ISSN 0378-5955
ELSEVIER
Hearing Research
This Issue Completes Volume 277
Volume 277, Issues 1–2, July 2011

Hearing Research 277 (2011) 192–203

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

A quantitative analysis of spectral mechanisms involved in auditory detection of coloration by a single wall reflection

Jörg M. Buchholz [a,b,*]

[a] National Acoustic Laboratories, 126 Greville St., Chatswood NSW 2067, Australia
[b] Department of Linguistics – Audiology, Macquarie University, North Ryde NSW 2109, Australia

ARTICLE INFO

Article history:
Received 27 November 2009
Received in revised form
20 December 2010
Accepted 4 January 2011
Available online 12 January 2011

ABSTRACT

Coloration detection thresholds (CDTs) were measured for a single reflection as a function of spectral content and reflection delay for diotic stimulus presentation. The direct sound was a 320-ms long burst of bandpass-filtered noise with varying lower and upper cut-off frequencies. The resulting threshold data revealed that: (1) sensitivity decreases with decreasing bandwidth and increasing reflection delay and (2) high-frequency components contribute less to detection than low-frequency components. The auditory processes that may be involved in coloration detection (CD) are discussed in terms of a spectrum-based auditory model, which is conceptually similar to the pattern-transformation model of pitch (Wightman, 1973). Hence, the model derives an auto-correlation function of the input stimulus by applying a frequency analysis to an auditory representation of the power spectrum. It was found that, to successfully describe the quantitative behavior of the CDT data, three important mechanisms need to be included: (1) auditory bandpass filters with a narrower bandwidth than classic Gammatone filters, the increase in spectral resolution was here linked to cochlear suppression, (2) a spectral contrast enhancement process that reflects neural inhibition mechanisms, and (3) integration of information across auditory frequency bands.

1. Introduction

Any sound presented inside a room is accompanied by a large number of reflections stemming from surrounding surfaces. According to the precedence effect, the auditory system integrates the direct (or original) sound and the various reflections into a single auditory event located in the direction of the direct sound (e.g., Blauert, 1997; Litovsky et al., 1999). However, the reflections introduce spectral, temporal, and spatial modifications to the sound, changing perceptual qualities such as its loudness, spatial extent, or timbre. Considering early reflections (i.e., reflections that arrive at the receiver within 50–80 ms after the direct sound), a reflection mainly introduces a sensation of pitch or “coloration” (e.g., Bilsen and Ritsma, 1969/70).

Coloration is of particular importance for the acoustic design of performance spaces (e.g., concert halls) as well as for technical applications such as hearing aids or multi-channel audio systems. In all these applications, coloration is introduced by sound traveling along different paths to the listener, degrading overall sound quality. In order to evaluate and improve the quality of such applications, it is important to better understand the auditory mechanisms underlying coloration perception. Moreover, a better understanding of the auditory processing of early reflections in general is also important for improving modern speech and audio technologies, including automatic speech recognizers, adaptive beamformers, or cocktail party processors. These technologies often fail when operating in reverberant environments. In contrast, humans have usually no problem communicating in such adverse conditions, the presence of early reflections even enhancing auditory speech intelligibility (e.g., Haas, 1951; Bradley et al., 2003).

Several studies have systematically investigated the monaural detection of coloration produced by a single reflection (e.g., Buchholz, 2007). Atal et al. (1962), for example, have measured the coloration detection threshold (CDT), i.e. the level of a test reflection relative to the direct sound level at which coloration becomes just audible. For a broadband noise input they found that, for delays above about 3–5 ms, the CDT increases (sensitivity decreases) with

Abbreviations: ACF, auto-correlation function; BP, bandpass; CD, coloration detection; CDT, coloration detection threshold; ERB, equivalent rectangular bandwidth; FIR, Finite impulse response; LP, Lowpass; SCE, Spectral contrast enhancement; TQ, Threshold in quiet.

* National Acoustic Laboratories, 126 Greville St., Chatswood NSW 2067, Australia. Tel.: +61 2 9412 6909; fax: +61 2 9411 8273.
E-mail address: Jorg.Buchholz@nal.gov.au.

