On the Audibility of Comb Filter Distortions

(Zur Hörbarkeit von kammfilterartigen Verzerrungen)

Stefan Brunner, Hans- Joachim Maempel*, Stefan Weinzierl**

* Technische Universität Berlin, Fachgebiet Kommunikationswissenschaft/Audiokommunikation,
brunnerberlin@web.de, hans-joachim.maempel@tu-berlin.de, stefan.weinzierl@tu-berlin.de

Abstract

Superpositions of delayed and undelayed versions of the same signal can occur at different stages of the audio transmission chain. Sometimes it is a deliberate measure to provide audio material with certain spatial or timbral qualities. Often it is a result of multiple microphone signals, sound reflections on walls or latencies in digital signal processing leading to comb-filter-shaped, linear distortions. The measurement of a hearing threshold for this type of distortion with its dependance on reflection delay, relative level and the type of audio content can be the basis for boundaries in everyday recording practice below which undesired timbral distortions can be neglected. Therefore, a listening test was conducted to determine the just noticeable difference for three stimulus categories (speech, a snare drum roll and a piano phrase) and different time delays between direct and delayed signal from 0.1 ms to 15 ms, equivalent to 0.03 - 5.15 m of sound path difference. The results show that comb-filter distortions can still be audible if the level of the first reflection is more than 20 dB lower than the level of the direct sound.

1. Subject

Comb-filter-shaped linear distortions are a result of audio systems that superimpose coherent or nearly-coherent signals, i.e. a signal and a delayed, filtered, or scaled version of itself. In a time-discrete signal flow diagram these factors could be denoted as a delay by n samples, a transfer function R(z), and a scale factor a.

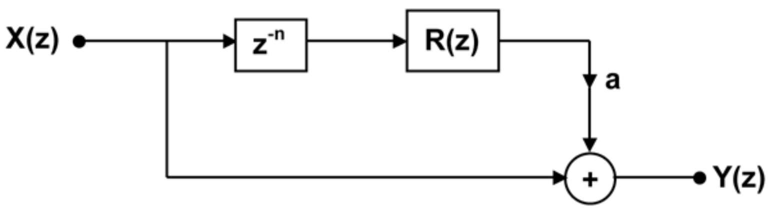


Fig. 1: Audio system with a comb-filter-shaped transfer function

Frequent occasions where distortions of this kind occur are electrical/digital delays where we can assume that R(z) = 1. The scale factor a represents the relative Level ΔL of the delayed signal in the mix with ΔL = 20log(a). Summing up the channels of non-coincident stereophonic signals will lead to a modulated comb-filter response depending on the time difference different acoustical sources are coded with on the stereo channels. When it comes to an electrical superposition of different microphone signals (such as main and spot microphones) or an acoustical superposition of sound reflections on surfaces we have to take into account the transfer function R(z) applied on the delayed signal, usually with an attenuation of high frequencies. For the simplest case of R(z) = 1 the transfer function in Fig. 1 is

$$H(z) = 1 + a \cdot e^{-jn\Omega} \tag{1}$$

For a = 1 (no attenuation) the familiar comb-filter-effect is given by the shape of the gain

$$|H(z)| = 2\left|\cos(\frac{n\Omega}{2})\right| \tag{2}$$

with its characteristic +6 dB peaks at frequencies

$$f_{peak} = 2\pi\frac{k}{n} = \frac{k}{\Delta t} \tag{3}$$

and extinctions by destructive interference at frequencies

$$f_{dip} = \pi\frac{(2k-1)}{n} = \frac{(2k-1)}{2\Delta t} \tag{4}$$

with the delay time $\Delta t = n / f_S$, the sampling frequency f_S and k = 1,2,3,..., denoting the order of peaks and dips (Fig. 2).

For a < 1 (attenuated reflection) the transfer function becomes less distorted approaching |H(z)| = 1 for a→ 0 (Fig. 3).

