COLORATION IN ROOM IMPULSE RESPONSES

Per Rubak

Aalborg University
Department of Communication Technology
Fredrik Bajers Vej 7, 9220 Aalborg, Denmark
pr@kom.aau.dk

ABSTRACT

A literature review concerning perception of coloration is presented. In the paper we also present a review of general psycho acoustic background material concerning auditory modeling including perception of timbre and amplitude modulation. Furthermore, some classical room acoustic measures for echo perception are considered. Finally, some of the simple engineering measures to rate coloration is evaluated and an auditory signal processing model including a combined time-frequency detection criterion is presented.

1. INTRODUCTION

Strong reflections can have a detrimental effect on the quality of the reproduced sound. The direct sound added to one strong reflection results in a spectrum with peaks and valleys (cosine modulated). This rippled spectrum can give an unpleasant "***coloration of the sound"***. For delay times up to about 25 ms, the coloration is mainly due to the regular frequency-response variations (perception in the frequency domain). For longer delay times the "coloration" changes character, the sound becomes "rough" or "fluttering" and the perception is mainly related to the time domain. If the delay time of a reflection exceeds a certain value (about 50 ms for speech in an auditorium), the hearing system is able to resolve two distinct sound events, first the direct sound and subsequently a distinct echo of the original sound. A hand clap between to parallel walls can give a "flutter echo", where we hear a "fluttering" with a regular repetition rate. The strategy employed by a subject to judge the coloration can be based on psychoacoustic attributes like pitch, timbre, flutter and loudness.Generally the perceptual attributes can be divided in frequency domain perception (i.e. timbre related) and time domain perception (i.e. flutter). At the moment there is a great deal of confusion regarding verbal descriptors for **"time-domain coloration"** (whooshing, fluttering, echo disturbance, rattle, wavering etc.). Blesser (2001) describes temporal flutter in the following way: *Temporal flutter originates from defects during the spreading process (reverberation). If there is energy gaps or peaks, or if the spreading is periodic or too slow, the reverberation will be perceived as having temporal flutter*. In this paper we will use the term **temporal coloration** for coloration related to time domain perception/effects. The definition of timbre and pitch according to The American Standard of Acoustical Terminology (1960) is:

*"**Timbre** is that attribute of cochlear sensation in terms of which a listener judge that two sounds similar presented and having the same loudness and pitch are dissimilar". "**Pitch** is that attribute of cochlear sensation in terms of which sounds may be ordered on a scale extending from low to high".*

Salomons (1995) proposed the following definitions:

*The **"color"** of a sound signal is that attribute of cochlear sensation in terms of which a listener can judge that two sounds similar presented and having the same loudness are dissimilar; it thus comprises the timbre,* *rhythm sensation as well as the pitch of the signal. The **"coloration"** of a signal is the audible distortion, which alters the (natural) color of the sound*

The color of a signal is a quality of the signal, which may be changed by the surroundings (i.e. a room) in which the sound is produced or by a reproduction system. This change in color is called "coloration". Examples of everyday coloration are, e.g. the "hollow" quality of sound in a bathroom, or a change in the equalizer setting of an audio amplifier.

2. REVIEW OF LITERATURE

Atal, Schroeder & Kuttruff (1962) considered perception of white noise added to a delayed version of the same signal (a single reflection). They proposed two models to predict the measured thresholds based on short-time analysis (weighed autocorrelation function and power spectrum). The two models (one time domain and the other frequency domain) were equally suited to model the experimental results.
Bilsen (1968) derived a modified time weighing function and related the autocorrelation criterion to the "temporal-diffusion index" (originally defined by Kuttruff to characterize the randomness of a room impulse response).
Barron (1971) and Barron & Marshall (1981) investigated the subjective effects of a single lateral reflection in an attempt to understand the importance of early reflections in concert halls. The following effects were observed: loudness change, localization effects, tone coloration, echo disturbance and spatial impression. The effect of *spatial impression* was identified as the predominant subjective effect of a lateral reflection (for reflection delays between 10 and 80 ms). *Tone coloration* was also observed as perceptual effects of a single reflection. For lateral reflections, tone coloration was especially noticeable for delays between 10 and 20 ms, and for frontal reflections, tone coloration was substantial for delay times about 5 ms. They reported: "*the tone of the music appeared to sharpen, especially the violin tone*". They suggested that the effect was produced by comb filtering (interference between the direct sound and the reflection). They also noticed that coloration is mainly a monaural effect; the effect becomes less noticeable as the direct sound and the reflection sources are separated laterally.
Zurek (1979) investigated binaural echo suppression (decoloration). With echo delays less than 5-10 ms, thresholds for the diotic (same signal to both ears) echo signal were about 10 dB lower than for the dichotic (interaural difference) echo signal, thus showing a substantial binaural echo suppression. He proposed a model in which the monaural representations of spectral magnitude are non-linear compressed prior to being combined centrally.
Berkley (1980) introduced a more general approach based on application of a room simulation program providing a collection of well-controlled realistic room responses with different room acoustic parameters. Using multidimensional scaling of difference judgment, he concluded that perception of reverberation is mainly based on a two-dimensional perceptual space. The two components were 1: *Coloration* (related to *spectral deviation*) and 2: *Echo* (related to *reverberation*). *Spectral deviation* is a measure of the frequency response roughness (spectral modulations). Berkley suggested that **spectral deviation** of room transfer functions could predict ***spectral coloration***. More specifically, Berkley showed that the variance of the frequency response irregularities (on a log amplitude scale) was well correlated to the subjective perception of this "coloration-component" (*Spectral coloration*). Berkley did not propose any simple engineering metric to rate the degree of *temporal coloration* (for example echo disturbance).
Ando and Alrutz (1982) proposed a coloration criterion based on the autocorrelation function (ACF). The most preferred value of the initial time delay gap (between direct sound and the first reflection) has been found to be related to the effective length of the signal's ACF. The coloration threshold values depend strongly on the signal type, therefore it was tempting to investigate weather there is a relation between the coloration threshold and the ACF for the signal. They measured the coloration thresholds for band pass noise and compared the results with an ACF model.