0378-5955/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2011.01.002

increasing reflection delay. For delays below 3–5 ms, the threshold increased again and above about 80 ms no coloration could be heard. A similar increase in CDT with increasing reflection delay was shown by Zurek (1979) and Salomons (1995). Salomons additionally considered the case of applying a 180° phase shift to the reflection, which resulted in a threshold increase of about 3 dB, independent of the reflection delay. Ando and Alrutz (1982) measured the CDT as a function of the reflection delay using Gaussian noise as input that was bandpass filtered at different centre frequencies. They observed that the increase in threshold with increasing reflection delay was steeper the higher the considered frequency range was.

Different criteria (or models) have been proposed to quantitatively describe auditory detection of coloration (e.g., Atal et al., 1962; Ando and Alrutz, 1982; Kates, 1985; Salomons, 1995). These criteria have either been based on the auto-correlation function (ACF) $\varphi(\tau)$ of the room impulse response $h(t)$ or the corresponding power transfer function $|H(\omega)|^2$ (i.e., the Fourier transform of the ACF). For a single reflection, approximated by a delayed and attenuated copy of the direct sound, these functions can be given by:

$$h(t) = \delta(t) + g_T\delta(t - d_T) \quad (1)$$

$$\varphi(\tau) = \left(1 + g_T^2\right)\delta(\tau) + g_T\delta(\tau + d_T) + g_T\delta(\tau - d_T) \quad (2)$$

$$|H(\omega)|^2 = 1 + g_T^2 + 2g_T\cos(\omega d_T) \quad (3)$$

With ω the angular frequency $\omega = 2\pi f$, t the time, τ the auto-correlation lag, g_T the reflection gain, d_T the reflection delay, and δ the delta impulse function. Hence, a single reflection introduces: (1) a delta impulse to the ACF at the negative and positive reflection delay and (2) a spectral ripple to the power spectrum with a ripple density that is proportional to the reflection delay. Within the different models, the CDT is typically assumed to be related to either the value of the ACF at the reflection delay, i.e. $\varphi(\pm d_T)$, or to the (maximal) depth of the spectral ripple inherent in the power spectrum. To account for the observed increase in the CDT with increasing reflection delay, Atal et al. (1962) applied an exponential weighting function to the ACF. This weighting function was assumed to reflect the short-time analysis performed by the auditory system. Ando and Alrutz (1982) considered the envelope of the ACF at the reflection delay to account for their coloration detection data for bandpass filtered noise. Kates (1985) and Salomons (1995) considered further aspects of auditory processing in their coloration criteria, such as the limited resolution of auditory frequency analysis in the inner ear or the synchrony of the firing patterns in the auditory nerve. However, all of these coloration detection (CD) models are based on engineering approaches that do not claim any physiological relevance. It should be highlighted that the term coloration is here used to describe the percept of spectral patterns that are produced by wall reflections and that are regular (or periodic) on a linear frequency scale (see Eq. (2)). This is in contrast to coloration perception in general, which would also include regular patterns on a logarithmic (or equivalent rectangular bandwidth; ERB$_N$) frequency scale (e.g., Eddins and Bero, 2007) or arbitrarily shaped spectra (Green, 1988).

A phenomenon that is closely related to coloration perception is the perception of ripple noise pitch (e.g., Bilsen and Ritsma, 1970; Yost, 1982) or iterated ripple noise pitch (e.g., Yost, 1996). Ripple noise pitch refers to the pitch heard when a single early reflection is added to a noise stimulus and thus, producing the same stimulus as used in coloration detection. The difference is only due to the listener's task in the corresponding experiments, which in ripple noise pitch perception is a discrimination task and not a detection task. Although these two tasks are obviously different, and thus results cannot be directly compared, it still seems to be useful to consider the auditory mechanisms that may underlie ripple noise pitch perception. Hence, classic pitch perception models should be taken into account when modeling CD. These models can be grouped into spectral models and temporal models (or a combination of the two). Spectral models perform a matching of a regular pattern, which is learned and stored at a central stage of the auditory system, to an auditory spectral representation of the current stimulus (e.g., Goldstein, 1973; Wightman, 1973; Terhardt, 1974; Cohen et al., 1995). When applying sinusoidal patterns as internally-stored patterns, the pattern-matching process is equivalent to the derivation of the auto-correlation function via a spectral analysis of the power spectrum, i.e., applying the Wiener–Khintchine relation (e.g., Hartmann, 1998). Since these models rely on the assumption that the auditory system is able to resolve the features in the spectral patterns, they are limited by the spectral resolution of the auditory system. Temporal models analyze periodicities in the time domain and are not inherently limited by auditory spectral resolution. Most temporal models assume an auto-correlation analysis after cochlear filtering and hair-cell transduction (e.g., Licklider, 1951; Meddis and Hewitt, 1991a,b). A ripple noise stimulus produces a peak in the auto-correlation function at a lag equal to the reflection delay. Given that these models perform a temporal analysis, they rely on the preservation of the waveform's temporal fine structure, which in the auditory nerve is limited to frequencies below about 5 kHz (e.g., Pickles, 2008). Since models that follow either theory can successfully describe a large number of pitch phenomena, there is no consensus in the literature as to which theory is most correct or complete (for an extensive review of pitch perception and models, see Plack et al., 2005).

