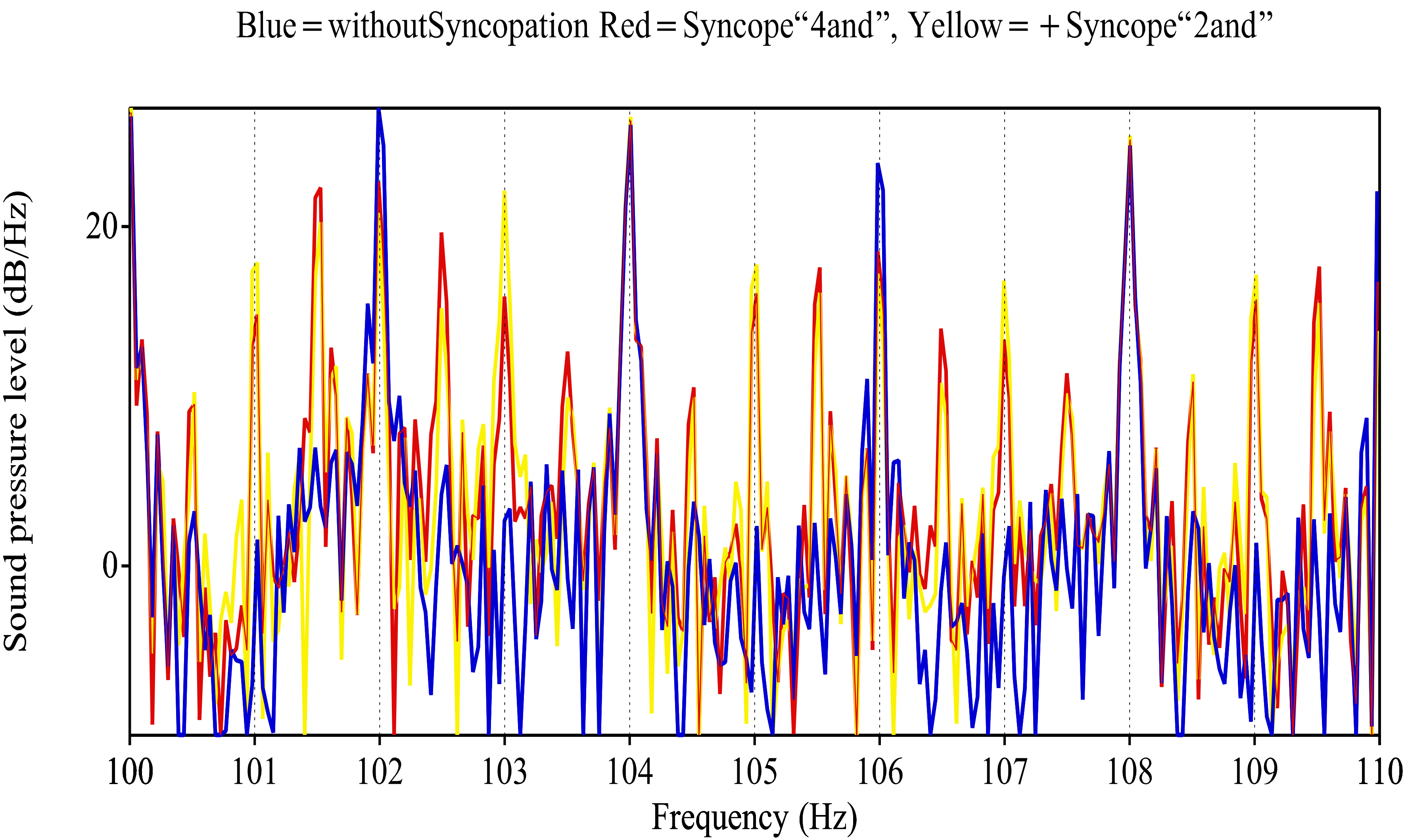




POWER CEPSTRUM



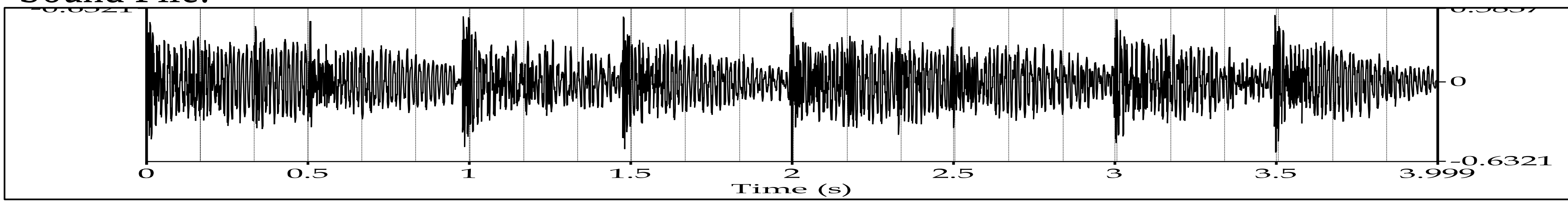
SPECTRUM (Zoom-In)