$A^2 = k/|\varphi(\tau)|$ at $\tau = t_1$

where A is the threshold, $|\varphi(\tau)|$ is the envelope of the ACF, t_1 is the delay time of the reflection, and k is a constant. The constant k has to be frequency dependent to account for the measured thresholds and the model was only valid up to a critical delay time $\tau_{c.}$ The model is interesting because signal dependence is included through the ACF of the signal.
Olive and Toole (1989) investigated perception of reflections with emphasis on small listening rooms and sound reproduction. The main results are reported below.
Salomons (1995) measured thresholds for perception of coloration for single and multiple reflections and modified the two criteria proposed by Atal et al by including an auditory filter bank.
The thresholds for *correlated cosine noise* were compared with Atal et al (1962), Bilsen et al (1970) and Zurek (1979). Most of the thresholds agree very well (difference within ± 1.5 dB). Thresholds for *uncorrelated cosine noise* was compared with the results obtained by Zurek (1979); again good agreement was found (average difference: 1.4 ± 2.3 dB). *Comparison of thresholds for correlated with uncorrelated cosine noise* shows that the average difference is 3.9 ±1.9 dB (delay times from 2-20 ms).

Multiple reflections: When adding more than one repetition (reflection) to white noise, the thresholds depend on the number, the relative gain, and on the delay times of repetitions. The model based on the inner spectrum (Auditory filter bank) predicts the measured thresholds quite well for reflection delays up to about 30 ms.
Bech (1995 & 96)] presented a substantial amount of listening experiments to evaluate the contribution of individual early reflections to the timbre of reproduced sound in listening rooms. As timbre is close related to coloration, the results are valuable as reference material to evaluate signal-processing models for objective measures of timbre/coloration. Some of the results are presented below.
Meynial and Vuichard (1999) proposed a method based on the analysis of the distribution of the frequency response modulus. They employ a measured room impulse response where the initial part (Early reflections) is removed, as the Rayleigh distribution is only valid for the late *statistical* part of the reverberation process. Furthermore, a smoothing process is applied. The proposed *objective index E* is based on the quadratic error between backward integration of the theoretical Rayleigh distribution and the measured (modified) distribution. Results obtained from objective and subjective evaluation of coloration with measured impulse responses show that the objective indices can distinguish diffusive from non-diffusive ordinary rooms. The new objective measure is also a relevant measure of coloration in active rooms (Equipped with *active reverberation systems*).

Recently, **Brüggen** (2001) extended the work described by Berkley (1980) and proposed another and more advanced measure of spectral deviation based on Bilsen's "timbral-difference".
It was intended to employ stimuli that simulate natural listening conditions with the flexibility to separate monoral and binaural processes. This was achieved by using a room simulation system combined with different receiver configurations including an omni directional microphone, a dummy head and a stereo microphone configuration. A large number of rooms and source-receiver placements were simulated. Reflection-induced coloration and the relation to the interaural properties was the main goal. Custom-made semantic differentials were applied to direct the subject's attention to timbral variations. Application of musical excerpts as source signals appeared to cause substantial difficulties. *Therefore the main part of the experiments employed speech signals*.
Factor analysis: Two main components were found. These two components explain approx. 63% of the variance in the data and it is likely that these component are comparable to Berkley's two dimensions found with the paired comparison method. Specifically dimension one is related to spectral variation in the stimuli whereas the second component is largely related to temporal (and spatial) diffusivity. An analysis of variance indicates that the interaction of channel and source could not be neglected but the relative small influence of

the source itself indicates that the channel has a dominant influence (It should be noted that most of the results are obtained using speech signals).

Brüggen employed a measure equivalent to Plomp's "*timbral difference*", Plomp (1976). This prediction does not include interaural differences. Therefore, the prediction of "component one" was optimized with respect to diotic stimuli. It turned out that 50 (overlapping) bands only covering the frequency range 100-2000 Hz provided a reasonable match to the experimental results. It is likely that this rather narrow frequency range can be explained by the so-called "*dominance region.* Comparable observations were also reported by Bech (1996), and Olive and Toole (1989).
The predictions for dichotic stimuli (dummy head) reveal that binaural decoloration does not provide a fixed, general advantage but depends on the channel. A ***better-ear*** effect can usually be observed, and in some cases this effect give a bigger advantage.