In summary, spectral and temporal pitch models and CD models can both be based on the concept of auto-correlation function. However, the different model approaches significantly differ in their motivation, strategies, realization, limitations, and physiological relevance. Given the very extensive literature on pitch models, the development of CD models may significantly benefit from taking pitch models into account.

In this study, the monaural auditory mechanisms underlying CD were investigated by combining psychoacoustical experiments with a quantitative auditory detection model. Reviewing the literature on CD, it became clear that the influence of spectral content on the CDT was not fully understood. Hence, in the first part of the present study, a psychoacoustical experiment was conducted to investigate the CDT for a single reflection as a function of spectral content and reflection delay. The spectral content was modified by applying a bandpass filter to a noise input stimulus and then systematically varying the upper and lower cut-off frequencies of this bandpass filter. In the second part of the study, the auditory processes addressed by the experimental data were investigated by developing a spectrum-based auditory CD model. Within the proposed model, it was assumed that the auditory system performs a frequency analysis of an auditory representation of the power spectrum to derive the auto-correlation function of the input stimulus (applying the Wiener–Khintchine relation). Hence, similar across-frequency mechanisms were considered as in the pattern-transformation model of pitch (Wightman, 1973) or the coloration detection model proposed by Salomons (1995). Although a model based on spectral processing was used here, the idea was not to question the general applicability of models based on temporal processing. The goal was rather: (1) to evaluate how far a purely spectral approach can account for the experimental CDT data and (2) to understand what processing steps are required for such a model to be successful. The analysis highlights quantitative limitations of a pure spectral approach and provides an indication of when temporal mechanisms need to be taken into account. In contrast to existing CD models, which consider the (ideal) room impulse response, the present model was based on the actual stimulus waveforms. Moreover, the model is a quantitative detection model that is applied as an artificial observer, i.e., the model simulates a listener in the CD experiments. In this way, most of the stimulus variability involved in the actual experiments with human subjects was taken into account. Such a detection model is conceptually very different from existing pitch models, which are mainly pitch identification models with limited applicability to quantitatively describing experimental data on pitch strength or coloration detection.

2. Methods

CDTs were measured as a function of spectral content and reflection delay. The resulting data (Section 3) served as the basis for the evaluation of the CD model proposed in Section 4.

2.1. Subjects

One female (SB) and two male (JB, PK) subjects aged between 30 and 35 took part in the experiments. All three subjects had normal hearing, according to a pure tone audiogram, and had at least 10 h of training. Subjects SB and PK were paid for their participation on an hourly basis and JB was the author.

2.2. Stimuli

The stimuli were composed of a direct sound and a test reflection (the signal), diotically presented via headphones (Sennheiser HD580). The headphone was equalized to a flat spectrum measured in a Brüel & Kjaer artificial ear 4153. The direct sound was a 320-ms long Gaussian noise and the test reflection was a delayed and attenuated copy of the direct sound (as described in Eq. (1)). The offset of the test reflection (i.e., the part of the reflection that stands out after the direct sound) was truncated to avoid offset-listening effects (see Buchholz, 2007). The spectral content of the stimulus was modified by bandpass filtering (8th order butterworth) the entire stimulus and varying the lower (f_1) and upper (f_2) -6 dB cut-off frequencies of this bandpass filter. Two different conditions were considered:

1. The lower cut-off frequency was fixed at $f_1 = 0.1$ kHz and the upper cut-off frequency was varied: $f_2 = 0.5, 1, 1.5, 2, 3, 5$ kHz. The reflection delay was fixed at $d_T = 2, 4,$ or 8 ms.
2. The upper cut-off frequency was fixed at $f_2 = 5$ kHz and the lower cut-off frequency was varied: $f_1 = 0.1, 0.5, 1, 2, 3$ kHz. The reflection delay was fixed at $d_T = 2, 4,$ or 8 ms.

