BNAM 2021
Baltic-Nordic Acoustics Meeting
3rd-5th of May, Oslo, Norway

CEPSTRUM; A "FORGOTTEN" ANALYSIS?

Tor Halmrast
a) Cowi, Oslo, TRHT@cowi.com
b) University of Oslo, Dept. of Musicology, torhalm@online.no

"Every" sound/vibration is the result of an impulse trying to move "something". When this "something" cannot be moved further, it gives a reflection back. This goes for any oscillation, from musical instruments to the stock market. Mixing a signal with a delayed reflection is the "recipe" for building a filter, both in in electronics and in room acoustics. The paper will give an introduction to how a reflection in the time domain gives a Comb Filter in the frequency domain (and possible perceived coloration in room acoustics). When adding feedback, we gradually get a musical instrument (or the annoying feedback in a PA sound system). This general "recipe" could be described as *"The Filter That Explains It All"*. For increasing feedback, the comb filter gradually changes to an «upside down"-shape, in the transition towards higher Harmonicity. The word **Ceps**trum is of course a deliberate mis-spelling of **Spec**trum. In the paper we will see how Cepstrum analysis can be used to detect any «rhythmical behaviour» in the frequency domain, giving information regarding Echoes/Delay Time, Timbre/Coloration, Pitch and Rhythm: From Elvis´ Slap Back Echo to reflections in a concert hall. From rhythm analysis of a simple march to complex poly-rhythmic.

1 Introduction

This paper is an overview of the similarity between how sound is actually produced (in musical instruments, in nature, and in machines) and how the spectrum is changed/filtered later. In fact, "Everything you can send a signal through is a Filter" [1] also the instrument itself. The original impact is a broadband noise; continuous blowing, a click, or making an abrupt "kink" on a string when picking it. The reflections from the ends of a pipe or a string gives the filtering that characterises the instrument as sound source. When the sound from this instrument is combined with reflections from surfaces, etc., the frequency content (timbre) changes; so, the room is a new filter. Even our ear is a (or two) filter(s). **"Everything is Filter"**, and the "reflections" happening in this filter can be analysed using **Cepstrum.**

2 Sound=Reflections

2.1 TTT: Things Take Time

"All" sound is based on reflections, both sound production and sound manipulations. A steady wind itself does not produce sound, but when it reaches a tree, the branches/leaves are bent with the wind, until they cannot move any further, and then they are reflected backwards, and a new movement forward is started. (Of course, the wind might make sound even if there are no trees or objects present, but then the "reflections" must be due to some un-linearity in temperature or pressure in different layers of the ear, so that the different layers give "reflections" between each other; refraction, as in turbulent air).

[1] *Of course, there are some exceptions: for instance, a distortion pedal adds harmonics to a signal, and is thus not a filter.*

In analogue electronics, the current is “reflected” and moves back and forth between the capacitors and inductors, both for a signal generator/oscillator and a filter, and in digital filters we calculate the sum or difference between incoming (sometimes also outcoming) samples in order to create filters. All is based on reflections, so **Thing Take Time (TTT)**, meaning that **There is no actual NOW in sound**! (Perhaps an infinitely short click, but even with the highest sampling rate, the Dirac pulse has the length of one sample). In **a flute**, the broadband pressure-blow is transported down the pipe/tube, and reflected at the end/opening (or a tone hole), ready for one more turn. This goes on as a feed-back as long as you blow. For **a string,** the “kink” you make by picking or hammering the string, moves to each end of the string and is reflected. After some turns of such reflections, the filtering removes almost everything but the frequencies corresponding to the string/tube length, (fundamental plus overtones). So: **Sound production is, by itself, reflections/filtering**. When this instrument (or a noisy machine) as a sound source is **put in front of a wall**, the listener will hear both the direct sound and a delayed reflection, giving a **Comb Filter**. If there are several walls, the comb filters add up. If all close reflections have the same time delay, the comb filters add up to be clearly perceived. If the delays are more random, they might “cancel each other out”. **Rhythm** might be looked upon as “long reflections in time”. Even if the piece of music is totally “free jazz” with no perceived rhythm, there must be at least one “reflection” in time, even if the delay is randomly changing all the time. (For our perception of both rhythm and pitch, there must be several equal periods).

2.2 Comb Filters. One reflection

“All” sounds start with a broadband noise (either a continuous more or less White Noise, or a short burst, more or less like a Dirac pulse). When adding a reflection, some frequencies arrive in phase, so that they add up, and some frequencies arrive out-of-phase meaning that the output is reduced. The shape of this filter looks like a comb (when shown using a linear frequency axis), and is thus called a **Comb Filter**.

