How Spectra influence the Audibility of Comb Filter Coloration

Marius Tøndel Eliassen

Master's thesis at the Department of Musicology, Faculty of Humanities, University of Oslo

Spring 2020

Abstract

This paper is concerned with the spectra of reflections - to what degree the spectra of reflections influence the audibility of comb filter coloration. The research question is investigated from the perspective of the high-frequency absorption typical in listening rooms. A listening experiment is conducted to observe how filtering of reflections modelling the absorption characteristics of a 20 mm thick rock wool absorbent, influences coloration detection thresholds in two categories of broadband sounds; band-limited white noise and orchestrated classical music. The results suggest that the absorption characteristics employed reduce comb filter coloration in the test signals. The results are supported by the spectral analysis of the signals following the experiment. However, to what extent the coloration is reduced after filtering of reflections is also dependent on the source signal; if “vital” spectral energy is removed from the reflection under investigation the reflection will have less impact and the resulting timbre will be closer to the timbre of the original signal.

Acknowledgments

Working with this master's thesis has been both exciting, interesting, enjoyable, and instructive, but also much more challenging than expected.
First, much time was spent on finding an appropriate research question feasible within this field. However, the main challenge has been the psychoacoustic experiments where I intended to recruit about twenty respondents to make it as reliable as possible. After sending about eighty emails to students and employees at the Department of Musicology and the Academy of Music and also putting up information posters on boards at the same places, I only received two answers.
Then I asked my mother, who is a teacher at Mailand upper secondary school in Lørenskog, to help me with the recruiting and actually, she managed to find enough teachers willing to participate. Unfortunately, just before starting the listening experiments the school was closed because of the corona virus, which resulted in some teachers wanting to postpone, some withdrawing and luckily, some completing.
I would first like to thank my supervisor Tor Halmrast for his excellent inspiration, guidance, patience and good support throughout this process. I would also like to thank my mother, Marit Tøndel Eliassen, who has been a great source of support and has encouraged me all the time despite some problems occurring on the way. Without helpful and willing teachers at Mailand upper secondary school it would not have been possible to realize the experiments.

You are very much appreciated - thank you!

June 2020

Marius Tøndel Eliassen

Contents

1 Introduction

1.1 Background and research question

When listening to speech or music in an ordinary-sized room, not only do we hear the direct part of the sound reaching our ears - we also hear reflected sound. Sound waves that are not reaching our ears directly will reach various surfaces in the room and be reflected from them. Some of these reflections will be directed towards us, which may result in various perceptual effects.

We may perceive one of these effects as a *coloration of sound*. While subjective, Salomons' (1995:2) explanation of the term is straightforward; *"..the coloration of a sound signal is the audible distortion which alters the (natural) color of the sound"* (for a more in-depth discussion of this term and its relation to other relevant terms see, "clarification of terms", 1.2).

The topic of coloration is debatable. In larger listening rooms like concert halls, the coloration from the reverberation in the room may give a subjective impression of depth and "airiness" to the music. This type of coloration is often desirable - at least in rooms where music is performed, such as in concert halls. However, in smaller rooms coloration of sound (due to reflections) is usually detrimental. The short distances to room boundaries and other surfaces in these rooms facilitate reflections with relatively short arrival times at the listener's ears.

When the reflections are shorter than a substantial part of the direct signal they will [1]*superimpose* on the direct signal creating [2]*acoustic interference*. The result is a periodically rippled spectrum which looks like the teeth of a comb. This response is, therefore, called a *comb filter effect*.

Investigations on the effects of reflected sounds have been ongoing for several decades.

[1] place, or set over, above, or on something else

[2] see section 2.1.1

Listening tests can be conducted in several ways; 1) In the laboratory and 2) In a real listening room.
In the laboratory, or in a laboratory type setting, the reflected part of the sounds is simulated. If the experimenter wants to study the interaction between the various components in an entire sound field, a form of [3]*sound field synthesis* can be implemented. This technique also allows the experimenter to isolate the different components in the sound field, to study the interaction between the various components more closely. Such listening tests allow the experimenter better manipulative control over the various variables involved compared to listening tests in actual listening rooms (where individual components of sound fields to a lesser degree can be manipulated).

In actual (listening) rooms reflected sounds tend to be spectrally modified. Olive & Toole (1989:11) note that this spectral modification is often overlooked in listening tests with artificially created sound fields, or individual reflections.

There is a clear need to investigate, to a greater extent than what has been reported so far, the effect of spectral modification on coloration.

This sowed the seed for what I want to investigate.

However, while spectral modification (of reflections) may lead to a perceptual change, it does not necessarily reduce the audibility of the coloration. As such, it is not the perceptual change itself that is of primary interest in this thesis. We ask - how much spectral limiting is needed before the coloration is reduced, to being insignificant?

The main interest in this thesis will then be to further investigate, how the spectra of reflections influence the audibility of coloration.

[3] see method chapter, section 3.2

The research question is the following:

To what degree do the spectra of reflections influence the detectability of comb filter coloration?

A major contributor to the spectral modification of reflections is absorption.

Therefore, the research question will be investigated from the perspective of the typical high-frequency absorption of reflections occurring in listening rooms.

A listening experiment is conducted to investigate how the absorption characteristics (modelled) of a typical absorbent influence the detection thresholds for the (comb filter) coloration.

Several practical questions are also raised prior to the listening experiment.

We also discuss, largely through analyzing the spectra of the various signals used for the experiment, how the coloration perception in each particular case depends on the relationship between the frequencies of the comb filter and the spectral content in original signal.

1.2 Clarification of terms

In this section I will try to give a more thorough explanation of some of the key terms used, that may be concerning the research question.

I have already given a short description of what is meant by the term *comb filter*, and in relation to that, its perceptual effect, coloration (a more thorough description of comb filter theory will be described in the theory chapter, section 2.1).

However, the more precise description of the perceptual effect of comb filtering is a change in *timbre*. So, when speaking of coloration what comb filter effects are concerned, coloration means *timbre change*. The question then is, what is *timbre*?

Cited in Rubak (2004:1), The American Standard of Acoustical Terminology has proposed the following definition of timbre: *"..the attribute in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar"*.

Kleiner (2014:216) notes that *timbre* is our subjective perception of spectral content. More specific, the timbre of a sound signal concerns its frequency structure and their inter-strength-relationship (frequency and balance between various parts in the spectrum). If the frequency structure and/or balance of these is somewhat changed *to a degree that is audible*, we should observe a timbre *change*. A subjective term, where its analogous physical term is (frequency) spectrum (Everest & Pohlmann 2015:56).

Of the two other relatable terms mentioned; *loudness* is our *perception* of how loud a sound is (not *actual* measured level), and *pitch* is our perception of where a sound can be ordered on a scale from low to high. Pitch is related to *frequency*, but the relationship is not linear.

Salomons (1995:1-2) has, with the definitions of pitch and timbre in mind, proposed the following definition for the color of a sound signal; *"The color of a sound signal is that attribute in terms of which a listener can judge that two sounds similarly presented and having the same loudness are dissimilar; it thus comprises the timbre, rhythm sensation as well as the pitch of the signal"*.

2 Theory

I will in this next section give an overview of the relevant theory within the fields of which the topic of this thesis resides.

2.1 Acoustics

2.1.1 Comb filters

A brief explanation of how comb filtering occurs, as well as its audible effect, was given in the introduction. In the following section we will have a closer look at the theory concerning comb filters.

A comb filter arises due to the superimposition of coherent [3] or nearly-coherent signals; i.e. a signal and a delayed, filtered, or scaled version of itself (Brunner 2007:1). This causes a change in the frequency response, or transfer function [4], of the signal (that is, in the new signal formed) (Everest & Pohlmann 2015:137).

Zölzer (2002:63) notes a comb filter has two tuning parameters; the size, or amount of delay *t*, and the relative amplitude of the delayed signal to that of the reference signal. The transfer function, as well as the difference equation of a FIR (finite impulse response) [5] comb filter, can be given as:

$$y(n) = x(n) + gx(n - M) \tag{2.1}$$

$$\text{with} \quad M = t / f_n \tag{2.2}$$

$$H(z) = 1 + gz^{-M} \tag{2.3}$$

Zölzer (2002:64) notes, the filters' time response consists of the direct and the delayed version of the combined signal. For positive values of *g*, the comb filter amplifies all

frequencies that are multiples of 1/t and attenuates all frequencies that exist in between. Consequently, the gain varies between 1 + *y* and 1 - *g*.

The transfer function (frequency response) of the filter will thus look like a comb (Figure 2.1) which is the reason for the name *comb filter*. Thus, which frequencies the comb filter amplifies and attenuates depends on the time delay, *t*, between the direct and delayed version.

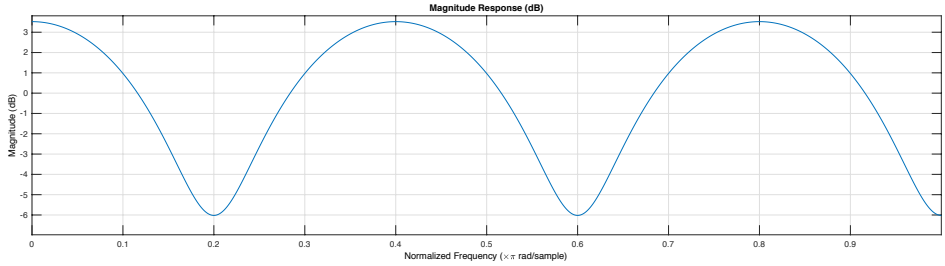


Figure 2.1. *The impulse response of a 5 ms FIR comb filter. The superimposing of coherent or nearly-coherent sound signals produces a comb filtered frequency response due to one signal being delayed in relation to the other. The resulting frequency response looks like the teeth of a comb, which is the reason for the name. A linear frequency scale is used.*

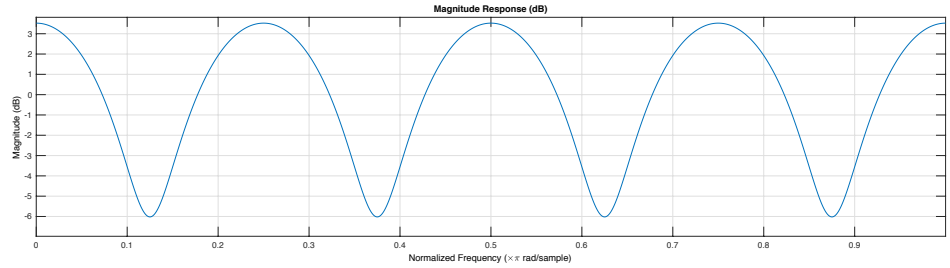


Figure 2.2. *A larger time offset in the delayed signal increases the number of interference events, and the peaks and dips are spaced more closely together.*

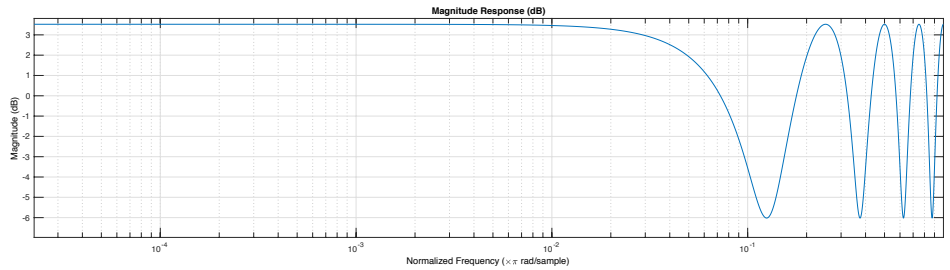


Figure 2.3. *The same comb filter as in figure 2.2. where a logarithmic scale is used.*

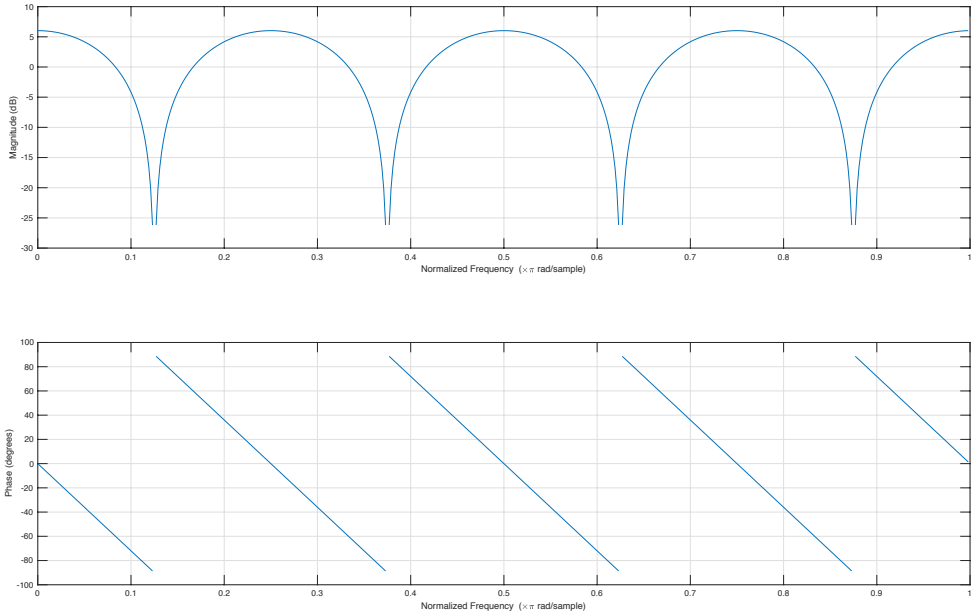


Figure 2.4. *Increasing the gain of the amplitude ratios in the filter changes the magnitude response and shape of the comb filter peaks and dips.*

By utilizing simple maths we can estimate the response of a comb filter. The first attenuation frequency occurs where $n = 1$ in $f = n/(2t)$, and t equals the delay time in seconds (Everest & Pohlmann 2015:141). Each successive dip will occur at odd multiples of n in $f = n/(2t)$, meaning 3, 5, 7, etc. Peaks in frequency response, on the other hand, occur at $f = 1/t$, with successive peaks at $f = n/t$, where n = 1, 2, 3, 4, 5, etc.

The smaller the delay time the higher up in frequency the first attenuation frequency moves. For very small values of t, usually below 1 milliseconds, the effect will be that of a low pass filter (Kleiner 2014:243). Figure 2.2 shows the response of a 1 ms comb filter, giving the effect of a low pass filter.

The distance between peaks or nulls is $1/t$ (Everest & Pohlmann 2015:150). As we then understand, the bigger the time delay, the smaller the distance between adjacent peaks and nulls. Increasing the time delay on the delayed part will thus also increase the number of interference events proportionally. Figure 2.3 shows a comb filter where the number of interference frequencies has doubled to that of figure 2.1. When the delay time becomes large, closely spaced peaks and narrow nulls might even out the response, making the effect of comb filtering less noticeable (Everest & Pohlmann 2015:143). When the delay time

becomes substantial the ear will resolve the effect of delaying a signal in relation to its original, as two completely separate events. Halmrast (2020:3) suggests as a rule of thumb the border for echo perception is roughly around 50 ms. The reason is, a 50 ms comb filter gives a comb peak-to-peak bandwidth of 20 Hz, which is generally known to be lowest frequency audible for humans. However, it is reasonable to believe the outer border for comb filter effects in practice is lower than 50 ms - especially in listening rooms (which reasons will be mentioned in section 2.2.1).

Comb filtering can only occur in distributed energy signals (Everest & Pohlmann 2015:139). Thus, pure tones (sine waves) cannot produce comb filtering. The combing of sine waves only result in changes in volume.

However, we can employ sine waves to demonstrate the effect of amplitude ratios on comb filter peak height and null depth. Everest & Pohlmann (2015:139) note that adding two sine waves of the same frequency, amplitude, in phase, doubles the amplitude of the new sine wave formed relative to the original combing waves. This doubling equates to an increase of 6 dB relative to either component by itself. The amplitude of null frequencies, on the other hand, will be at a theoretical minimum of minus infinity as the sine waves cancel at phase opposition, Everest & Pohlmann (2015:151) note. Figure 2.4 illustrates the magnitude response of a theoretical comb filter, where the gain of the delayed sound component is equal to the gain of the original.

2.1.2 Comb filtering and signal type - noise and music

As noted, comb filtering only occurs in signals with distributed spectra (Everest & Pohlmann 2015:139). Sound signals with distributed spectra are music, speech, and noise.

In literature concerning the type of signal on the perception of delayed sounds, we make the most apparent distinction between discontinuous (impulsive) and continuous sounds. Typical examples of the former are clicks and short pulses, whereas noise signals are prime examples of the latter.

Olive & Toole (1989:13) note that discontinuous and continuous sounds produce fundamentally different absolute thresholds as a function of reflection delay time. Kleiner (2014:243) noted that, when the sound is longer than a substantial part of its repetitions, the perception will be spectral. However, if the sound is much shorter than its repetitions, the effect comes in the time-domain.

As implied, the *spectral content* of the combing signals is substantial for coloration perception. Two aspects concerning spectral content are 1) [4]*bandwidth*, 2) [5]*spectral density*. Sound signals can vary considerably in their bandwidth; from the single frequency sinusoid to broadband white and pink noise signals.

Theoretically, without having taken into account other signal-related factors, it is reasonable to assume that spectrally dense (rich) signals are more exposed to comb filtering than signals of a low spectral density. The reason is, when the [6]*partials* of a signal are tightly packed, there is more information for the comb filter to change.

We can assume the same regarding bandwidth and comb filtering, as with spectral density; when the signal has a broad frequency range there is naturally a higher probability that comb filtering will affect the signal.

Rubak (2004:11) notes that white noise is well suited for revealing coloration in the frequency domain. It is reasonable to assume this observation inherits from the following;

1) Noise signals produce continuous spectra (Zwicker & Fastl 2006:3). Spectral continuity should make coloration easier to detect due to comb filtering being a steady-state phenomenon. Also, the energy distribution of white noise is uniform throughout the spectrum (Everest & Pohlmann 2015:85). Due to the non-linearities of the human ear (discussed in 2.3.2), white noise will appear more intense in the high-frequencies compared to the low-

[4] describes the frequency range passed by a signal (Everest & Pohlmann 2015:596).

[5] describes the closeness of the frequency components within a signal

[6] another term for the individual frequency components within a signal

frequencies (which also should be an advantage when spectrally limiting the high frequencies).

2) Noise signals are spectrally dense. Also, the spectral density of its long term spectrum is independent of frequency. (Zwicker& Fastl 2006:3).

3) White noise is a wide bandwidth signal. Zwicker& Fastl (2006:3) mention it is often practical to band limit the noise to 20-20 000 Hz (there is little sense in passing frequencies above 20 000 Hz, as they are outside of the human hearing range).

When we use white noise as input signal and add one reflection delay giving (comb filter) coloration, the resulting spectrum is often called *harmonic cosine noise* (cosine modulated). The spectrum is said to be periodic rippled. The resulting spectrum will essentially look like what is already illustrated in figure 2.1, 2.3, 2.4, respectively, dependent on the delay time and strength of the added reflection.

The primary perceptual effect of comb filtering in noise signals is that of so-called [7]*residual pitch* or *repetition pitch* (Brunner et al. 2007:2). The reason why we observe a sensation of pitch is that noise signals contain little or no tonal information. According to Rubak (2004:11), the dominant spectral region for pitch perception in cosine noise is around 4/T, where T is the delay time of the reflection. This proposed dominance region thus corresponds to the fourth comb reinforcement frequency.
Of course, the color sensation of comb filtered noise changes with the delay time (Salomons 1995:3). For delay times smaller than 1,0 ms, the sensation is only a high hiss, which becomes lower with increasing delay time, Salomons (1995:3) notes. Towards 1 ms, the hiss turns into the previously noted pitch sensation, which becomes very distinct for delay times from 1-20 ms, making it possible to play a melody by varying the time delay (Salomons 1995:3). Above 20 ms, the color sensation becomes that of a rattle sensation, also called infra pitch. A further increase in delay time and the frequency of the rattling rhythm decreases. Salomons (1995:4) notes, at these delay times, the coloration is no longer spectral, and the pitch sensation thus disappears.

[7] a well defined strong pitch sensation in complex signals

In music signals, on the other hand, the primary perceptual effect of comb filtering is timbral differences (Brunner et al. 2007:7). Apparently, the term *tone coloration* is also used (Barron 1971; Barron et al. 1981), cited in Rubak (2004:2).

Compared to noise, music signals are often highly transient (Everest & Pohlmann 2015:137). Loud passages follow faint passages and vice-versa (Zwicker & Fastl 2006). According to Everest & Pohlmann (2015:137), comb filtering may have limited application to music if only considering the transients of music and the fact that comb filtering is a steady-state phenomenon. Thus, the coloration perception might change rapidly from one point in time to the next. Also, the different complexities of various music signals are substantial: "*from the near sine wave-like simplicity of a single instrument to the highly intricate tonalities of a symphonic orchestra, in which each instrument varies its tonal texture according to the notes"*, Everest & Pohlmann (2015:75) note.

We might view music signals as an intermediate between a continuos noise signal and a discontinuous click or pulse sound. Kleiner (2014:205) notes that, from the perspective of acoustics, we can consider music as wide bandwidth signals modulated by low-frequency signals. Reverberation in rooms will diminish the modulation depth, making them more continuous.

Consequently, the threshold for detecting comb filter coloration in music signals should be highly variable, and are much dependent on the content of the signal relative to the delay time of the reflection. It becomes clear that for coloration to be clearly perceptible (in music signals), the interference frequencies of the comb filter must coincide with "vital" frequency-areas of the signal; the comb filter must affect frequency-areas that our ear notices as substantial for recognizing timbre of the signal - otherwise we cannot perceive the effect clearly.

We can perhaps criticize prior studies for the lack of discussion on this matter.

However, apparently, Olive & Toole found in an experiment, cited in Everest & Pohlmann (2015:107), that classical music produced reflection hearing thresholds close to that produced by pink noise. Thus, one suggested that pink noise is a reasonable surrogate for (noisy) classical music in measurements. As such, it would be reasonable to assume that a complex,

fairly continuous music signal should be able to produce reflection hearing thresholds closer to noise signals. This agrees with the assumption made concerning spectral content and coloration perception mentioned.

The findings of Brunner et al. (2007:6), Kuhl, and several others, support the assumptions regarding spectral content and the audibility of comb filter coloration; in general, respondents are particularly sensitive to spectral changes in noisy signals.

2.2 Room Acoustics

2.2.1 Comb filtering due to reflections

Salomons (1995:2) notes that reflections are often the predominant cause of coloration. In rooms, reflections producing comb filtering come from plane, nearby surfaces. We call these *specular* reflections because they behave the same as light reflects from a mirror; the incidence angle equals the angle of the reflection (Everest & Pohlmann 2015:97).

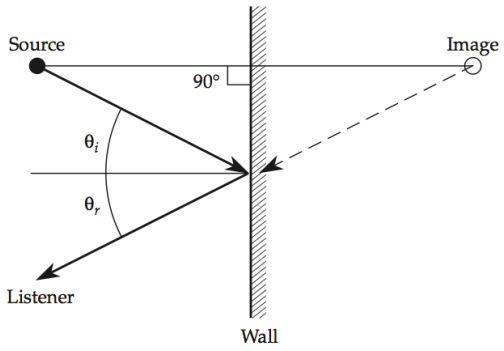


Figure 2.5. *Specular reflection creating a virtual mirror image of the sound source. Angle of incidence equals the angle of the reflection. (From Everest, F.A. & Pohlmann, K.C., Master Handbook of Acoustics, McGraw-Hill, 2015.)*

Multiple reflections are created when a sound strikes more than one surface. Also, images of images are created when the reflection from one surface strikes another.

In the introduction we mentioned why comb filtering mainly poses a problem in small rooms (and not in larger rooms such as concert halls). Short distances to room boundaries and other surfaces in small rooms facilitate reflections with early arrival times at the listener's ears, which leads to acoustic interference between direct sound and its reflections.

Acoustic interference (due to reflections) might also occur in larger rooms like concert halls (if the reflections are sufficiently short). However, in small rooms, practically all first-order reflections arrive within a time-frame to produce coloration in the frequency-domain. Kleiner (2014:243) notices sounds coming within a window of approximately 30 ms will be processed as one signal. Thus, the coloration results in the frequency-domain. However, in small rooms, most first-order reflections are much shorter; many will arrive within a time-frame of 10 ms after the direct sound. Salomons (1995:4) mentions what happens when multiple reflections are present; if the reflections are regularly spaced in time, spectral peaks in the colored signal sharpen and the coloration becomes more noticeable. On the other hand, when the reflections arrive less regularly spaced, they tend to cancel each other and the coloration effect becomes less noticeable. However, in small rooms several of the shortest reflections might end up arriving more or less at the same time. When this happens, they will add up, creating a clear coloration perception Halmrast (2020) notes. However, due to that multiple reflections tend to cancel each other rather than reinforce, a single, strong reflection might be of more concern in producing coloration than many spread out in time.

In theory, a substantial number of factors might affect the audibility of comb filter coloration. In the next sections, we will look at the most relevant factors regarding this, and which may concern the research question presented initially.

2.2.2 Absorption

The materials of walls and ceilings seldom have the acoustic properties to treat reflections completely. Acoustic treatment will, therefore, in most situations be necessary to control the sound transmission.

We categorize acoustic materials according to their acoustic properties. Dependent on such properties materials can both reflect, absorb, and diffuse sound.

We will only deal with absorption because we investigate the research question from the perspective of absorption.

We can group absorbers into *natural absorbers* and *added absorbers* (Kleiner 2013:165). Natural absorbers exist naturally within, or as a part of the room. Examples of natural absorbers are walls, ceilings, carpets, floors, drapes and furniture.
On the other hand, added absorbers are made to control sound transmission. Kleiner (2013:165) notes that we employ *added absorbers* to control reverberation time and other acoustic properties, that is, early reflections.

All sound absorbers function principally in the same way; they dissipate sound energy (in the form of heat). When sound waves strike the surface of a material, some sound energy is reflected, and some is absorbed by the material. Because the material will be denser than air, the sound that is absorbed will be directed downwards (Everest & Pohlmann 2015:183). Kleiner (2014:121) notes that the energy loss (dissipation) by absorption is proportional to the particle velocity of the incoming wave.

While we can group natural and added absorbers, as previously mentioned, it is perhaps better to draw a line between *resonant* and *non-resonant* absorbers, according to Kleiner (2013:165). We primarily use *resonant absorbers* for the absorption of frequencies below 400 Hz (Kleiner 2013:177). They usually consist of a surface membrane with an inner air cavity. The theory of resonant absorbers will not be covered here due to it does not concern the research question.