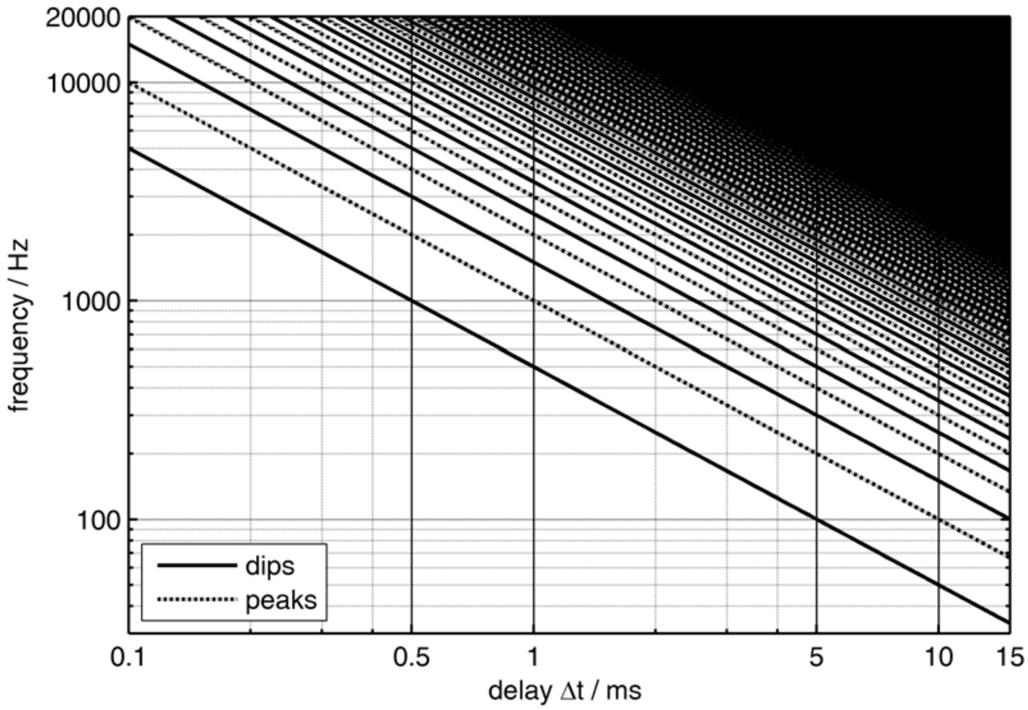


Fig. 2: Position of peaks and dips in a comb-filter-shaped transfer function for different delay times Δt

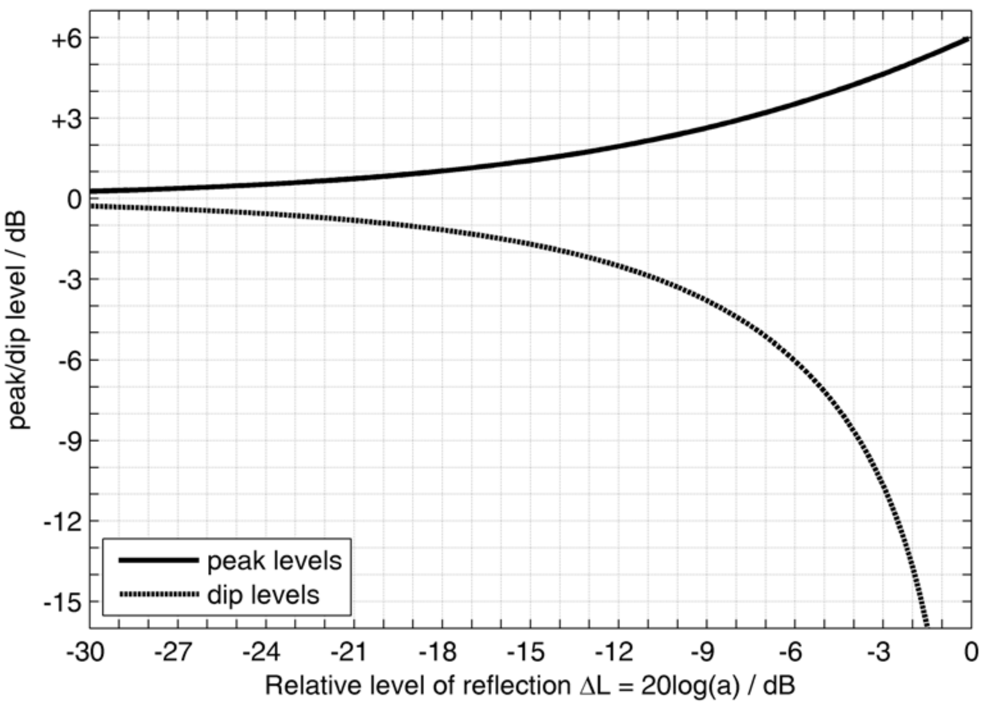


Fig. 3: Altitude of peaks and dips in the comb-filter-shaped transfer function due to reflections with relative level ΔL.