3. ROOM- AND PSYCHOACOUSTIC BACKGROUND

3.1. Room acoustic background

After a certain time t_{mix} the number of reflections per second is high (say above 4000 per second) and we are not able to resolve the individual reflections, and a stochastic model can describe the sound field, see the figure below. But the early reflections arriving before t_{mix} are very important because they represent the personality of the room. The time distribution is very important for the room acoustic quality, a "wrong" distribution can cause coloration.

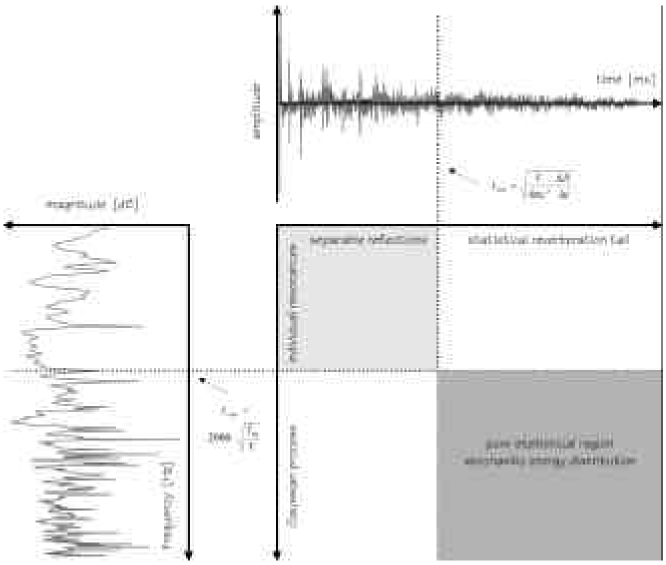

Figure 1: The combined time-frequency domain split in regions of particular interest.

The general stochastic model is invalid for frequencies below the **"Schroeder frequency"** given by, f_c = 2000 sqrt(RT/V). In this frequency range the eigentone density is low and the response is determined by a limited number of egentones with no or moderate overlap (up to a few overlapping egentones). In addition, the eigentones have typical low damping. Frequency response variations in the order of +15 dB to –30 dB can typical are observed. For small rooms (e.g. a typical living room) the Schroeder frequency is in the order of 200 Hz and music reproduction will suffer in the low frequency region. Substantially distortion (coloration) can be observed in this case. For a large performance space like a concert hall f_c is typical low enough (in the order of 20 Hz) to avoid this type of coloration.

When the eigentone density is sufficiently high and properly distributed (see below), the output spectrum will generally be a perceptually fair (after filtering through the auditory filter bank) “reproduction” of the input spectrum. For large performance spaces like concert halls the Schroeder frequency is typical below 20 Hz and coloration caused by insufficient eigentone density will normally not take place. However, we can still have coloration due to interaction with strong early reflections.

3.2. Perception of reflections

An important issue is: Under what condition is a reflection perceivable at all, regardless of the specific detection mechanism? Numerous researchers have investigated this threshold of absolute perceptibility. Another important question is: Under which conditions will the presence of a reflection be perceived as a disturbance?

Hass (1972) observed that a reflection (added to the direct sound) would change the *loudness*, *sound quality*, *liveness* and *body* of the reproduced speech. Also a “*pleasant broadening of the sound source*” was observed. The main goal was investigation of speech quality in rooms. Thresholds for *echo disturbance* in different acoustical environments were obtained. *At a delay time of 80 ms, only 20% of the test subjects were disturbed by a reflection with a relative level of –3 dB* (RT of the listening room was 0.8 sec).
According to Kuttruff (1991) the *absolute threshold for speech* (frontal incidence for direct sound and reflected sound) is approximately given by:
$\Delta L \cong -0.6\, t_0 - 8$ dB, (t_0 in ms).

Our hearing is less sensitive to echoes in music than in speech. For music the slope is much lower, approx. –0.13 dB/ms as maximum. The reason could be that speech is more impulsive than music and short pulses are very suitable to reveal *temporal coloration* (for example flutter echoes). *The thresholds are also dependent on the direction of incidence and are substantially lower for lateral reflections compared to frontal reflections*.
An overview of perceptual effects of combined sound fields is given in Kuttruff (1991), chapter VII.

Perception of a single reflection: The following is mainly based on Olive and Toole (1989).

Anechoic listening: The absolute threshold for one reflection, presented anechoically, will depend on: *delay time, direction of incidence, and signal type*. Olive and Tool determined the threshold for *pink noise and short pulses*. Reflections from median plane and horizontal plane have practically identical thresholds, whereas a reflection from the same direction as the primary source has thresholds up to 10 dB higher values (only determined for pink noise). Thresholds for pulses and pink noise are nearly identical (approx. –22 dB) for delay times around 5 ms, for shorter delay times pink noise is the most sensitive test signal, and for delay times above 10 ms the pulse signal is much more sensitive. Olive and Toole also present a *comparison of thresholds for a lot of different signals*. For a 20 ms delay time, the thresholds span from about -25 dB (Castanets) to –10 dB (Concerto Grosso, Händel).