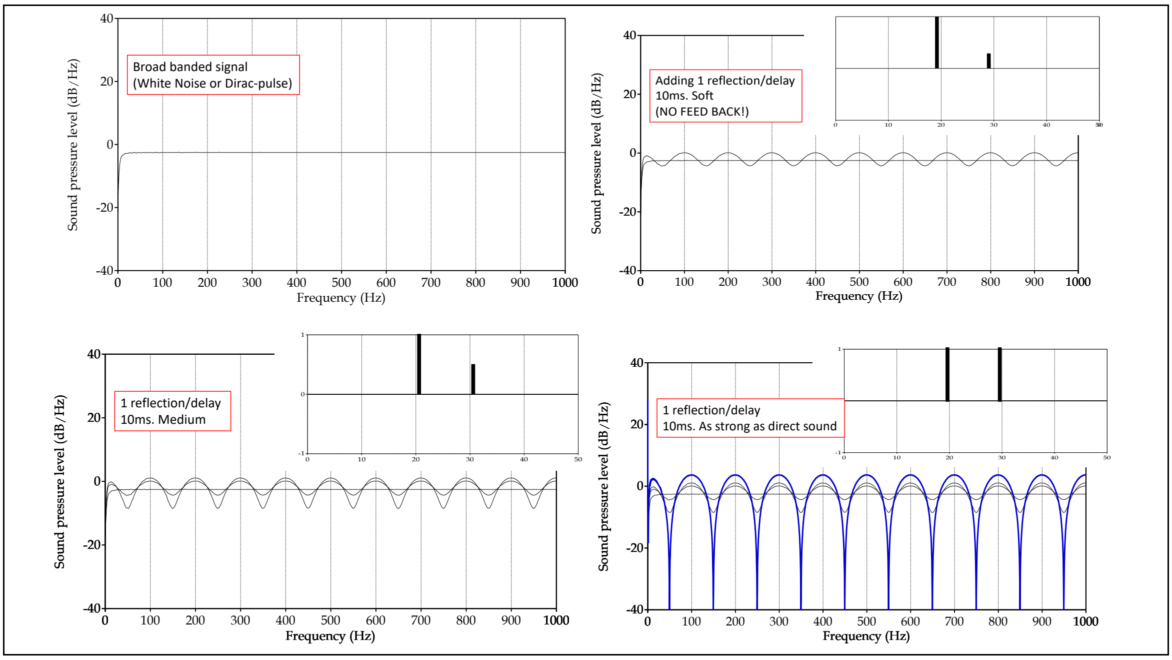


Figure 1: Broad banded signal with reflection of gradually increasing strength, giving a Comb Filter

In fig. 1, the delay is chosen to be Δt =10 ms, giving that the difference between the teeth is CBTB (Comb-Between-Teeth-Bandwidth, see *Halmrast [1]*) of 1/Δt = 100 Hz. This is the kind of comb filter you will get if you are standing 1.715 m in front of a totally reflecting (and very big) wall, and make a broad banded noise with your mouth. (Assuming the speed of sound to be 343 m/s, the total distance to/from the wall is 2 x 1.715 m = 3.43 m, giving Δt =10 ms). If the delay is very long, the CBTB is very small, and the reflection is not perceived in the frequency domain but as an echo in the time domain. We might notice that the “rule of thumb” for perceiving echo: Δt = 50 ms, (PS! for European languages?) corresponds to CBTB = 20 Hz, which fits nicely with what is commonly known to be the lowest frequency perceived as sound (1/50ms = 20Hz). If the delay is very short, the CBTB is very long, and the filter is just a low pass filter with the first dip at a very high frequency. If the CBTB is approximately the size of Critical Bandwidth we perceive a clear coloration of timbre, which might be called “Box-klangfarbe” (see *Halmrast [1]*).

(PS! It is commonly believed that a digital filter that takes the mean value of 2 and 2 samples is a low pass filter. Actually, it is a Comb Filter with the first dip at the Nyquist frequency).

2.3 Adding Feed-Back

So far, we investigated one single reflection. If we gradually add some feed-back to the system we get the following typical type of "frequency response": (Δt =5 ms).