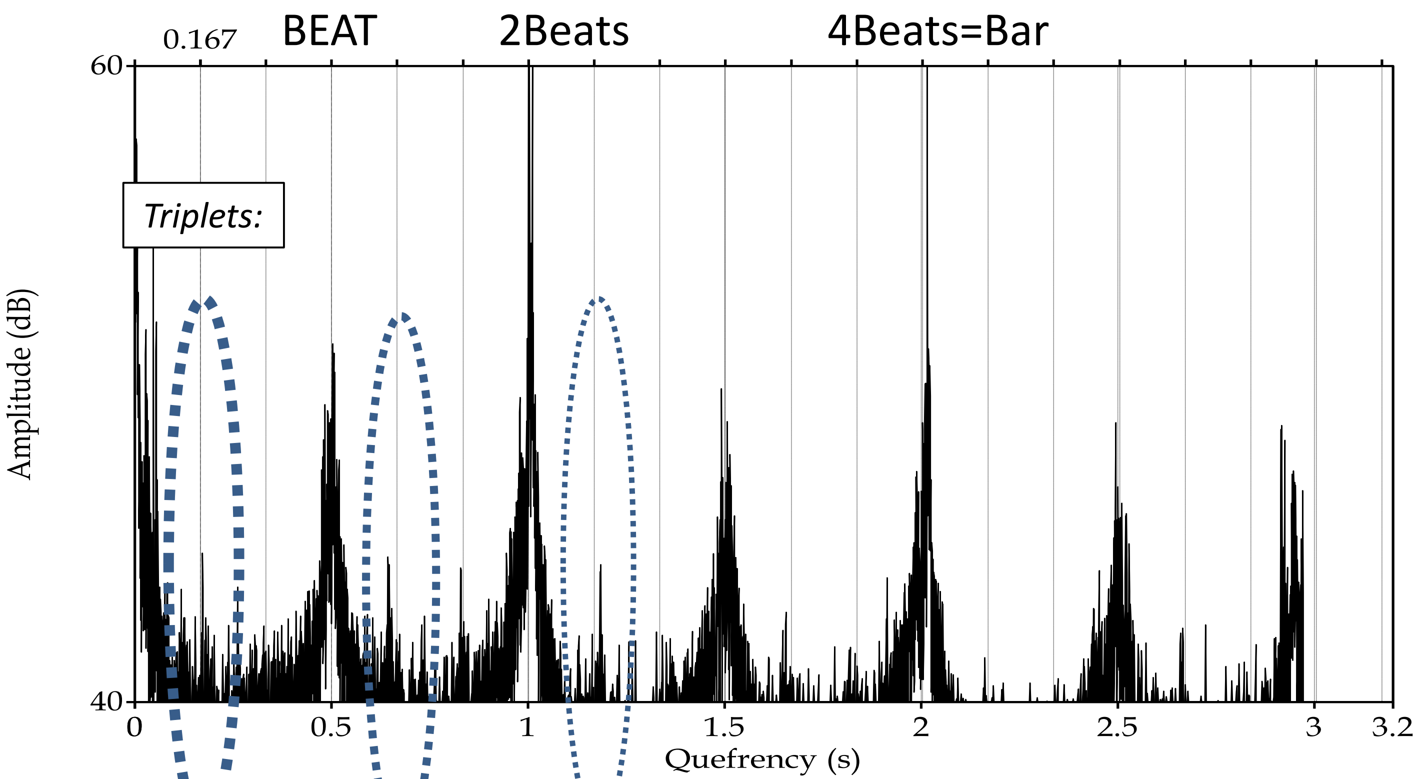
DRUMS

DRUMS_120bmpSTART_Triol:
Sound File:

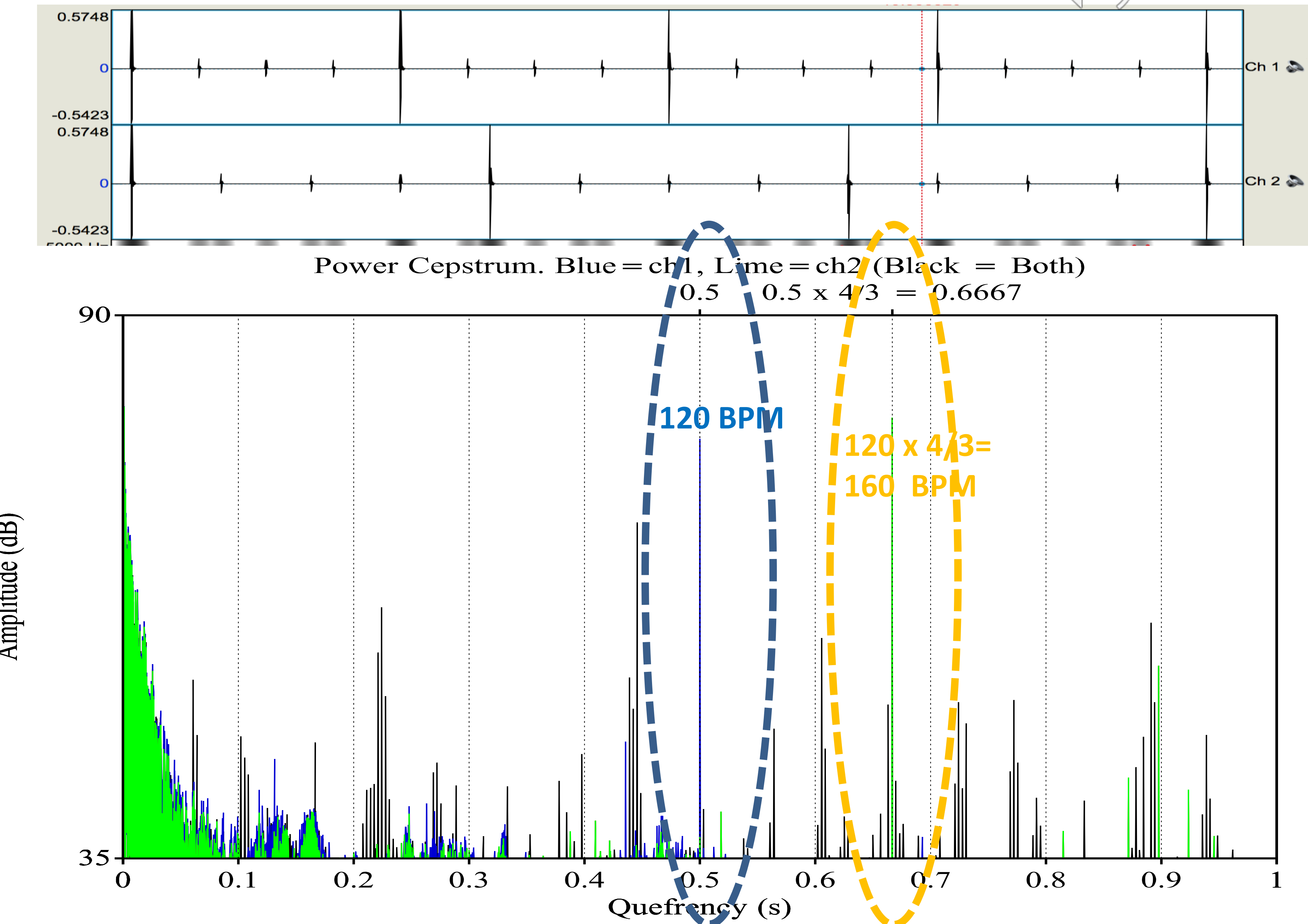
(Sound File: «Drums» from NM Stillstand)