Influence of room reverberation: Olive and Toole measured the absolute threshold and the image-shift threshold of a single (artificial) lateral reflection in a standard IEC listening room (and for the IEC room with additional damping). *The image-shift thresholds were consistently well above the absolute thresholds (In the IEC room, 7 dB above absolute threshold).* Up to a reflection time delay about 30 ms, the absolute thresholds (for speech) changed only about 6 dB when going from anechoic listening to listening in the IEC room. But for delay times above about 30 ms the thresholds for the reflections increase substantially when going from anechoic condition to normal room condition.
The increase in thresholds at longer delay times is likely to be related to the presence of a high echo density, caused by the reverberation in the room. Room reverberation will mask the temporal cues.
For longer delay times (above≈25 ms) the thresholds are mainly based on *temporal coloration.* The "continuous" signal (speech) builds up a more or less continuous background of reflections in the room, causing a substantial time-domain masking (delay times above 25 ms). Notice that the *mixing time* is approx. 25 ms in a typical living room.

Perception of reflections in a living room: *The following is based on Bech (1995 & 96).*
Two basic questions were investigated: Which early reflections are sufficiently strong to contribute individually to overall timbre and: How much must the level of an individual reflection change to produce a change in the overall timbre of the sound field? Bech employed electro acoustic simulation in a big anechoic room. The sound field in a typical listening room was simulated. The simulation included direct sound, 17 individual reflections and the reverberant field (Lexicon PCM70 room simulator). For experimental details and psycho acoustic procedure, see Bech (1995 & 1996). Pink noise and speech was employed as test signals.

Main results and conclusions for frequency independent reflections, Bech (1995): *In general*, the threshold for perception of an individual early reflection (in the presence of the complete sound field) was higher than the "natural level" in a typical listening room. Therefore these reflections will not individually contribute to the timbre, but collectively. Only three reflections i.e. floor, ceiling and left wall have a threshold above the natural level. The results show that only the first-order ceiling and floor reflections are likely to contribute individually to the timbre of a speech signal. For a noise signal additional reflections from the wall to the left of the listener will likely contribute individually to the timbre. *Timbre threshold* was generally well defined and unambiguous.
The influence of the reverberant field: The threshold values were measured for selected reflections without the reverberation. Removal of the reverberant field causes the threshold values to decrease by 2-5 dB. These results are in agreement with the results of Olive and Toole (1989). This suggests that the contribution of individual reflections to the timbre of the sound field will increase with less reverberation.

Main results and conclusions for frequency dependent reflections, Bech (1996): In these experiments the electro acoustic simulation is closer to the acoustic properties of a real listening room. *For noise signals* it was confirmed that reflection 1 (1st floor reflection) is likely to contribute individually to the timbre.
The filtering had no significant effect on the thresholds for the speech signal. A possible explanation is that the spectrum for speech has its main energy in the frequency range up to about 500 Hz. Another explanation could be that the subjects employed loudness difference instead of timbre difference.

3.3. Psycho acoustic background

A brief introduction to modeling of temporal resolution/summation is given. A common used model consists of initial band pass filtering (auditory filters) followed by a non-linear device representing transduction from BM vibration to neural activity. The output from the non-linear device is feed through *a "smoothing" device, often referred to as a temporal integrator* (it is assumed to represent a relatively central process). The output from the smoothing device is fed to a decision device. In one implementation of such models, the nonlinearity is a half-wave rectifier and the smoothing devise is a low pass filter Viemeister (1979). The low pass filtered signal is close related to the amplitude envelope of the band pass filter output signal. Alternatively the smoothing effect can be modeled *as a temporal integrator, which sums the energy within a certain time interval or "window"*. Considering temporal summation (or integration) we can pick up pretty different time constants (or window length) for the auditory system depending on the input signal, experimental conditions and the psycho acoustic task. Apparently, the auditory processor uses different time constants in different parts of the auditory system, depending on the specific tasks. The shape of the ears *temporal window* was investigated by Plack and Moore (1990). The threshold data could be fitted to a temporal window modeled by two roex functions. *They found an ERD (equivalent rectangular duration) about 8 ms in the mid-frequency range, increasing to about 12 ms at 300 Hz (the values are level dependent).* In subjective evaluation of room impulse responses Niese employed a "time constant" of about 15 ms. for delay times larger than about 20 ms, the reflections produce a rhythm sensation due to temporal recognition using pattern recognition.

Dau and co-workers (1996 & 97) has proposed a highly sophisticated non-linear model of the auditory periphery. They include a modulation filter bank. This model is also interesting in relation to perception of ***temporal coloration***, but we suppose that there is a lot of work to be done before a general model to predict coloration could be established.

4. ANALYSIS OF IMPULSE RESPONSES

A room impulse response h(t) is normally divided in three parts: direct sound, early reflections and the late diffuse reverberation. The acoustic quality of a room is determined by this sequence of reflections described by level, delay time and direction of incidence. Measuring a number of room impulse responses for different source and receiver positions normally captures this information. Analysis of the room impulse response (IR) is a very important tool providing information about delay times and levels for the early (separable) reflections and general information on the energy distribution in different time intervals. The directional information is lost in this one-dimensional representation where the room is modeled by a black box described by the impulse response (normally captured by an omni directional microphone). Directional information can be obtained by using a dummy head or a directional microphone. The early reflection part (up to about 200 ms in concert halls, and about 20 ms for a typical listening room) of the sound decay process is of particular importance, because this part represents the fingerprint of the room, which is responsible for a substantial part of the subjective impression.