The direct sound level was set to 25 dB spectrum level. Hence, the overall sound pressure level varied with stimulus bandwidth. All stimuli were digitally generated at a sampling frequency of 44.1 kHz. The experiments were run on a PC with a high quality sound card using MATLAB. The listeners were seated in a double-walled sound-attenuating booth.

2.3. Procedures

CDTs were measured using a three-interval, three-alternative forced-choice procedure. Each interval contained a different sample of the direct sound and one randomly chosen interval additionally contained the test reflection. The intervals were separated by 500 ms of silence. The listener's task was to pick the interval containing the test reflection. The weighted up-down procedure (Kaernbach, 1991) was used to track the 79% point on the psychometric function. To limit level cues influencing the measurements, level roving of ±2 dB was introduced. The participants were instructed to attend solely to coloration cues. The level of the reflection relative to the direct sound (i.e., the reflection gain $g_{T,dB} = 20 \cdot \log_{10}(g_T)$) at the beginning of each run was set to 0 dB, which produced the highest possible coloration strength. The step-size was gradually reduced to a final step-size of 1 dB. Using this final step size, 10 reversals were measured and the mean value and standard deviation of the reflection level over these 10 reversals were calculated. At least three runs per subject were obtained for each condition. This research has been approved by the Copenhagen city council ethics committee (approval No: KA04159g).

3. Results

3.1. Variation of lower cut-off frequency f_1

The results are shown in Fig. 1 as a function of the lower cut-off frequency f_1, with the reflection delay d_T as parameter. The data show rather large standard deviations, reflecting the difficulty of the coloration detection task.

Although the general behavior of the CDT data is similar across subjects, strong inter-subject differences in overall sensitivity can be observed. The CDT generally increased (i.e., sensitivity decreased) with increasing lower cut-off frequency f_1, except for cut-off frequencies $f_1 \leq 1000$ Hz and $d_T = 2$ ms and $d_T = 4$ ms, where the CDT was roughly independent of cut-off frequency. Moreover, the CDT increased with increasing reflection delay, the increase being pronounced at higher cut-off frequencies f_1. For frequencies $f_1 \geq 3000$ Hz subjects JB and SB could not hear any coloration. Subject PK was able to detect the coloration cue for $f_1 = 3000$ Hz and $d_T = 2$ ms and $d_T = 4$ ms, but showed a significant drop in sensitivity for these two conditions. The CDT increase with increasing reflection delay d_T is in agreement with previous research on detection of coloration in broadband noise, at least for delays $d_T \geq 3$ ms (e.g., Atal et al., 1962).

3.2. Upper cut-off frequency f_2

The individual and mean data when the upper cut-off frequency f_2 was varied are shown in Fig. 2. The CDT decreased with increasing upper cut-off frequency up to a certain frequency f_{max}, above which the thresholds stayed approximately constant. This frequency f_{max} was almost independent of reflection delay d_T. The thresholds increased with increasing reflection delay d_T, at least for $d_T \geq 4$ ms, and this increase was more pronounced for higher cut-off frequencies f_2. Thus the delay dependency of the CDT increases with increasing f_1 and increasing f_2. This suggests that, with increasing delay the contribution of high-frequency components to overall CD decreases. This agrees in principle with the results of Ando and Alrutz (1982), who observed that, for bandpass filtered noise, the increase in CDT with increasing reflection delay is faster for higher bandpass center frequencies.