A coloration of sound material due to comb-filter-distortions is audible as long as the reflexion is not perceived as a seperate sound event which happens at delay times above 30 ms, depending on content and relative level [1]. Due to the harmonic structure of the comb-filter-shaped frequency response the coloration can lead to the perception of a so-called residual pitch or repetition pitch which occurs mainly if the source sound itself contains little or no tonal information [2,3].

Considering the variety of occasions where comb-filter-distortions can occur within the audio transmission chain it would be highly desirable to establish a threshold of audibility for linear distortions of this kind. This threshold would also serve as a lower boundary for the sensitivity to frequency response irregularities in general.

Research on this topic was done by Anazawa et al. [4] who found a minimum in the threshold of audibility of added reflections at time offsets between 1 and 2 ms and relative levels of 20 dB using white noise, vocal, chorus and trumpet signals. In experiments with pink noise conducted by Kuhl [5] a minimum in the threshold was found at a relative level of 17 dB at 0.2 ms time offset and 13 dB at time offsets from 5 to 15 ms. Müller [6] found threshold values for level differences between first reflection and direct sound of 17 dB for white noise and 12 dB for a speech signal. Dickreiter [7] states that level differences of 10 dB between direct sound and first reflection will suffice to keep the distortions inaudible. Testing the audibility of single-peak linear distortions Bücklein found spectrals peaks in general easier to detect than dips of the same level difference [8]. In the context of headphone equalisation Ramsteiner and Spikofski found a just noticeable difference for one-third octave filter peaks in the frequency response with around 1 dB level difference for frequencies above 500 Hz [9].

Since none of the cited references used a “state of the art”-method to investigate just noticeable differences between audio signals and regarding the relevance of the subject it seemed worthwhile to conduct a new experiment using a generally accepted test procedure and artificially generated stimuli.

2. Listening test

The stimuli were generated using a MATLAB script. Three respresentative types of audio signals were used: 1) a piano phrase, 2) a snare drum roll and 3) a phrase of a male speaker, all taken from the SQUAM reference CD published by the European Broadcasting Union.

A pretest was conducted to find out the optimum duration of the stimuli to achieve the highest sensitivity for differences between original and filtered sounds in a 3AFC situation (see below).