Based on the impulse response h(t) sound energy ratios, like the *Clarity* index C_x, have been applied to quantify and predict the acoustical quality of concert halls. The Clarity index C_x is defined as: the ratio between the early energy in the room impulse response up to the time limit t=x and all the late energy succeeding the time limit t=x. In speech related analysis, the Clarity index C_{50} is useful and C_{80} is commonly used in music related analysis (in small rooms C_{15} is likely to be a useful parameter). A more detailed analysis of this part can be achieved by plotting C_x for each 5 ms up to 200 ms. This procedure called ELR is proposed by Marshall (1996), and he calculate the early-to-late energy ratios (based on the energy-time curve ETC) in octave bands. Marshall also compares the measured/calculated ELR curve with some theoretical ELR values for pure exponential decay.

High quality reverberation, including artificial reverberation, should also have short-time quasi-stationarity. The Schroeder reverberator built by cascading of all-pass filters was not "colorless" as expected because the short-time frequency response became colored in the late part of the impulse response. Kohlrausch (reported in Schroeder (1984**)** invented an objective measurement technique to reveal this deficit. The method was based on a running short-time frequency analysis of the impulse response multiplied by a factor increasing exponential with time (to remove the exponential decay). Kohlrausch applied a 30 ms window length corresponding to a 30 Hz frequency resolution. A window length in the same order of magnitude is common in speech analysis.

Autocorrelation analysis: The autocorrelation function is another important time domain tool to analyze room impulse responses. The degree of randomness can be measured and Kuttruff (1991) proposed an index, **temporal diffusion,** defined as:

$\Delta = \varphi_s(0) / \max \varphi_s(\tau \neq 0)$,

where $\varphi_s(\tau)$ is the autocorrelation function of the impulse response. One of the main criteria employed below is based on *temporal diffusion*.

Srodecki (1994) has investigated reverberation decay quality in rooms by using the autocorrelation function and cepstrum analysis. The measurement of the autocorrelation function is useful for the estimation of the number, time distribution and energy of strong reflections (or series of reflections) perceived subjectively as coloration or echo disturbance. It also allows determination of the degree of sound dispersion in a room

Several JTFA methods could be applied for analysis of room impulse responses. The **cumulative spectrum decay** (Waterfall plot) is an important tool, with several applications within room- and electro acoustics. A review of impulse response analysis techniques and a software platform for analysis of room impulse responses is presented in Johansen & Rubak (1999 & 2000).

5. COLORATION CRITERIA

5.1. Introduction

Atal Schroeder and Kuttruff (Atal et al. 1962) investigated the coloration caused by one strong reflection interacting with the direct sound, employing white noise as input signal. They proposed two objective criteria to predict the perception of coloration, a spectral criterion (A0) and an autocorrelation criterion (B0). Both criteria (described below) are based on short-time analysis. When one reflection is added to the direct sound, the spectrum is periodic rippled (modulated with a cosine). For white noise as input signal and addition of one reflection the resulting spectrum is often called *harmonic cosine noise* (system impulse response: $\delta(t) + a\ \delta(t - t_d)$. In most literature concerning cosine noise, the threshold of coloration is expressed as the "threshold gain" of the reflection, in dB.

A short-time autocorrelation function was defined:

$\varphi_s(\tau) = \varphi(\tau)\ \rho(\tau)$

where $\varphi(\tau)$ is the original long-term autocorrelation function, $\rho(\tau)$ the weighing function and $\varphi_s(\tau)$ is called the short-time autocorrelation function. The **short-time power spectrum** is given by the Fourier-transform of the short-time autocorrelation function:
$P_s = F[\varphi_s(\tau)]$

The definition of the A0 criterion is: coloration is perceptible if the level difference between max-value and min-value of the short-time power spectrum exceeds a threshold A0.

At threshold we have the condition: $A0 = 10 \log \{\max P_s(f) / \min P_s(f)\}$

The power spectrum, for one reflection (a=relative amplitude and T=delay time) is given by:
$P_s(f) = 1 + a^2 \rho(T) \cos(2\pi fT)$

Application of the A0 criterion requires a determination of the maximum modulation of the power spectrum, given by $a^2 \rho(T)$. Notice that the weighing function ρ is frequency independent, for the Atal et al. model

The B0 criterion states that: coloration is perceptible if the ratio of the max value of the short-time autocorrelation function for any nonzero delay to its value at zero delay exceeds a threshold B0.

At threshold we have the condition:$B0 = \max \varphi_s(\tau \neq 0) / \varphi_s(0)$

The optimal values for the A0 and B0 criterion was determined to be: **A0=1.1 dB** and **B0 = 0.063**

For delay times below approx. 10 ms, the A0 and B0 criteria cannot predict the coloration thresholds with satisfying accuracy. Bilsen (1968) proposed a modified autocorrelation weighing function ρ', to provide a better fit for short delay times.

5.2. Salomon's revised A0 & B0 criteria

Salomon (1995) extended the method to incorporate the auditory filter bank and reformulated the A0/B0 criteria. The auditory filters have a substantial influence on how the signal spectrum is perceived. The perception of coloration will obviously be based on filtering of the original signal with auditory filters. Several researchers have investigated the shape of the auditory filters; a review is given in Patterson and Moore (1986). Salomons (1995) applies the auditory filter shapes developed by Patterson (described in Patterson and Moore 1986).