4. CD model description

The proposed CD model is based on the pattern-transformation model proposed by Wightman (1973), which represents one possible implementation of the general concept inherent in most spectrum-based pitch models (e.g., Goldstein, 1973; Terhardt, 1974; or Cohen et al., 1995). According to this classic approach, an auditory-filtered power spectrum is calculated from the coloration (or pitch) stimulus waveform, which is then processed by a spectral pattern analysis stage to derive an auto-correlation function. The main novelty of the proposed CD model lies in extending this basic

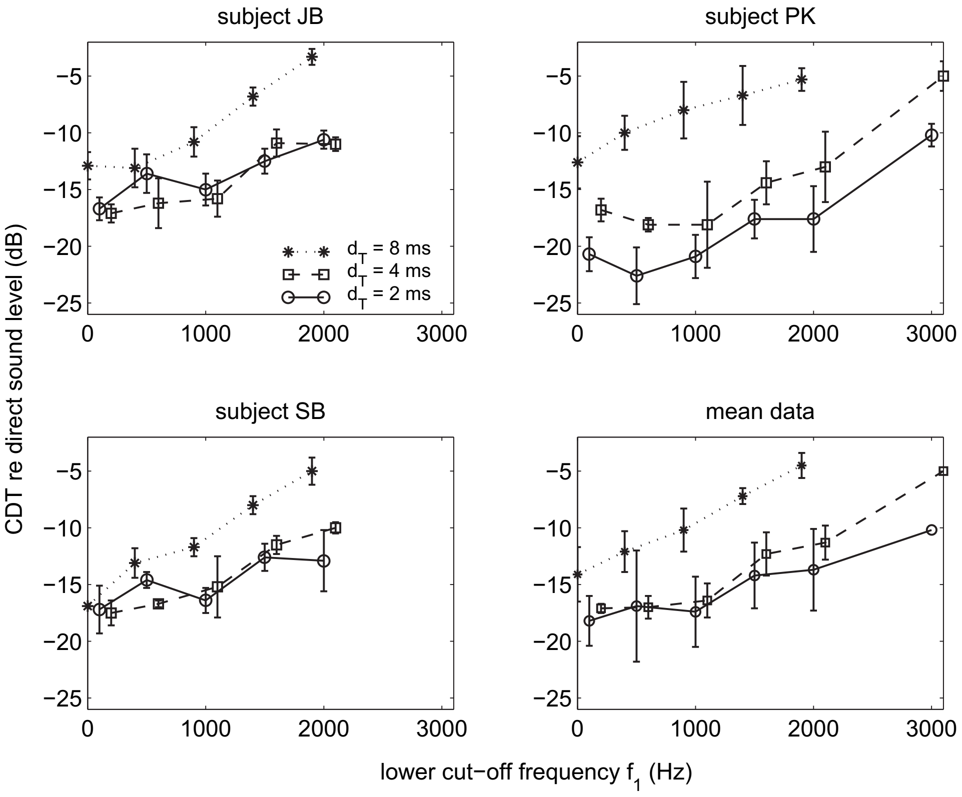

Fig. 1. CDTs as a function of lower cut-off frequency f_1, with the reflection delay d_T as parameter. The upper cut-off frequency was fixed at $f_2 = 5$ kHz. The upper two panels and the bottom-left panel show individual data, and the bottom-right panel shows the mean data. The error bars indicate ±1 standard deviation either across runs (upper two and bottom-left panel) or across subjects (bottom-right panel). The data points for the different reflection delays were slightly shifted horizontally to improve readability.