We often refer to *non-resonant* absorbers as *porous absorbers*. Porous materials commonly consist of very thin fibers that are typically compressed and glued together to form a plate of a specific thickness (Kleiner 2014:127). When a sound wave strikes the surface of a porous material, sound energy dissipates through tiny pores (Everest & Pohlmann 2015:191). The air friction on the surface of the fibers converts the acoustic energy into heat (Kleiner 2014:127).

More energy dissipates as the wave travels further into the boundary layers.
Porous absorbers include materials such as mineral wool, textiles, and carpets.

Porous absorbers are generally proficient at absorbing high-frequency sound energy, and poor at absorbing low-frequency sound energy (Everest & Pohlmann 2015:191). The effectiveness of absorption depends on such factors as material thickness, the airspace behind, and material density of the absorbent (Everest & Pohlmann 2015:191).

To absorb well, the porous material must be of a thickness that is comparable to the wavelength of the sound (Everest et al.). When one places a porous absorbent on the surface of a solid wall, the majority of the energy loss occurs at a quarter wavelength distance from the wall behind, Kleiner (2014:121) notes. Thus, thicker panels, sometimes with a rear air cavity, are often preferred over thinner panels (Everest & Pohlmann 2015:192). However, an argument that thicker panels are the best absorbents, holds primarily for low frequencies only. Figure 7.2 p 170 Kleiner (2013). E.g., above 500 Hz, increasing the absorbent thickness from two inch to four inch has little effect, but considerable effect below 500 Hz, as one increases thickness, (Everest & Pohlmann 2015:195).

The distance from the absorbent to the reflective surface behind can be just as important as material thickness, for the effectiveness of absorption.

By placing the porous material a quarter wavelength from the wall, or at odd multiples of the quarter wavelength, it will achieve maximum absorption at the corresponding frequency because the particle velocity is maximum at the absorbent and the greatest frictional losses will occur, (Everest & Pohlmann 2015:195).

A particular type of porous material is glass-fiber. This material can be of both high-density, as well as low-density materials. Semirigid boards of glass-fiber excel in sound absorption and are thus widely used for room treatment.

Absorption coefficients rate a material's effectiveness in absorbing sound (Everest & Pohlmann 2015:184). The absorption coefficient is angle dependent; it varies according to the incident angle of sound striking the surface. Therefore, absorption coefficients are often averaged over all incident angles.

The sound absorption coefficient alfa is defined as:

$$\alpha = \frac{W_{abs}}{W_{inc}}$$

where W_{inc} is the power of the impinging sound and W_{abs} is the power removed by the absorber (Kleiner 2013:166). By multiplying the absorption coefficient by the surface area of the material exposed to sound, we obtain the sound absorption *A* provided by a the surface area of a material (Everest & Pohlmann 2015:185).

As such:

$$A = S\alpha$$

where $A =$ absorption units, sabine or metric sabins

$S =$ surface area, ft^2 or m^2

$\alpha =$ absorption coefficient

Absorption coefficients also vary according to frequency. In general, the effectiveness of absorption increases with increasing frequency. The six standard frequencies are 125; 250; 500; 1,000; 2,000; and 4,000 Hz (Everest & Pohlmann 2015:185).

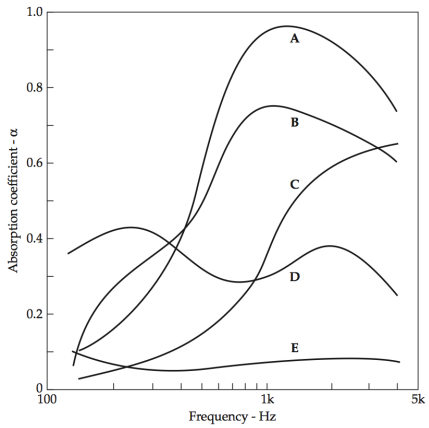


Figure 2.6 *Absorption coefficients of common porous materials A, B, and C show the typical characteristic of good high-frequency absorption and poor low-frequency absorption. A) High-grade acoustical tile. B) Medium weight (14 oz/yd^2) velour draped to half area. C) Heavy carpet on concrete without padding. D) Coarse concrete blocks, unpainted. E) Coarse concrete blocks, painted. (From Everest, F.A. & Pohlmann, K.C., Master Handbook of Acoustics, McGraw-Hill, 2015.)*

2.2.3 The effect of spectrum of the reflections

The effect of spectrum (of reflections) on the detectability of comb filter coloration is the topic under investigation in this thesis. We mentioned in the introduction chapter why this has relevance; room boundaries and other surfaces in most rooms produce reflections that spectrally deviate from that of the direct sound. Sounds become both reflected, diffused, and absorbed because of the different acoustic properties of different surface materials in the rooms, in addition to the relationship between frequency wavelengths and surface area reflections are impinging. Spectral deviations typically occur in the form of attenuation of high-frequencies due to absorption.

However, it appears that the effect of spectrum of reflections on their perception is not something that previously has been of much interest. As such, Olive & Toole (1989:11) called for more extensive research on this topic.

Olive & Toole (1989:11) investigated the effect of spectral modification of a single [8]*lateral reflection* on its detectability. They determined absolute thresholds (for spectral differences) and image shift thresholds (localization) as a function of its spectrum.

The single test reflection was created artificially in an anechoic environment. The reflection was low pass filtered at various frequencies to simulate the effect of high-frequency sound absorption in rooms (for details regarding appropriate methods, see method chapter). Figure 2.7 shows the results of this investigation.

The reflection incident angle was 50° azimuth and 0° elevation relative to the median plane of the listener, which is consistent with the event of a reflection from a sidewall. Olive & Toole (1989:11) employed test signals ranging from short pulses of various durations to speech, to pink noise. The test reflection was lowpass filtered at frequencies of 20 kHz, 4.4 kHz, 1.2 kHz, and 600 Hz. The time offset for the reflection was fixed at 4 ms.

It is reasonable to assume that, after some dramatic lowpass filtering, we should observe a change, in the form of an increase of threshold values. I.e. the effect of the reflection should

[8] sideways reflection

be harder to detect. However, as Figure 2.7 shows, except for the 5/s and 20/s pulses, lowpass filtering produced only a marginal increase of detection threshold values for the reflection.

On the other hand, while low pass filtering produced very little change in reflection hearing thresholds, perceptual changes were nevertheless observed as one attenuated the high frequencies of the reflection.

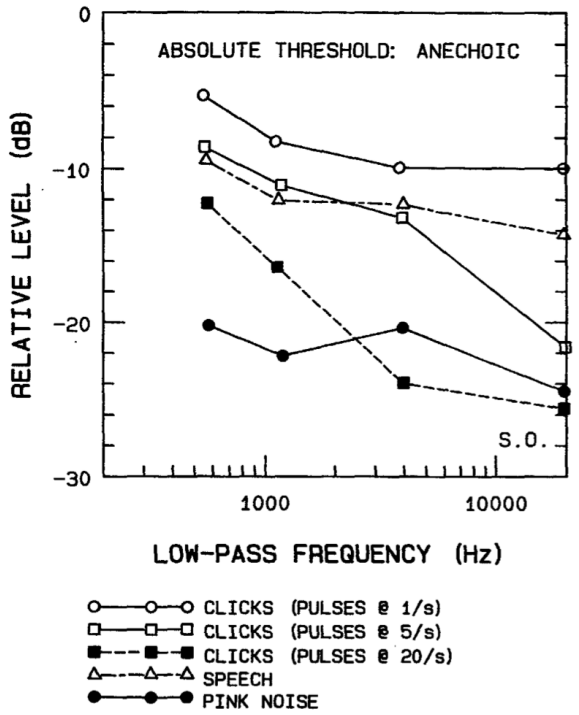


Figure 2.7 *Absolute thresholds for a lateral reflection (*50° *H,* 0° *V) that has been low pass filtered at 20 kHz, 4.4 kHz, 1.2 kHz, and 600 Hz. The reflections were all delayed by 4 ms and the signals ranged from discontinuous pulses 1/s to continuous pink noise. (From Olive S.E., & Toole F.E., The Detection Of Reflections in Typical Rooms, Audio Engineering Society, 1989.)*

The observed supports the assumption made in the introduction: spectral limiting of some magnitude does not necessarily produce a significant change in reflection detectability - but one can observe perceptual changes nevertheless.

However, Olive & Toole (1989:11) underlined that more research is needed. It is also not clear whether the detection was solely based on *just noticeable differences* in timbre, as is of interest in this thesis. Olive & Toole (1989:3) noted: "*in general, the absolute threshold was of interest. In this, listeners were instructed to respond to any audible change in the nature of the sound itself or of the sound field"*.

The most comprehensive research until date concerning the effect of spectrum of reflections is likely that by Bech (1995). He investigated the influence of filtering on hearing thresholds for individual reflections affecting timbre of reproduced sound, in the presence of a complete sound field (simulated). The experiment had a similar layout to that of Bech (1994).
Bech (1995) investigated the effect of spectrum of reflections by filtering the transfer function of individual test reflections. He derived the filtering from measured frequency-dependent absorption coefficients from the surfaces of a typical domestic room. Bech (1995) also took the off-axis frequency response of the modelled loudspeaker into account in the filtering of each reflection's transfer function. Also, the absorption coefficients were angle-dependent. As such, Bech (1995) could perhaps observe the cumulative effect (if any) of both frequency-dependent absorption as well as loudspeaker off-axis frequency response on threshold values of reflections affecting timbre.
Figure 2.8 shows the resulting transfer functions of selected reflections in the experiment conducted by Bech (1995). Figure 2.9 and 2.10 from Bech (1995) show the established threshold values for each reflection with and without filtering, for pink noise and speech, respectively.
Figure 2.9 reveals that for pink noise, the introduced filtering only changed the hearing thresholds for reflections No. 5, 7, 9 significantly (in the form of a significant increase in threshold values). Figure 2.10, on the other hand, reveals that the introduced filtering had hardly any effect at all on hearing thresholds of the reflections, when speech was employed as the test signal.

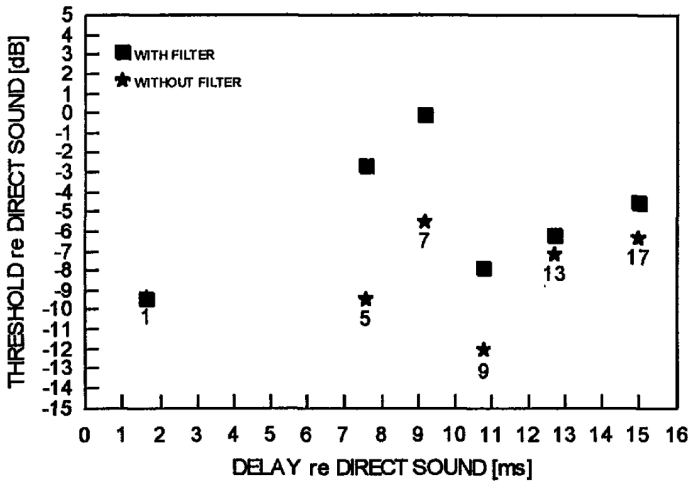


Figure 2.8 *Magnitude response of the filter transfer functions implemented for selected individual reflections (solid lines) from the sound field experiment conducted by Bech (1995). The reflections had different time-offsets and incident angles relative to the direct sound and listening position. The dashed lines represent the response of the same reflections from an earlier experiment conducted by Bech (1994), where the test reflections had had the same spectrum (frequency independent attenuation) as the direct signal. (From Bech S., Perception of Reproduced Sound: Audibility of Individual Reflections in a Complete Sound Field, II, Audio Engineering Society, 1995).*

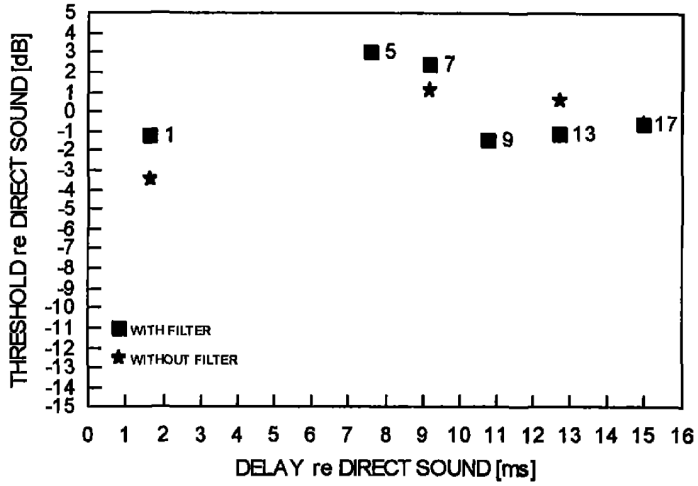


Figure 2.9 *Detection thresholds for a pink noise signal for reflections numbered 1, 5, 7, 9, 13, and 17 with and without filtering. Confidence intervals (95%) are ± 0.96 dB with filtering*

and ± 0.98 dB without filtering. The confidence intervals are based on the variance between blocks and the mean values on the same subjects for both experiments and 400 trials per subject. (After Bech S., Perception of Reproduced Sound: Audibility of Individual Reflections in a Complete Sound Field, II, Audio Engineering Society, 1995).

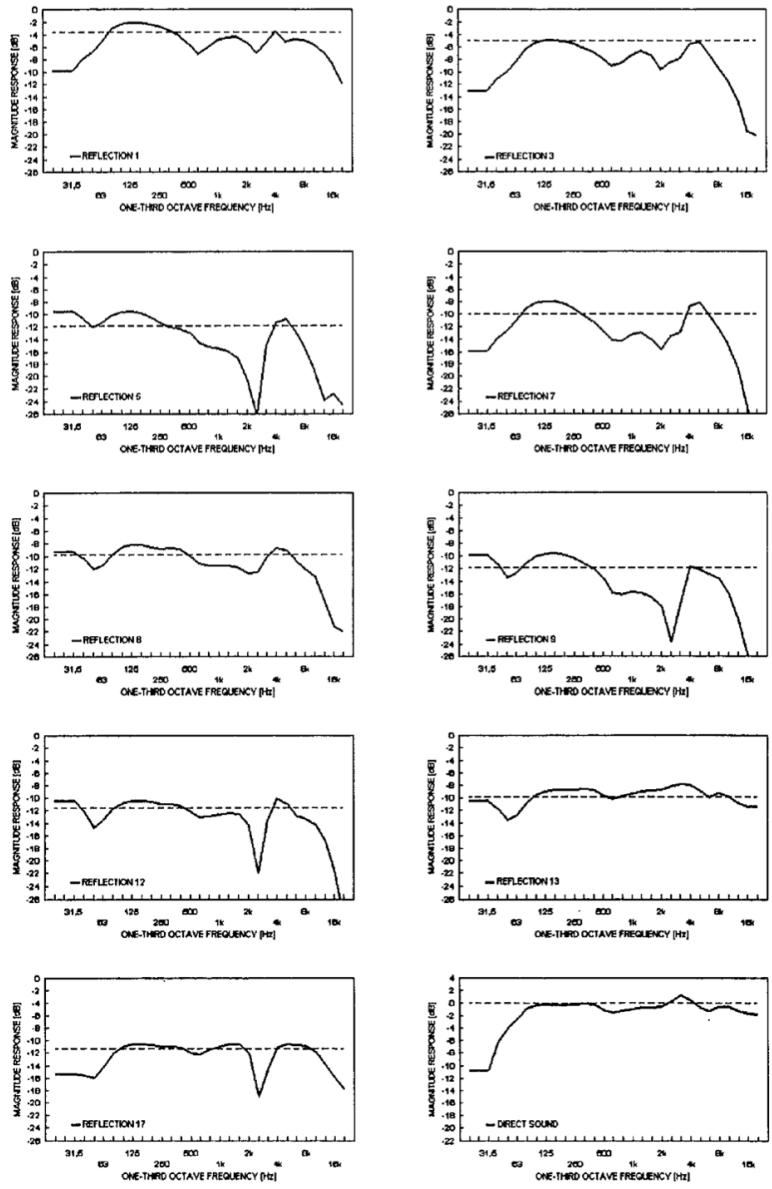


Figure 2.10 *Detection thresholds for a speech signal for reflections numbered 1, 5, 7, 9, 13, and 17 with filtering and reflections 1, 7, 13, and 17 without. Confidence intervals (95%) are ± 0.98 dB with filtering and ± 1.07 dB without filtering. The confidence intervals are based on the variance between blocks and the mean values on the same subjects for both experiments and 300 trials per subject. (After Bech S., Perception of Reproduced Sound:*

Audibility of Individual Reflections in a Complete Sound Field, II, Audio Engineering Society, 1995).

Bech (1995) proposed that the significant elevation in thresholds values observed for only some reflections (No 5, 7, 9) was due to the removal of energy in vital mid- and high-frequency regions by the filter transfer functions of those reflections. He suggested that the reflection hearing thresholds were determined by the spectral changes in the dominant range from 500 Hz to 2 kHz, mainly.

However, the effect of filtering was only observed for the pink noise signal (not in speech), as can be seen in figure 2.9 (noise) and 2.10 (speech). Bech suggested this observation was either due to 1) the lack of energy in the speech signal in frequency ranges affected by the filtering, or 2) the subjects responded to a loudness cue instead of a timbral cue for the speech signal.

Nevertheless, the conclusion from the study of Bech (1995) was that the introduced filtering of the individual transfer functions had a significant effect on reflections where spectral changes occurred in the mid- and high-frequency ranges.

We should take some precautions when attempting to make direct comparisons between the study by Bech (1995) and that by Olive & Toole (1989). By the look of it, Bech's study is perhaps the more realistic; most notably, taking into account the filtering derived from actual frequency-dependent absorption coefficients.

1) While the filtering introduced in these two studies may give the same effect, the method employed is different 2) Bech (1995) established the individual threshold values in the presence of a complete sound field. The study by Olive & Toole (1989), on the other hand, was conducted in an anechoic environment without any other reflections than the test reflection present. As such, we observe, as expected, an elevation in established threshold values from that by Bech (1995) relative to that by Olive & Toole (1989). Then, we also have some other factors that may or may not affect the results, such as 1) The time offset employed for the test reflections. 4) The incident angles of the test reflections.

Burgtorf (1961), cited in Buchholz et al. (2001:3), also investigated the effect of spectral differences between direct sound and test reflection. He concluded that the masking effect of the direct sound is most apparent if the spectral distribution of the direct sound and the test reflection coincide.

For the audibility of comb filter coloration it is reasonable to expect an inverse pattern, as proposed in the section concerning signal type; the coloration perception should be clearest when a signal and the test reflection has overlapping frequency spectra. As such, this is also in agreement with the finding of Bech (1995).

Brunner et al. (2007:6) notice that when the delayed signal undergoes filtering *"the peaks and dips in the frequency response will be less pronounced and the established threshold can be regarded as a lower boundary"*. The peaks and dips of the comb filter will be less pronounced because there is less information for the comb filter to change when one limits the bandwidth of the delayed signal.

However, it remains to investigate the extent of spectral limiting that needs to take place before we see a significant change in the established thresholds.

2.2.4 Other factors that influence coloration perception

While they are not the study in this thesis, we also need to consider factors other than absorption, both room-related and non room-related, that influence coloration perception. Especially so, when the perspective of the research question is *reproduced sound*.

The reason why we need to consider these factors is that they have importance for the choices we make before and during the investigation.

1) Loudspeakers. There are, in particular, two aspects of loudspeakers that determine the impact of reflections on the sound. These are:

a) *Position and configuration*. The position of speakers matters on the influence of reflections on the timbre of reproduced sound. In particular, the distance to the front wall is essential.

Dammerud (2013:23) notices (as will be mentioned further down) that we, in particular, should avoid distances of 0,3 - 2,2 meters to the front wall. The reason to avoid such distances is, they will give interference, in the form of reduced bass (due to comb filtering) in the range from 39-286 Hz.

The *configuration* also matters. In stereo listening, the two loudspeakers and the listener should ideally form an equilateral triangle (Kleiner 2014:334). In an equilateral triangle the sides are of equal length; the distance is the same between speakers as is between speakers and listeners' ears. The equilateral triangle results in a 30-degree angle between the speakers and creates a phantom sound source between them (Kleiner 2014:334). Nearfield-listening also helps minimize the impact of reflections.

We should also place speakers symmetrically according to the boundaries of the room. While symmetrical speaker set up first and foremost concerns, and is paramount, for the stereo stage, it might also affect timbre (of reproduced sound). However, symmetry is not possible in all rooms. Kleiner (2014:334) notes, in asymmetrical rooms where the walls on the left and right of the listener have different acoustic properties, the stereo stage may become biased towards the wall that reflects the most.

b) *Directional characteristics*. The directional characteristics of a speaker will also affect the impact of reflections (and so also on coloration). E.g., a loudspeaker that is omnidirectional at all frequencies will radiate sound (equally) towards front, sides, and back at all the frequencies reproduced by the speaker. Dependent on the distance to room boundaries, the reproduced sound may be affected by comb filtering from the front wall and perhaps sidewalls, due to little level difference between direct- and reflected sound. However, the typical loudspeaker in use is omnidirectional only at low frequencies, which then primary poses a problem for comb filtering in low-frequencies due to short distance to the front wall, as mentioned.

Also, due to high-frequencies being very directional, the tweeters should be placed at approximately ear level.

2) Reflection incident angle. Having not taken loudspeaker directivity into account, we generally find that the audibility of a reflection decreases when it arrives from the same direction as the source signal. The reason is the *masking effect*[1] provided by the direct signal; it will be greatest when the source signal arrives from the same angle as the reflection. According to Buchholz et al. (2001:2), when the reflection comes from the same direction as the source, the detection threshold for the reflection may increase by as much as 10 dB relative to other directions.

According to Kleiner (2014:292), the average listening room typically has a volume in the range of 25 to 75 m^3. In rectangular rooms with plane, rigid walls, floor, and ceilings, strong interference (comb filtering) often occurs when floor- and-ceiling reflections arrive from angles close to the median plane of the listener (Kleiner 2014:297). Bech (1994) found in one of his sound field experiments that only first-order floor and ceiling reflections contributed on an individual basis to the timbre (color) of reproduced sound, with speech as a test signal.

However, the audible coloration from repeated ceiling-floor reflections diminishes when the arrival angle relative to the median plane decreases for higher-order reflections, according to Kleiner (2014:297). Also, carpets (on floors) may help to reduce floor-reflections somewhat.

Reflections from surfaces located between loudspeakers and the listening position may also lead to interference (comb filtering) at the listening position. In the studio control room the mixing desk is one such surface. However, Dammerud (2013:22) notes that we can often avoid mix desk reflections by placing speakers and desk accordingly, as well as proper angling of the desk surface.

There are several reasons why reflections close to median plane may lead to strong interference (comb filtering) at the listening position: 1) Our ears may not be able to separate reflections close to median plane from the direct sound by binaural means (Kleiner 2014:297). 2) Floor- and ceiling reflections often coincide and result in an approximate doubling of reflection level (Kleiner 2014:297). 3) For mix desk reflections the path length difference between reflections and direct sound will be minimal. Thus, the level difference

between reflections and direct sound will also be small, resulting in a strong comb filter at the listening position (Dammerud 2013:22-23).

In loudspeaker playback, front wall reflections mainly lead to comb filtering in the low-frequencies. The reason is the omnidirectionality of most speakers in this frequency-area (Dammerud 2013:23).

Comb filtering from front wall reflections will be perceptible as either more bass (a bass lift) or less bass (attenuation). The response is dependent on the phase between the low-frequency direct- and reflected sound, as previously mentioned.

Dammerud (2013:22) notes that it is often difficult to avoid comb filtering from front wall reflections due to the difficulty in absorbing low frequencies. The best way to deal with comb filtering from front wall reflections may, therefore, be either: 1) Place speakers right up against the front wall. Placing speakers next to the front wall will push the first attenuation frequency of the comb filter sufficiently high up in frequency where the speakers are more directional. The level difference between reflection and direct sound then increases, and the comb filter weakens. 2) Place speakers at a substantial distance from the front wall where the level of front wall reflections is sufficiently weak. However, only at distances exceeding two meters from the front wall, level differences (between direct and reflected sound) are big enough to sufficiently weaken the comb filter, Dammerud (2013:23) notices.

In the studio control room environment we usually treat the rear wall with diffusive material (Kleiner 2014). Diffusion on the back wall helps in avoiding comb filtering (between direct sound and rear wall reflections), as well as creating a “livelier” and “bigger” sounding room. In the private room we typically place sofas (or other furniture) by the rear wall. Perhaps we also hang carpets on this wall. Both furniture and carpets will help to tame strong reflections leading to coloration from the back wall.

However, Dammerud (2013:23) notices that due to the usual substantial path length difference between direct sound and back wall reflections (which are usually larger than reflections from other surfaces of the room) the resulting comb filter may be weak (than from

other surfaces of the room). Regardless, one should not write off comb filtering from the back wall.

On the other hand, sidewalls are often left untreated, especially in private listening rooms. The primary effect of strong side wall reflections is an alteration of phantom center (shifts the phantom center towards the side if reflections are not symmetrical) and a change in auditory source width, Kleiner (2014). For speech signals, in particular, a broadening of the source is expected with strong side wall reflections.

However, strong sidewall reflections might also give coloration at the listening position. Bech (1994) found in one of his sound field experiments that when the source signal was broadband (noise), sidewall reflections individually contributed to the timbre of reproduced sound (in addition to floor and ceiling reflections). The finding that sidewall reflections contributed to timbre for broadband sounds and not for a narrowband sound (speech) on an individual basis also confirms the observation from other researchers; broadband signals make for easier spectral recognition than narrowband sounds.

The findings of Bech (1994) are evidence that sidewall reflections may be of concern giving unpleasant coloration of the reproduced sound. Also, Dammerud (2013:23) notices this.

However, in the event of a single sidewall reflection (only from one side), our hearing will be able to separate it from the source by binaural means, dependent also on the presentation of the source signal. As such, we might not be able to hear any coloration. In addition to that, if the reflection is strong it will have a primary effect on localization, shifting the image towards the side of the reflection, as mentioned.

However, the perception might change if two simultaneous reflections of equal level occur; one from left and one from the right sidewall, “mirroring” each other. We might then observe the following: 1) Even though the situation is not entirely the same as in coinciding floor- and ceiling reflections (as previously outlined), the two sidewall reflections might in sum create a stronger coloration perception. 2) Because they arrive approximately at the same time at the listeners’ ears, one on the left, the other on the right (and with the reservation that they have the same spectrum), our hearing may not be able to separate them as two separate reflection

events.