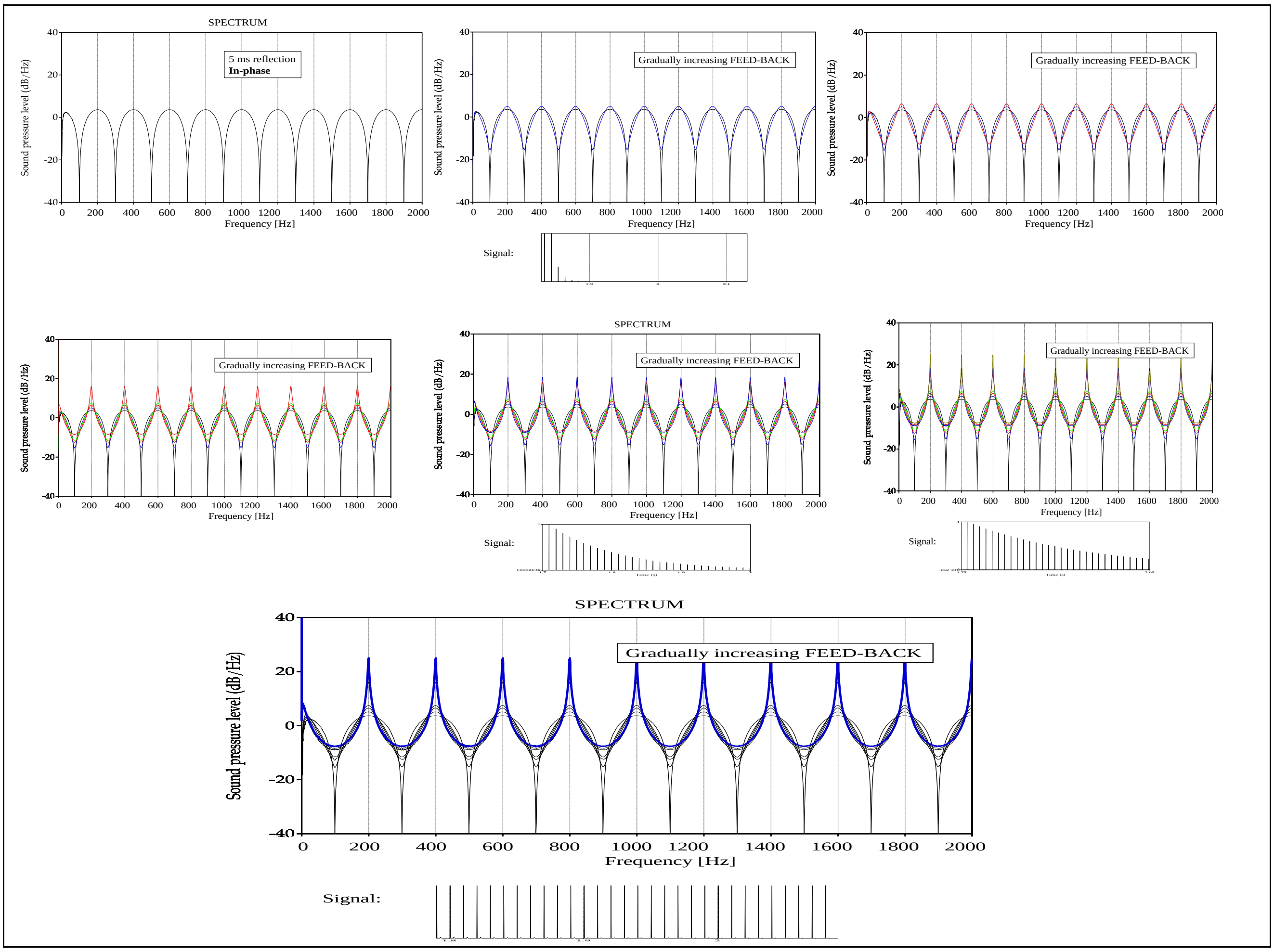


Figure 2: Comb Filter with increasing feed-back (Δt =5 ms, CBTB=200 Hz)

With "full feed-back" we see that the shape of the comb filter has turned upside/down, giving sharper peaks and smoother "dips". We have transferred the coloration from a single reflection into a musical instrument with a fundamental and overtones! The following figure shows the spectrogram of this evolution, and to the right, the typical comb-filters are overlaid.

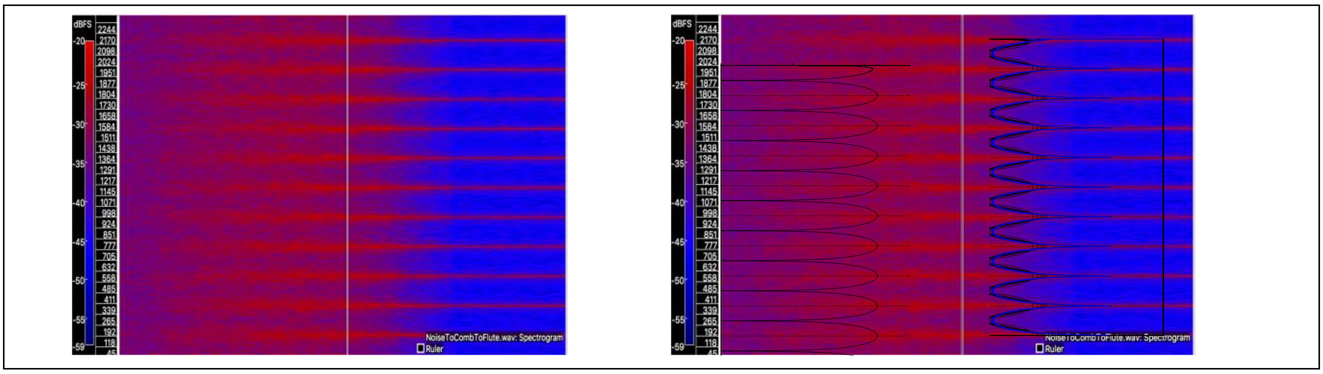


Figure 3: Spectrogram of Comb Filter with increasing feed-back (Δt =5 ms/CBTB=200 Hz)

2.4 Reflection In/Out-of Phase

In room acoustics, the reflection from a plane/hard/big surface is (mostly) In-Phase. In wood instruments etc., the reflection might be either In- or Out of Phase, depending on the boundary conditions. The following figure shows that if the reflection is Out-of-Phase, the comb filter shifts ½ CBTB, and this means that the fundamental frequency is the half (one octave lower), and that only the odd partials of this will be present. This shows the common difference between organ pipes open in both ends compared to open in one end and closed in the other.