The new definition of the A0-criterion is: Coloration is perceptible if the maximum modulation depth (the level difference between maxima and minima) of the inner spectrum (after filtering through the auditory filters) exceeds a certain threshold.

The modulation depth is calculated from the global maximum and minimum of the inner spectrum. This criterion will give a god prediction of spectral coloration caused by reflections with delay time differences smaller than about 20 ms.

The A0 criterion fitted to Salomons experimental threshold values is given by:
$A0=1.3 \pm 0.3$ dB, uncor. harmonic cosine noise
$A0 = 1.7 \pm 0.3$ dB, correlated harmonic cosine noise

The Salomon A0 criterion provides satisfactory prediction up to about 20 ms delay times, above this value the coloration perception change from mainly spectral perception to mainly temporal perception.

Autocorrelation criterion (B0): The new B0 criterion is based on the autocorrelation $\varphi_c(\tau)$ of the internal auditory spectrum. The autocorrelation function is calculated as the inverse Fourier-transform of the power spectrum. Now the peaks ($\tau \neq 0$) in the autocorrelation function are not discrete any more. The value $\varphi_c(T)$ is replaced with the area under the peak (at τ=T) in the autocorrelation curve, where the integration is extended to include the significant part of the dominant peak, here labeled τ=T.

$B0 = \int_{\tau=T} [\varphi_c(\tau)\, d\tau] / \int_{\tau=0} [\varphi_c(\tau)\, d\tau]$

B0 criterion as stated by Salomons: Coloration is perceptible if the area of the autocorrelation peak belonging to the delay time of the most dominant reflection exceeds a certain threshold B0, normalized on the area of the peak at zero delay time.

The B0 criterion fitted to Salomons experimental threshold values is given by:
B0=0.072 ±0.014, uncor. harmonic cosine noise
B0=0.098 ± 0.011, correlated harmonic cosine noise

The two different criteria predict almost the same threshold values. Rating of reflections with delay times larger than about 20 ms require a different model, based on temporal analysis.

Threshold prediction for multiple reflections: In Salomons (1995) the theory is further extended to multiple reflections. She also derived analytical expressions in this more general case. The dominant autocorrelation peak is still defined as the peak having the largest area. With multiple reflections there are several overlapping peaks. This complicates the determination of the area of the dominant peak. Salomons derived a method to estimate the area of a peak in the autocorrelation function:

$$\int \varphi_c(\tau)\, d\tau \approx c\, [\, \tau\, \varphi_c(\tau)]_{max}$$

The B0 criterion is now formulated:

$$B0 \approx a^2\, [\, \tau\, \varphi_c(\tau)]_{max}$$

This method is much more convenient for practical applications.Also for the case with multiple reflections, the predictions of the coloration thresholds are satisfying and generally there is no substantial difference between both criteria. The conclusion is that both the A0 and B0 criteria will give a fair prediction of the coloration.

5.3. Discussion

Considering the investigations until now, it is obvious that **Spectral coloration** is related to spectral irregularities (modulation in the frequency domain) and that the criteria should be based on the inner spectrum (after filtering through the auditory filter bank). Broadband signals like white noise (or pink noise) are well suited to reveal spectral coloration. The perceptual window length is about 25 ms. The optimal window shape is not quite established. For delay times longer than about 25 ms, the spectral modulation, after auditory filtering, is too small to account for the measured coloration threshold values.

This is in accordance with our expectations because the subjects now describe their sensation as a perception of rattle, roughness or fluttering e.t.c (temporal structures). The time structure in the signal is now perceived as the main cue. For longer delay times criteria based on temporal analysis have to be employed. Our knowledge concerning the temporal processing in the auditory system is not too coherent at the moment and the verbal descriptors for temporal coloration are to some degree confusing (see sec. 2.1). More psychacoustic research concerning perception of time structure and temporal masking is needed.

In room acoustics several procedures based on the energy-time curve, ETC, have been investigated. Several methods have been proposed for prediction of echo-disturbance. One of the well-reputed methods is developed by **Dietsch-Kraak** (reported by Kuttruff 1991, p 188). But also criteria based on autocorrelation analysis have been employed with success; for example temporal diffusion index Δ as defined by Kuttruff (1991). The definition of Δ is in principle equivalent to the B0 criterion defined by Atal et al; the only difference is the length (and shape) of the time window.

6. A NEW COLORATION DETECTION PROCEDURE

6.1 Spectral coloration

Spectral coloration (timbre related) is mainly based on the spectral modulation caused by the early part of the room impulse response (first 30 ms). We propose a half-hanning window with a 30 ms length as a simple engineering approach. White noise is a suitable signal for revealing frequency domain coloration, and the procedure is simplified because the autocorrelation function of white noise is a delta function. In this case the spectral modulation is given directly by the power spectrum of the room impulse response (here the first 30 ms). The windowed impulse response is filtered through an auditory filter bank.

The "inner spectrum" is evaluated by using Salomons's modified A0 criterion (or alternatively Berkley's or Brüggens spectral deviation measures).