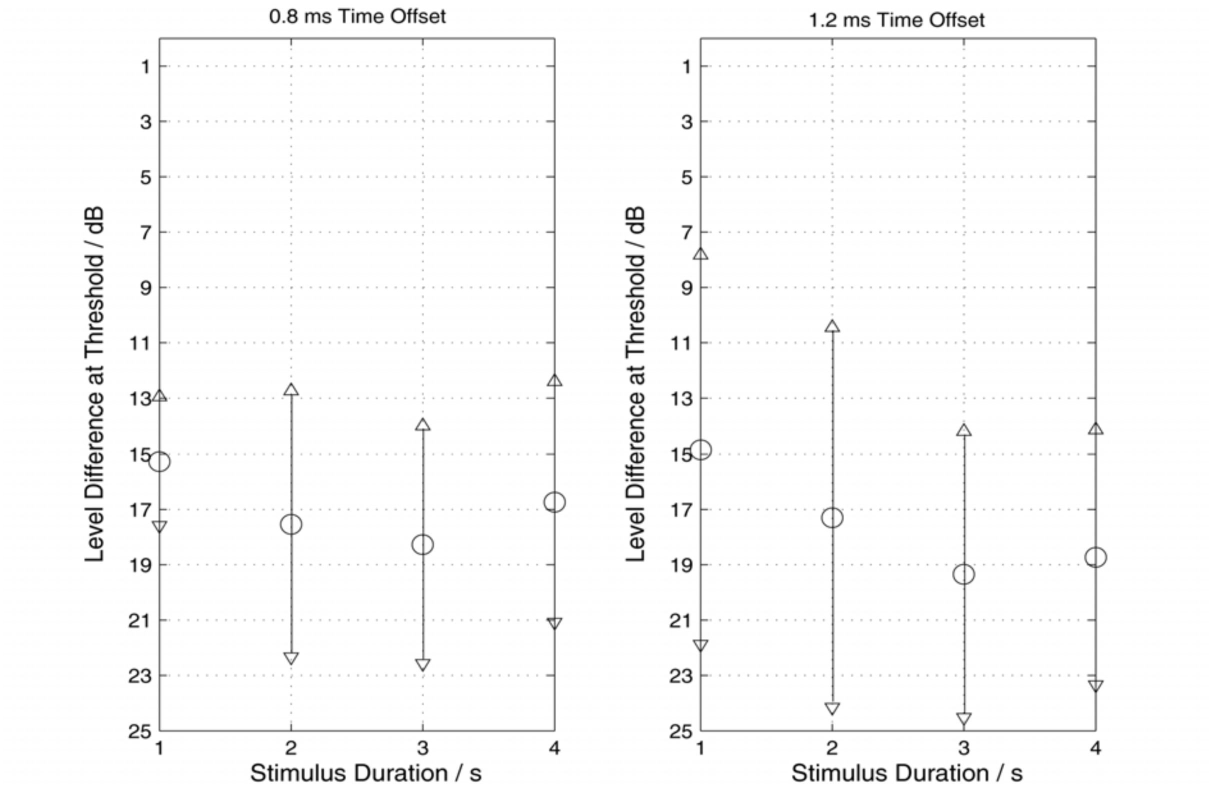


Fig. 4: Just noticeable difference between original and comb-filtered signal showing the level difference between direct and delayed signal for time offsets of 0.8 and 1.2 ms (snare drum signal, means and standard deviations).

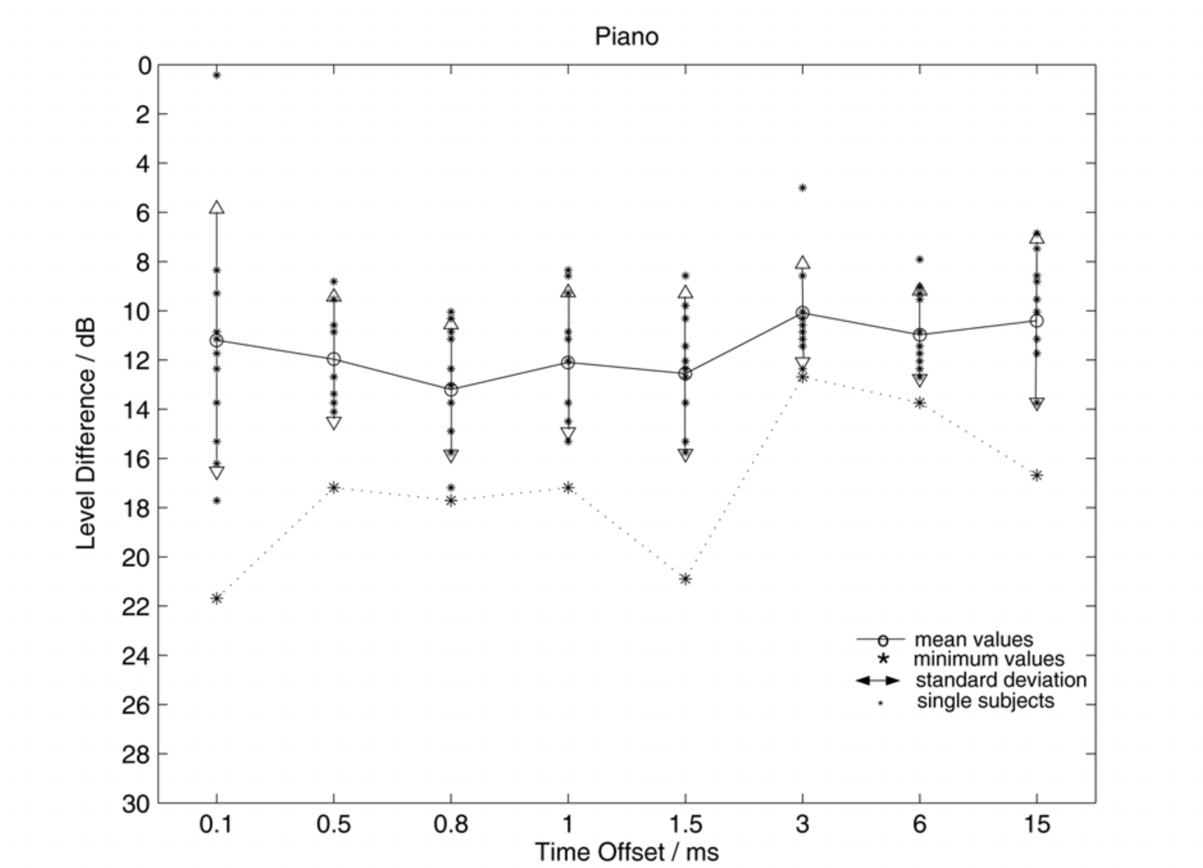
An optimal stimulus duration was found out to be 3 seconds, where a minimum in the just noticeable difference for time offsets of 0.8 and 1.2 ms and snare drum content was detected (Fig. 4).