Dependent also on other factors such as the type of source signal, and its bandwidth, the situation outlined above might lead to comb filter coloration at the listening position.

3) Reflection delay time. In section 2.1.1 we discussed the effect of delay time on the audibility of comb filter effects. We showed, in section 2.1.1, how increasing the amount of delay of the repetition will affect the comb filter response. In section 2.3.8 we discuss how we can evaluate the audibility by seeing the comb filter peak-to-peak frequency in relation to our ears' *critical bandwidth*.

As noted, the perception will be spectral as long as the direct sound component is longer than a substantial part of its repetitions. According to Halmrast (2020:2), the border for echo perception might be as high as 50 ms, as noted. The reason is that the distance between two comb peaks at this time-offset is 1/50 = 20 Hz, which is known to be the lowest perceivable frequency for humans.

Of course, in listening rooms the time window for the perception of comb filter effects would be much shorter. Kleiner (2013:243) mentions 30 ms as the time-window for spectral recognition. However, this window would be more or less dependent on the total room size. Specifically, in small rooms, large reflections lose [9]*coherence* to the direct signal, becoming part of the diffuse-field, such that no comb filtering can occur (Brunner et al. 2007:6). The loss of coherence comes from the reflections being spectrally modified by the surfaces of the rooms.

Another factor that will decrease the audibility of comb filter effects in rooms is the attenuation of reflections due to distance. When the path length difference between the direct- and reflected signal is substantial, the comb filter produced will be weak, and as such, the effect will be negligible.

4) Other reflections and reverberation. Because we will not investigate this factor in this thesis, this section will be short. However, it does matter when seeing the research

[9] two signals are perfectly coherent if their frequencies and waveform are identical and their phase relationship is constant

question in perspective. In a normal room there is usually no such thing as one single sound field component (reflection) occurring (Kleiner 2013:115). We need to consider what happens when multiple reflections are present.

We mentioned previously what happens when multiple reflections that may produce comb filter coloration on an individual basis, are present. If the reflections occur at more or less the same time, i.e. their arrival times are very close, they are expected to add up, producing a clear coloration perception. On the other hand, if they are irregularly spaced, they tend to cancel each other, which will reduce their effect, as mentioned.

Also, we need to consider that not all reflections happening in a room contribute to the timbre of the sound. In many situations it is reasonable to assume that they will mask the presence of each other, especially if the room is highly reflective.

5) Presentation level. Research shows that the presentation level of the direct signal is of importance for the audibility of its reflections. According to Buchholz et al. (2001:3), for noise signals, the sensitivity for the presence of a reflection increases linearly with direct signal level, which implies an easier detection of reflections in loud sounds than in soft sounds.

6) Individual differences among listeners. The ability to detect perceptual effects of reflections may vary greatly among listeners in listening tests. Brunner et al. (2007) found the hearing thresholds for reflections to vary as much as 10 dB between individuals. In this regard, we should perhaps expect to see a distinction in performance between experienced listeners and non-experienced listeners. However, both Olive & Toole (1989), Schubert (1969), and Brunner et al. (2007) did a somewhat remarkable observation in that prior experience in critical listening had little effect on the results. The range of thresholds was as large among experienced listeners, as non-experienced listeners. Olive & Toole (1989) suggested that this observation has to do with critical listening tasks are such a focused activity that once we are focused on the listening, prior experience matters less.

Bech (1994) conducted an extensive training program for his listeners before his main threshold experiments. However, Bech (1994) did not report the effect of training on

performance, other than the reason for such training - to ensure the hearing thresholds had reached an asymptotic level for each individual.

2.3 Human perception of sound

In short, *psychoacoustics* is the study of our hearing system as a receiver of acoustical information (Zwicker & Fastl 2006). It combines the study of auditory physiology and acoustics to determine the relationship between a sound's characteristics and the auditory sensation it provokes (Lorenzi 2016). According to Everest & Pohlmann (2015:39), psychoacoustics can (thus) in many ways be viewed as the human basis for the entire field of audio engineering.

2.3.1 The hearing system

The ability of the hearing system to receive information is determined both by the qualitative relation between sound and impression and by the quantitative relation between acoustical stimuli and hearing sensations (Zwicker & Fastl 2006:preface). I.e. we need qualitative "input" to assess the information, but our ears are also dependent on physical magnitude - a physical stimulus only leads to a hearing sensation if its physical magnitude triggers a sensation in the hearing organ (Zwicker & Fastl 2006:11).

The *threshold of hearing* lies at the minimum sound pressure level that triggers a hearing sensation (Everest & Pohlmann 2015:40). The *threshold of hearing* varies with frequency. For an average hearing person the frequency range (of the hearing area) extends from 20 Hz to 20 kHz. Within the hearing area human ears are most sensitive to the frequencies from 2 - 5 kHz. According to Zwicker & Fastl (2006:20), for the average hearing person, the threshold for sensing stimuli in the frequency region of 2-5 kHz may even reach below 0 dB. According to Everest & Pohlmann (2015:41), a peak in sensitivity is reached around 3 kHz.

In the frequency area of 5 - 12 kHz hearing sensations vary individually to a great extent. Peaks and valleys in the response curve are evident. Above some 12 kHz the threshold for

hearing sensations increases rapidly. Zwicker & Fastl (2006:20) note that the actual upper limit is very dependent on both age and whether or not the listener is previously exposed to sound levels that induce hearing loss.

At the other end of the spectrum the threshold for hearing sensations also increases rapidly with lower frequency. At 50 Hz the threshold for sensing a stimulus reaches 40 dB (Zwicker & Fastl 2006:20).

Various sound signals have different frequency distributions. Out of all broadband sound signals white noise has the greatest frequency distribution within the human hearing range. It contains all frequencies within this range. Second, comes music signals. According to Zwicker & Fastl (2006), music signals occupy mainly the frequencies from 40 Hz to 10 kHz. Speech signals, on the other hand, has a range from approximately 100 Hz to 7 kHz.

Complex sound signals like speech and music have potentially a substantial dynamic range. According to Zwicker & Fastl (2006),the dynamic range of music starts at sound pressure levels below 20 dB and reaches levels above 95 dB. Low-level sounds in the vital mid-frequency area of music are thus well above the hearing threshold.

At the opposite end (from the hearing threshold) lies *the threshold of feeling* - the level where a tickling sensation is feels. According to Everest & Pohlmann (2015:48), at 3kHz this occurs at sound pressure levels of about 110 dB. Levels above this cause pain in the ears. The risk of permanent hearing damage at these levels is high.

Whereas the physical magnitude of a sound, the sound pressure, is measured in unit Pascal (Pa), the perception of how loud the sound, the *perceived loudness*, is described in unit *sone*.

There is a non-linear relationship between sound pressure and perceived loudness. E.g., a level increase of 6 dB yields an approximate doubling of a sound's physical magnitude. However, an approximate 10 dB level increase is needed, for the perception of a level doubling (Everest & Pohlmann 2015:49). Likewise, to perceive the level to be cut in half, a 10 dB decrease is necessary.

While the frequency response of the human ear is non-linear, the curves tend to flatten as one approaches higher sound pressure levels. The *loudness level* (LL), is thus relevant. It is measured in phons and is defined as the sound pressure level of a reference signal at 1 kHz, where it is perceived equally as loud as a test signal. *Equal-loudness contours* describe how sound pressure levels vary for equal loudness level (Kleiner 2013:54-44). They can be viewed as an inversion of the frequency response curves of the human ear.

Our hearing system consists of three main parts, or subsystems: ears, auditory nerves, and brain (Kleiner 2013:47). The process of sound perception starts with a sound *stimulus* striking our eardrums. The sound stimulus sets in motion mechanical movements that result in electrical discharges sent to the brain. The brain recognizes and interprets those discharges, creating a sensation that we call sound.

2.3.2 Ear anatomy

The human ear comprises three main parts; the outer ear, middle ear, and inner ear (Everest & Pohlmann 2015:40). The outer ear consists of the visible part, pinna, as well as the ear canal and eardrum. The primary function of the pinna is to collect and direct the sound energy into the ear canal (Kleiner 2013:49).

The shadow action of the head, pinna, and torso provides directional cues about a sound. In sound field listening the frequency response of sounds at our ears changes according to the incident angle. The reason is that sound waves of different frequencies have varying directional properties. E.g., low frequencies, due to their long wavelengths, will diffract (bend) more readily around the head and torso than treble- frequencies. The frequency response of the sound at opposite ear side from where it struck will thus be severely altered, primarily because of lack of frequencies above approximately 1 kHz.

The pinna is especially effective in differentiating sounds from the frontal- and backplane. According to Kleiner (2013:49), the sound level difference from rear and front directions can, at frequencies above 2 kHz, exceed 10 dB.

The (outer) ear canal is on average 2,5 cm long and of 0,7 cm width (Kleiner 2013:51). It shares acoustical similarities with an organ pipe - closed at one end and open at the other (Everest & Pohlmann 2015:41). The non-linear frequency response of the hearing organ, with its peak in sensitivity around 3 kHz, is primarily due to the dimensions and shape of the ear canal. The peak in sensitivity mentioned at around 3 kHz corresponds to the resonance that occurs at 1/4 wavelengths of the ear canal's total length (Zwicker & Fastl 2006:24).

From an evolutional viewpoint this is likely not a coincidence - the vital speech frequencies lie within the frequency area of 2000 - 3000 Hz.

The middle ear comprises three middle ear bones; the malleus, the incus, and the stapes. They permit a mechanical contact between the eardrum and inner ear.
The first of the three middle ear bones, the malleus (hammer), is attached to the eardrum. The stapes are part of the oval window, which is in contact with the fluid of the inner ear. Initially, an impedance mismatch exists for energy to be transferred efficiently from the eardrum to the inner ear. The middle ear bones act, therefore, as a mechanical impedance transformer, such that the power from the eardrum effectively transfers to the inner ear (Kleiner 2013).

The inner ear comprises the cochlea and semicircular canals. When a sound stimulus strikes the eardrum, the power of this sound energy transmits to the fluid of the inner ear. The fluid of the inner ear starts to vibrate and the basilar membrane, which stretches from the (inner) ear canal to the inner ear, responds to this vibration - it starts to move. Resonance peaks arise at different spots along the basilar membrane, according to frequency. Everest et al. (2015) note that the complex signals of music and speech lead to many momentary peaks (resonances) which constantly shift in strength and position along the basilar membrane. The peaks (resonances) have different tuning curves according to the intense of the sound. For low sound levels, the curves sharpen, whereas for higher levels the curves broaden. However, Kleiner (2013) notes that the very behavior (movement) of the basilar membrane is not frequency selective enough to account for the frequency resolution of human hearing fully. As nerve signals propagate to the brain, further signal processing is thus needed.

2.3.3 Psychoacoustic hearing thresholds

In psychoacoustics, the "difference threshold" or just "threshold", refers to the step size of a stimulus ΔA_s that leads to a difference in a hearing sensation, ΔB_s. The *absolute hearing threshold* of sensation, illustrated in Figure 2.11, is defined as the level at which one can *barely* recognize the existence of a stimulus (Kleiner 2014:212). However, the absolute threshold may vary according to the circumstances (Zwicker & Fastl 2006:12). Not to mention, the detection may produce different thresholds if starting from above (heard), relative to from below (not heard), Kleiner (2014:212) notes.

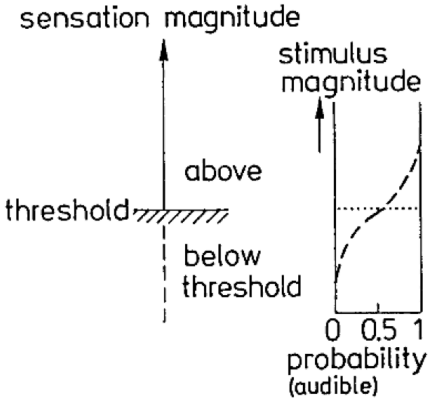


Figure 2.11 *Determination of absolute threshold, i.e. the stimulus magnitude for which the corresponding sensation or sensation increment is audible with 50% probability. (From Zwicker, E. & Fastl, H., Psychoacoustics, Springer-Verlag, 2006.)*

Regardless, below the threshold, the stimulus will not lead to any sensation.

Another idea is that of detecting the just noticeable difference between two stimuli. The threshold at which one can detect a just noticeable difference is called the *just noticeable difference threshold* (JND). One can base the detection on differences in loudness, or any other dimension. In the determination, one often choose a 50% probability level as the threshold. According to Kleiner (2014:212), one defines the JND as the smallest detectable difference between two paired sensory stimuli.

2.3.4 Spectral masking and critical bands

Masking occurs when a sound masks (blocks) the perception of another sound (Kleiner 2013:64). Masking plays a role in many aspects of everyday life; In a conversation in noisy environments to the perception of sounds in music (Zwicker & Fastl 2006:61).

The masking of a tone by noise is a well-known experiment in psychoacoustics. The task is simple; The tone, noise, or both, must be adjusted in level, such that the tone becomes just audible (in the presence of the noise).

The above is an example of a quantitative experiment. As such, the *masked threshold* is the sound pressure level required for a test signal (tone) to be *just audible* in the presence of the noise, which is the masker (Zwicker & Fastl 2006:61).

The level by which a sound needs to be increased to be audible in the presence of a masker is called the *masking level*. The masking level is thus an expression of the *threshold shift* necessary for a sound to (still) be audible in the presence of a masker (Kleiner 2013:64).

Simultaneous masking occurs when the masker is present during the entire duration of the test signal (like in the above example). However, masking effects can also take place when the signals are not presented simultaneously (Zwicker & Fastl 2006:61). *Pre-stimulus masking*, shortened “pre-masking”, can occur when the test signal is a short transient sound, which starts right before the masker stimulus is introduced. Likewise, when the test signal starts right after the termination of the masker stimulus, *post-stimulus masking*, shortened “post-masking”, can occur.

Naturally, the phenomenon of masking also concerns the field of room acoustics. Here, masking effects can appear in different ways. E.g. in a listening room, a recording might be played back through a pair of loudspeakers. However, reflections in the room might blur the sound, making low-level details in the recording difficult to hear. In this case, we can view the reflections as a *partial masker* of the sound coming from the loudspeakers. Zwicker & Fastl (2006:61) note, if the masked signal is a test tone and the masker is increased steadily,

there is a continuous transition between an audible (unmasked) test signal and one that is completely masked. Thus, we also observe *partial masking* in addition to *total masking*.

Masking effects are mostly active upward in frequency (Kleiner 2013:64). Also, masking becomes the strongest when the masker and the masked signal are close to one another in frequency, according to Kleiner (2013:64).

Results from experiments with noises as maskers indicate that masked thresholds rise with increasing spectral density and bandwidth of the masker signal (Zwicker & Fastl 2006:63).

The observation concerning masker-bandwidth has to a large degree to do with the critical bands of the ear. According to Kleiner (2013:65), there are 32 different-width critical bands in the ear. While the width of the critical bands is relatively constant below some 500 Hz, they become much broader at higher frequencies (Everest & Pohlmann 2015:145). At 1 kHz, the critical bandwidth is about 160 Hz, which is approximately the width of a 1/3-octave width bandpass filter (Kleiner 2013:65).

According to Zwicker & Fastl (2006:64), masked thresholds did not rise with masker-bandwidths broader than the critical bandwidth of the ear. The only effect observed after increasing the bandwidth further, was a change in the subjective loudness of the masker.

However, unlike relatively steady-state signals such as noises, music and speech elicit a strong temporal structure. I.e., loud passages follow faint passages and vice-versa (Zwicker 2006:78). In music and speech premasking and postmasking effects thus become especially relevant (in addition to simultaneous masking).

2.3.5 Binaural hearing and localization

Binaural hearing means listening with both ears. Using both ears, we can obtain a sensation of the direction from where a sound comes (Zwicker & Fastl 2006:293). This sensation of direction is challenging to produce listening monaurally (with one ear). The reason for this

localization ability using both ears is our hearing systems' ability to perform cross-correlation analysis of the signals from the two cochleae (Kleiner 2013:67).

If a sound source is not located directly in front of us, the signal on one side will be delayed relative to the other (Zwicker & Fastl 2006:293). The *interaural time* is the time difference of the signal reaching one ear relative to the other. Our hearing uses the interaural time difference of signals at the ears to determine the angle of incidence of a sound signal relative to the median plane (Kleiner 2014:230). According to Zwicker & Fastl (2006:293), one achieves the maximum possible interaural delay when the sound source locates 90° from a frontal incidence. Thus the maximum possible interaural delay between ears will be 0,6 ms, according to Zwicker & Fastl (2006:293), but there are individual differences.

Our head also causes inter-ear level differences leading to spectral deviations which help in the localization of sounds. According to Kleiner (2013:68), the inter-ear level differences above 3-4 kHz are sufficient for localization.

Localization ability in the median plane is weak due to both ears receiving essentially the same signal (Kleiner 2013:68). In such situations immediate head movements will improve the localization ability, increasing the interaural time and inter-ear level differences of a sound.

The head-related transfer functions (HRTFs) are important as they describe the mentioned properties of sound at our ears as a function of the incident sound field (Kleiner 2014:190). Figure 2.12 shows examples of measured frequency responses of HRTFs as a function of the lateral (sideways) angle.

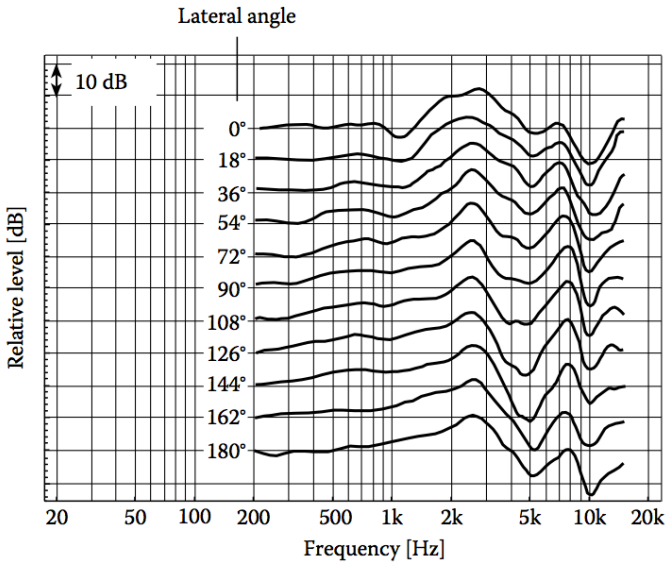


Figure 2.12 *Measured frequency responses of HRTFs as a function of the lateral angle (After Ando, Y., Concert Hall Acoustics, Springer, Softcover reprint of the original, 1st ed., 1985 edn.,2011). Cited in Kleiner (2014:195).*

2.3.6 Psychoacoustic dimensions - Timbre and Pitch

Auditory experience is impossible to describe using physical terms. The various psychoacoustic dimensions responsible for the audio experience are interrelated.

Cited in Rubak (2004:1), The American Standard of Acoustical Terminology proposed the following definition of timbre: *"..the attribute in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar"*.

As noted, *timbre* is the subjective impression of spectral content (Kleiner 2014:216). Everest & Pohlmann (2015:56) note that we mainly use timbre to describe the tone of musical instruments. Thus it is common to describe timbre as *tonal color* (Kleiner 2014:216).

In sounds with broad frequency ranges, timbre depends on the relative spectral balance (Kleiner 2014:216). The analogous physical counterpart of timbre is thus (frequency) *spectrum* (Everest & Pohlmann 2015:56).

However, timbre is also affected by sound pressure (loudness). We can relate this to the equal-loudness curves; at higher sound pressure levels the frequency response of the human

ear flattens, as mentioned previously. Thus, one expects timbre to change slightly with sound pressure.

Cited in Rubak (2004:1), The American Standard of Acoustical Terminology proposed the following definition of pitch: *".. the attribute in terms of which sounds may be ordered on a scale extending from low to high"*. A musical instrument produces various pitches.

However, the perception of *pitch* is also dependent on sound pressure (loudness); When the level of low-frequency sounds increases, the pitch goes down. For high-frequencies, the opposite happens; when the level increases, the pitch goes up.

2.3.7 Coloration

As noted, Salomons (1995) proposed the following definition of coloration: *"..the coloration of a sound signal is the audible distortion which alters the (natural) color of the sound"*.

Literature often refers to coloration as a change in tonal color, i.e. a change in timbre.

However, we do not exclusively restrict coloration to the frequency-domain. According to Rubak (2004:1), coloration concerns both *pitch*, *timbre*, *flutter* and *loudness*. Flutter, which is a type of echo effect, occurs when a signal is shorter than a substantial part of its reflections and occurs typically from very short sounds such as a handclap occurring between hard parallel walls.

In practice, we can thus divide coloration into timbre-related perception in the frequency-domain, as well as flutter echoes in the time-domain.
As such, Rubak (2004:1) employs the term *spectral coloration* for the perception of coloration in the frequency-domain (timbre-related), and the term *temporal coloration* for time-domain perception.

2.3.8 Comb filtering and critical bands

One should also consider how the frequency processing in the ear might affect the perception of comb filters. Halmrast (2020:2) notes, for longer time delays the peak-to-peak comb filter

frequency becomes particularly small. At substantial delay times, where the peak-to-peak comb frequency is particularly small, the audible effect will be that of an echo, as mentioned previously.

On the other hand, at a particular small delay time, where the peak-to-peak comb frequency is substantial, the (first) attenuation frequency of the comb filter will appear so high up in the spectrum the audible effect will be that of a low pass filter (Halmrast 2020:2).

According to Kleiner (2013:65), the frequency processing (analysis) of the ear is equivalent to an adaptive filter bank with a large number of parallel frequency-selective filters. If we could visualize these filters, they would look like bandpass filters, i.e. “critical bands”. According to Kleiner (2013:65), there are 32 different-width critical bands in the ear. The width of these critical bands varies according to frequency, research shows (Everest & Pohlmann 2015:144).

While the width of the critical bands is relatively constant below some 500 Hz (Kleiner 2013:65), they become much broader at higher frequencies (Everest & Pohlmann 2015:145). At 1 kHz the critical bandwidth is approximately 128 Hz (Everest & Pohlmann 2015:144), which is about the width of a 1/3-octave bandpass filter (Kleiner 2013:65).

As previously mentioned, 1 kHz approaches the frequency-area where we are most sensitive. Thus, it is reasonable to consider the overall audibility of sounds according to this frequency area.

A peak-to-peak comb filter frequency of 125 Hz corresponds to a time delay of about eight milliseconds (Everest & Pohlmann 2015:144). Therefore, at 1 kHz, two comb peaks fall within a critical band.

Halmrast (2020:2) calls the coloration effect one might perceive when the peak-to-peak comb filter frequency is approximately the size of the ear’s critical bandwidth (at 1 kHz), *“Box-klangfarbe”* (after the German “Klangfarbe”, which means timbre). The reason is that this type of coloration gives a rather “boxed” sound quality, typically of what one might perceive in boxed-shaped listening rooms.

2.3.9 Binaural decoloration

Binaural hearing affects the ability to hear comb filter coloration for laterally separated sounds. When direct- and reflected sound are separated laterally, (comb filter) coloration becomes less noticeable (Barron & Marshall 1981), cited in Rubak (2004:2). The reason for coloration becomes less noticeable is due to a mechanism called *binaural decoloration*, or *binaural* unmasking (Kleiner 2014:244).

The *inter-ear time difference* (or interaural delay cf. Zwicker & Fastl 2006:293) between sounds increases as one separates them laterally. The *inter-ear delay* is the time difference of a signal reaching one ear relative to the other. As previously mentioned, our hearing uses the inter-ear time difference of signals at the ears to determine the angle of incidence of a sound signal relative to the median plane (Kleiner 2014:230). Thus, coloration perception decreases due to *decorrelation* of the signals (as they are separated horizontally). The decorrelation diffuses the coloration perception. The maximum possible interaural delay occurs when the sound source is located 90° from frontal incidence, according to Zwicker & Fastl (2006:293).

Also, the shielding effect by the head (Salomons 1995:4) and filtering by signals reflected by the pinna, head, and torso provide cues that contribute to the binaural decoloration mechanism.

Kleiner (2014:244) notes that binaural decoloration becomes especially relevant in multichannel sound reproduction. If two channels receive different signals (stereophonic listening), our ears can separate their direction. However, if the two channels receive the same signal (monophonic recording presented diotically), our ears are not able to separate the signals, and the listening essentially mimics *monaural* (one ear) listening. In this latter situation (diotic listening) the binaural decoloration mechanism of the ear, therefore, does not work.

However, rooms with diffuse sound fields reduce the perception of comb filter coloration, due to removing regularity in the impulse response, as previously mentioned.

On the other hand, the sound field of small rooms are hardly diffusive (because of the small room size). Thus, binaural decoloration will not be as prominent as in diffuse rooms. As such, comb filtering from sidewalls is still relevant.

2.3.10 Determining audibility of comb filter effects by threshold estimation

A natural way of determining the audibility of a reflection is to consider its amplitude (Buchholz et al. 2001:2). The Reflection Masked Threshold (RMT) is defined by Burgtorf (1961) as *"the amplitude threshold below which a human listener is unable to perceive the effect of single reflections, multiple reflections and reverberation. The perceived effect can include a variation in all possible sound attributes such as loudness, spatiality, localization, coloration, timbre, temporal structure, etc."*.

We may not be able to hear a low level reflection because it is masked by the presence of the direct sound, other reflections, or background noise. Buchholz et al. (2001:2) note, *"by increasing the amplitude of a reflection the Reflection Masked Threshold (RMT) is reached, and the reflection becomes audible and its effect is manifested as variation in timbre (coloration), loudness, or spatiality, although it is still temporally fused with the direct sound"*.

We quantify the amplitude threshold where perceptual effects become audible, in unit dB. In this regard, we should note that the reflection level measured at threshold, is not the *actual* dB level, but the *relative* level of the reflection to that of the direct sound.

The reflection level where *just noticeable changes* in timbre become audible is the threshold of interest (when investigating coloration). In literature this threshold is often referred to as the *absolute threshold* because it is the first threshold we reach when gradually increasing the gain of the reflection from zero (no sound). The smallest decrease in the amplitude of the reflection (from this threshold), and coloration becomes again inaudible. A further increase in reflection level from the absolute threshold and we may perceive an increase in loudness and in turn also an (increased) sense of spatiality (broadening of auditory image), also dependent on the type of source signal. If the test reflection is separated laterally from the source, we may eventually perceive changes in localization, where the auditory image shifts towards the

side of the reflection. The threshold for the perception of image shift is called the *image shift threshold*. A further increase in the amplitude of the reflection and we reach the *echo threshold* (ET). The reflection then becomes audible as a separate audio event (Buchholz et al. 2001:2).