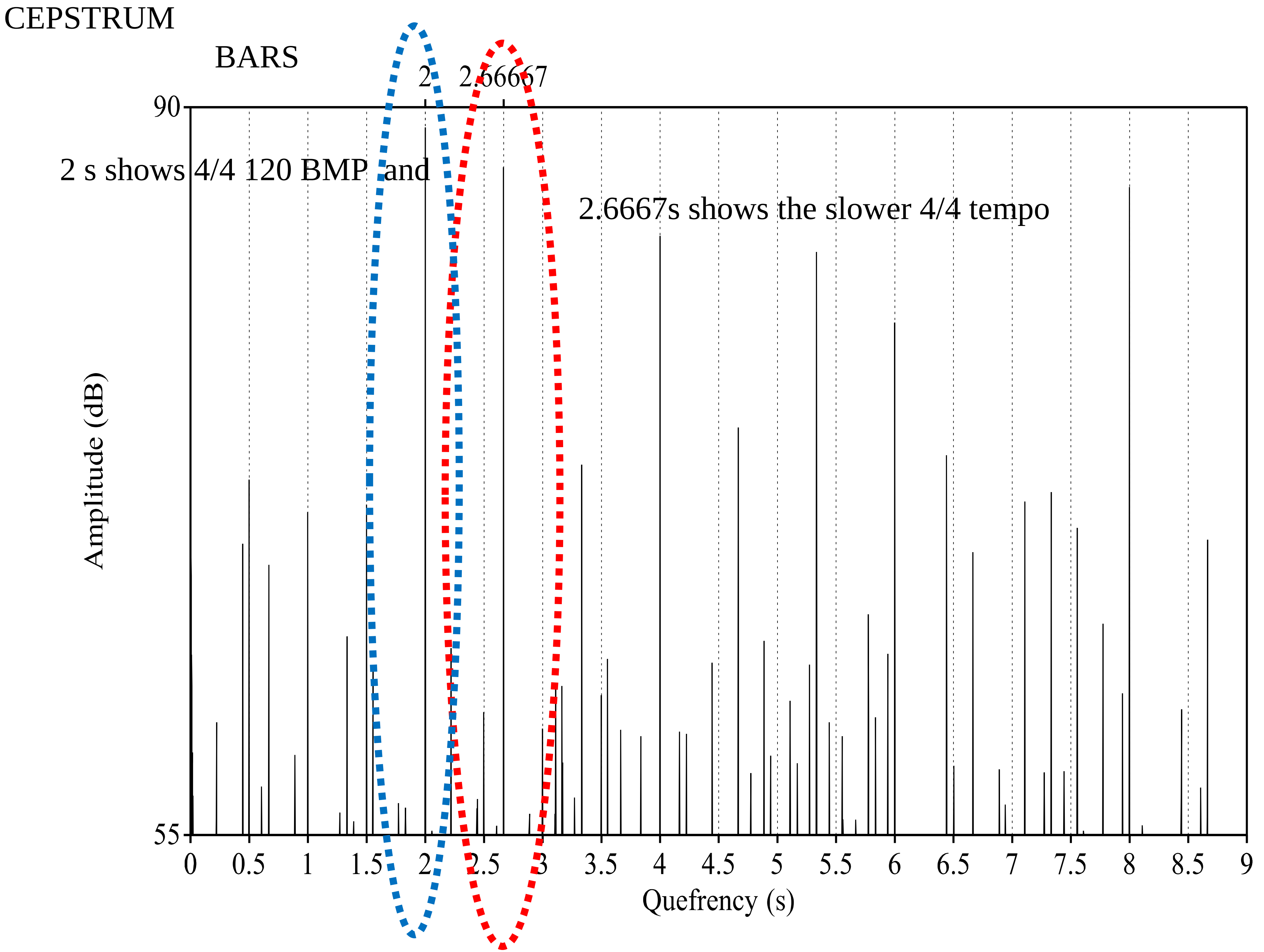


SMOOTHED POWER CEPSTRUM:



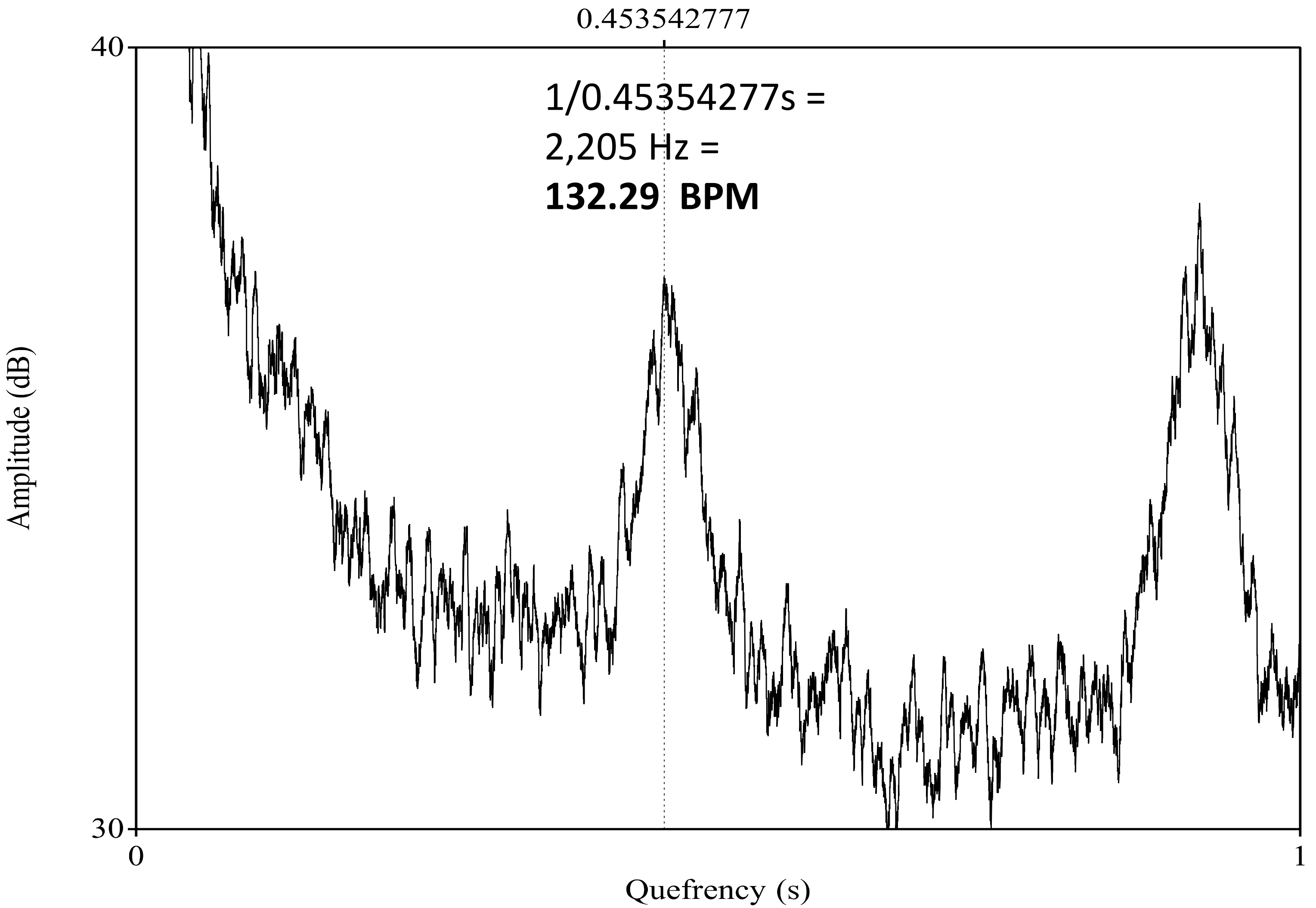
POLY-RHYTHMS : Both 4/4, but Tempo 3 against 4




CEPSTRUM
BARS
2
2.66667
90
55
2 s shows 4/4 120 BMP and
2.6667s shows the slower 4/4 tempo
Amplitude (dB)
Quefrency (s)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9

“Du Kan Godt Få Sitte In´te Mæ Leif”
(**Got To Get You Into My Life**) Lennon-McCartney
(arr. + lead guitar: TH for Ballade)