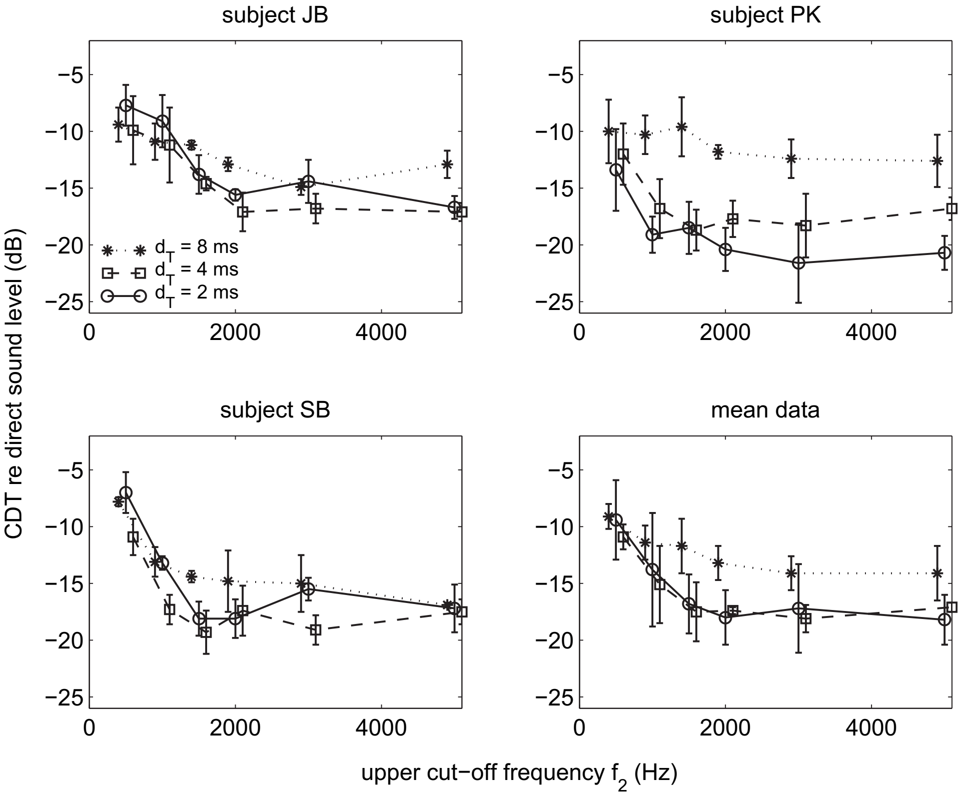

Fig. 2. As Fig. 2, but showing CDTs as a function of upper cut-off frequency f_2. The lower cut-off frequency was fixed at $f_1 = 100$ Hz.

concept to provide a quantitative model that mimics a listener participating in a CDT experiment, including the involved stimulus variability (i.e., realizing an artificial observer as described, for instance, by Dau et al., 1996). Such approach allows direct quantitative comparisons between experimental CDT data (described in Section 3) and corresponding CDT model predictions. In order to successfully describe the quantitative behavior of the CDT data, a spectral contrast enhancement function had to be applied to the auditory-filtered power spectrum, which was inspired by the peripheral-weighting model proposed by Yost (1982). Finally, overall sensitivity of the CD model had to be limited by the addition of auditory-internal (or central) noise. The general structure of the proposed CD model is illustrated in Fig. 3 and the important processing details are described in the following sections.

The CD model consists of two main stages, a peripheral stage and a central processing stage. The peripheral stage refers to the "hardwired" signal processing of the auditory periphery and the central processing stage refers to decision processes occurring at higher levels of the auditory system.

4.1. Peripheral processing stage

According to Fig. 3, the stimulus is passed through a linear bandpass filter H_{ME}, which approximates the transfer function of the middle ear and is realized by a first-order bandpass filter with -3 dB cut-off frequencies of 1 kHz and 4 kHz (e.g., Breebaart et al., 2001). The signal is then analyzed by a Gammatone bandpass filterbank (e.g., Patterson et al., 1988), which simulates the frequency analysis performed on the basilar membrane. In each frequency channel i, a squaring operation is applied followed by a non-leaky temporal integrator W, which calculates the sum over the entire stimulus interval. The output forms the (long-term) power spectrum $u_x(i)$ of the input signal. This power spectrum is then processed by a spectral contrast enhancement (SCE) mechanism that emphasizes changes in the spectrum. The principle processing of the peripheral stage is similar to the peripheral-weighting described by Yost (1982), although both the implementation details and the quantitative behavior are very different.