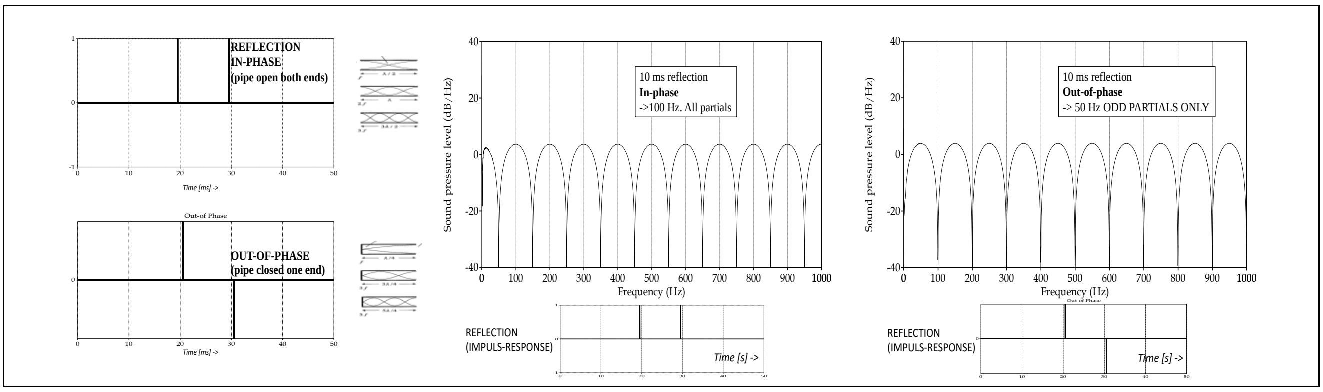
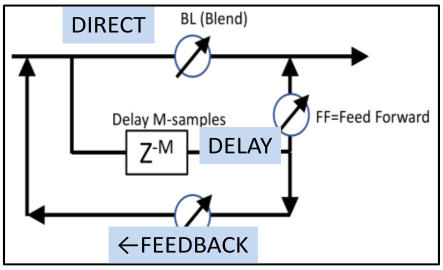


Figure 4: Comb Filter for reflection In/Out-of-phase (Δt =10 ms, CBTB=100 Hz)

2.5 "The Filter that describes it All"

The following typical filter describes all the possibilities.



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

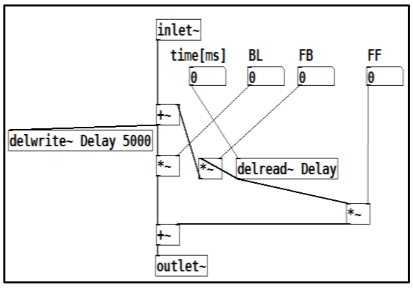
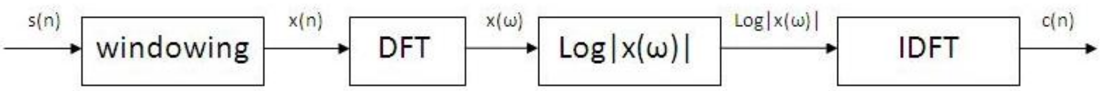


Figure 5: "The filter that describes it all": Block diagram, Overview of possibilities and Pure Data (Pd)-patch

3 Cepstrum

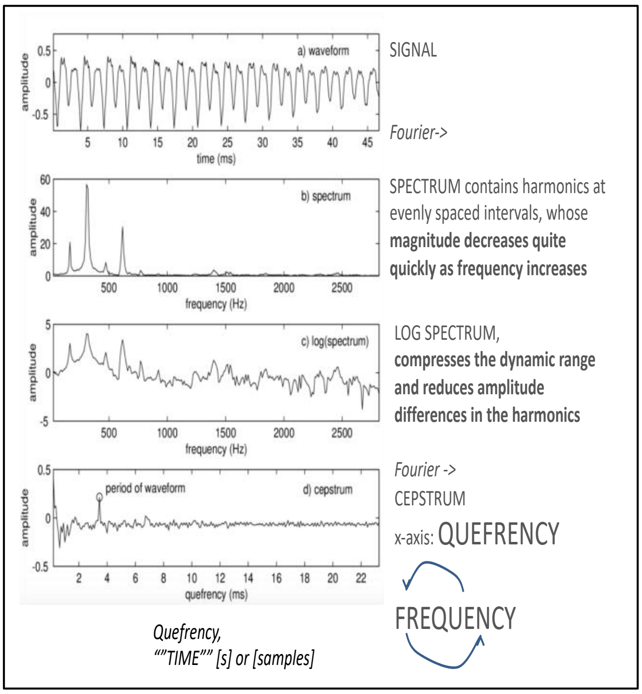
The word **Ceps**trum is made by reversing the first four letters in **Spec**trum. In non-academic terms it is often called the "spectrum of the spectrum". That gives an indication on how it works: The Spectrum gives information about "rhythmical behaviour in the time domain", and Cepstrum gives information about any "rhythmical behaviour in the frequency domain", meaning checking if there is a common divider between peaks (and/or "dips") in the spectrum. A harmonic sound having a fundamental and overtones is of course an example of such "rhythmical behaviour" in the frequency domain. The scope of this paper is not to go into the details of the signal processing, but Cepstrum is defined as the "Inverse Fourier Transform of the Logarithmic Magnitude Spectrum, indicated in the following block diagram:



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

Figure 6: Block diagram of Cepstrum

One might think that first taking the Fourier transform of a signal and then the inverse Fourier transform just brings you back to the signal, but the clue is that the inverse transform is taken for the Logarithmic Magnitude Spectrum, not the Spectrum itself. There are many types of Cepstrums: *Complex, Real, Power, Phase Cepstrum* etc. We will mainly deal with the Power Cepstrum (and sometimes the smoothing of this). The **Power Cepstrum** is the squared magnitude of the inverse Fourier transform of the logarithm of the squared magnitude of the Fourier transform of a signal. As shown in the following figure, the x-axis for the Cepstrum is given in seconds, but it is not actually a time scale, and the parameter is called **Quefrency**, by "scrabbling" the word Frequency, because we analyse "repetitions" in the frequency domain.