Salomons (1995:5) notes that when investigating timbre (coloration), we must be aware that dichotic (stereo) signals may result in *lateralization*; the shift of auditory image from the middle of the head to one side. Salomons (1995:5) notes that the image shift is created by either applying a difference in level or phase between the signals. E.g., in a laboratory-setting, adding a strong lateral test reflection to a source signal presented in the center of auditory image, we should expect lateralization to happen. Lateralization is unwanted when investigating timbre (coloration) because it confuses the perception, making coloration harder to judge.
Also Bech (1994:1717) mentions the possible confusion of timbre and localization in dichotic signals. In the method chapter, section 3.4.3, we will look at the approaches we can use to avoid confusion (in listening tests) between timbre and localization.

3 Method

3.1 Starting point for the research

The research question is - *to what degree do the spectra of reflections influence the detectability of comb filter coloration?*

As mentioned, the small listening room was of interest (of reasons mentioned in section 1 and 2.2.1, in particular). In small rooms, when the purpose is music listening, we (mostly) listen to *reproduced sound*; we seldom listen to live-sound sources. Therefore, we should first start by looking into how researchers study the perception of reproduced sound.

According to Bech (1990), we can either study the perception of reproduced sound in a) real listening rooms or b) *simulated sound fields*. However, in real sound fields, i.e. listening rooms, the listener would be subject to uncontrolled sound field components due to reflections from additional walls and boundaries. Also, background noise would be a possible problem (Kleiner 2013:111). In real listening rooms we also have limited control over the variables under investigation.
Simulation, on the other hand, gives the experimenter the required manipulative control over the variables under investigation. Thus, Bech (1990) notes, the experimenter should employ some form of simulation technique of the room acoustics.

3.2 Simulation techniques

According to Bech (1990), we can divide the study of the perception of reproduced sound through simulated sound fields into two groups according to the simulation principle:

A. Computer simulation of the sound field and reproduction via headphones.

B. Electroacoustic simulation in an anechoic, or semi-anechoic chamber.

Both of which are variants of a technique known as *sound field synthesis* (Kleiner 2013:111). Also, several variants of these simulation techniques may exist.
However, regardless of the simulation technique, the approach we choose would take the basis in *geometrical acoustics*.

3.2.1 Geometrical acoustics

In geometrical acoustics we consider sound waves as rays reflected by surfaces of the room just as a mirror reflects light (Bech 1990). Thus, we assume plane, rigid surfaces leading to geometrical reflections. Dalenbäck (2018) notes that the use of *geometrical* reflection is another term for *specular* reflection, which thus indicates the original meaning of geometrical acoustics.

We separate between two methods for prediction in geometrical acoustics: the ray tracing method (RTM) and the mirror image source method (MISM).

In an electroacoustic simulation we consider the reflecting surfaces as mirrors and model the sound field in the room by the so called *image source principle* (Bech 1990). This principle states that if we consider all mirror images of the sound source as separate sound sources switched on simultaneously together with the original sound source, then these image sources will combine into a sound field which is identical to the sound field in the real room (Bech 1990).

3.2.2 Electroacoustic simulation in an anechoic chamber

The classic system for sound field synthesis is a large loudspeaker array set up in an anechoic chamber (Kleiner 2013:111). In such systems the direct sound and early reflections are usually represented individually by single loudspeakers and the reverberant field by a group of loudspeakers (Bech 1995). The idea is to create an *illusion of presence* similar to being in a real sound field. Also, additional equipment, such as delay and reverberation units are necessary (Kleiner 2013:112).

However, there are both practical and acoustical challenges with such experiments. The practical challenges are that they require a substantial number of loudspeakers if each early reflection is to be represented individually by a single loudspeaker.

The acoustical challenges would be that speakers cabinets themselves also reflect sound, and this would be more problematic the more speakers in the room.

Therefore, to make the experiment more practical, we should limit the number of loudspeakers.

In sound field experiments, Bech (1990) thus suggests considering the following points when attempting to limit the number of loudspeakers:

1. Subjects limited ability to localize sound sources in space

2. Threshold of perception for timbre change with regard to direct sound in combination with a single reflection

3. The auditory systems' limited time resolution

Our ability to localize sound sources in space is both dependent on incident angle and signal spectrum, as having been accounted for in section 2.3.5. Our hearing has difficulty identifying the location of specific image sound sources (reflections) within a sound field. In this regard, Bech (1990) suggests that image sound sources (reflections) positioned within this *localization blur* - that is, in areas of the sound field where our hearing has trouble identifying the location of sounds - should be represented by a single loudspeaker.

Also, to reduce the number of loudspeakers Bech (1990) suggests not to model (by loudspeakers) those reflections that have their natural levels below threshold values determined for single reflections. On the other hand, not modelling the reflections below threshold assumes that the measured threshold values for a single test reflection represent the lowest attainable threshold values also in the case where the direct sound combines with more than one reflection, Bech (1990) notes.

As previously mentioned, we usually model the reverberant field by a group of loudspeakers. By using an equation for the calculation of a room's mean impulse density, we can find when the onset of subjective reverberation - i.e. the time when subjects start perceiving the sound field as a continuous whole. According to Bech (1990), in a small to medium sized room the onset of subjective reverberation starts approximately 20 ms after the arrival of direct sound (of course, the time will vary according to the volume of the room, as mentioned in section 2.2.4, 3)). Therefore, it is not necessary to model individual reflections arriving after this. This unencessity then implies that subjects are not able to detect changes in reflection density after around 20 ms relative to the arrival of the direct sound.

Bech (1990) proposes that we can then divide the sound field of a normal domestic room into three components:

1. The direct sound
2. Reflections arriving before 20-25 ms relative to arrival of the direct sound at the listening position.
3. Reflections arriving after 20-25 ms relative to the arrival of the direct sound.

Bech (1990) underlines that such an electroacoustic simulation implicates the following:

A. A single source representation of the direct sound.

B. The image sources with time offsets relative to the direct sound of less than 20-25 ms and levels higher than detection threshold values are represented either individually or in combination with other images (reflections) according to the degree of localization blur.

C. Image sources (reflections) with time offsets of 21 ms or more relative to the arrival of the direct sound are represented by a separate system that simulates the reverberant part of the sound field.

Bech (1990) also notes we should aim for a reproduction system that is as neutral as possible such that we can implement any required changes in the sound field through signal processing. The following implications exist with regard to the design and position of the loudspeaker representing the direct sound:

1. The on-axis free field response should be independent of frequency in the broadest possible frequency range.
2. The perceived distance to the direct sound source should be independent of the physical distance between participant and loudspeaker.

The last step in the process of setting up an electroacoustic simulation is to verify that the experimental setup meets specific acoustical room criteria, as well as conveying a satisfactory room impression to the listener, according to Bech (1990). The verification contains both objective and subjective measurements.
The purpose of the objective measurements is to ensure that we meet specific room acoustic conditions. The objective measures include measurement of reverberation time and the steady-state transfer function from the sound source to the listening position.
The purpose of the subjective measurements is that the simulated sound field gives a similar impression to that of the room we simulate.
According to Bech (1990), the subjective measurements include listening tests conducted in real rooms as well as listing tests based on [10]*dummy head* recordings.

[10] See bibliography p. 116

3.2.3 Computer simulation

We use geometrical acoustics also as the basis for computer modelling. However, computer simulation of the sound field in traditional sense means to calculate the *impulse response* of the modelled room. We then convolve the test material with the calculated impulse response, and the result is presented via headphones to the respondent (Bech 1990). We call the process of providing audible, either through physical or mathematical modelling, the sound field of a source in space, in such a way as to simulate the binaural listening experiment at a given position in the modelled space, *auralization* (Everest & Pohlmann 2015:559).

Commercially available software for simulation and auralization of room acoustics are programs such as CATT-Acoustic and Odeon.
In general, the simulation starts with defining the room geometry and the location of sound sources. We then assign frequency-dependent material properties to room surfaces, as well as frequency-dependent source directivities to sound sources. Everest & Pohlmann (2015:562) note that we define each source, which could be either a natural source or a loudspeaker, by its octave-band directivity at minimum 125, 250, and 500 Hz, and 1, 2, and 4 kHz. So-called directivity balloons describe the source directivity in each octave-band.

Then we can either use ray tracing or the image method, as mentioned, to track down each reflection path and thus generating a room *echogram*, which describes the reflection path history of the reflections as a function of time.

In the final processing before auralization can take place we convolve the impulse responses obtained from the source- and receiver-positions in the room (echogram) with the *head-related transfer functions* (HRTF) for each ear (Kleiner 2014:403). Thus each reflection is transformed into a binaural impulse response (BIR) for each ear. When convolving each of the reflections in the echogram with its corresponding HRTF, the result is the left and right room impulse responses (Everest & Pohlmann 2015:564).

We can then convolve the desired test signal, preferably anechoic recorded or generated electronically, with the obtained binaural room impulse responses.
Kleiner (2014:403) notes that there are mainly two presentation methods we use for

auralization. The first is binaural playback and the other is surround system playback such as Ambisonics. In binaural playback we can use either headphones or loudspeakers using cross-talk-cancellation.

Also, we can apply headphone equalization to the final output stage to compensate for any deviations from linear frequency response in headphone sound reproduction.

3.2.4 Simulation strategies

Generally, there are several ways to conduct a listening experiment with artificially created reflections. In addition to the reflection under the investigation, we could include: 1) Only the direct sound. 2) The direct sound and one or several room reflections. 3) The entire sound field.

We may consider 2) and 3) typical sound field experiments. In these, we present the room reflections, i.e. the reflections that are not under investigation according to their natural (obtained via measurements or calculated on the computer) levels and the correct relationship between incident angle and time delay relative to the direct sound. Only the reflection under investigation will have a variable level when we want to determine audibility by considering its amplitude.

Of course, 2) and 3) are closer to the situation in a real listening room, due to taking into account the presence of other reflections. Also, in 2) and 3), there is a good chance that the room reflections, dependent on their levels, will contribute to the masking of the reflection under investigation, such that we can consider the obtained threshold value for the reflection under investigation as higher boundary (and thus closer to the actual threshold).

While we can consider simulated sound field experiments more realistic, the question is whether they are ideal when investigating the influence of spectrum of a single reflection (on the audibility of its effect). We may initially study the influence of spectrum of a reflection on its effect, 1) with only the direct sound and the reflection under investigation present. Then, when we have observed the effect of the factor of interest, we may expand the experiment as in 2) and 3), conducting a sound field experiment.

3.2.5 Point of view

Absorption is a major contributor to the spectral modification of reflections. Therefore, the research question was going to be investigated from the viewpoint of absorption. The effect of (high-frequency) absorption on the audibility of coloration is a very viable topic also from a practical point of view. We may ask:

Does high-frequency absorption reduce comb filter coloration significantly?

Initially, the plan was to investigate the practical question asked above, taking base in the (absorption) characteristics of two types of porous absorbers. 1) A material that only absorbed frequencies high up in the spectrum, and very little (e.g. a curtain). 2) A material with a predominant effect in the high-frequencies of sounds, but also reaching much lower down in the spectrum (e.g. mineral wool).

The investigation would take place in the form of one or several listening tests, in which a group of listeners had to be recruited.

Due to unforeseen events (which are mentioned in the acknowledgements) I had to swamp the idea of including the absorption characteristics of several materials in the investigation. The reason was that in the event of including several materials, this would make the experiment longer, which was not justifiable during this period. Thus, I had to limit the overall length of the experiment.

We have now looked at the two main categories for simulation. It was clear, due to the information given in section 3.1, that simulation was the best option for a listening experiment. Of course, this also meant that absorption, or rather the characteristics of absorption (the filtering effect), had to be modelled. In the next section we will, therefore, look into how we can do this.

3.2.6 Modelling sound absorption

Olive & Toole (1989) used basic lowpass filtering in their experiment investigating the effect of spectral limiting of a single reflection on its detection threshold. However, as their aim was not to model the absorption characteristics of a specific material, such basic lowpass filtering was sufficient enough.

We need a more accurate way of modelling absorption characteristics than what basic lowpass filtering provides. A natural way would be to match the response of a digital filter to available response data. According to Huopaniemi et al. (1) the response data can either be in the form of an impulse response or transfer function, or magnitude response (absorption coefficients).

I decided to use the absorption coefficients of a 20 mm thick mineral wool absorbent, which data I found publicized at rockfon.no, and is given in table 3.1, as base for the modelling. I first converted its absorption coefficients to magnitude data given in table 3.1, using the equation:

$$d = 20\log(1 - C) \tag{3.1}$$

where d indicates the attenuation in dB and C indicates the absorption coefficient

	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Abs. coeff.	0.05	0.25	0.70	0.95	1.00	1.00
Magnitude data (dB)	-0.44	-2.49	-10.45	-26.02	- infinity	- infinity

Table 3.1 *Absorption coefficients of 20 mm thick rock wool, converted to magnitude data. The calculated dB attenuation was used as the basis for the design of a filter that could be used to approximate the absorption of the mineral wool absorbent.*

I decided to use MatLab for the filter creation, as the program offers several functions to design filters. By using its designfilt function, I arrived at a filter that, if not a complete match, gave a reasonable approximation of the absorption magnitude data, at least in the three first octave bands (table 3.2). I adjusted filter parameters such as passband-frequency, stopband-frequency, and stopband attenuation, to get a closer fit (the Matlab code used is included in the Matlab-script files, see appendix 2).

	125 Hz	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**
Absorption magnitude data (dB)	-0.44	-2.49	-10.45	-26.02	- infinity	- infinity
Filter magnitude data (dB)	-0.32	-1.94	-10.00	-21.65	-33.74	-46.13

Table 3.2 *A Comparison of the absorption magnitude data attained from absorption coefficients, and the created filter magnitude data.*

The magnitude response of the three first octave bands of the created filter is shown in figure 3.1.

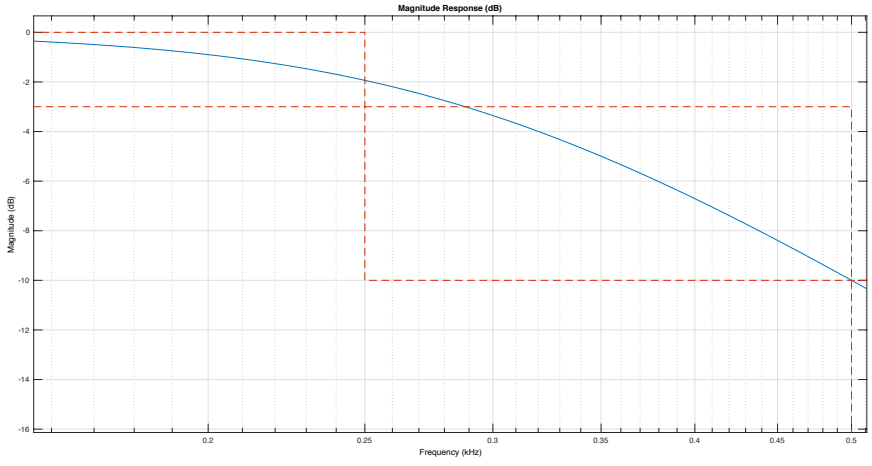


Figure 3.1 *Magnitude response of a filter created from taking base in the absorption data from a 20 mm thick mineral wool absorbent given in table 3.2. The x-axis shows the frequency of the three first octave bands, the y-axis gives the magnitude of the filter.*

3.2.7 Choosing a simulation approach

After having considered the two main categories of sound field simulations, the decision fell on using the computer.

The main factor that excluded the electroacoustic variant from an experiment was the challenge in finding a listening room that met the criteria of an anechoic chamber. Kleiner (2013:111) notes, for a room to be characterized as anechoic, the background noise levels should be less than 15-20 dB, and the walls should be nearly nonreflective over some limiting frequency, e.g. 100 Hz. Also, the walls should have a sound absorption coefficient larger than 0.99.

Another challenge with the electroacoustic variant is how the listening should occur. The most practical would be to have the respondents, in turn, seated in the room of the experiment, such as in the experiments by Olive & Toole (1989), and Bech (1994) and (1995).

The alternative to listening in-situ would be to make binaural recordings of the listening room sound field and have the respondents listen to the recordings afterwards over headphones. However, the latter then requires the appropriate recording equipment to make the recordings. As noted, the recordings would have to be binaural, simulating the way the listener perceives the sound at the listening position in the room. Such recordings are probably best done with a manikin that models the head and torso, with one microphone placed in each ear of the manikin. Such recording equipment is usually expensive.

The criteria for loudspeakers, on the other hand, could be met, especially since I had excluded the necessity of a larger sound field experiment.

Before choosing computer software I had to decide on a room which dimensions I could use as the basis for a model. After considering several listening rooms, I decided to use the dimensions of the recording studio control room at the Department of Musicology, University of Oslo (UiO), as the basis for the model. By using a laser distance meter, I acquired its rough dimensions and the position of the main speakers as well as the listening position. The

approximate position of the speakers was measured from the speakers to the nearest boundaries, giving the propagation paths and incident angles for the direct sound and reflections relative to the listening position.

Figure 3.2 is a plan view of the recording studio control room at the Department of Musicology, UiO. The dimensions, as given in figure 3.2, are quite typical for listening rooms. The detailed surfaces and the actual room acoustics were, however, not used.

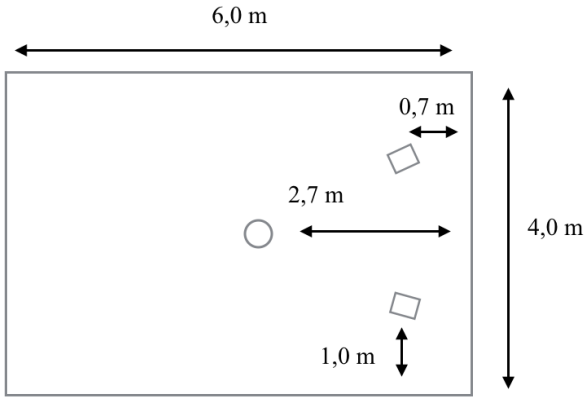


Figure 3.2 *Plan view of the recording studio control room that was used as the basis for a geometrical model, determining the propagation paths for direct sound and first-order reflections.The height of the room was approximately 2,7 m. The loudspeakers were aiming towards the listening position in an equilateral triangle. The speakers' tweeters located approximately 1,3 m above floor.*

To determine a room echogram, we use the principle of geometrical acoustics. After considering several software options, the decision fell on a demo of CATT-acoustic v.9.1 for setting up a simple basic room model to predict low-order early reflections.

Licenses for full version room modelling software are expensive. Therefore, I chose to use this program only for the echogram part. Also, because my thesis is not about testing commercial room acoustic programs, I wanted to try a different route.

While the research question was going to be investigated from the viewpoint of absorption, as noted, I included the directivity data of a typical studio monitor (Genelec 8030C) in the CATT-prediction.

Because reproduced sound (in listening rooms) was of interest, speaker directivity was important to take into account (for the echogram part) because it would help in determining which reflections, based on their predicted strength, to base on the investigation.

In CATT-acoustic, I used three separate files: 1) A master geometry-file, which defined the room geometry, as well as the surface material properties (absorption coefficients) for each room boundary. 2) A source-file, which defined the number of sources and their positions. 3) A receiver-file, which defined the number of receivers and their positions. The receiver (ear of the listener) was defined at the same level as the sources (loudspeakers), which was set to 1,3 meters above the floor.

Figure 3.3 shows plane views of the resulting room geometry as well as its source- and receiver-positions.

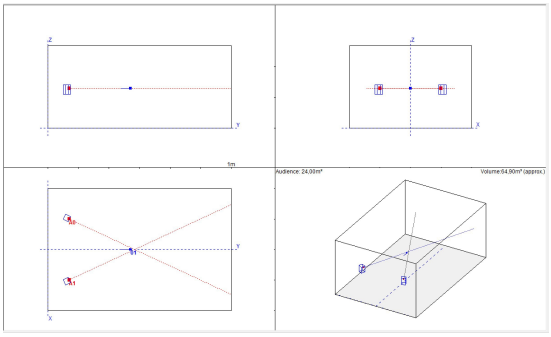


Figure 3.3 *Plane views showing the room geometry with the corresponding source- and-receiver positions that were used to calculate the echogram.*

Having defined the geometry as well as the source- and receiver-positions, the echogram was generated using the Image Source Model option in TUCT, which is an extension that comes with the main program. I set the prediction to only include first-order reflections, as the main interest here was first to determine which reflections, based on levels, incident angles and arrival times at the listening position, which were candidates for a listening experiment.

Figure 3.4 shows the calculated echogram and propagation paths for the direct sound and reflections resulting from the Image Source model.

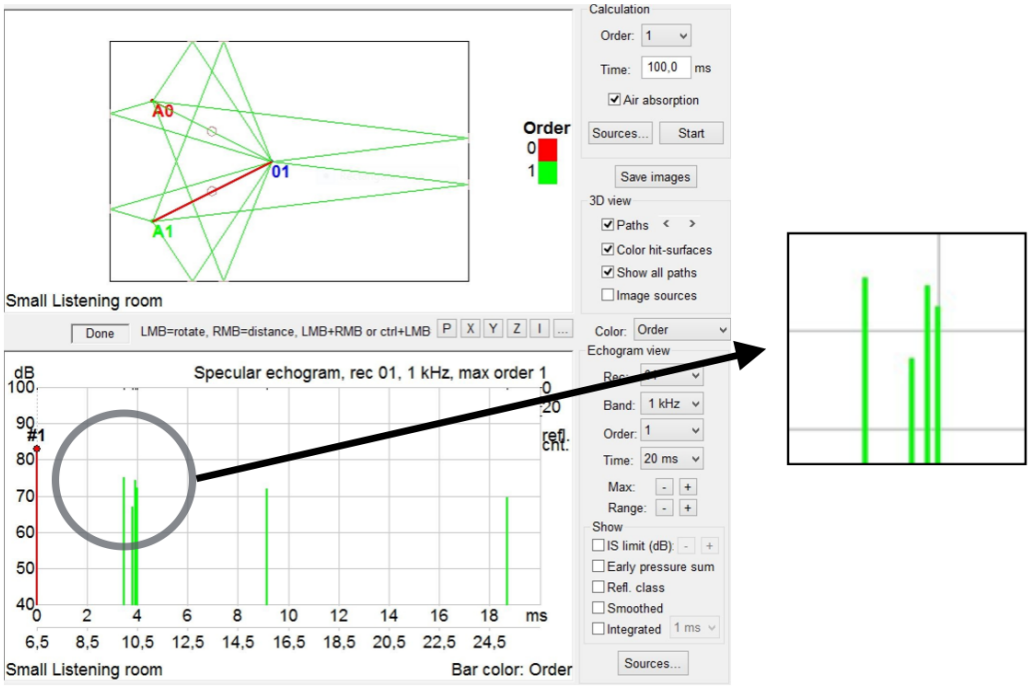


Figure 3.4 *Image Source Model in TUCT, calculated for only first-order specular reflections. All first-order reflections calculated had an arrival time at the listening position within 20 ms after the direct sound incident.*

Table 3.3 presents the results of the predictions from the Image Source Model. The angle of incidence of the direct sound relative to the listening position is presented, as well as the calculated reflection time offsets and attenuation relative to the direct sound, and reflection incident angles relative to the listening position. We see from table 3.3 the prediction gave a total of twelve first-order reflections.

Delay (ms)	Att.(dB) x 2	Azimuth	Elevation	Reflection no.	Surface
0,00	0,0	-30/30	0		left+right speaker
3,20	10	-30/30	-53	1+2	floor
3,80	18,3	-18/18	0	3+4	front wall
3,90	13,6	-60/60	0	5+6	side walls
3,90	11	-30/30	55	7+8	ceiling
9,00	13,6	-70/70	0	9+10	side walls

Delay (ms)	Att.(dB) x 2	Azimuth	Elevation	Reflection no.	Surface
18,70	15,7	-173/173	0	11+12	back wall

Table 3.3 *A presentation of source positions relative to the listening position, reflection time offsets and attenuation relative to the direct sound sources, and their incident angles relative to the listening position. Both loudspeaker directivity and distance to sound sources were taken into account in the calculations of the reflection levels relative to the direct sound.*

As we see from table 3.3, all the predicted reflections, except from the two coming from the rear wall, arrive within a window of 10 ms after the direct sound. We see that the floor reflections (1+2) and ceiling reflections (7+8) will potentially be the strongest, followed by reflections 5+6 and 9+10 from the side walls.

An option would be to use CATT also to model sound absorption, which I did not do of reasons mentioned. If I had chosen to use CATT also to model absorption, I would need to assign frequency-dependent absorption to the walls, floor, and ceiling, such that sound absorption could be modelled properly.

3.2.8 Binaural processing in Pure Data (Pd)

As noted, in virtual sound fields on the computer, the listening must take place over headphones. If the listening should bear any resemblance of sound field listening, we must find a way to simulate the properties of sound at the ears as a function of the incident sound field (Kleiner 2014).

I found Pure Data (Pd), which is an open-source programming language, a good solution for the preliminary phase to the experiment. Preliminary in the sense that I made the decision not to use Pure Data as the environment for the listening experiment. The reason was that I found MatLab to be the best option for running the experimental procedure. For details regarding the procedure, see section 3.4.4.

The purpose of using Pd was then to “auralize” the echogram first predicted in CATT. Thus, I had to make a patch that modelled the propagation paths of sources (loudspeakers) and

reflections from the echogram, relative to the listening position. Thus, each path had to be processed by its HRTF (which are mentioned further down). The paths representing the reflections had to be delayed according to the model. The positioning of sound field components was done by HRTFs, according to the model. Figure 3.5 shows the main window in the Pd-patch used.