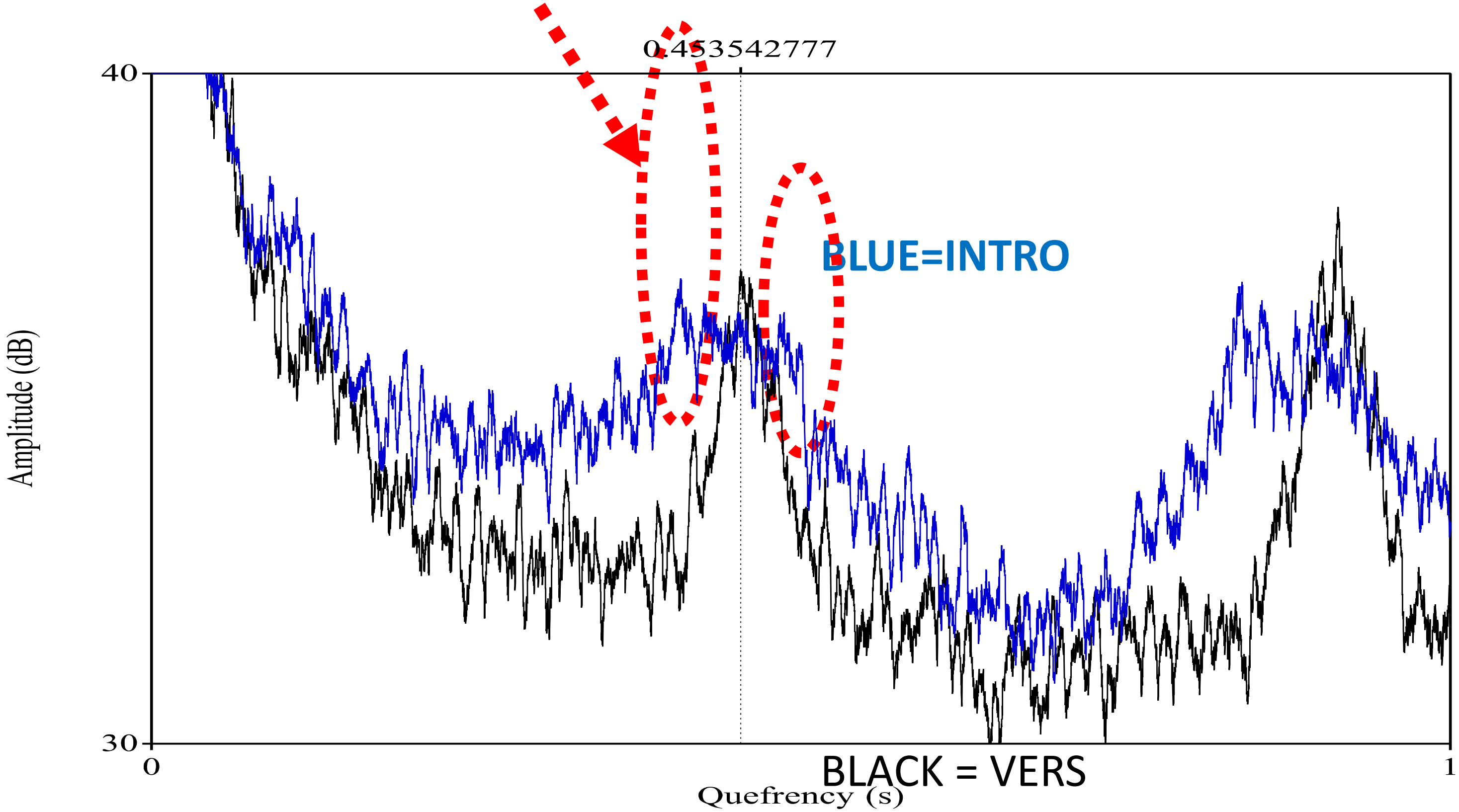
POWER CEPSTRUM (smoothed)