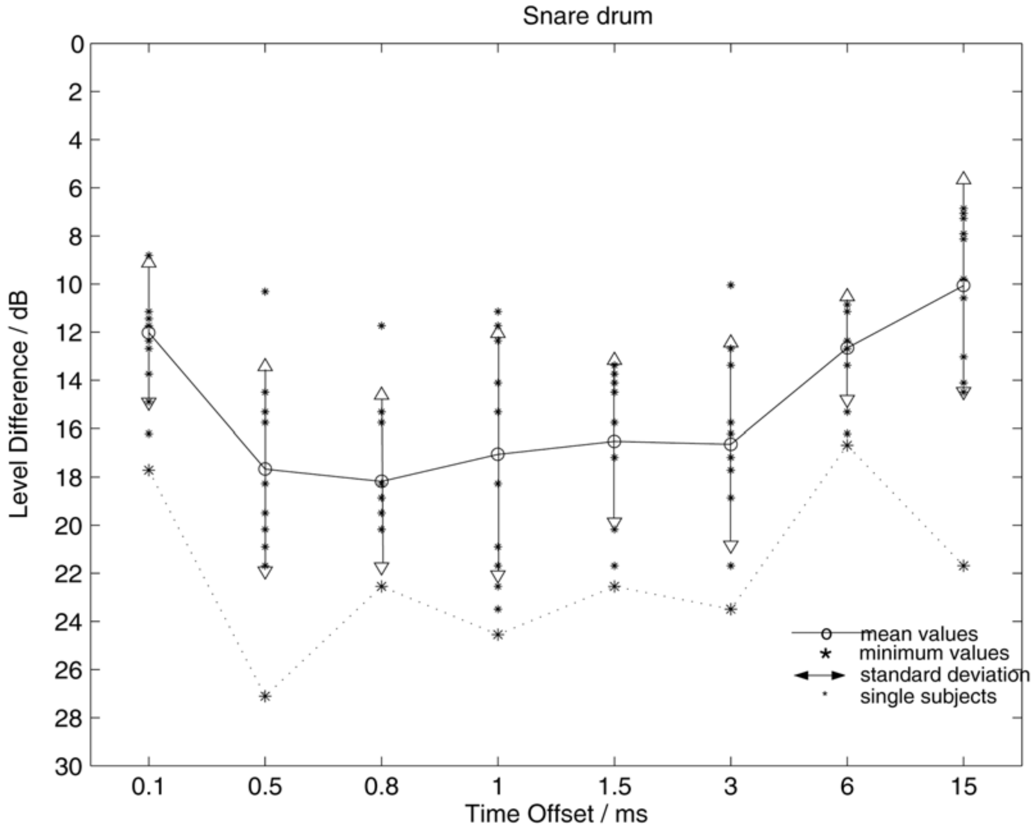
The main listening test started with an instruction and a training session making the subjects familiar with the test signals and the procedure. The monophonic stimuli were presented via headphones (Stax SR 202). The volume could be changed during the training phase to a comfortable level and was held constant during the rest of the test. Time offsets were tested between 0.1 and 15 ms. An 3-Down-1-Up adaptive tracking procedure was used to approximate the threshold for the just noticeable difference in a three alternative forced choice (3AFC) test. The subject was asked to detect the stimulus which he/she thought was manipulated using a graphical user interface. The initial values for each test procedure were set considerably higher than the threshold values determined in a pretest to allow an extra training in each test procedure. The frequency response of the comb filter was defined by the amplitude above 0 dB at peak frequencies. For optimal stimulus placement the step size of 0.5 dB was reduced to 0.25 dB after the first reversal that led to a “down” group. The resulting differences in loudness were compensated for using an algorithm based on Zwicker´s loudness calculation method [10]. Following a recommendation by Leek [11] the average of the last four turnaround points was taken as the resulting threshold value.