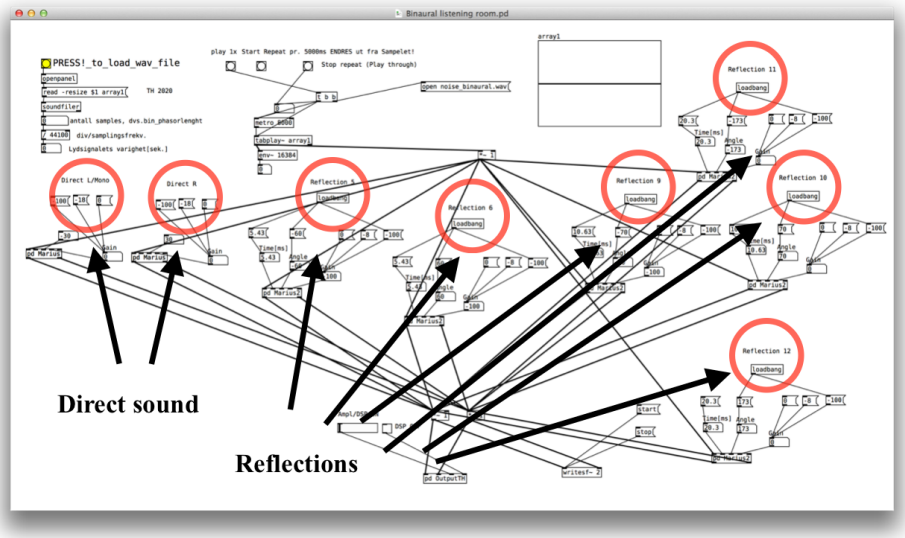


Figure 3.5 *The main patch window in Pd used for the binaural processing of stimuli signals. Red circles indicate the sound sources (direct sound) and the paths representing the reflections.*

Figure 3.6 gives a simplified block diagram of the signal paths inside the patch, from input to output. As can be seen from this block diagram (figure 3.6, the stimuli I used as input were mono (the reason for using mono signals is explained in section 3.4.3). I should note the block diagram (figure 3.6) only shows the path for each sound source and two reflections. The paths representing the left and right sources (speakers), were positioned inside their corresponding HRTF, -30 degrees and 30 degrees, respectively, according to the model (table 3.1).

Each reflection path was time delayed by an internal Pure Data object before it went through the HRTF (binaural filter).

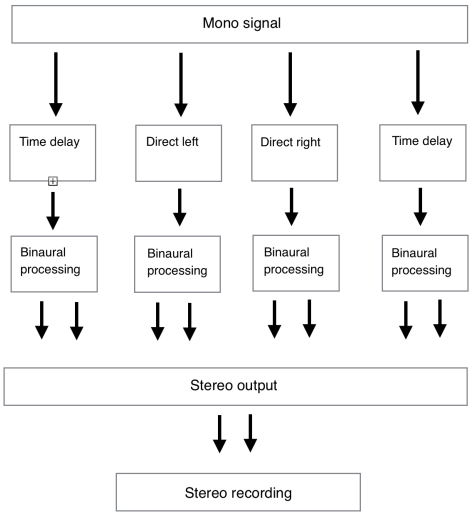


Figure 3.6 *A simplified block diagram of the processing of virtual sound field components in Pure Data.*

The HRTF (binaural filter) used was available as a download from www.soundhack.com.

The developer has written the following description of the filter, cited from its manual: *"+binaural is a filter which places a sound at a specific position around the listener's head. It does this by using filters which simulate the filtering effect of the head and outer ear for sounds at all angles. The filters in +binaural are optimized for headphones"*.

The filter had two main controls. With the angle slider, the position of the sound in the horizontal plane could be varied, from -180 degrees to 180 degrees. 0 degrees equalled straight ahead (in front of the listener), whereas 90 degrees were to the right, -90 degrees to the left, as well as 180 and -180, which were both directly behind the listener. The second control was the control for gain, which could go from -12 to 12 dB, however, in the patch, this control could be manipulated to whatever position, say -100 dB (no signal) to 0 dB.

Also, there were two binaural filter sets to choose from, in which I chose the one derived from Gardner and Martin's KEMAR *dummy head* measurements at the MIT Media Lab. For further information of these filters, see (x, literature).

Because the filters used were obtained from the measurement of a dummy head (and not from a real person), we consider them generic. Thus, they are so-called non-individualized HRTFs.

Prior research has shown that non-individualized HRTFs do not give the same performance as individualized (obtained from real persons) HRTFs (Mehra et al. 2016). What is meant by performance in this regard is, one believes the non-individualized versions do not give the same realism as the latter, particularly with regards to localization.

Naturally, in this experiment I had neither the opportunity nor the means to obtain individualized HRTFs for the respondents that I was going to recruit for the experiment. When individualized HRTFs are not an option, generic filters are the best solution. The reason is that a dummy head should represent the “average” head dimensions from a selection of individuals. The alternative in this situation would be to use one individuals’ HRTFs for all participants, but this individual might then not represent the average from the selection.

Zahorik (2009) also found evidence from his investigation on perceptual relevant parameters in virtual listening, suggesting that individualized HRTFs offer little benefit over non-individualized HRTFs for applications with fixed source directions, which is the case in my investigation.

Basically, the HRTF inputted a mono signal, applied its processing, and outputted a stereo signal. In the event of processing a stereo signal, the stereo signal would first need to be split in left and right, and then processed in parallel (two instances of the same filter). However, all stimuli signals that were eventually used for the experiment were derived from a mono signal. I will account for this in a later section.

Figure 3.7 compares the impulse response of a theoretical 3,9 ms comb filter (black), to the impulse response of a 3,9 ms comb filter, processed through the HRTFs in Pure Data. The theoretical comb filtered response (black) was obtained by adding a delayed (3,9 ms) dirac pulse to the original signal. For the binaural comb filtered response (red), two delayed dirac pulses, -60/60 degrees azimuth, were added to the source signals, -30/30 degrees azimuth.

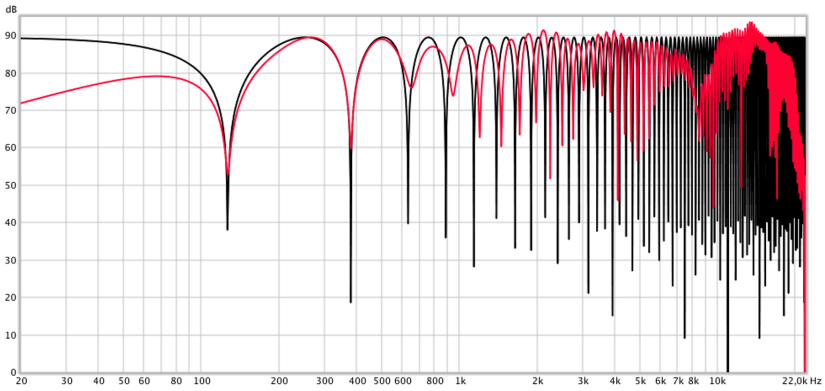


Figure 3.7 *Black: the impulse response of a theoretical 3,9 ms comb filter. Red: Binaural comb filtered impulse response.*

As such, figure 3.7 should illustrate what the binaural filters were doing to the response of the signals that were processed through them.

As noted, I could not conduct the listening experiment directly in Pure Data. I had to record the output from the processing in Pure Data, so that I could import the material into the application (MatLab) where the experiment could run from.
I used the following approach to record the stimuli: The signals representing the sources, i.e. left and right speaker, were recorded to one stereo file, and each "pair" of reflections, according to the model (table 3.1), was recorded to one stereo file. The reason for recording two simultaneous reflections is explained in section 3.4.3. As such, this approach allowed the stimulus variable (pair of reflections) to be level adjusted in realtime by the application running the procedure.

The stimuli were recorded to disk at the sample rate of 44100 kHz, and a bit depth of 32-bit floating-point.

I did not do the modelling of sound absorption by filtering in Pd. The approach used to model the sound absorption was explained in section 3.2.6.

As such, the only processing that took place in Pd was the binaural processing and delaying of the stimuli that represented the reflections.

After having recorded the material to audio files (wav) in Pure Data, I edited the files in Audacity. The editing consisted of 1) Removing space which did not contain any signal in the audio files, and making sure the files aligned correctly. E.g., it was particularly important to verify the transient onset in the files representing the reflected sound components were delayed by the correct amount, i.e. according to the predicted model, relative to the transient onset in the files representing the direct sound components. I verified this by visually inspecting the files as well as auditioning the audible result of the sum of direct and reflected components, which created a comb filter effect. 2) The files containing the reflected sound components and direct sound components had to be of approximate equal loudness, such that the files could be level balanced according to the procedure. This was done by applying root-mean-square (RMS) normalizing to all the files, equalizing their levels, as well as verifying levels by listening.

The next step then was to look into relevant experimental procedures for a listening experiment investigating the research question.

3.3 Experimental procedures

To what degree do the spectra of reflections influence the detectability of comb filter coloration?

In theory many procedures might be applicable for investigating this research question. In this section we look at those procedures, in the view of the experimenter, that are most applicable to an experiment investigating the research question, and, that is well documented in literature.

As noted, in psychoacoustics, we investigate the *audibility* or *detectability* of a stimulus variable, by threshold estimation. When investigating the audibility of comb filter coloration due to a reflection, the natural way is to consider the amplitude of the reflection, as noted. I.e. reflection level would then be the experimental variable. By this approach, we adjust its level

until coloration becomes *just audible* in the signal - the latter would, thus, be a combination of the direct sound and reflection, or possibly, a combination of several reflections, the reflection under investigation, and the direct signal.

As noted, the amplitude threshold at which the effect of a reflection becomes audible, is called the Reflection Masked Threshold (RMT). In the detection of coloration we may refer to this as the absolute threshold, or just the detection threshold, as noted.

Several experimental procedures for threshold determination may be applicable. In the following sections these procedures are presented.

3.3.1 Method of Adjustment

In this approach, Zwicker & Fastl (2006:8-9) note, the subject has control over the stimulus. I.e. he or she manually varies the level of the test signal until it becomes just audible. This can be done by means of e.g. a fader or potentiometer. This procedure, or a variation of it, was used in the experiments by Olive & Toole (1988) and (1989). Apparently, it was also used by Seraphim (1961).

In the experiment by Olive & Toole (1989), the reflection level was controlled by means of a potentiometer. The description of the procedure was as follows: *"By means of a potentiometer, the listener was required to adjust the level of the test reflection until it was at the detection threshold, somewhere between the conditions of "just audible" and "just not audible". A non-linear multi-turn potentiometer was used so that knob position would not be a reliable clue to the listener. To confirm what the target and background sounds were, the listener could, at will, switch to the signal without the test reflection, or to the signal with the test reflection at maximum level (10 dB above the level of the direct sound). No time limit was imposed on this adjustment"*.

When judging timbre by this method, the detection threshold for coloration lies at the reflection level that produces a *just audible change* in timbre, from that of the original signal without the reflection present.

At first glance, this method appears to be a good contender. Its strength is perhaps that it does not appear to be particularly time consuming. The determination of a threshold, based solely on the listeners' own decision, may in many cases, go quite fast. Also, the listener has the freedom to follow his or her own pace.

As noted, this method leaves it solely to the listeners to determine the threshold. While one should not underestimate the abilities of listeners, this method does not account for, to a degree at least, the psychological aspect of listening. I.e. that we sometimes may be "fooled" by our own perception, and thus, the established threshold *may* produce inaccurate results. Thus, it may be reasonable to assume, using this method there is a higher risk that the established threshold is not at the *actual* threshold for that particular listener.

On the other hand, Olive & Toole (1989) reported that the reproducibility of the results in their experiments was high. Based on the results from an unknown amount of trials, with only isolated exceptions, the listeners responded throughout with standard deviations typically in the range of 1 to 3 dB.

3.3.2 Simple yes/no

In a yes/no task we present the subject with a set of isolated stimuli differing in level from below to above the expected threshold. In each trial we present one stimulus to the subject, which is then asked whether he or she has detected the stimulus, yes or no.

The main problem with yes/no tasks is that due to the subject's responses are self-reported, bias might occur (Green 1993), cited in Soranzo et al. (2014). E.g. a subject might give a positive response (yes), i.e. a *false alarm*, even in absence of any stimulus, Soranzo et al. (2014) note.

While no literature reports the use of the yes/no procedure in experiments investigating the audibility of reflections, it is nevertheless conceivable to be applicable to an experiment investigating coloration from reflections. In that case we would then present the reflection at various levels to the subject, ranging from well below to above the expected threshold, together with the direct signal which would be presented at a fixed level. The subject's task

would then be to report, after each new presentation, whether he or she has detected coloration due to the presence of the reflection, with a yes or no.

3.3.3 Alternative-Forced-Choice (AFC)

In this procedure the subject is presented with several intervals. E.g. in a two-interval alternative forced choice the subject is presented with two intervals and has to decide whether the stimulus variable is present in the first or second interval (Zwicker & Fastl 2006:10). Dependent also on the number of intervals, one stimulus variable changes its level across several trials, whereas the others (standards) are fixed. The difference between standards and variable ranges from below to above the expected detection or discrimination threshold, and subjects are asked to report which the variable stimulus was (Soranzo et al. 2014). Also, the listeners are often provided with feedback, informing them about correct and incorrect answers (Zwicker & Fastl 2006:10).

E.g. in a 3-interval alternative forced choice (3I - 3AFC), to estimate the detection threshold of a tone within noise, Soranzo et al. (2014) note that we present three noise intervals in succession and only one includes the target tone. Subjects' task would then be indicated which interval contained the tone, at the end of each trial.

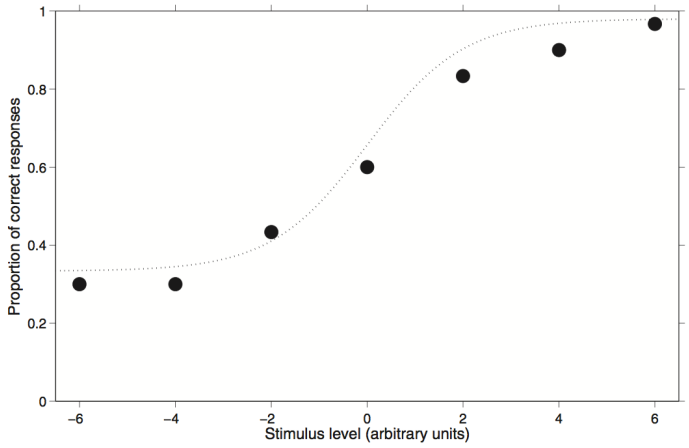


Figure 3.8 *Hypothetical results of a 3 AFC task. The dotted curve interpolating the subject's data points is the psychometric function (From Soranzo A. & Grassi M., Psychoacoustics: a comprehensive Matlab toolbox for auditory testing. Frontiers in Psychology, 5 (712).*

In alternative forced choice procedures, the threshold is statistically determined from the *psychometric* function, which tracks the relationship between a subject's performance and the level of a stimulus variable. The threshold lies at an arbitrary point of the psychometric function, defined as *p-target* (i.e. probability target), Soranzo et al. (2014) note. In threshold determination based on the psychometric function, the procedure searches for the stimulus level eliciting the p-target, i.e. the proportion of correct responses. E.g. in 2 AFC procedures the p-target should be 75% of correct responses due to the proportion of correct responses spans from 0 to 100%.

The alternative forced choice (nAFC) procedure is widely reported in use by experimenters, investigating the audibility of the effects of reflections. In such experiments it is then most often paired with an *adaptive staircase* according to literature reports.

3.3.4 Adaptive staircase

In adaptive procedures, such as the staircase, the stimulus presented to a subject in each trial, is dependent on the previous answers by the subject. As such, we call them *adaptive*, because they adapt to the subject's performance.
Adaptive procedures maximize the ratio between the stimuli presented close to the threshold and those presented far from threshold (Watson & Fitzhugh 1990). As such, they are more efficient than non-adaptive procedures (Soranzo et al. 2014), because they quickly approach threshold. Soranzo et al. (2014) note that their efficiency makes them generally preferred over non-adaptive procedures, especially when estimating just threshold, and not the entire psychometric function.
In staircase procedures the stimulus variable adjusts its level gradually according to the subject's response. In (adaptive) staircase procedures, a run typically starts from above (Soranzo et al. 2014). The stimulus variable is then presented at its maximal level, and decreases until threshold is reached. Similarly, if from below, the stimulus variable is presented at the lowest possible level, e.g. no signal, and increases until threshold is reached.

Different staircase procedures differ in how they act upon the answers given by the subject.

In *the Method of Limits* the run terminates after just one reversal (incorrect response), which thus defines the threshold. As such, the threshold estimation should be quick, but the procedure does not account for *attention lapses* during runs (Soranzo et al. 2014). I.e. the chance is high that a subject, at some point during a run gives a negative response (stimulus variable not heard) even though the stimulus was well above threshold for that particular subject. As such, the procedure might be viewed as unreliable, producing incorrect thresholds. According to Soranzo et al. (2014), the Method of Limits tracks 50% of the psychometric function.

In a *simple up/down* task a run does not terminate after the first reversal (wrong answer) as in the Method of Limits, but proceeds until a pre-set number of reversals is reached. When the respondent gives a correct answer, the level of the stimulus variable, delta g, decreases. Similarly, when the subject returns a wrong answer, the level of the stimulus variable increases. According to Soranzo et al. (2014), the simple up/down method, tracks 50% of the psychometric function.

In a *transformed up/down* as defined by Levitt (1971), the variable stimulus moves down towards threshold only after minimum two positive responses in a row, whereas it moves up after just one negative response. Levitt (1971) suggests moving down after the subject returns *n* positive responses (e.g. two) and moving up after the subject returns one negative response. The probability of moving down, towards threshold, becomes p^2 whereas the probability of moving up, away from the threshold, is either 1-p (i.e., one negative response only) or p (1-p); i.e. one positive response followed by one negative response.

Soranzo et al. (2014) summarize with the following equations:

$$p^2 = p(1-p) + (1-p) = 1 - p^2$$

$$p = \sqrt{1/2} = 0.707$$

Thus, a 2-down 1-up method tracks 70,7% of the psychometric function. Similarly, a 3-down 1-up will track $\sqrt[3]{1/2} = 0.794$, thus 79,4% of the psychometric function.

However, while the transformed up/down procedure is widely used (Soranzo et al. 2014), according to Leek (2001) the 2-down 1-up is not reliable, especially when utilized in a 2AFC task. Thus, while a 3-down 1-up procedure might be long, it would be a better option what reliability is concerned, especially when utilized in a 2AFC task.

According to Soranzo et al. (2014), we can change the stimulus level in two ways; either by addition/subtraction or by multiplication/division.

Soranzo et al. (2014) note, the simplest way is to reduce/increase by subtracting/adding a fixed amount every time a subject returns a positive/negative response. The fixed increment by which the stimulus level is increased/reduced, is called step size.

E.g. a in a simple 1-up 1-down task, where the step size is 1 dB, the stimulus level decreases by 1 dB for every correct response, and increases by the same 1 dB for every incorrect response. On the other hand, in a transformed 1-up 2-down task with a step size of 1 dB, the subject must give two correct responses in a row before the stimulus level drops by the step size.

To make the transformed up-down procedure more efficient (less time consuming), it is convenient, in the initial phase, to use a larger step size, approaching the threshold relatively quicky, and a smaller one when the subject is closer to threshold, for fine estimation.

Soranzo et al. (2014) note that we usually calculate the threshold by averaging the various thresholds collected during the runs. In the case with the simple and the transformed up-down procedure, we calculate the threshold by averaging either arithmetically or geometrically the various thresholds at the reversal points. Alternatively, we can also use the median.

The adaptive staircase alternative forced choice (AFC) procedure was reportedly used by Salomons (1995), Bech (1994:1995), Brunner et al. (2007), and possibly others, in experiments estimating thresholds for coloration perception. Bech (1994) reported the following concerning the general procedure of his sound field experiments, investigating the

contribution of individual reflections on timbre of reproduced sound: *"The adaptive staircase two alternative forced choice procedure was used. The level of the reflection under investigation was varied adaptively (two down/one up) to estimate that level, which would produce 70,7% correct responses (Levitt 1971). The step size initially was 4 dB was reduced to 2 (TD experiments) or 1 dB (jnd experiments) after three reversals. Typically, 10-15 reversals would occur during each 50-trial block. For each block the threshold was estimated as the average of the levels at the midpoints of runs 4, 6, 8, etc. The reported TD or JND are averages across subjects for eight (noise) or six (speech) 50-trial block per listener"*.

3.3.5 Remarks

It becomes clear that when considering reflection level (amplitude) the experimental variable, we need to consider the spectrum of the reflection as the experimental parameter. By this approach, using one of the procedures mentioned, we must divide the experiment into separate blocks (runs). E.g. in the first experimental run, we may determine the absolute threshold for coloration without filtering (of the reflection under investigation), and in the next run, with filtering (of the reflection under investigation). We can then, after the completion of the experiment, compare the established thresholds for each subject with and without filtering.

This approach was used by Bech (1995) in his experiment investigating the influence of filtering individual transfer functions of reflections on their detection thresholds, as accounted for in the theory chapter.

The alternative to this approach would be to estimate thresholds for multiple parameter values, i.e. not just with and without filtering. E.g., if investigating the effect of multiple absorbers, we may establish thresholds for several different filter transfer functions, or if using the plain low pass filter approach, establish thresholds at multiple lowpass cutoff-frequencies. However, dependent also on the procedure chosen, these approaches would naturally make the experiment longer, increasing the number of runs, compared to distinguishing just two data points (with and without filtering).

The approach of determining thresholds at multiple cutoff-frequencies was reportedly used by Olive & Toole (1989) in their experiment, previously accounted for in the theory chapter.

Alternatively, we may consider the spectrum of the reflection the experimental variable, using one of the procedures mentioned. However, this method would probably be sensible only if we use simple basic lowpass filtering.

We must then present the reflection at a fixed level, and we then first have to determine the level at which the reflection should be fixed. In this regard, the easiest solution maybe to present the reflection under investigation at the same level as the direct sound.

However, practically no reflections will have such high natural levels relative to that of the direct sound. A more natural approach would, therefore, be to first establish detection thresholds for coloration without filtering, considering the amplitude of the reflection as the variable. Then, in a new experiment, with the reflection under investigation fixed at the established amplitude threshold, apply filtering to the reflection, from less to more dramatic, to establish a new kind of threshold - the cutoff frequency at which the reflection (its effect) becomes inaudible. This approach then automatically assumes that the reflection would be harder to detect after filtering, after first having established a threshold considering the reflection level the variable.

3.3.6 Choosing a procedure

As The Method of Adjustment and the Adaptive Staircase Alternative Forced Choice (AFC) were the only procedures reportedly used by other researchers in prior experiments (investigating audibility of reflections), the selection was narrowed down to one of these.

Also, I concluded that it would be best to consider the amplitude of the reflection the variable, as none of the alternatives are documented by examples in literature.

Of the two procedures mentioned, The Method of Adjustment is seemingly the fastest approach for threshold estimation. Also, in this procedure it is entirely up to the respondent to

decide what level is the threshold, as opposed to the adaptive staircase AFC, where the threshold is statistically determined from the answers given by the respondent.

A thorough evaluation of the two procedures was done. The decision fell on the adaptive staircase alternative forced choice as the procedure for the listening experiment (for details, see section 3.4.4).

The reason for choosing this procedure was that it seemed to be a standardized procedure based on the reports from prior experiments. Also, in a pilot experiment I found it to be the most reliable (three down/one up variant).

However, one downside was that it appeared more time consuming than the alternative, the Method of Adjustment. E.g. in the pilot experiment, the time determining one threshold value, thus completing one run, took 6-10 minutes using the three- down/one up variant. To see this in perspective, considering I was looking into the possibility of investigating several incident angles from the echogram and also using various test signals, the experiment would be time-consuming. On top of that, Zwicker & Fastl (2006:14) also note, it is advisable to perform several runs of the same type, because the results may vary if performing the same experiment several times. Fortunately, from the pilot experiment, I found the three-down/one up variant reliable in this regard. I.e. performing the same run several times, the standard deviations were within the range of 1 - 2,8. As such, the number of runs of the same type could be held to a minimum (two or three maybe).

Of course, if attempting to shorten the experiment duration, there was also the option to choose the shorter two-down/one up 2AFC. While this constellation was used in, e.g. the experiments by Bech (1994) and (1995), according to Leek (2001) the two-down/one up variant is not reliable, especially in a 2AFC task, as noted.

Therefore, deciding on the three-down/one up 2AFC, which was also my initial intention, I had to reconsider the number of reflections to include. After all, to investigate the research question, it was unnecessary to include all the incident angles from the model (when estimating thresholds for coloration for one reflection at a time). Because the main interest was the effect of high-frequency absorption (on coloration) from the perspective of the listening room (reproduced sound), I had excluded front wall reflections (predicted in section 3.2.7) from the investigation (due to loudspeaker directivity). Furthermore, based on the listening in Pure Data, I concluded that I also could exclude several other incident angles

from the investigation. The reason was, these reflections also lacked high-frequencies after the processing by the HRTFs.

The reflections that had the most preserved high-frequencies (after the processing by the HRTFs) were those arriving from the sides. I.e., the reflections no. 5+6 and 9+10 (table 3.1).

As such, it made sense to include these only for the experiment.

Therefore, all things concerned, the following practical questions then became relevant for the listening experiment:

1) How noticeable is comb filter coloration from sidewalls in listening rooms?

2) Does the absorption characteristics (filtering effect) of a relatively thin absorbent reduce the audibility of the coloration significantly?

However, while the results of the experiment cannot answer the former directly, it may help here to compare the established thresholds from the experiment with the natural predicted reflection levels. We can obtain the natural predicted levels of the sidewall reflections by looking at the echogram from CATT.

The latter question, on the other hand, assumes that the created filter gives a reasonable approximation of the absorption (filtering effect) provided of the material. While the modelling approach used in 3.2.5 did not match the absorption characteristics of the material 100%, it gave a reasonable fit to the data, at least in the three first octave bands (table 3.3.). As such, it will suffice in the investigation.

3.4 The experiment

3.4.1 Recruitment of respondents

Initially, the plan was to recruit 5-8 participants and repeat the experiment several times, thereby strengthen the durability of the experiment, in a statistical sense. However, after some advice, I found it better to recruit a larger selection of participants, making it more durable without having to repeat the experiment several times. The strategy of running the same experiment multiple times, was meant to account for both interindividual, as well as intraindividual differences among various participants, as noticed by Zwicker & Fastl (2006:14).

However, recruiting a larger selection of respondents is probably the better strategy, attempting to increase the reproducibility of the results, compared to choosing a smaller selection of respondents.

I opted then, instead, involving around 20 participants for the experiment, in which the recruitment form is included in the appendix.

However, finding 20 individuals willing to volunteer for this type of experiment proved difficult. Part of this, we must probably attribute to the events occurring at the time of planning the experiment, as noticed.

Nevertheless, 15 individuals agreed to participate in the listening experiment. Of these, two had to cancel their participation. Thus,13 were left, and of these, 10 were males and 3 were females. Ideally, the gender distribution could have been more even, but the resulting distribution was not deliberate. The participants were in ages from 25 to 70.