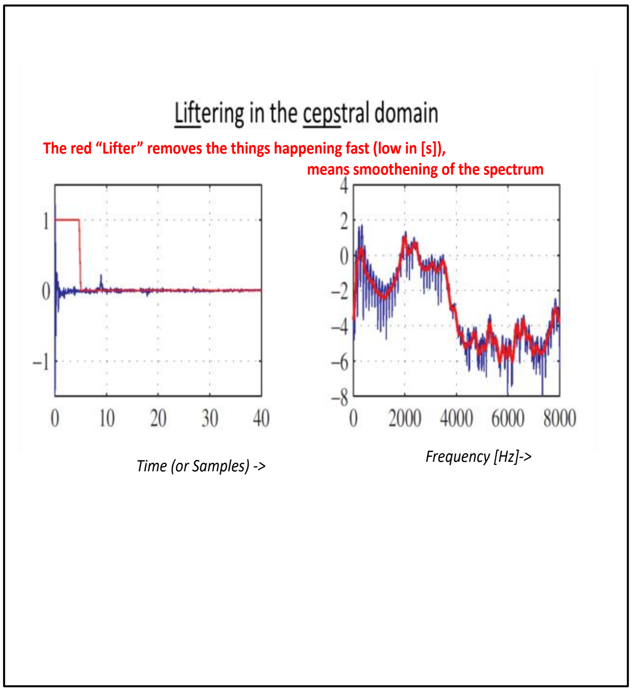


Figure 7: Left: The path from signal to Cepstrum
Right: Liftering is Filtering in the Cepstrum Domain

We can look closer into specific regions of quefrency by **Liftering**, which is a kind of Filtering in the Quefrency domain, as shown in the left pane of fig.7. In this paper we will not go into details of Liftering, but instead look into different regions of values for quefrency directly, in part 4.

Cepstrum is used for many signal processing issues: **Echo Detection**, **Pitch Detection, Rhythm Detection**, **Separation of Source and Filter** (example voice pitch and formant-filters) etc. The last issue shows an interesting side of cepstrum: We know that we can Convolve a signal from a Source with a Filter to get an output. Examples of Source/Filter might be Voice Chords/Mouth (formants) or Musical Instrument/Room. In Cepstrum analysis we go the other way around; we have the output, and we want to separate the source and the filter along the quefrency axis, so **Cepstrum=De-Convolution.** Other types of analysis using Cepstrum is **MFCC** Mel Filter Cepstral Coefficients, which combines the advantages of the cepstrum with a frequency scale based on critical bands and is used for Speech Recognition, **LPCC,** Linear Prediction Cepstral Coefficients, and **SIFT** (Simplified Inverse Filter Tracking algorithm), which encompasses the desirable properties of both autocorrelation and cepstral pitch analysis techniques.

4 Cepstrum-examples

Figure 8 shows a Dirac pulse with a reflection after Δt = 137 ms; Time domain, Spectrum and Cepstrum. (We will see later why 137 ms is chosen). In the left part we see that the Spectrum has a «rhythmic behaviour», because the peaks are separated by 7.3 Hz (CBTB). In the left pane we see that the peaks in the cepstrum are repeated not just by a quefrency of 7.3 ms., but also by 2x, 3x, 4x….. this value. These «overtones» has no direct meaning in the time domain. (This shows that the quefrency scale is not an actual «time scale» even if it is given in seconds). For the simple detection of the delay time, we skip these «overtones» in the cepstrum, and we see that we have detected the delay time to be 137 ms (which we for this example of course knew, since we introduced this time delay). Later we shall see that the «overtones» might give additional information about the structure of the spectrum, but surely it is not «overtones in time»!

The lower left pane in fig. 8 shows overview of the quefrency ranges. We must remember that low values of Quefrency correspond to high frequencies in the spectrum, so for quick oscillations like the fundamentals (pitch) of speech and music, we need to zoom in on the very first part of the figure, and for the overtones of music we need to zoom in even further.