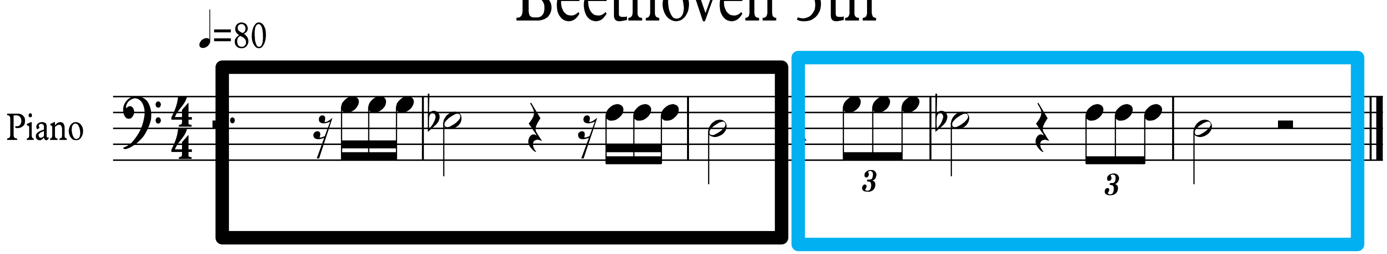
The INTRO has less distinct 4/4,

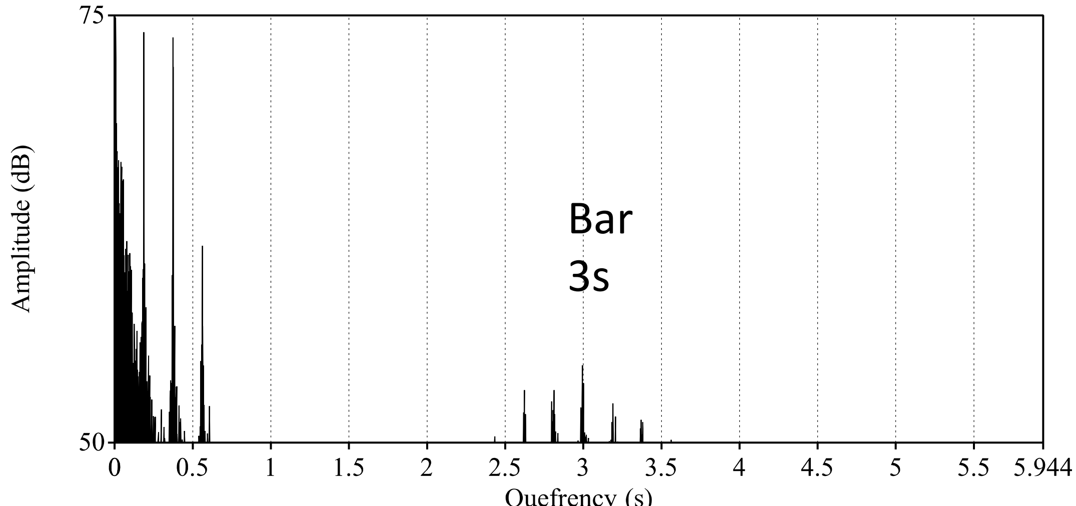
and we see mote of the off-beats and syncopations in the Cepstrum:

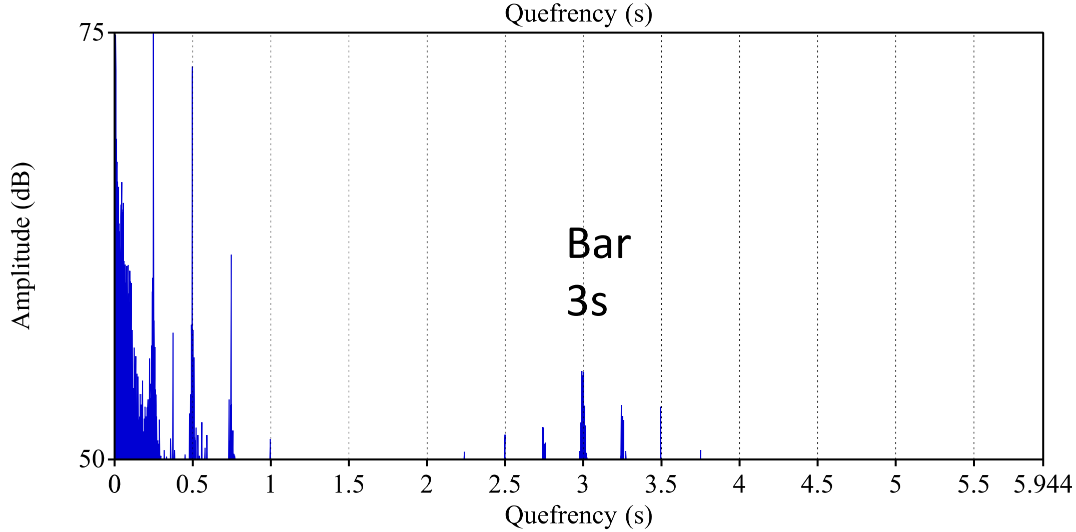
POWER CEPSTRUM (smoothed)