Of these, two individuals located at the Department of Musicology, University of Oslo - a student and an employee. The 13 others were all teachers at Mailand upper secondary school, and none of these had any previous experience with such listening tasks.

3.4.2 Test signals

I chose two different stimuli signals for the experiment. Generated in Audacity, the first stimulus was a 1-second sample of band limited white noise (20 Hz - 16 kHz). The sample was lowpass-filtered at 16 kHz, removing some hiss from the high-frequencies of the signal.

The second stimulus was a 3,2 seconds sample of orchestrated music taken from the European Broadcast Union, Subjective Quality assessment Material (EBU-SQAM), download version. The sample was taken from track 65 (Orchestra, Strauss). Information on the recording of this track was not given in the manual, but for more information regarding the assessment material in general, see appendix 1.

Both the noise and music sample were edited in Audacity and bounced to disk at a sample rate of 44100 kHz, 32-bit floating-point.

Based on evaluation through both listening and a pilot experiment, the duration of 1 second for the noise appeared sufficient for judging the timbre of this noise sample, after having tested both shorter and longer duration intervals.

On the other hand, taking into account the transient sounds in the music sample, it had to be of longer duration such that the respondents had better time to judge its timbre. In this regard, I found the duration of 3,2 seconds to be the shortest acceptable, while still being able to judge the timbre of the signal properly.

As the music sample came from a stereo track, the channels of this track were summed to mono in Audacity *before* the processing in Pure Data. The reason for using mono signals as source (before processing), is accounted for in the experimental setup.

As noted, after the initial editing in Audacity, the stimuli were processed through the HRTFs in Pure Data. Further editing was done in Audacity before they were imported into Matlab, the application that ran the procedure.

The choice of using (band limited) white noise as a test signal in a listening experiment investigating the audibility of coloration, was accounted for in section 2.1.2. As such, a

thorough argumentation for this choice should not be necessary for this section. However, we might add to what has already been said in section 2.1.2, the following. It is reasonable to assume that white noise should be receptive to various kinds of filtering due to the factors mentioned in 2.1.2. As such, if we assume, on a general basis, that spectral limiting of high-frequencies of a reflection makes coloration harder to detect in broadband signals, we should observe this trend in white noise due to its wide bandwidth. As such, white noise should be a good indicator of whether spectral limiting of the reflection will make coloration less audible.

However, white noise might not represent the type of signals we listen to normally, such as speech and music. As mentioned, it has a long-time power spectrum independent of frequency. Therefore, we should also include a more realistic signal, and the choice fell on a sample of classical orchestrated music. In general, the slow modulation characteristics of classical music are more suited to reveal timbral differences than other genres that are more transient (Kleiner 2013:119).

3.4.3 Experimental setup

The experiment was set up on a Macbook Pro laptop, running MatLab (I account for the general procedure in the section 3.4.4).
The experiment took place at two separate locations. For the two respondents from the Department of Musicology, University of Oslo (UiO), the experiment was held here. For the rest of the respondents, the experiment was held at Mailand upper secondary school.

At both locations, the room where the experiment was held, was quiet. Also, no noticeable noise came from adjacent rooms. No other individuals than the respondent were present in the room during each listening task, and it was made sure no unforeseen events could occur, interrupting the respondent during the listening.

The listening took place over headphones of make AKG k240 mk2, which is high-quality open-back studio headphones. The choice of open-back was due to the apparent better soundstage on the former relative to the latter, which may or may not improve such aspects as externalization in virtual sound field listening.

The stimuli signals were played back via a Motu M2 audio interface at a sample rate of 44100 Hz, 24-bit floating-point. The headphones were connected to the headphone output of the audio interface, and each respondent could choose between three fixed playback levels. The reason for this was such that each individual could find a comfortable level. However, the level had to be set beforehand, and there was no option to change the level during the course of the experiment. As such, the respondents were advised to choose the playback level wisely.

As noted, the stimuli signals were monophonic, duplicated in a left/right configuration for the direct sound components and then pre-processed through HRTF (explained by the block diagram in figure 3.6). The pre-processing of signals through binaural filters (HRTFs) was to simulate sound field listening where the stimulus presents to the listener over two loudspeakers, and the listener is seated at a fixed position in front of the loudspeakers in the room.

However, because the source signal was mono, the listening configuration illustrated above essentially mimics a situation where we are listening to a set of loudspeakers receiving the same signal, as opposed to listening in stereo where the loudspeakers are receiving different signals.

The reason for choosing this presentation was that we judge timbre best when the source is a mono signal. Thus, coloration is also better perceived using this presentation, due to the binaural decoloration mechanism, as previously mentioned.

However, while this configuration is not the same as listening monaurally, it essentially gives the same result. The reason is that the two ears are receiving the same signal. However, we should note that in real sound field listening immediate head movements, as well as that the listening position seldom is fixed, contribute to the left and right ear receiving different signals.

Also, two simultaneous reflections were presented to the respondents (No 5+6 from table 3.3). One threshold value was thus determined for two reflections, not one one, as illustrated in most of the experiments in literature.

The reason for presenting both the direct sound and the stimulus variable in a pairwise configuration was to prevent confusion caused by the simultaneous perception of timbre and localization, as discussed in the theory chapter, and as mentioned in Bech (1994) and Salomons (1995). The alternative to this configuration would be to simulate only the left or right loudspeaker, as did Bech (1994) and (1995), and present only the reflection from the side of the speaker. However, listening to one channel only appear unnatural over headphones. Therefore, I decided to present the stimuli over two channels in the experiment.

Any differences in loudness between the stimuli representing direct and reflected sound components were compensated for beforehand, such that the established threshold values were correct.

3.4.4 General procedure

As noted, the experiment was conducted in Matlab (Mathworks inc.). The software package in Matlab that was used to run the procedure was psylab. The procedure was carried out by three different Matlab-scripts, which can be found in the appendix 2.

In all listening sessions the task of the respondents was to detect a change in the timbre of a band limited white noise signal or a music signal (3.4.2). The interpretation of timbre was discussed with the respondents beforehand. The instruction form given to the respondents can be seen in the appendix 1.

One psychoacoustic quantity was determined: the detection threshold for coloration (timbre change), corresponding to the just noticeable difference for two stimulus categories: a white noise signal and a music signal (section 3.4.2).

The experiment was split in two, carried out on separate days. The stimuli presentations on the two days were identical except that on the second day the audio file containing the stimulus variable (reflection pair) had undergone filtering. As such, on the first day thresholds were established without filtering, and on the second day thresholds with filtering of the stimulus variable (reflection pair) were established, for each of the two signals (section 3.4.2).

The stimuli variable (reflections under investigation) were processed through the filter created in section 3.2.6. As noted, it approximated the absorption characteristics of a 20 mm mineral wool absorbent (table 3.1). The filter processing was done in MatLab prior to the second day of the experiment. The filters' magnitude data and response were shown in table 3.3 and figure 3.1, respectively.

The whole experiment had a total length of approximately 1 1/2 hour over two days, for each respondent.

An adaptive staircase two-alternative forced-choice procedure (section 3.3.4 and 3.3.3) was used for threshold estimation. Thus, the respondents had to judge two intervals according to the 2AFC. These were the following:

1) The standard. The interval containing the direct sound only, i.e. only the direct sound field components according to the model were present.

2) The comparison. The interval formed by adding a variable level of the reflection pair under investigation to the standard.

Each of the days, the session consisted of in all four runs; two runs of noise and two runs of music, which took approximately 45 minutes to complete. The runs alternated between the noise and the music in which the noise was the starting signal. The respondents could choose to take a rest between each run if they wanted. However, none of them reported any fatigue from the listening.
The two runs for each stimulus category were identical in presentation, and the threshold for each stimulus was determined from calculating the mean value from the two runs. In general, the reproducibility of the results was high, with standard deviations within the range from 1-2,8. The latter is documented in the appendix.

Before each session, or between runs, the respondents could audition the various stimuli

intervals through a separate graphical user interface (figure 3.9) not part of the procedure. As such, the respondents could familiarize themselves with the timbre of the standard and comparison interval. Or alternatively, if they felt the need to recalibrate their ears as to what was the standard and what was the comparison interval. In this interface (figure 3.9) the stimulus variable in the comparison interval was equal to the level of the direct sound. The differences in loudness between the two intervals were compensated for using the same approach as described at the end of section 3.2.8.

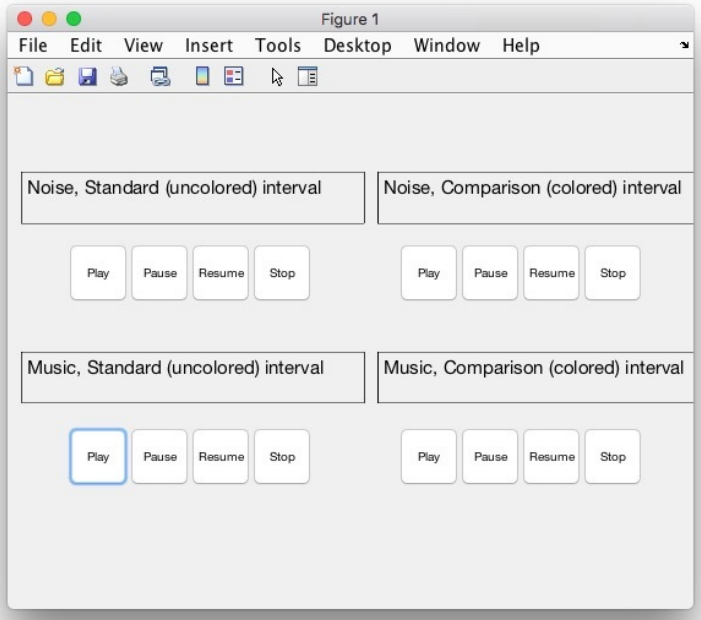
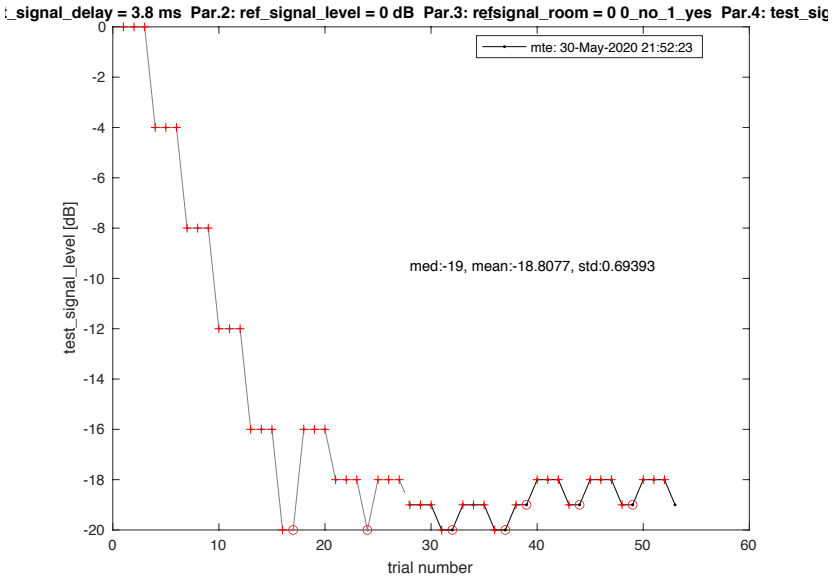


Figure 3.9 *The interface where each respondent could familiarize with each of the intervals presented during the experiment.*

At the start of a run, the initial level of the stimulus variable (reflection pair) was equal to that of the direct sound. The level of the stimulus variable was varied adaptively (three down/one up) to estimate the level which should produce 77,4% correct responses, according to Levitt (2001), as noted. The initial step size was set to 4 dB. The beginning of each run was a so-called familiarization phase, where the purpose was to familiarize the respondent with the procedure, and the trials during this phase were not included in the determination of the threshold. However, the step size during the familiarization phase was reduced such that the

respondent could carry the stimulus variable towards a threshold more or less rapidly. After this initial phase, the measurement phase began. According to Hansen (2017), at the start of the measurement phase the respondent should have approached a point relatively close to the threshold. At this point the stimulus variable changed only with the minimal step size, which was set to 1 dB. The threshold was then determined based on all values that the variable had taken on during the measurement phase.

Dependent on the type of stimulus and the answers given by the respondent, one run typically consisted of 50-80 trials, and it typically took from 7-10 reversals (section 3.3.3 and 3.3.4) to determine a threshold value completing one run (figure 3.10).



Figur 3.10 *The graph tracking all the responses given by a respondent during a run. A cross indicates a correct answer, a circle indicates an incorrect answer.*

Through an answer graphical user interface (figure 3.11) the respondents followed the procedure and gave their answers. The following question was asked to the respondents at the end of each paired interval presentation: which interval is colored?

The respondents could either give their answers by pressing a number key on the computer keyboard, or clicking inside one of the number boxes in the answer GUI (figure 3.11). E.g., if they gave their answers by pressing a number key - if they thought interval no. one contained the stimulus variable (reflection pair), they pressed the corresponding keyboard button. And vice-versa, if they thought the second interval in each presentation contained the stimulus variable, they pressed the corresponding number key. The respondents were also given feedback on the screen, indicating if their answer was correct or incorrect after each presentation.

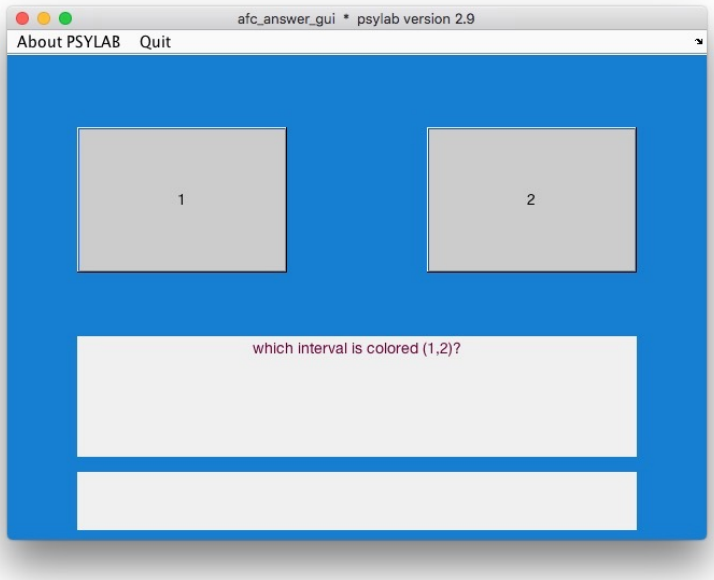


Figure 3.11 *The answer GUI.*

The comparison stimulus was present either in the first or second observation with equal probability. The other period contained the standard. Based on initial testing as well as a pilot experiment, the optimum inter stimulus interval was found to be 0.8 seconds. If this interval is too small, the respondent will have trouble comparing the timbre of the two intervals, and similarly if the inter stimuli interval is too long.

Different loudness between the intervals was compensated for in realtime by the procedure, such that the respondent could not use a loudness cue in the detection.

4 Results and analysis

This chapter is divided in the following sections. In the first section I present the collected data from the experiment. That is, the threshold values for each respondent as well as the mean threshold values obtained from these, for each of the two stimuli categories.

Then I will do a frequency analysis of test signals with emphasis on the comb filtered versions as they were presented to the respondents at the start of each run. I will discuss the established threshold values in the light of this frequency analysis. I will try to approach explanations as to why these results are observed and try to link them to existing theories in the field.

If applicable I will also compare the results to that of prior experiments that used the same or similar type of signals.

4.1 Collected data

Table 4.1 presents the established threshold values from the psychoacoustic experiment. Listed first are the threshold values produced on an individual basis. At the bottom of table 4.1, the mean threshold values are given. As previously mentioned, the established threshold values from each individual were calculated by averaging the thresholds from two identical runs. The standard deviations produced from these two runs were within the range of 2,8.

The main threshold values are listed at the bottom of table 4.1.

The data in table 4.1 was obtained from reflections 5+6 in the room model (table 3.1)

All the data from the experiment can be retrieved following one of the links in the appendix. The results from each run were automatically saved to files with the prefix “psydat”. These are in the ASCII-format. There is one psydat-file for every respondent. In addition to these files containing the results from each respondent, time and date are also available.

	Unfiltered dB		Filtered dB		Difference dB	
Signal type	**White noise**	**Music**	**White noise**	**Music**	**White noise**	**Music**
Respondent 1	-17.08	-11.66	0.91	-8.85	17.99	2.81
Respondent 2	-18.43	-13.45	-11.58	-11.76	6.85	1.69
Respondent 3	-18.38	-10.08	-2.93	0.77	15.45	10.85
Respondent 4	-20.49	-17.26	-5.06	-9.04	15.43	8.22
Respondent 5	-19.91	-8.63	-5.22	0.13	14.69	8.76
Respondent 6	-14.22	-12.29	-2.98	-10.78	11.24	1.51
Respondent 7	-7.69	-8.01	-2.94	-7.86	4.75	0.15
Respondent 8	-16.30	-6.69	-0.67	-1.16	15.63	5.53
Respondent 9	-19.81	-8.14	-7.02	-5.87	12.79	2.27
Respondent 10	-15.35	-5.97	-2.24	1.24	13.11	7.21
Respondent 11	-9.00	-3.25	0.62	1.54	9.62	4.79
Respondent 12	-10.84	-11.93	-0.32	-4.75	10.52	7.18
Respondent 13	-28.23	-20.9	-17.41	-20.92	10.82	-0.02
Mean thresholds	**-16.59**	**-10.55**	**-4.37**	**-5.94**	**12.22**	**4.61**

Table 4.1 *Gives the individual and mean established threshold values obtained from the psychoacoustic experiment. The values indicate the reflection level where spectral changes became just audible due to the presence of two identical, but opposite reflections in the comparison interval. Values are in unit decibel.*

As we see from table 4.1 there are quite a few inter-individual differences in the obtained results among the respondents. For the white noise signal with unfiltered reflections, the variance is calculated to a value of nearly 30. Because of the mentioned circumstances, none

except one of the participants had received any training beforehand. Also, the trained listener had moderately experience with critical listening tasks.

4.2 Detection cues

The respondents were instructed to detect spectral differences, i.e. preferably a change in timbre of the two stimuli categorizes, as explained in the method chapter, section 3.4.4.
The respondents were then asked after each of the two listening sessions, which spectral "cues" they used to detect the comb filtered intervals. For the white noise signal with unfiltered reflections all respondents named a distinct pitch sensation as the primary cue for the detection. However, after the second day where the reflections under investigation were filtered, the pitch sensation in the colored noise signal was reportedly gone. Instead, many reported of a slow low frequency modulation as the cue for detection.

For the music signal, on the other hand, all respondents noticed distinct timbral differences as the cue for detection. With the unfiltered reflections all had noticed that the comb filtered interval was substantially brighter than the original. Some also noticed the apparent thinner sound of the comb filtered version compared to the original. Several noticed the comb filtered signal sounded "unnatural", particularly at reflection levels close or equal to the direct sound.

4.3 Spectral analysis of test signals

We should first start by mapping the interference frequencies of the comb filter that was created from the superimposition of the original sound with the test reflections in the experiment. Then we can look at the spectra of the created comb filtered signals to see if it would be possible to identify these interference frequencies.
Following the equation for calculating comb filter interference frequencies given in the

theory, section 2.1.1, figure 4.1 gives the 23 first interference frequencies calculated from a 3,9 ms comb filter.

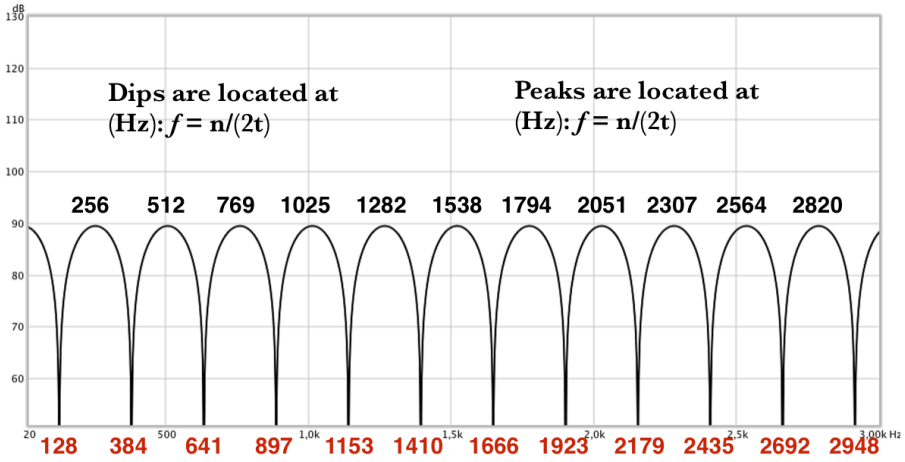


Figure 4.1 *Calculated interference frequencies of a 3,9 ms comb filter.*

Figure 4.2 and 4.3 show the frequency spectra of the comb filtered signals presented in the comparison intervals of the experiment. In these figures, the stimulus variables (reflections under investigation) were unfiltered and had the same level as the direct sound.

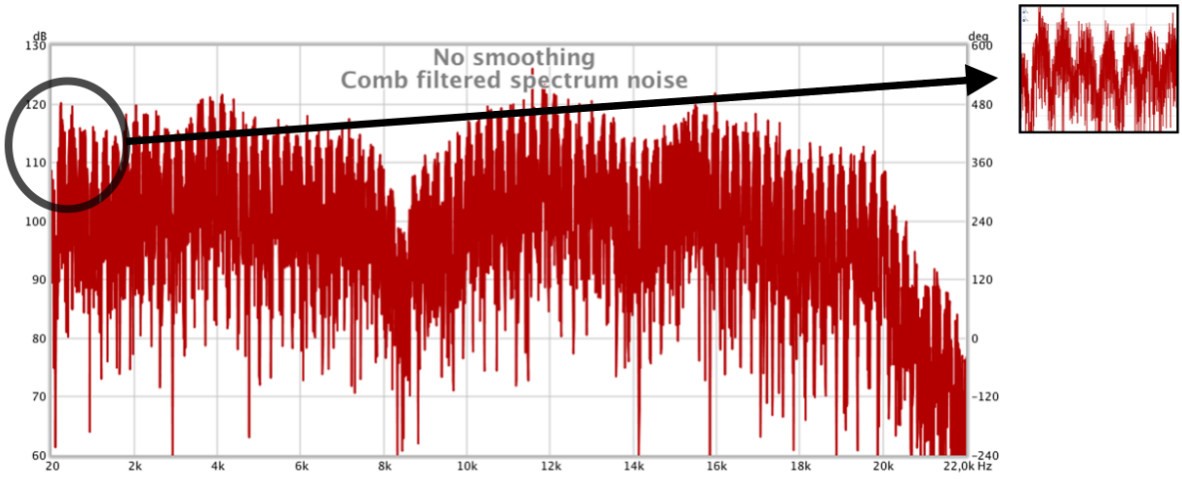


Figure 4.2 *The spectrum of the comb filtered white noise signal presented in the comparison interval. The amplitude of the reflection was equal to that of the direct sound. A linear frequency scale is used.*

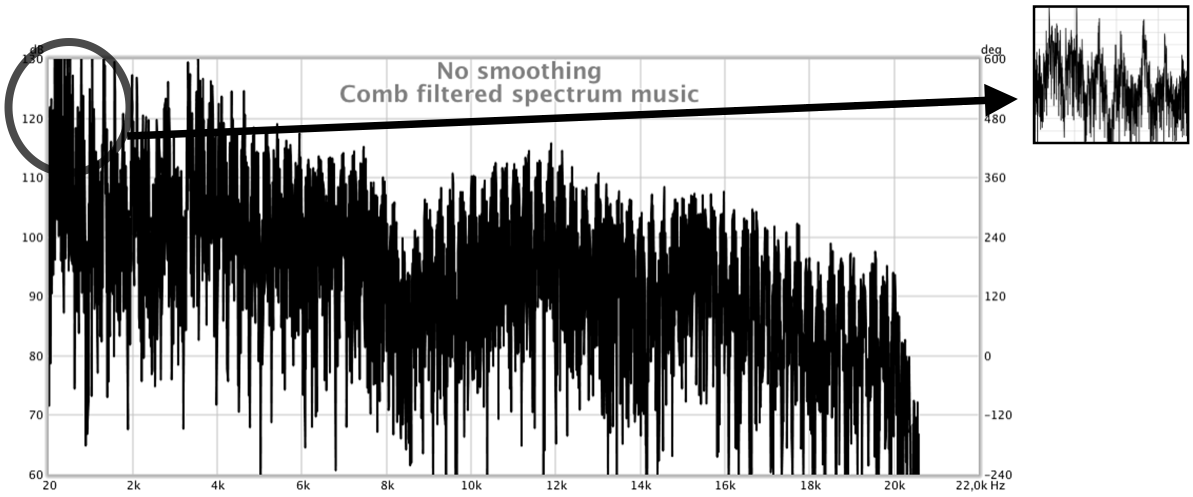


Figure 4.3 *The spectrum of the comb filtered music signal presented in the comparison interval. The amplitude of the reflection was equal to that of the direct sound. A linear frequency scale is used.*

Thus, figure 4.2 and 4.3 represent the spectra of the comparison intervals at the initial phase of a run, with unfiltered reflections.

A direct comparison between the spectra of the direct (standard) and comb filtered (comparison) intervals for each of the two categorizes of signals is given in figure 4.4 and 4.5, respectively. In figure 4.4 and 4.5, by overlaying their spectra, we see the effect of comb

filtering on the power spectra of the signals. As noted, a comb filtered signal will have increased loudness compared to the original signal.The differences in loudness between the standard (blue and gold, figure 4.4 and 4.5) and comparison intervals (red and black, figure 4.4 and 4.5) were compensated for in realtime by the procedure, eliminating the possibility that the respondents used a loudness cue in the detection. However, from figure 4.4 and 4.5, we may get the impression that the comb filtered versions (red and black) are (still) louder than the originals (blue and gold). On the other hand, all the files measured around -18 dB RMS, choosing "RMS" in the analyzer window of Audacity. In the comb filtered versions the frequency dips (which are not markedly visible in figure 4.4 and 4.5) may also counterbalance the peaks such that the overall loudness change would be less than what the figures suggest.

In the comb filtered noise signal (red), seen in figure 4.4, the relative balance of the low, mid, and high frequencies appears to stay more or less equal to the original signal (blue).