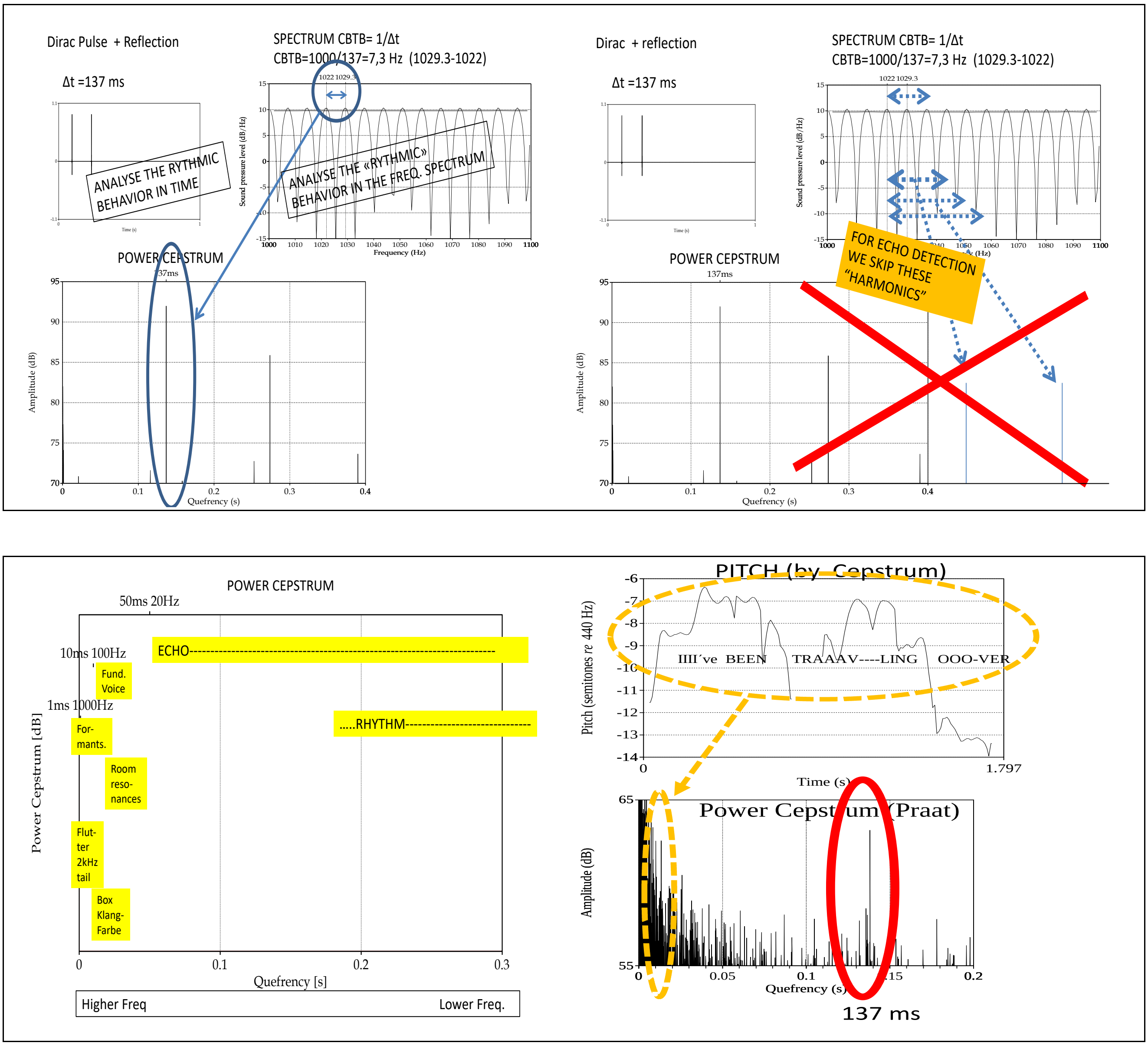


Figure 8: Upper pane: A Dirac pulse with a reflection after 137 ms: Signal, Spectrum and Cepstrum
Lower Pane: The typical ranges of Cepstrum, and
Pitch and Cepstrum of Elvis singing *"I´ve been travllin´ over….."*

When the signal is shorter than the delay, we do not need to use advanced methods like cepstrum to analyse "echoes", but for «running music", it is more difficult. The well-known sound effect called Slap Back Echo (a single tape-echo) was introduced by Sam Phillips on the first recordings of Elvis Presley, and there has been heavy discussion about the actual delay time of this echo (see *Halmrast [3]*). Figure 8, lower/right pane shows that Cepstrum analysis for the *«a cappela»* intro of *Tryin´ To Get To You*. For low quefrency values we have the pitch. As a check we also made a separate pitch analysis to secure that we do not mix pitch and echo in the cepstrum analysis). For even shorter quefrencies; we could find the overtones of the pitch. The pitch is of course changing over the time, so the lower part of the figure is of course «crowded". In the time region for perceiving echoes, we see a peak of 137 ms, which is the tape echo. A Cepstrum analysis of the whole "Noise-to-Instrument" transition from fig. 2 is shown in the upper, left pane in fig. 9. To the right we see the evolution of the cepstrum, indicating that the narrower the peaks in the comb filter of the spectrum gets, the stronger the "overtones" are shown in the cepstrum. This is shown also in the lower pane, left, which is a Cepstrogram, showing the evolution of the Cepstrum over time. We can conclude that the level of the "overtones" in the cepstrum is a measure of how narrow (and strong) the peaks in the spectrum are. This is confirmed in the lower, right pane, which shows the evolution of the **Harmonicity** of the sound on its way from white noise to a musical instrument. (We could also measure this using Spectral Entropy).