We propose to generalize these measures by using a *frequency weighing of the internal difference spectrum.* According to Kates (1985) coloration perception has a *dominant spectral region* below 1/T (T is the repetition time delay) and the greatest sensitivity occurs in the frequency region around 1/2T. Ando and Alruts (1982) measured the coloration threshold for band pass filtered white noise centered around 250 Hz, 1 kHz and 4 kHz. Their results support (to a certain degree) the dominant spectral region proposed by Kates. Considering pitch perception the dominant spectral region is around 4/T. Therefore it seems worth while to introduce a frequency weighing to account for a possible dominant frequency region; at least we are provided with additional free parameters to fit the model to the experimental threshold data.

Overview of the procedure:

1: Measurement or calculation of impulse response (preferable BIR)

2: FFT of the first 30 ms of impulse response, windowed by a half Hanning window

3: Transformation to auditory inner spectrum (ERB or BARK)

4: Calculation of spectral coloration index

6.2 Temporal coloration

Presently we are investigating the application of the autocorrelation function calculated for octave filtered impulse responses, and employing a modified "temporal diffusion index" (Kuttruff) to rate the *temporal coloration* (echo-disturbance or fluttering). This criterion is compared with the Dietsch-Kraak echo criterion. Octave bands in the frequency range from 125 Hz to 4 kHz is proposed.

Calculation procedure:

1: Octave filtering of impulse responses

2: Calculation of temporal *coloration index* based on the autocorrelation function, calculated for each octave band in the frequency range 125 Hz to 4kHz.

3: Dietsch-Kraak procedure for each octave band

6.3 Time-frequency analysis

Especially for testing artificial room impulse responses, we propose to include a modified Kohlrausch (see above) time-frequency spectrogram. Kohlrausch employed a short-time Fourier transform (window length about 30 ms) with constant absolute bandwidth (about 30 Hz). A short pulse excites the reverberator and the output is given a time dependent gain to compensate for the exponential decay.

We propose to employ an auditory filter bank instead of the short-time Fourier transform. A coloration index is calculated for each 30 ms time interval up to t=RT (reverberation time). In this way the evaluation of the impulse response is based on auditory filtering and objective measures.

Calculation procedure:

1: Calculation of *spectral coloration index* for consecutive frames (30 ms) of the room impulse response (from 0 to RT sec.)

2: Stationarity test of the running frames

This test examines the stationarity of the perceived running short-time auditory spectra during the reverberation decay process. Coloration during the reverberation decay can be observed for room transfer functions with a low density and improper distribution of the eigentones (poles in the z domain). Dominant low-frequency room modes (egentones with low damping) can be revealed in this test.

7. DISCUSSION AND CONCLUSION

At the moment it seems practical from an engineering point of view to employ signal dependent coloration criteria, as we otherwise need more advanced detection models. As an example, thresholds for echo disturbance is considerably higher for symphonic music than for speech.

Binaurally recorded (dummy head) impulse responses, BIR, will be the best starting point. Evaluation of each ear signal and selection of the worst case will be one possibility. Another possibility is to apply a spectral average of the two ear signals (this is applied in the Dietsch-Kraak procedure). The final goal will be to develop a binaural model for coloration perception that can account for binaural decoloration.

A comprehensive literature review is presented; focus is put on coloration threshold measurements and prediction hereof by objective criteria. We have covered artificial signals (headphone reproduced) and natural sound fields in rooms including early reflections and subsequent reverberation. Only a few investigations concerning rating of supra-threshold coloration perception are available. Binaural modeling of coloration perception is only covered through a few references.

New (preliminary) test procedures for objective calculation/rating of coloration are presented. The new procedures extend some of the existing evaluation methods.

A *spectral coloration measure* is based on the spectral modulation caused by the first 30 ms of the room impulse response. The spectral modulation is evaluated by using an auditory filter bank (Inner Auditory Spectrum). The measure is based on Salomons A0 criterion (or alternatively *Berkley's or Brüggens spectral deviation measures).*

Temporal coloration is evaluated through octave filtering of the impulse responses and subsequent evaluation based on autocorrelation analysis (temporal diffusion index) and compared with a modified Dietsch-Kraak procedure.

Time-frequency analysis: A running short-time spectrum of the room impulse response, based on auditory filters, is proposed to evaluate the “stationarity” of the room reverberation process. The method is a modification of the Kohlrausch reverb-print.

8. REFERENCES

[1] B. Atal et al., ”Perception of coloration in filtered gaussian noise-short-time spectral analysis by the ear. Proceedings 4th ICA, Copenhagen 1962.

[2] M. Barron and A. H. Marshall, ASpatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure,@ *J. Sound and Vibration*, **77**(2), pp. 211-232, 1981

[3] S. Bech. “Timbral aspects of reproduced sound in small rooms. I”. J. Acoust. Soc. Am. Vol. 97, No. 3, pp. 1717-26, 1995

[4] S. Bech. “Timbral aspects of reproduced sound in small rooms. II”. J. Acoust. Soc. Am. Vol. 99, No. 6, pp. 3539-49, 1996.

[5] D. A. Berkley. “Normal listeners in typical Rooms - Reverberation Perception, Simulation, and Reduction”. In: “Acoustical factors affecting hearing aid performance”. Pp. 3-24. G.A.Studebaker, Hochberg (eis.). University Park Press, Baltimore, 1980.