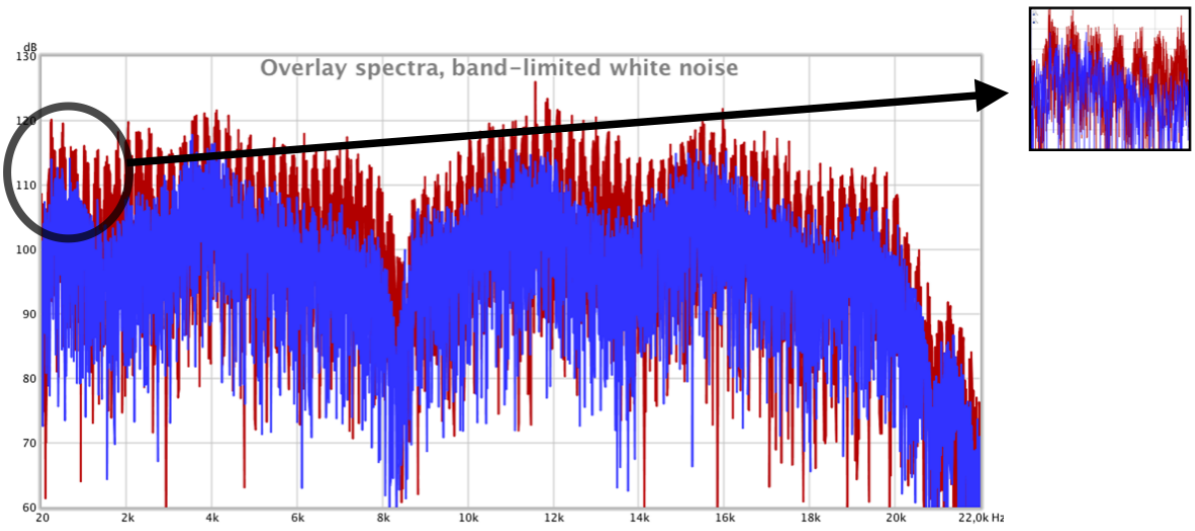


Figure 4.4 *A comparison of the comb filtered spectrum (red) to the original (blue) for the white noise signal. A linear frequency scale is used.*

For the music, on the other hand, figure 4.5 reveals that the comb filtered version (black) is substantially brighter than the original (gold). Specifically, the power in the frequencies of the

filtered signal (black) gradually increases towards the high-frequencies relative to the original signal (gold). This implies that the comb peaks are most effective towards the high frequencies for this signal. The reason may be the relationship between the interference frequencies of the comb filter and the spectral content of the music.

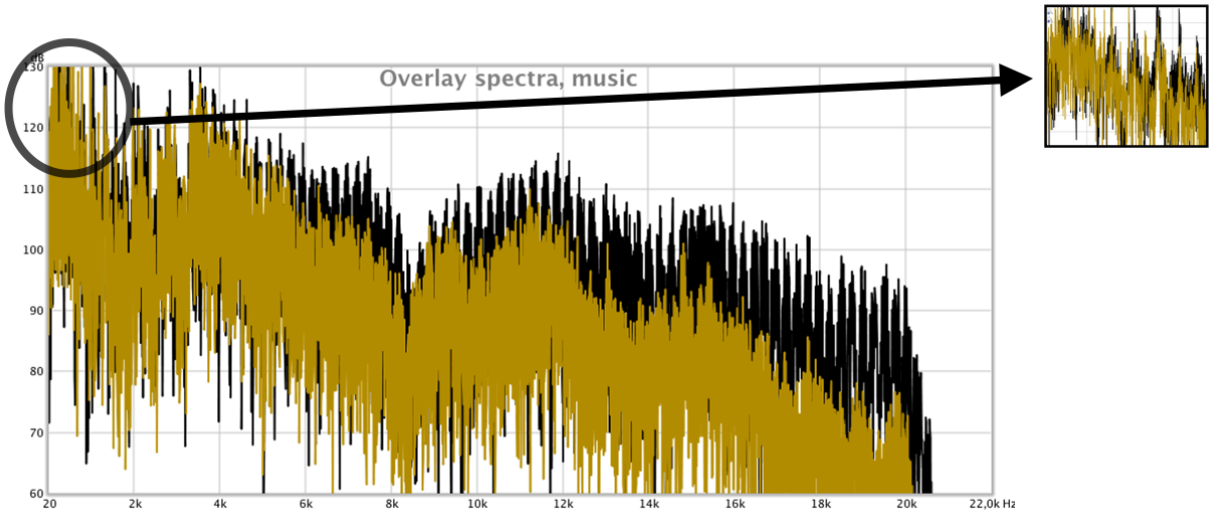


Figure 4.5 *A comparison of the comb filtered spectrum (black) to the original (gold), for the music signal. A linear frequency scale is used.*

To get a clearer view of comb peaks and dips we should zoom in on the frequency axes. In figure 4.6, we clearly see the characteristic shape of the filter, imprinted on the noise signal, with regularly spaced peaks and dips.

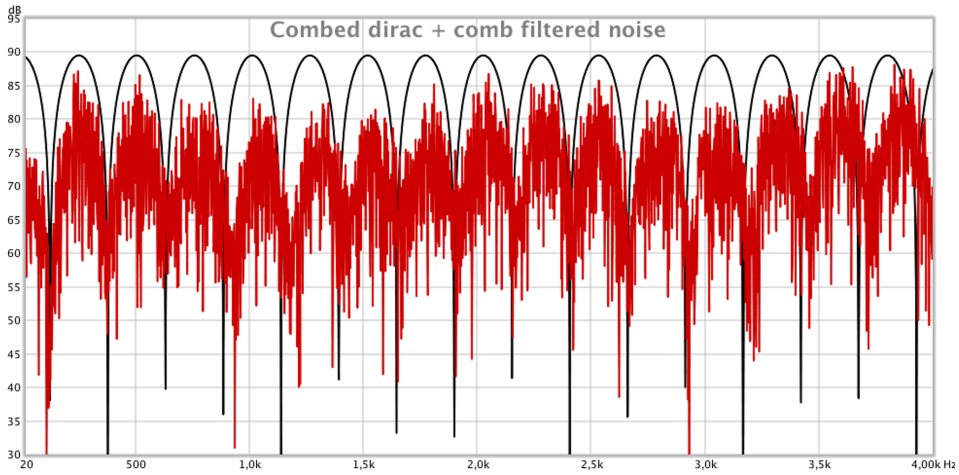


Figure 4.6 *Zooming in on the frequency axis we observe the comb filter shaped frequency response characteristic inherit in the noise (red), after combing of direct and reflected components in the comparison interval. The signal is compared to a theoretical comb filter generated from the superimposition of a delayed dirac pulse with its original (black). The frequency resolution goes from 20 to 4000 Hz.*

For the music signal, in figure 4.7, the filters' response is still visible but not as clearly defined as in the noise, due to the transients of the music signal.

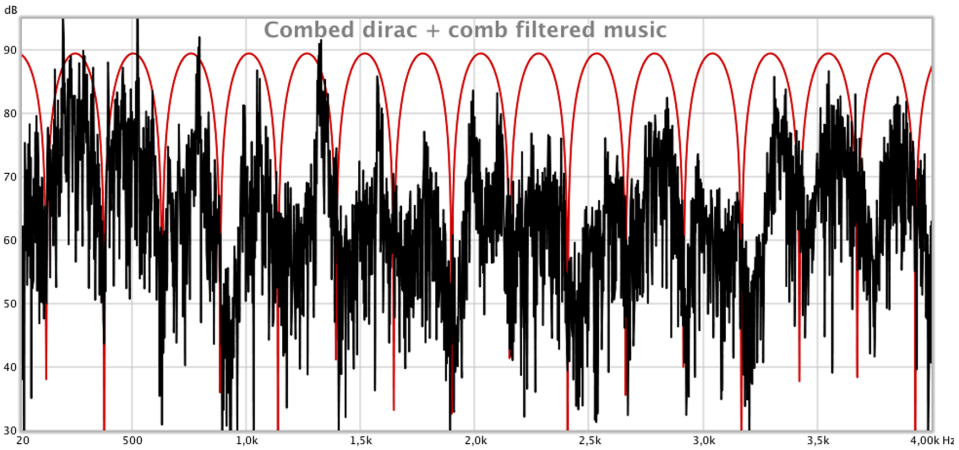


Figure 4.7 *Zooming in on the frequency axis we observe the comb filter shaped frequency response characteristic in the music (black), after combing of direct and reflected components in the comparison interval. The signal is compared to a theoretical comb filter generated from the superimposition of a delayed dirac pulse with its original (red). The frequency resolution goes from 20 to 4000 Hz.*

Zooming further in on the frequency axes makes it easier to identify the location of the affected frequencies. Figure 4.8 and 4.9 identify the seven first comb frequencies in the comparison intervals of the noise and music, given in figure 4.1, for the 3,9 ms comb filter.

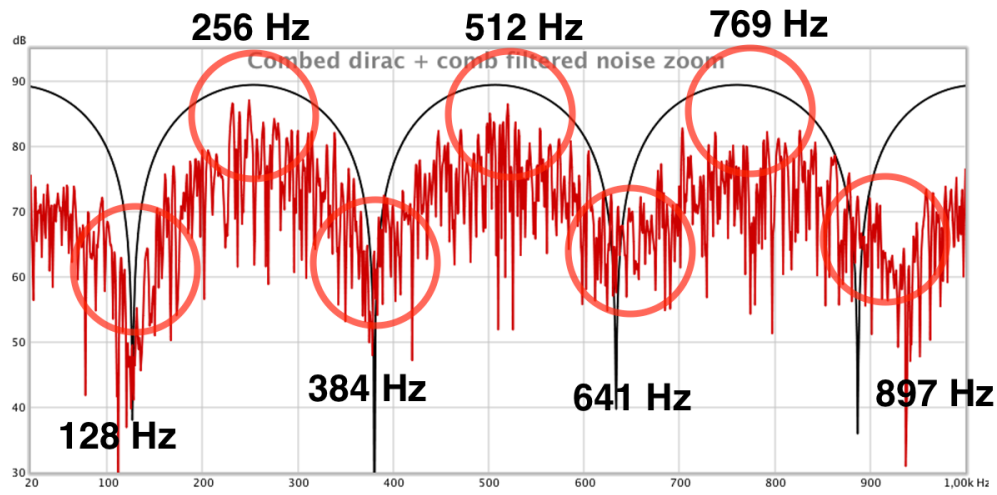


Figure 4.8 *The red circles identify the seven first comb frequencies in the band limited white noise comparison interval. Dips are located at 128 Hz, 384 Hz, 641 Hz, and 897 Hz, while peaks are located at 256 Hz, 512 Hz and 769 Hz, respectively. The frequency axis goes from 20 to 1000 Hz.*

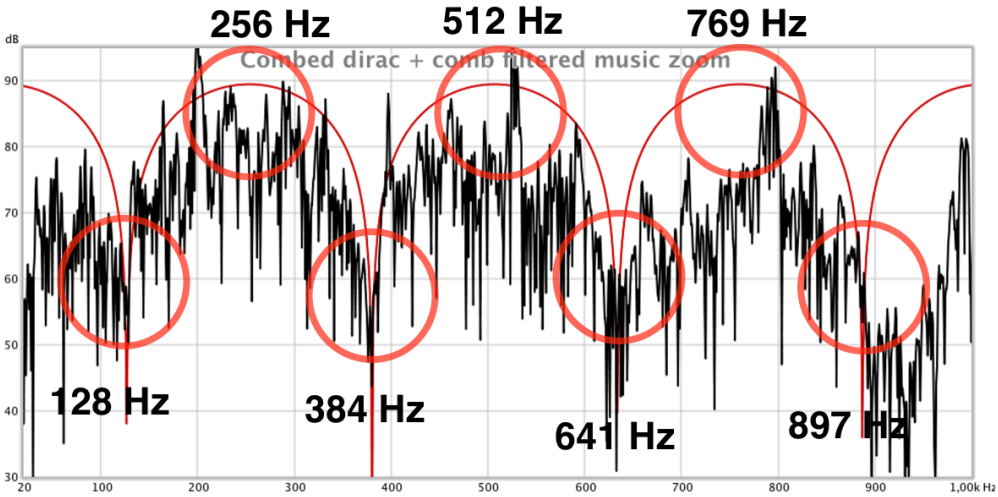


Figure 4.9 *The red circles identify the seven first comb frequencies in the music comparison interval. Dips are located at 128 Hz, 384 Hz, 641 Hz and 897 Hz, while peaks are located at 256 Hz, 512 Hz, and 769 Hz, respectively. The frequency axis goes from 20 to 1000 Hz.*

The next step would then be to look at the spectra of the comparison intervals when the stimulus variables (reflections under investigation) had undergone filtering.

Figure 4.10 and 4.11 compare the spectra of the comb filtered signals without filtering to those where the reflections had undergone filtering. The magnitude data and response of the filter applied to the reflections were shown in table 3.3 and figure 3.1, respectively.

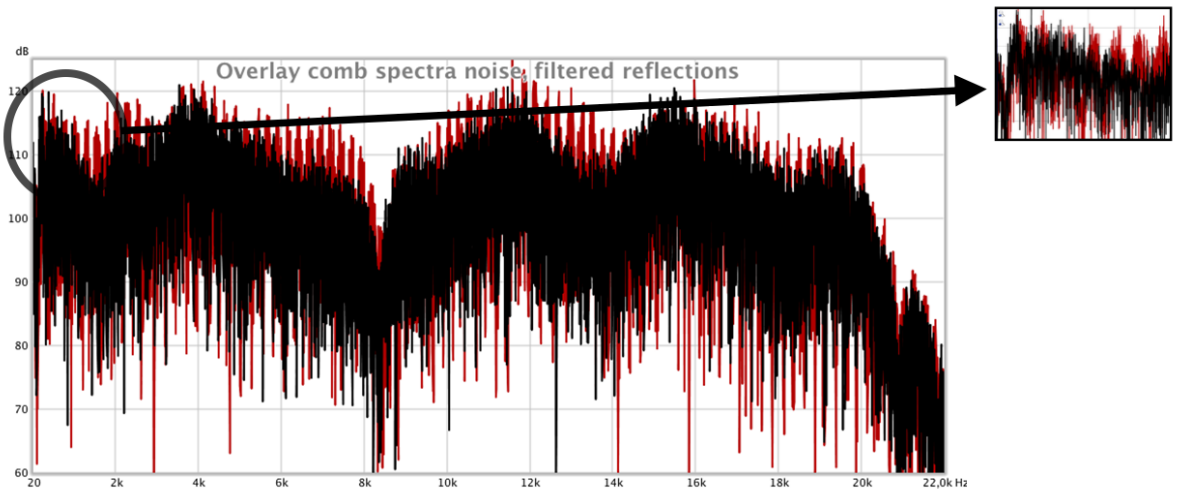


Figure 4.10 *A comparison of the power spectra of the two comparison intervals of noise; the comb spectrum with filtering applied to the reflection (black) looks very similar to the original spectrum in figure 4.4 (blue).*

In figure 4.10 and 4.11 the comparison intervals with filtered reflections, as presented to the respondents on the second day of the experiment, looks very similar to the spectra of the standard intervals in figure 4.4 (blue) and 4.5 (gold). This is a reasonable considering the magnitude response of the filter applied to the reflections (table 3.3 and figure 3.5). The lowpass filter applied to the reflections, which magnitude data and response were given in table 3.3 and in 3.5, as noted, implies that its cutoff-frequency is somewhere around 290 Hz. This leads to dramatic high-frequency attenuation even reaching down in the low middle

frequencies. As such, when combining the reflections that were pre-processed through this filter with the original direct signal, the result is comb filtering of low-frequencies, predominantly below the area of the cutoff-frequency of the filter.

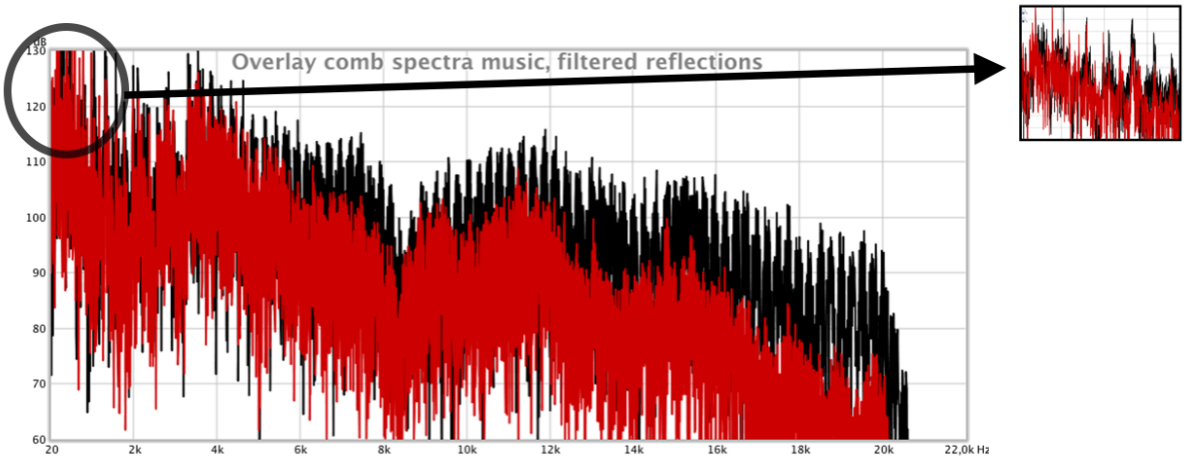


Figure 4.11 *A comparison of the power spectra of the two music comparison intervals; the comb spectrum with filtering applied to the reflection (red) looks very similar to the original signals' spectrum in figure 4.4 (blue).*

In figure 4.12, the frequency axis of the signals shown in figure 4.10, goes from 20 to 4000 Hz. Figure 4.12 shows that in the filtered signal (black), above 400 Hz, the frequency response characteristic of the comb filter is practically absent, as was expected.

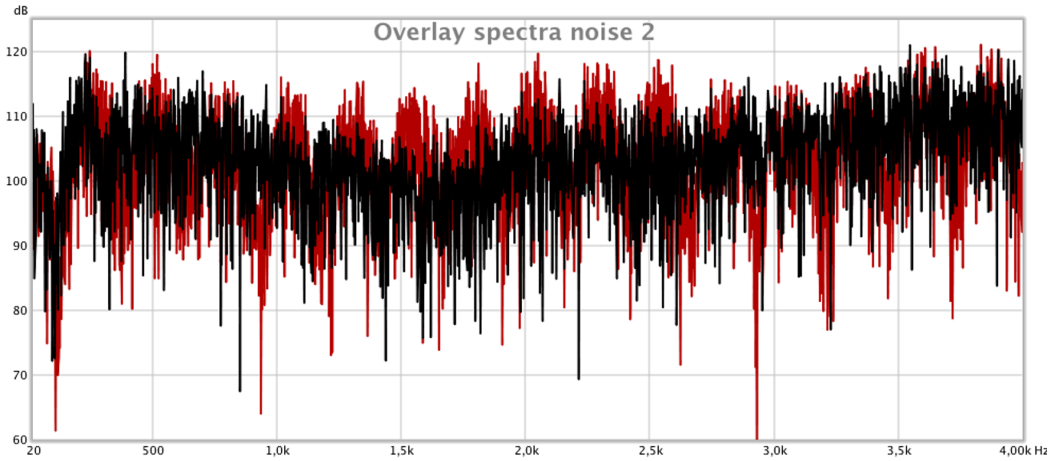


Figure 4.12 *A comparison between the two comb spectra of noise. Read: Comb filtered signal with unfiltered reflections. Black: Comb filtered signal with filtered reflections.*

In figure 4.13 (noise) and 4.14 (music) the comb spectra of the two test signals are further zoomed to get a clearer overview of the region below the cutoff-frequency of the lowpass filter.

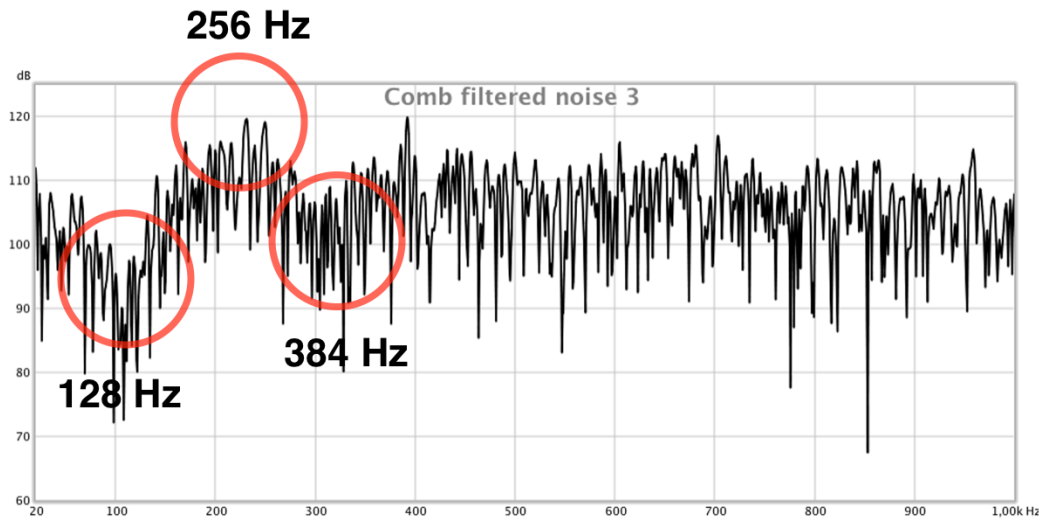


Figure 4.13 *Red circles mark regions of the comb frequencies visible after reflections were filtered in the band-limited white noise signal. The second dip (384 Hz) has become clearly shallower due to it locates above the cutoff-frequency of the lowpass filter.*

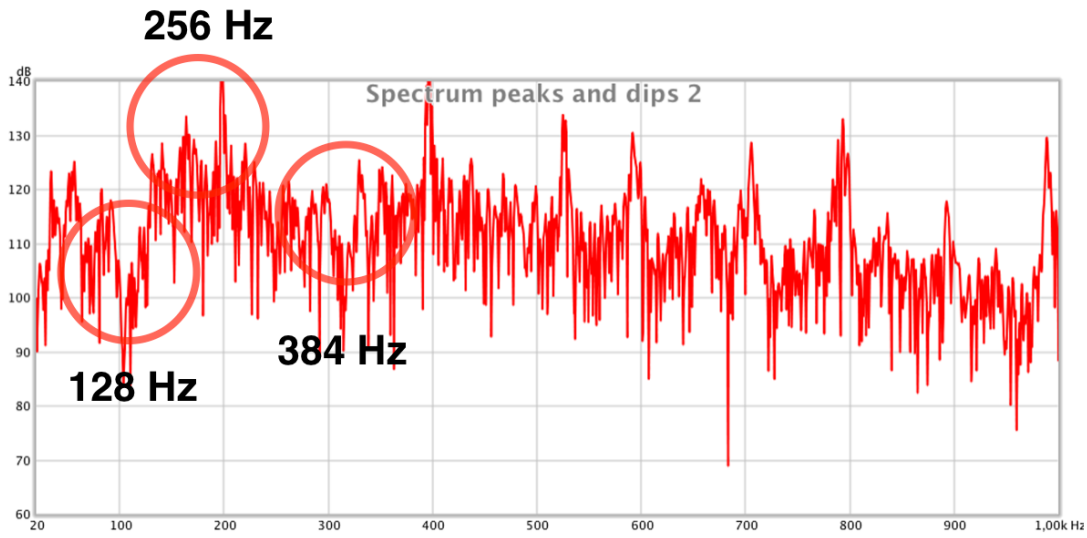


Figure 4.14 *Red circles mark regions of the comb frequencies visible after reflections were filtered in the music signal. The second dip (384 Hz) has become clearly shallower due to it locates above the cutoff-frequency of the lowpass filter.*

Through figure 4.13 and 4.14 we observe that the lowpass filter practically removed comb filtering above 400 Hz or so in the test signals. As such, there should be no surprise that we observe an elevation in established threshold values after lowpass filtering was applied to the reflections. However, while the lowpass filter produced a significant increase in established threshold values for the white noise signal, we only observe a marginal increase for the music signal (after lowpass filtering was applied to the reflections). To examine this we should first start by analyzing the tonal information in the music signal.

The score of the section investigated is included in the appendix 2. The orchestra in the music sample plays in the key of C major. The main melody line consists of notes E-F-G. If we disregard the initial few milliseconds of the sample (where the lowest note is an F2 apparently), the lowest note played appears to be a G2 at 98 Hz, played by the contrabasses (score). In the original (no comb filtering) version, (a) woodwind(s) or brass plays a C3 which sustains throughout. In the upper spectrogram in figure 4.15 and 4.16, 131 Hz is barely visible, which means the amplitude of this note is not particularly strong. However, from figure 4.1 we see that the first comb filter null frequency at 128 Hz practically coincided with the C3 at 131 Hz, removing it almost completely, as documented by the lower spectrogram in figure 4.15 and 4.16. Also, the second null frequency at 384 Hz nearly coincided with G4 at 392 Hz, reducing its amplitude, which is also visible in the lower spectrogram in figure 4.15 and 4.16. On the other hand, C4 at 262 Hz and C5 at 523 Hz, spaced one and two octaves above C3, respectively, are reinforced by the comb peaks at 256 Hz and 512 Hz. If looking closely at the spectrograms in figure 4.15 and 4.16, we may as well be able to see that C4 and C5 are reinforced in the comb filtered version (indicated by a stronger orange color). We also see from the power spectra in figure 4.5 that several of the overtones are reinforced by higher comb frequencies, making the comb filtered version brighter than the original.