The subjects, 36 students and professors of the TU Berlin and the UdK Berlin with experience in audio engineering were divided into three groups of twelve. Each group listened to one of the three stimulus types (piano, snare, speech). After the test the subjects were asked to describe the cues they used to detect the distortion. The whole test took about 45 minutes per subject.

3. Results

The thresholds for the three stimulus categories differ considerably. For piano and snare drum a minimum in the mean threshold of all subjects was found at 0.8 ms time offset. The mean level difference between direct sound and first reflection is 13.2 dB at the threshold for piano and 18.2 dB for snare drum, with single subjects reliably detected level differences as much as 21.5 dB (piano) and 27 dB (snare drum). The threshold curve for the speech signal shows a trend to an easier detection of the distortion with growing time offset. For a time offset of 15 ms the just audible difference was detected at a level difference of 16 dB (average) resp. 22 dB (single subject), requiring an extension for futere tests towards higher time offsets.





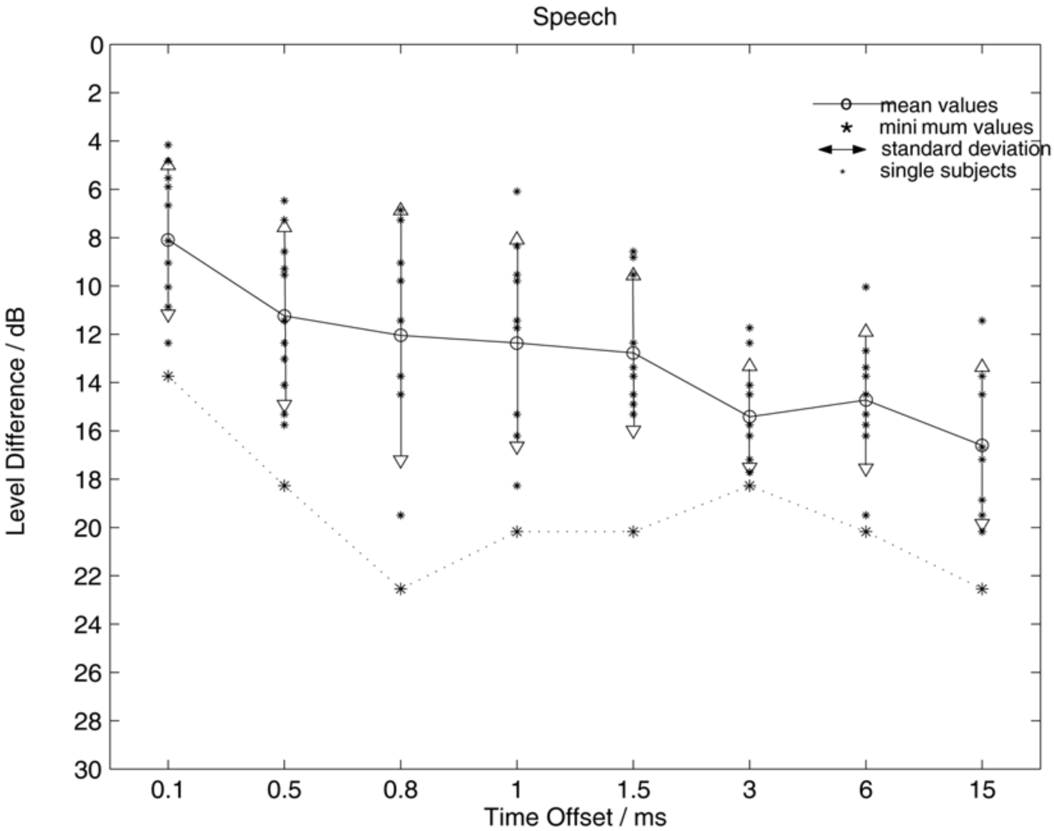


Fig. 5: Just noticeable difference between original and comb-filtered signal showing the level difference between direct and delayed signal for delay times (time offsets) between 0.1 and 15 ms.