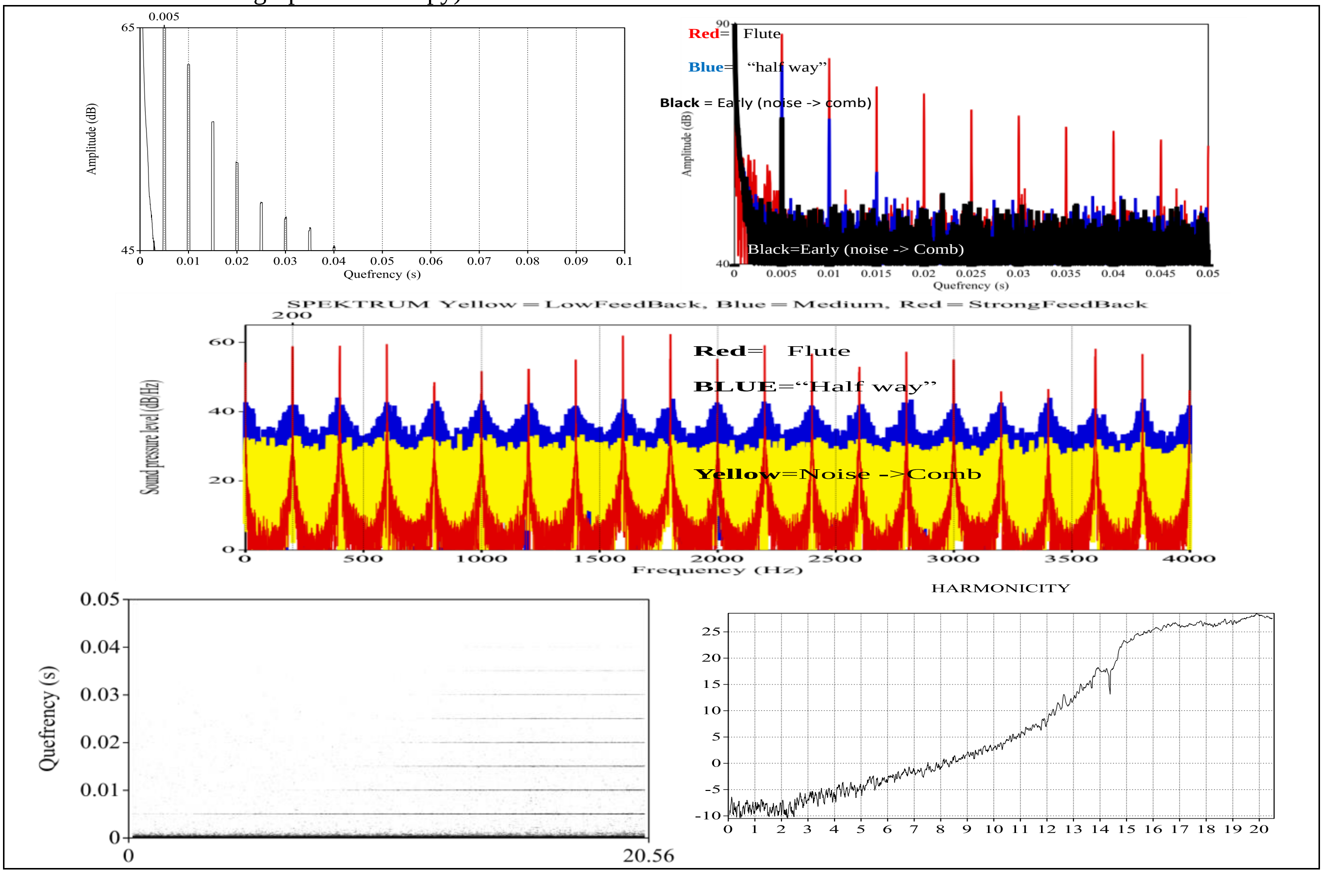


Figure 9: Cepstrum of Noise-Instrument("flute"). Upper: Cepstrum (overall and evolution over time)
Middle: Spectrum for different stages of the evolution
Lower: Smoothed Cepstrogram and Harmonicity (over the 21 sec. evolution)

Cepstrum analysis of a simple 4/4 **march rhythm**; MM=120 BPM (Beats per Minute). and slightly syncopated versions is shown in fig. 10. We clearly see the quefrency of 0.5 s, corresponding to 2 beats pr. second, which of course is correct for the tempo MM=120 For the syncopated versions we see gradually more influence of half of this (0.25 s) which corresponds to the 8th-note. For the syncopated versions we also see somewhat increasing values for 2 seconds, which is the indication of the bars (4 beats).

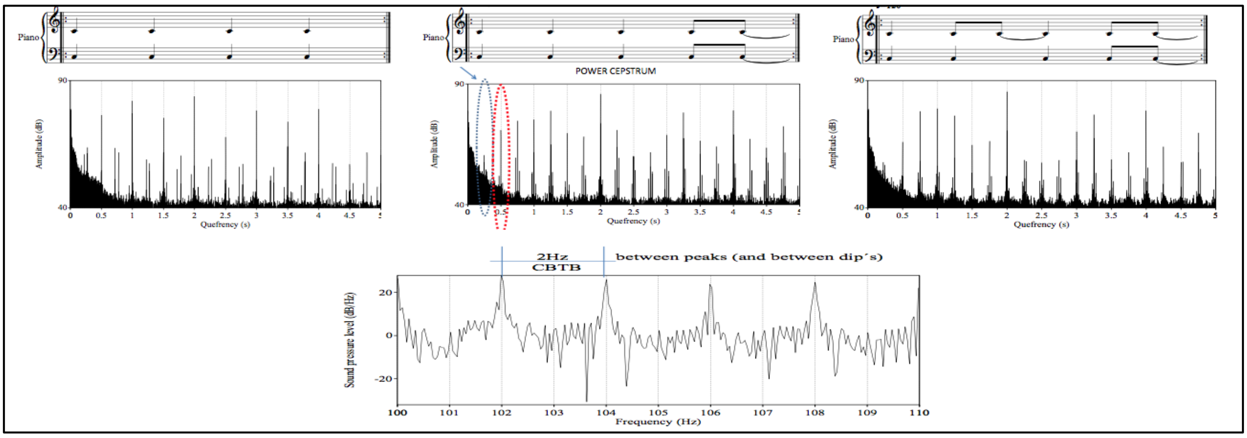


Figure 10: Cepstrum for rhythm detection. Musical notation and cepstrum for three 4/4 rhythms
Lower: Zoom-in of spectrum of the first 4/4 rhythm

The lower pane in fig. 10 shows the spectrum of the plain 4/4 rhythm (from the left, upper pane), and we see that there is actually a comb filter with a CBTB of 2 Hz. This is of course much, much lower than the Critical Bandwidth, so these narrow comb filters are not perceived in the frequency domain. Fig. 11 shows a poly-rhythm (3 against 4) and we see that both rhythms are detected by Cepstrum.