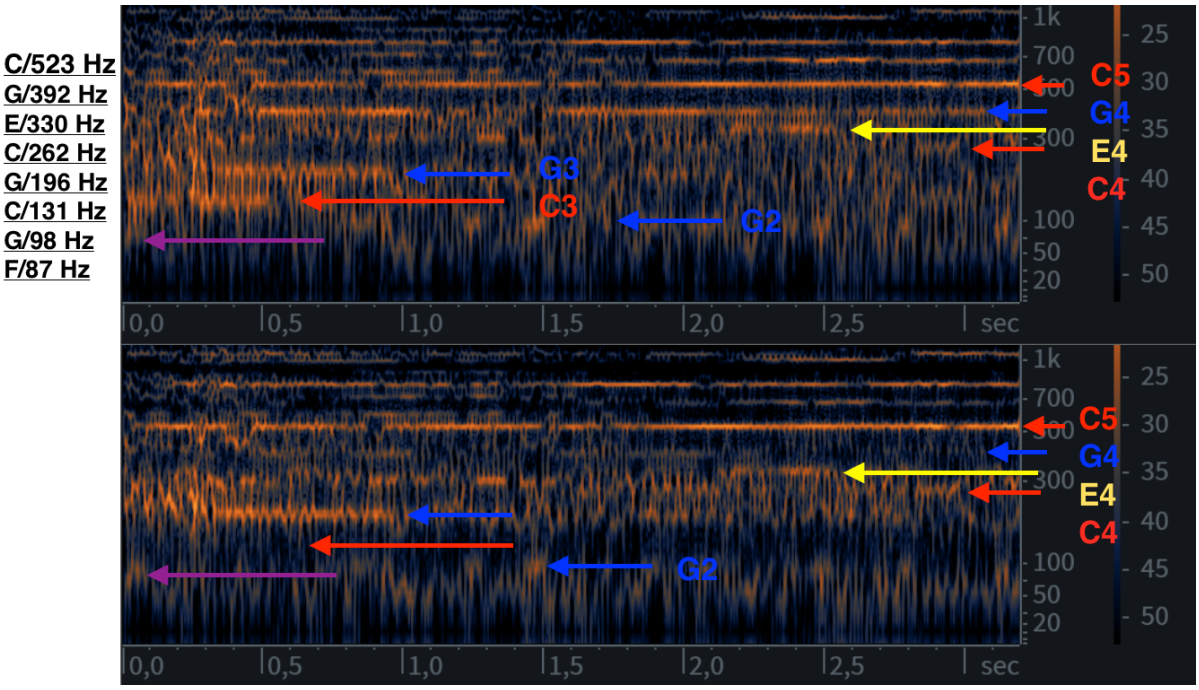


Figure 4.15 *Spectrogram of the original music signal (upper), and the comb filtered version (lower), in which the reflections were unfiltered. C3 at 131 Hz, indicated by a red arrow in both spectrograms, is clearly visible in the original version (upper), but appeared to be removed completely in the comb filtered version (lower). The reason for this is the first comb null frequency at 128 Hz. The view is set to logarithmic.*

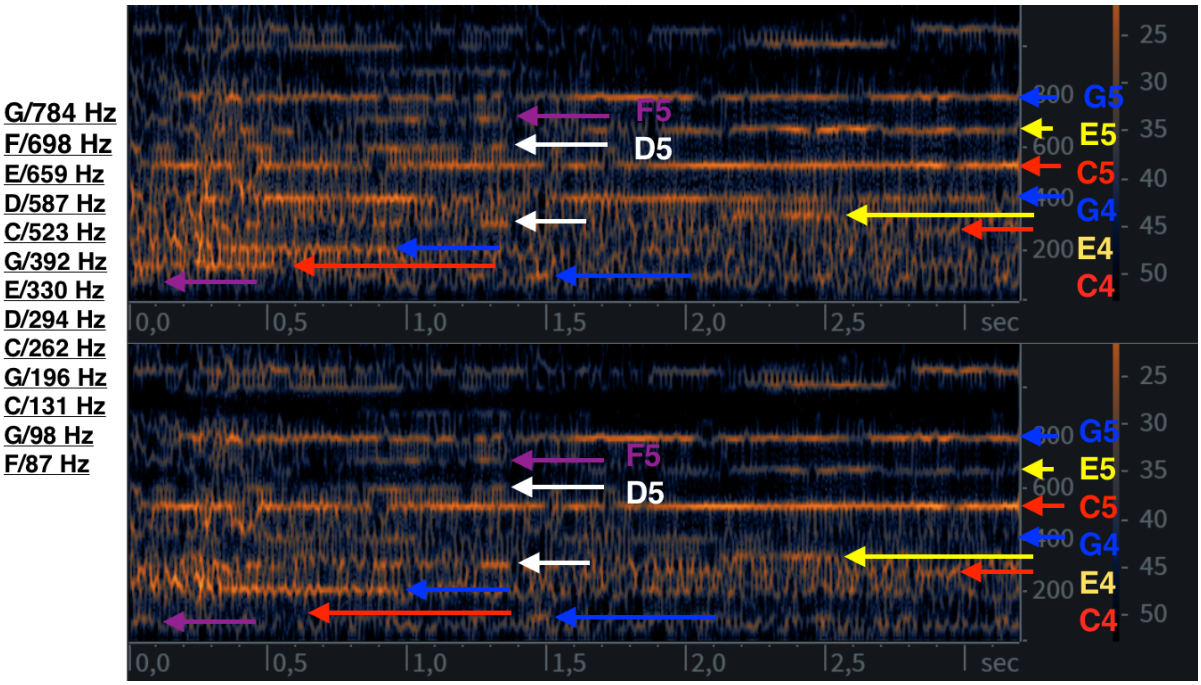


Figure 4.16 *Spectrogram of the original music signal (upper), and the comb filtered version (lower), in which the reflections were unfiltered. C3 at 131 Hz, indicated by the red line in both spectrograms, is clearly visible in the original version (upper), but appeared to be removed completely in the comb filtered version (lower). The reason for this is the first comb null frequency at 128 Hz. The view is set to linear.*

Therefore, a delay time of 3,9 ms gives a comb filter where several of the first comb frequencies coincided with several of the notes played by the orchestra, either removing or reinforcing them. As such, in sum, the timbre of the music signal was severely affected by the comb filter.

However, the fact that dramatic lowpass filtering only produced a marginal increase in threshold values for the music signal, must mean that the spectral information present below the lowpass cutoff-frequency was substantial for the detection.

Listening to the original version, it is not immediately recognizable what is the lowest note played. According to the score, a G2 at 98 Hz played by the double basses is seemingly the lowest note. As such, there should be little, if any, tonal information below 98 Hz. However, the timpanis create a layer of noise in the background occupying predominantly the low middle frequencies, which also reach down in the bass region. The noise layer might add to the perception of a “fullness” in the original version. On the other hand, listening to both

comb filtered versions of the music signal confirms that these versions are clearly “thinner” than the original. In the comb filtered version without lowpass filtering this thinning of sound may be further emphasized by the second comb dip at 384 Hz attenuating G4. However, the comb dip at 384 Hz was considerably reduced after lowpass filtering, as documented in figure 4.12. This means that the first comb dip at 128 Hz removing C3 is most certainly the predominant reason for this perception.

To test whether the removal of C3 is the predominant reason for this perception, and not the noise layer, we should either change the delay time of the test reflections or the key in which the music plays.

It was decided to transpose the music signal from C Major to D major to investigate if the new key gave a change in the mentioned perception. Because of the events mentioned, before and during the experimental period, I only had the opportunity to involve one of the respondents for this. Fortunately, the trained listener from the main experiment agreed on participating a second time. The original stereo-file of the music sample was transposed using the “change pitch” option in Audacity. The signal was then summed to mono and preprocessed the same way as the original test signals for the main experiment, previously accounted for.

The respondent received approximately the same amount of trials in preparation (training phase) for this experiment as were received in preparation for the main experiment. However, this time only two runs were conducted. The two runs were identical, and the reflections were filtered. The new threshold values produced by the trained listener in the key of D major were averaged, establishing an individual threshold value in the same way as in main experiment. The collected data from this version is found in the psydat-file for respondent 13 in the appendix.

In the key of D major, the trained listener produced a (mean) threshold value of -15,05 dB. This is an increase of 5,87 dB relative to the threshold value produced in the original key by the same respondent.

The respondent reported that he/she used the same cues for detection as the first time, only this time it proved “a bit more difficult”. Specifically, the “thinning” of sound was not as apparent as the first time, as the amplitude of the test reflections lowered.

Figure 4.17 shows spectrograms of the standard and comparison intervals in the key of D major.

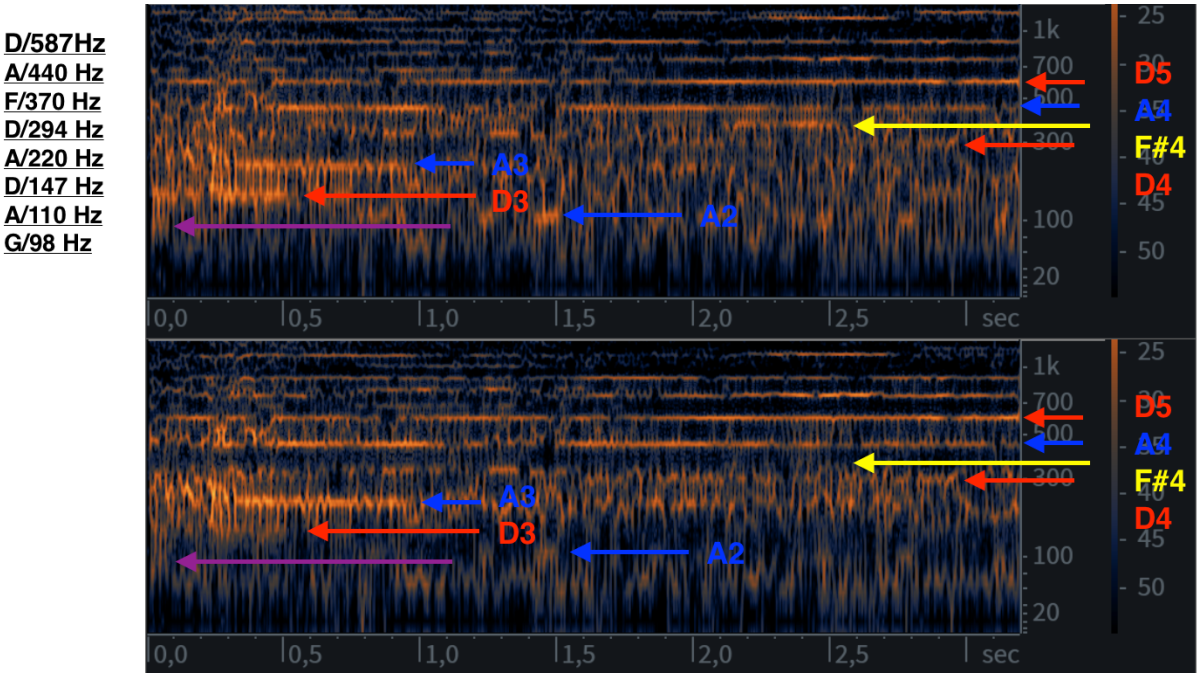


Figure 4.17 *Spectrogram of the transposed original music signal (upper) and transposed comb filtered version (lower), in which the reflections were unfiltered. The amplitude of D3 at 147 Hz, indicated by the red line in both spectrograms, is reduced due to the comb dip at 128 Hz but the 3,9 ms comb filters' overall effect on the timbre is not as strong in D major as in the original C major. The view is set to logarithmic.*

Because the test reflections in the transposed version had the same time offset as in the main experiment, we can interpret a raised threshold value towards that the coloration not being as "strong" in D major as in C major for this music sample. We also see from figure 4.17 that in the transposed comb filtered version, D3 at 147 Hz is only reduced, not removed by the comb dip at 128 Hz, as was the case with C3 in the original version.

On the other hand, for the white noise signal the lowpass filtering produced a significant increase in the established threshold values, relative to without lowpass filtering applied

(figure 4.1). This implies that the spectral information present above the lowpass cutoff-frequency was significant for the coloration detection (before lowpass filtering was applied).

Firstly, we notice that the white noise signal (figure 4.4 red and blue) has more and stronger high frequencies present than the music signal (figure 4.5 black and gold). I.e. we should not underestimate the importance of the relative balance between high- and low-frequencies for the detection.

However, the significant increase in the threshold values for white noise (after lowpass filtering of the test reflections) implies that the main detection cue for coloration was located above the cutoff-frequency of the filter.
To investigate this we should have a closer look at the apparent cue for detection in the white noise signal.

As noted, the main cue for comb filtering in the white noise signal was reportedly a distinct pitch sensation, and particularly so before the introduction of the lowpass filter.
According to Ritsma, Bilsen and Yost (1962;1969/70;1982), the dominant spectral region for pitch perception is located around the fourth harmonic. Adding a reflection to white noise producing so-called repetition pitch, according to Rubak (2004), the dominance region for the perception of pitch equates to $4/T$, where T is the delay time of the reflection, as mentioned in the theory chapter, section 2.1.2. Following this assumption, with a delay time of 3,9 ms, the dominance region for the perception of pitch should locate at $4/T = 1025$ Hz, which is at the fourth comb peak frequency. Thus, a logical explanation for the significant increase in threshold values (with filtered reflections) could be that the lowpass filter severely reduced, or completely removed the main cue for detection in the white noise signal.

A natural approach testing this assumption would be to conduct a new experiment where the test reflections have a larger time offset, producing a comb filter with a fourth peak frequency that locates below the cutoff-frequency of the lowpass filter at 290 Hz. Then we can observe whether the threshold values change or not. If filtering of the reflections with this time offset produces considerably lower threshold values, it may confirm the assumption that lowpass filtering removed the cue for coloration detection in the 3,9 ms comb filter. E.g., a delay time of 15 ms may be sufficient as it produces a comb filter with the fourth peak frequency located at 266 Hz, which is below the cutoff-frequency of the lowpass filter.

5 Discussion

5.1 Findings

The results (4.1) show a clear tendency that coloration became reduced after the filtering of test reflections. After having analyzed the various comb spectra it became clear why; the filtering of the test reflections practically removed comb filtering above 400 Hz or so in the combined signals.
Also, we have developed a theory why we observe after filtering a significant change in the coloration detection thresholds (raised mean thresholds) for white noise, but not for the music signal. For the music signal this change was only marginal (12,22 dB increase for noise, 4,61 dB increase for music). In this regard, we have shown how the coloration perception (especially so for music and other transient signals) is not only dependent on the delay time of the reflection and the relative strength of the combing signals (as shown in prior experiments), but also the relationship between the spectral content of the signals and the frequencies of the comb filter. This last observation is logical, but is not emphasized in prior research.

It is difficult to make direct comparisons between the results of this experiment to that of prior experiments of the same type. As such, we can only make assumptions. However, we believe the established mean threshold values (unfiltered reflections) are slightly higher than what could be expected. Especially so since we through the spectral analysis of the signals, as well as the reports from the trained listener, obtained the impression that the delay time used should result in strong coloration perception, which should equate to low detection threshold values.

There are many factors that can, in theory, have an effect on the results. The effect of training is one such factor. Considerations on the effect of training are summarized below.

5.2 Effect of training on performance

None except one individual had trained (extensively) before the experiments. The trained individual received an approximate amount of 320 trials of each stimulus category beforehand. Since not a larger portion of the respondents had trained beforehand, we cannot draw robust conclusions regarding the effect of training on performance. However, we appreciate that the trained listener (respondent 13 in table 4.1) produced noticeably lower threshold values than the rest of the selection (more so for the noise than for the music signal). Although this individual (respondent 13, table 4.1) did not improve his/her performance significantly during the training period, there was an improvement; he/she lowered his/her threshold values by a couple of dB throughout the training phase. On a general basis, if we can assume that each respondent with training would have been able to lower his/her threshold values by a couple of dB, the mean threshold values would have lowered (markedly) in the experiment.

As noted, the main interest with the listening experiment was to observe whether high-frequency absorption (modelled) reduced the coloration perception. I.e., to observe whether the threshold values changed significantly with filtering, which we observed for the noise signal. In this regard, we can only speculate if this change had decreased or increased if more listeners had trained beforehand. On the other hand, while the timbral differences between the two comb filtered versions of the music signal were substantial, the trained listener (respondent 13) produced practically the same threshold values (table 4.1) with and without filtering of the reflections (music signal). This implies that the trained listener used the same cue for the detection in both versions. Whether this pattern was a result of training or not is not clear. However, we may observe a similar pattern for some of those respondents that produced the lowest threshold values as well. This might imply that the observed (detection cue) is not an effect of training.

Bech (1994) reported that his listeners were exposed to an extensive training program before they participated in his sound field experiments. However, Bech (1994) did not report specifically to what extent the listeners performed better after training, other than the purpose to ensure that their thresholds had reached an asymptotic level. This suggests that there was

an effect. In this case, this is in agreement with the thresholds produced by the trained listener in the experiment relative to the thresholds produced by the other participants.

As noted earlier, Olive & Toole (1989) did a somewhat remarkable observation regarding previous experience and performance in listening tests, as did Brunner et al. (2007). Specifically, experienced critical listeners did not perform significantly better than unexperienced listeners. Olive & Toole (1989) suggested that the reason was that in the context of the experiment, the listening was such a focused activity that listening experience (or lack of experience) had less significance.
However, there might be a difference between an unexperienced listener, but having trained for a particular listening task and an experienced listener, in general, but not having trained for the same task. This means that the observations made by Olive & Toole (1989) and Brunner et al. (2007) do not say anything specifically about training for a particular task.

5.3 Considerations on validity, reliability, generalizability

Validity. There should not be any doubt about the validity of the results from the viewpoint of the procedure used (section 3.4.4). Based on literature reports, variants of the adaptive staircase alternative-forced-choice procedure seem standardized in psychoacoustic experiments of the kind as in this experiment. This was also substantial for the choice of procedure.

In terms of the modelling of absorption, variants of the approach used have also been utilized by other researchers, most notably Olive & Toole (1989) and Bech (1990;1994;1995). By looking at the response data from figure 3.2, lowpass filtering should give a reasonable approximation of the absorption effect in this particular case.

Reliability. The following factors were taken into account, attempting to increase reliability: 1) The selection of respondents. The initial purpose was to recruit a larger selection of participants (in a relative sense). While I did not obtain the amount of listeners I had foreseen

(around 20), the selection (13) proved large enough to see a tendency in the results.
2) The procedure. As noted, the adaptive staircase three-down/one up 2AFC task was chosen deliberately because it proved more reliable than the alternatives (mentioned in sections 3.3.4 and 3.3.6). As noted in section 3.4.4, two identical runs were conducted for each new stimulus presentation. Each of the individual threshold values was obtained by averaging the thresholds from two identical runs (identical in presentation). The standard deviations from these two runs were within the range of 2,8, as noted. The data from each of the respondents is available in the psydat-files, linked to in the appendix 2.

Generalizability. Of course, it is important to see the findings of the experiment in a larger context (i.e. real sound field listening). In this experiment, only a small part of the sound field was studied; the direct sound and two reflections, to observe the effect of (high-frequency) absorption on coloration.
As noted in section 3.2.4, the next logical step in the investigation would be to add room reflections to the virtual sound field, rendering more realistic room acoustic conditions. Of course, here we would also need to take frequency-dependent absorption characteristics of these other reflections into account, for a realistic simulation.

In terms of the absorption, factors that were not taken into account in the experiment were: 1) Angle-dependency (angle dependent absorption characteristics). 2 Air-absorption. However, of these points the latter, in particular, would be minuscule and not worth consideration (from the perspective of small room acoustics).

When looking at the effect of one particular factor (absorption) on coloration, it is reasonable to omit other factors that might reduce the clarity of the results (which was briefly discussed in section 3.2.4). E.g., we may assume that the inclusion of loudspeaker directivity in the simulation would have caused confusion. The reason is that off-axis loudspeaker response essentially gives much the same effect as (high-frequency) absorption, as noted in section 2.2.4.

6 Conclusion

The research question was - *To what degree do the spectra of reflections influence the detectability of comb filter coloration?*

The research question is complex and requires naturally further investigation than what was possible during the research period of this thesis, to fully answer. However, we can make statements based on the results from the listening experiment and spectral analysis of test material, as well as the theory. The research question was investigated from the perspective of the high-frequency absorption typical in listening rooms. Coloration thresholds for two categories of broadband sounds were established; band-limited white noise (20 Hz-16 kHz) and a sample of orchestrated classical music. The results (threshold values) from the listening experiment, given in table 4.1, suggest that the absorption characteristics of a 20 mm thick rock wool absorbent reduce the detectability of comb filter coloration in sound signals. This is also supported by the spectral analysis of the comb filtered signals with and without filtering of the reflections.

Few if any prior studies have been concerned with comparing the spectra of the signals to the interference frequencies of the comb filter. The small investigation in sequel to the main experiment confirmed the assumption that the coloration perception is as dependent on the spectral content of the signal itself as the delay time; the comb filter must affect frequency regions within the signals that our hearing notices as substantial for the timbre of the signals, for the timbral changes to be perceived clearly.

The absorption provided by a (even) thinner rock wool panel than what was used in the experiment of this thesis (see table table 3.1 for magnitude data), would have led to less dramatic filtering of the reflections. It is reasonable to assume less dramatic filtering (a lowpass filter with a higher cutoff-frequency) than what was used in this experiment (section 3.2.6) would lead to less overall change in the detection threshold values (for white noise, in particular). However, Bech (1995) found, as noted in section 2.2.3, that hearing thresholds for

coloration (timbre change) changed significantly only for those reflections where the filtering removed spectral energy in the vital mid- and high-frequencies (500 to 2 kHz).

For white noise it is reasonable to assume that the coloration perception is dependent on the presence of a spectral dominance region for pitch perception (mentioned in section 4.3) for a given delay time of the reflection. Specifically, when the dominance region for pitch perception is removed or partially removed by filtering, we propose here that the main cue for coloration is severely reduced. As noted in section 4.3, one approach testing this assumption would be to conduct a new experiment with a large enough reflection delay time so that the proposed dominance region relocates below the cutoff-frequency of the lowpass filter. If the coloration perception becomes stronger when changing the reflection delay time according to the approach above, it would largely confirm the theory of this dominance region.

Therefore, while it seems trivial, from the findings by both Bech (1995) and the findings in this thesis, we are able to conclude; if spectral energy "vital" for the coloration perception is removed from the reflection by filtering, the timbre of the combined signal (direct signal + test reflection) will be closer to that of the original direct signal (i.e. the coloration perception will be weak).

As noted in section 3.2.5, the initial plan was to take base in the absorption characteristics of at least two types of absorbers, or the same type of absorber with varying thicknesses.

Thus, the next step in the investigation of the research question (from the perspective of absorption) would be to perform the same type of experiment (section 3.4.4) several times, applying various filter responses to the test reflections each time. Also, a natural continuation would be to model more realistic room acoustic conditions, as sketched in section 3.2.4. However, arguments for not setting up a more realistic room model investigating the research question, were briefly discussed in sections 2.2.4 point 4), 3.1 and 3.2.4, in particular.

Bibliography

Barron, M. (1981), “Spatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure,” J. Sound and Vibration, 77(2), pp 211-232.

Bech, S. (1990), “Electroacoustic Simulation of Listening Room Acoustics; Psychoacoustic Design Criteria”. In Audio Engineering Society.

Bech, S. (1994), “Timbral aspects of Reproduced Sound in Small Rooms”. Audio Engineering Society.

Bech, S. (1995), “Perception of Reproduced Sound: Audibility of Individual reflections in a complete sound field, II”. Audio Engineering Society.

Bilsen, F. & Ritsma, R. (1970), “Some Parameters Influencing the Perceptibility of Pitch”. The Journal of the Acoustical Society of America.

Brunner, S., Maempel, H. & Weinzierl, S. (2007), “On the audibility of Comb Filter distortions”. Audio Engineering Society.

Buchholz, J., Mourjopoulos, J. & Blauert, J. (2001), “Room Masking: Understanding and Modelling the Masking of Room Reflections”. Audio Engineering Society.

Burgtorf, W. & Oehlschlägel, H. (1961), “Untersuchungen über die richtungsabhängige Wahrnehmbarkeit verzögerter schallsignale”. Physikalisches Institut der Universität Göttingen.

Dalenbäck, B. (2018), “Whitepaper: What is Geometrical Acoustics (GA)?”. CATT acoustics.

Dammerud, J. (2013), “Romakustikk”. Nordisk Institutt for Scene og Studio.

Everest, F. & Pohlmann, K. (2015), “Master Handbook of Acoustics”. New York: McGraw-Hill.

Green, D. (1993), “A maximum-likelihood method for estimating thresholds in a yes-no-task”. J. Acoust. Soc. Am. 93, 2096-2105.

Halmrast, T. (2020), “Cepstrum; a “forgotten” analysis?”. BNAM, Baltic-Nordic Acoustics Meeting.

Huopaniemi, J., Savioja, L. & Karjalainen, M., “Modeling of reflections and air absorption in acoustical spaces - a digital fitler design approach”. Helsinki University of Technology. Laboratory of Acoustics and Audio Signal Processing.

Kates, J. (1985), “A Central Spectrum Model for the perception of Coloration in filtered Gaussian Noise”. J. Acoust. Soc. Am. 77, 1529.

Kleiner, K. (2014), “ Acoustics of Small Rooms”. London: CRC press.

Leek, M. (2001), “Adaptive procedures in psychophysical research”. Percept. Psychophysics. 63, 1279-1292.

Levitt, H. (1971). “Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467-477.

Lorenzi, A. (2016), “Psychoacoustics”, http://www.cochlea.eu/en/sound/psychoacoustics [retrieved 02.03.20]

Toole, F. & Olive, S. (1988), “The Detection of Reflections in typical Rooms”. Audio Engineering Society.

Toole, F. & Olive, S. (1989), “The Modification of Timbre by Resonances: Perception and Measurement”. National Research Council, Divisions of Physics, Ottawa.

Rubak, P. (2004), “ Coloration in Natural and Artificial Room Impulse Responses”. Audio Engineering Society.

Salomons, A. (1995), “Coloration and Binaural Decoloration of sound due to reflections”. Thesis/dissertation.

Schubert, P. (1969), “ Die Wahrnehmbarkeit von Rückwürfen bei Musik”, Zeitschr. Hochfrequenz. u. Electroakust., vol. 78. pp 230-245.

Soranzo, A. & Grassi, M. (2014), “Psychoacoustics: a comprehensive Matlab toolbox for auditory testing”.

Watson, A. & Fitzhugh, A. (1990), “ The method of constant stimuli is inefficient". Percept. Psychophys. 47, 87-91.

Zahorik, P. (2009), “Perceptual relevant parameters for virtual listening simulation of small room acoustics”. J. Acoust. Soc. Am., Vol. 126, No. 2.

Zwicker, E. & Fastl, H. (2006), “Psychoacoustics: Facts and models”. Berlin: Springer-Verlag.

“DAFx Digital Audio Effects”, (2002), (ed. Zölzer, U.) *Delays*, Wiley & Sons.

Product data for the absorber material in which absorption characteristics were modelled:

https://www.rockfon.no/produkter/rockfon-blanka/?selectedCat=produktdatablad%20himlingsplater%20og%20veggabsorbenter

HRTF used for binaural processing:

https://www.soundhack.com/downloads/binaural/

Information on the HRTFS used for binaural processing:

http://alumni.media.mit.edu/~kdm/hrtfdoc/hrtfdoc.html

NB! See zip-file for thesis material