Beethoven 5th







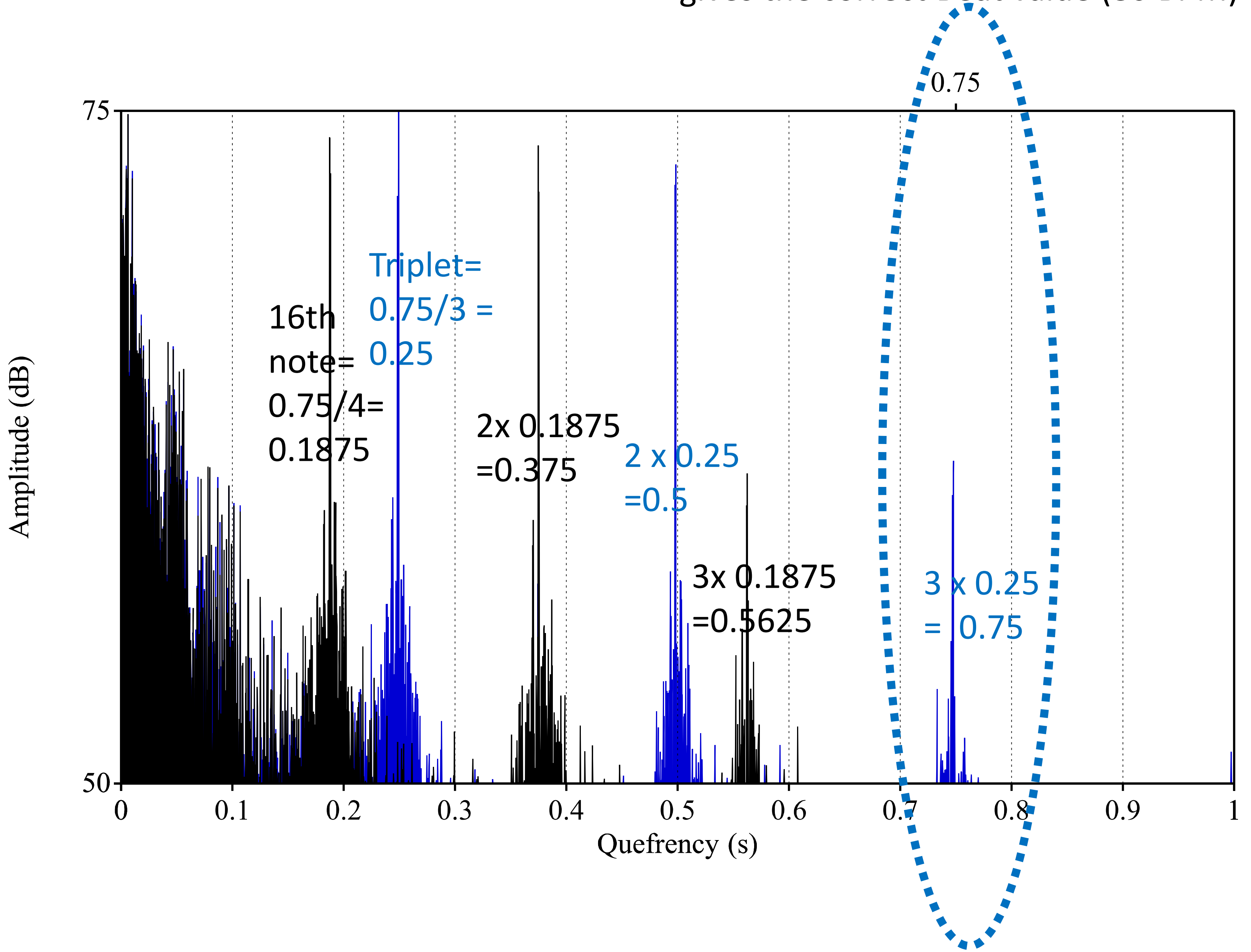
80 BMP=1.3333Hz=0,75 s

4/4: Bar= 4x0.75 = 3s

16^{th} = 0.75/4=0.1875

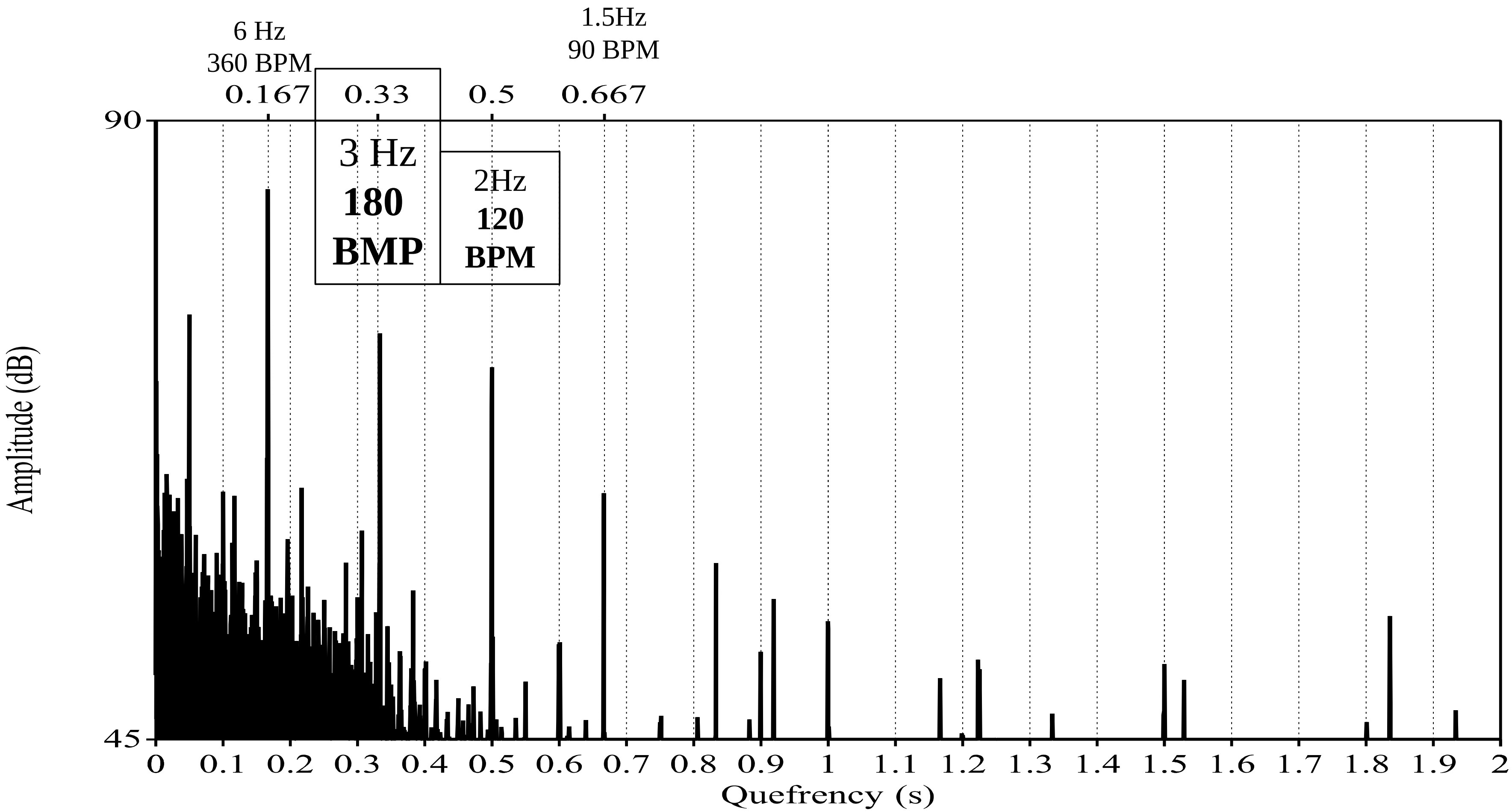
Triplet= 0.75/3 = 0.25

Only the **"false" version**, (with triplets) gives the correct Beat value (80 BPM)

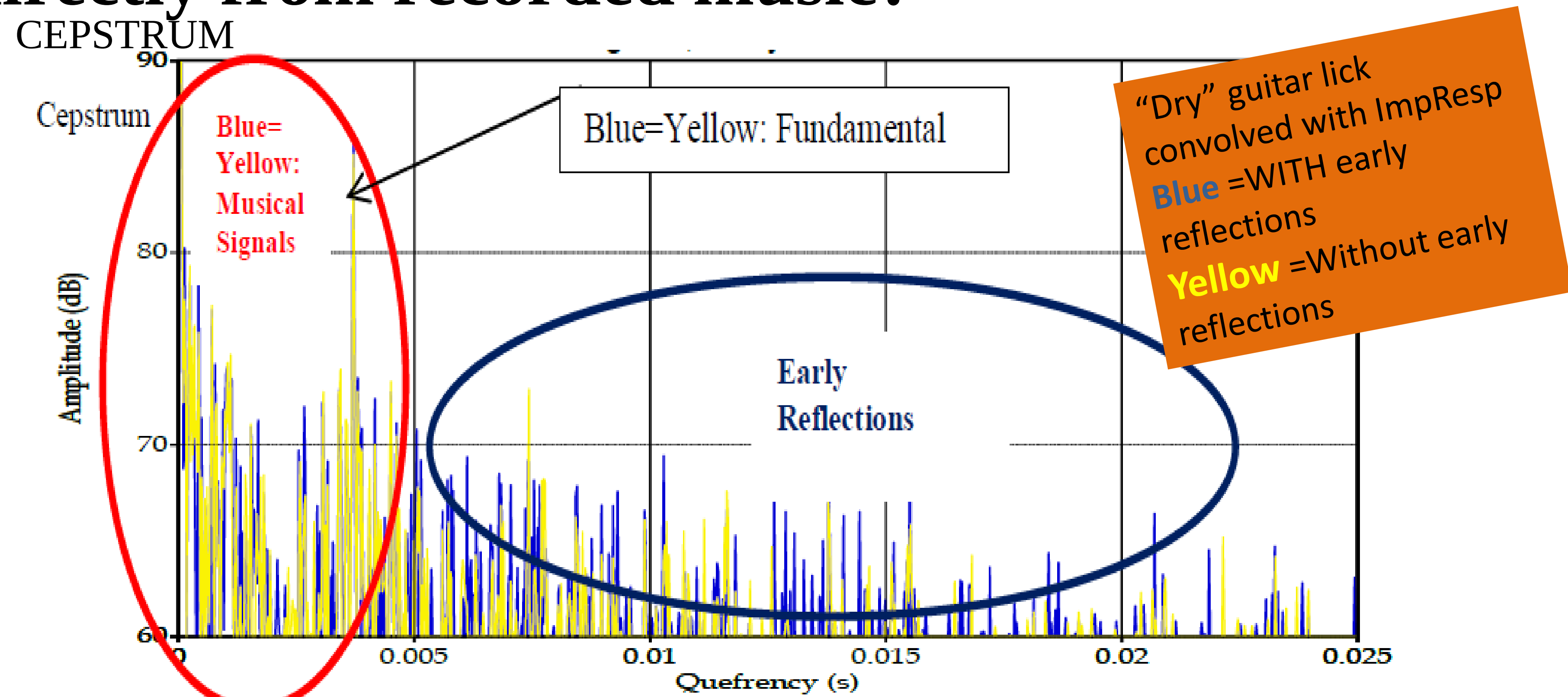


Tor Halmrast: “FLUTR”
25 ch electroacoustic
(4 Quadro + 20 Ambisonics + sub

POWER CEPSTRUM



Can we detect EARLY REFLECTIONS/ "INTIMACY" directly from recorded music?



Simple test (and we knew the answer), but:
With info about the lowest fundamental(s) of the music recorded, it should be possible to get an indication of any early reflections directly from the recorded music.

Free download from:
www.tor.halmrast.no

Some Chapters are translated to English