4. Discussion

The present study proved that listeners are considerably more sensitive to frequency response irregularities than suggested by previous investigations. For certain audio material and under good listening conditions differences are audible when a reflection with a level difference of 18 dB (average of all listeners) respectively 27 dB (single listeners) is added. This corresponds to peaks in the comb-filter curve (Fig. 2) of 1 dB (average) and less than 0.4 dB (single listeners). Kuhl's finding was confirmed that listeners are particularly sensitive for changes in timbre when dealing with noisy signals [5]. Maximum sensitivity for added reflections and broadband content is reached at time offsets between 0.1 and 1 ms, depending on content and listener, while sensitivity with speech content rises with ascending time offset and a minimum was not found below 15 ms.

Applied to the above mentioned occasions where comb-filter-shaped distortions can occur in the audio transmission chain we deduce that an electrical/digital superposition of identical but delayed content can be audible even if the delayed signal is as much as 27 dB lower than the direct signal. In situations where the delayed signal undergoes filtering when travelling through air oder being reflected on walls the peaks and dips in the frequency response will be less pronounced and the established threshold can be regarded as a lower boundary. With acoustical superpositions of sound waves a comb filter distortion will only occur for the free-field part of the sound since the diffuse-field portion looses time-coherence, thus no comb-filter distortion will occur.

Even in those situations we can assume that for content with broadband frequency distribution such as the pink-noise-like snare drum signal a distortion will be best audible for time offsets between 0.5 and 3 ms (Fig. 5) corresponding to sound path distances between 17 cm and 1 m. These sound path distances between microphones or at reflecting surfaces should be particularly avoided.

Further research has to be done on the perceived sound quality of added reflections. Subjects named timbral differences to be the primary cue for the piano signal, some named the perception of a residual pitch as the primary cue for the snare drum signal, whereas for the speech signal the primary detection cue was an impression of spaciousness even for small time offsets.

An interesting side observation is the suggested optimal signal duration of 3 s for the recognition of frequence response irregularities in direct AB comparisons – a situation that regularly occurs in sound engineering practice. The broad range of threshold values even for trained listeners was unexpected, as well as the fact that the distribution for experienced sound engineers was not much different from non-professionals among the 36 tested subjects. Hence, audio professionals should always be aware that some of their clients might be more sensitive to distortions than they are themselves.

5. References

[1] Haas, H.: “The Influence of a Single Echo on the Audibility of Speech“, JAES 20/2 (1972), 146-159

[2] Bilsen, F.A.; Ritsma, R.J.: “Repetition Pitch and ist Implication for Hearing Theory”, Acustica 22/2 (1969/70), 63-73

[3] Hartmann, W.M.: “Signals, Sound and Sensation”, Springer-Verlag, New York 1998, p.364

[4] Anazawa, T.; Takahashi, Y.; Clegg, A.H.: “Digital Time- Coherent Recording Technique”, 83th AES Convention, New York 1987, Preprint 2493

[5] Kuhl, W.: „Unterschiedliche Bedingungen beim Hören in einem Raum und bei elektro-akustischen Übertragungen“, Rundfunktechnische Mitteilungen 13/5 (1969), 205-208

[6] Müller, L.: „Zur Klangfärbung durch kurze Zeitreflexionen bei Rauschen, Sprache und Musik“, 6. Internationaler Kongress für Akustik, Tokio 1968, E-61 - E-64

[7] Dickreiter, M.: „Mikrofonaufnahmetechnik“, 3. Aufl., S. Hirzel-Verlag Stuttgart, Leipzig 2003

[8] Bücklein, R.: “The audibility of Frequency Response Irregularities”, JAES 29/3 (1981), 126-131

[9] Ramsteiner, A.; Spikofski, G.: „Ermittlung von Wahrnehmbarkeitsschwellen für Klang-farbenunterschiede unter Verwendung eines diffusfeldentzerrten Kopfhörers“, Fortschritte der Akustik – DAGA 1987, pp.581

[10] DIN 45631: Berechnung des Lautstärkepegels und der Lautheit aus dem Geräusch-spektrum, Verfahren nach E. Zwicker, 1991

[11] Leek, M.: Adaptive procedures in psychophysical research, Perception & Psychophysics 63/8 (2001), 1279-1292