[6] F. A. Bilsen, ”Thresholds of Perception of Repetition Pitch. Conclusions Concerning coloration in Room Acoustics and Correlation in the Hearing Organ”, ACUSTICA Vol. 19, pp. 27-32 (1967/68)

[7] F.A. Bilsen and R.J. Ritsma. “Some parameters influencing the perceptibility of pitch”, J. Acoust. Soc. Am. Vol. 47, No. 2, pp. 469-75, 1970.

[8] G. von Bismark. ”Timbre of steady sound: A factorial investigation of its verbal attributes”, Acustica 30, 146-59, 1974

[9] G. von Bismark. ” Sharpness as an attribute of the timbre of steady sounds”, Acustica **30**, 159-72, 1974

[10] B. Blesser. ”An Interdisciplinary Synthesis of Reverberation Viewpoints”, *J. Audio Eng. Soc.*, **49** (10), 2001 October

[11] M. Brüggen. “Coloration and Binaural Decoloration in Natural Environments”. Acta acustica – ACUSTICA Vol. 87, pp. 400-406, 2001.

[12] T. Dau et al. ”Modeling modulation perception: modulation low pass filter or modulation filter bank?”, In Psychoacoustics, Speech and Hearing Aids, ed. B. Kollmeier., 1996.

[13] H. Hass, AThe Influence of a Single Echo on the Audibility of Speech,@ *J. Audio Eng. Soc.*, **20** (2), 1972 March

[14] L. G. Johansen and P. Rubak, AA New Framework for Unified Analysis and Evaluation of loudspeakers and listening rooms,@ *a preprint of the Audio Eng. Soc. 108th Conv.*, Paris 2000

[15] L. G. Johansen and P. Rubak, AA New Software Platform for Room Impulse Response Analysis, *proc. of Internoise*, InterNoise99, Fort Lauderdale 1999

[16] J. M. Kates. “A Perceptual Criterion for Loudspeaker Evaluation”. J. Audio Eng. Soc., Vol. 32, No. 12, December 1984.

[17] J. M. Kates. ”A central spectrum model for the perception of coloration in filtered Gaussian noise”, J. Acoust. Soc. Am. Vol. 77, No. 4, pp. 1529-34, 1985

[18] H. Kuttruff, *Room Acoustics*, Applied Science Publishers Ltd. 1991

[19] L. G. Marshall. "An Analysis Procedure for Room Acoustics and Sound Amplification Systems Based on the Early-to-Late Energy Ratio", J. Audio Eng. Soc., Vol. 44, No. 5, May 1996.

[20] X. Meynial and O. Vuichard, AObjective Measure of Sound Coloration in Rooms,@*Acta Acustica*, **85**, pp. 101-107, 1999

[21] B. C. J. Moore *et al.*, AThe Shape of the Ear=s Temporal Window,@ *J. Acoust. Soc. Am.*, **83**(3), 1102-16, 1988 March

[22] J. A. Moorer, AAbout this Reverberation Business,@ *Computer Music J.*, **3**(2), 1979

[23] S. E. Olive and F. E. Toole, AThe Detection of Reflections in Typical Rooms,@ *J. Audio Eng. Soc.*, **7**(7/8), 1989 July/August

[24] R. D. Patterson and B. C. J. Moore. "Auditory filters and excitation patterns as representations of frequency resolution", In: Frequency selectivity in hearing, ed. B.C.J.Moore, Academic Press, London 1986

[25] R. Plomp. "Timbre as a multidimensional attribute of complex tones", in Frequency Analysis and Periodicity Detection in Hearing, ed. R. Plomp and G.F. Smoorenburg, 1970

[26] C.J. Plack and B.C.J. Moore. "Temporal window shape as a function of frequency and level", *J. Acoust. Soc. Am.*, **87**, 2178-87, 1990

[27] L.C.W. Pols. "Perceptual space of vowel-like sounds and its correlation with frequency spectrum", in Frequency Analysis and Periodicity Detection in Hearing, ed. R. Plomp and G.F. Smoorenburg, 1970

[28] A. M. Salomons. "Coloration and binaural decoloration of sound due to reflections",Thesis, Delft University, 1995

[29] M. R. Schroeder. "Progress in Architectural Acoustics and Artificial Reverberation: Concert Hall Acoustics and Number Theory", *J. Audio Eng. Soc.*, **32**(4), 1984 April

[30] K. Srodecki. "Evaluation of the Reverberation Decay Quality in Rooms Using the Autocorrelation Function and the Cepstrum Analysis", ACUSTICA Vol. 80, pp. 216-25 , 1994.

[31] F. E. Toole and S. E. Olive, AThe Modification of Timbre by Resonances: Perception and Measurement,@ *J. Audio Eng. Soc.*, **36**(3), 1988 March

[32] P. M. Zurek. "Measurements of binaural echo suppression", J. Acoust. Soc. Am. Vol. 66, No. 6, pp. 1750-57, 1979 December

[33] M. Vorländer, AObjective Characterization of Sound Fields in Small Rooms,@ *proc. of the Audio Eng. Soc. 15th Int. Conf.*, Copenhagen 1998