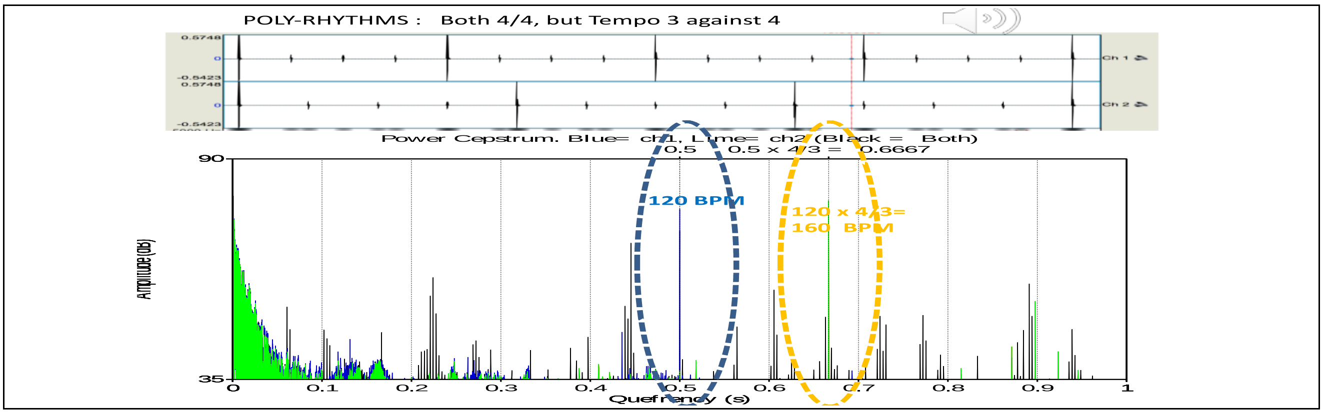


Figure 11: Cepstrum for poly rhythm detection. (3 against 4)

Cepstrum can of course be used for detection of reflections in room acoustics. However, as room acoustic measurements are often done by measuring the Impulse Response (which includes "no signal"), we often do not need to take the trouble to measure Cepstrum, but cepstrum might give more detailed information, also by investigating the strength of the "overtones" in the cepstrum. When analysing actual recordings of running music in a room, we can benefit from Cepstrum analysis, in order to find distinct echoes. Cepstrum analysis can also show if a recording includes short/close reflections. Such analysis would be interesting in order to investigate the often misunderstood issue of "perceived intimacy". Fig. 12 shows cepstrum analysis of a piece of music with/without close reflections. The version with the close reflections is shown in blue, yellow is without close reflections. (For the lower quefrency values (shorter delays; pitch and overtones), the blue and the yellow are the same, so only the yellow is shown).

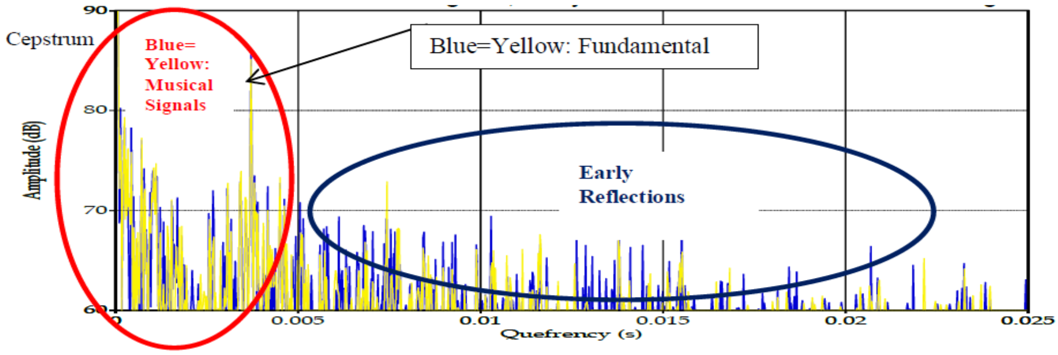


Figure 12: Cepstrum of a piece of music without close reflections (yellow) and with close reflections (blue)

5 Conclusions

In this paper we have shown how delayed reflections and repetitions in the time domain give comb filters in the frequency domain, and that the time delay can be found by Cepstrum analysis. Fast repetitions like for a human voice or musical instruments, and even faster repetitions as for the timbre/overtones are shown in the lowest quefrency region of the cepstrum, echoes in the middle, and rhythm for the highest quefrency values.

References

[1] T. Halmrast: Orchestral Timbre: Comb-filter Coloration From Reflections, *Journal of Sound and Vibration* (2000) 232(1), 53-59, and several other papers, see https://tor.halmrast.no

[2] A.V. Oppenheim: “From Frequency to Quefrency: A History of the Cepstrum», *IEEE Signal Processing Magazine* (2004) 21(5):95 – 106

[3] T. Halmrast: Sam Phillips´ Slap Back Echo; Luckily in Mono. In J.-O. Gullö (Ed.), *Proceedings of the 12th Art of Record Production Conference Mono: Stereo: Multi* (2017) (pp. 137-154). Stockholm: Royal College of Music (KMH) & Art of Record Production. Sam Phillips´ Slap Back Echo; Luckily in Mono (diva-portal.org)