4.1.1. Auditory bandpass filterbank

The auditory filterbank was realized by Gammatone filters (e.g., Patterson et al., 1988) with a time-discrete impulse response $h(n)$ given by:

$$h(n) = A(nT)^{\nu-1}e^{-2\pi nTb_{\mathrm{CD}}}\cos(2\pi f_0 nT), \tag{4}$$

with n the time index, T the sampling interval, f_0 the centre frequency of the filter, A a scaling factor to ensure a filter response of 0 dB at $f = f_0$, ν the order of the filter which was chosen to be $\nu = 4$, and b_{CD} a parameter determining the filter bandwidth. All parameters were realized as described by Patterson et al. (1988) except the bandwidth parameter b_{CD}, which was given by:

$$b_{\mathrm{CD}} = b_{\mathrm{GT}}\left(1 - 0.375\left(1 + \left(\frac{f_0}{4000}\right)^8\right)^{-1}\right), \tag{5}$$

with the centre frequency f_0 in Hertz and b_{GT} the classical bandwidth parameter given by Patterson et al. (1988) or Glasberg and Moore (1990) and which for nu = 4 resembles the ERB. As shown in Fig. 4, the bandwidth of the CD model filters (solid line) is similar to the classic Gammatone filters (dashed line) for center frequencies above about 5 kHz (i.e., $b_{\mathrm{CD}} \approx b_{\mathrm{GT}}$). For lower center frequencies the CD model filters are about 1.6 times narrower than the classical filters ($b_{\mathrm{CD}} \approx b_{\mathrm{GT}}/1.6$), approximating the auditory filters described by Oxenham and Shera (2003, dotted line). In contrast to the classical Gammatone filters, which were derived from notched-noise simultaneous masking (SM) data, Oxenham and Shera derived their narrower filters from notched-noise forward masking (FM) data. Pickles, (2008) and Shera et al. (2002) argue that FM-based measures are more appropriate for specifying auditory frequency resolution than SM-based measures (see Section 5.1).

With reference to the CD model approach the increased frequency selectivity was necessary to account for the high sensitivity observed in the CD data for reflection delays $d_{\mathrm{T}} \geq 4$ ms (see Section 4.4). This is illustrated in Fig. 5, where the normalized power spectrum of a single reflection with a delay of $d_{\mathrm{T}} = 8$ ms and gain $g_{\mathrm{T,dB}} = -3$ dB is shown after BP filtering with (1) classic Gammatone filters (dashed line) and (2) the filters used in the CD model (solid line). For the classic Gammatone filters, significant spectral ripples can be observed only for frequencies up to about 1.2 kHz. In the CD model, the frequency range is extended to more than 2 kHz.

The filterbank was realized by finite impulse response (FIR) filters with a length of 1024 samples at a sampling frequency of $f_s = 16$ kHz. The center frequencies f_0 of the filters were linearly spaced on an $\mathrm{ERB_N}$-number scale z (Glasberg and Moore, 1990). The entire filterbank employed 18 filters per $\mathrm{ERB_N}$, which for the considered frequency range of $f_0 = 50$–7000 Hz resulted in a total of 533 filters.

4.1.2. Spectral contrast enhancement

The output $u_y(i)$ of the spectral contrast enhancement (SCE) stage was calculated by convolving the auditory representation of the power spectrum $u_x(i)$ with the function $h_{\mathrm{SCE}}(i)$, which was given

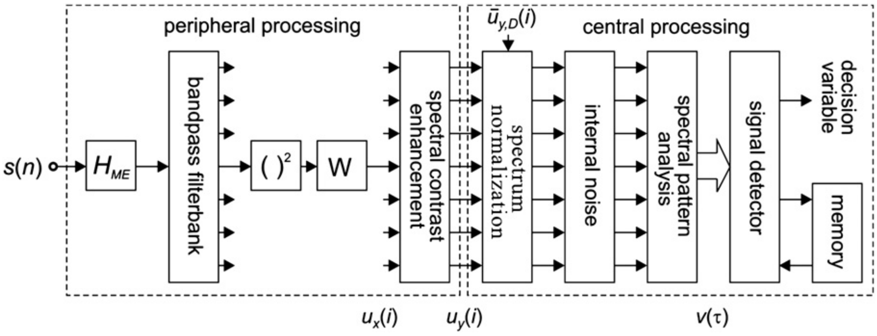

Fig. 3. Block diagram of the proposed CD model. The input signal $s(n)$ is passed through a linear middle-ear filter H_{ME}, a bandpass filterbank, a squaring operation, and a temporal integrator W. The resulting power spectrum $u_x(i)$ is then processed by a spectral contrast enhancement mechanism, resulting in an auditory power spectrum $u_y(i)$. This spectrum is then analyzed by a central processing unit consisting of a spectrum normalization, addition of "internal" noise, spectral pattern analysis, and a signal detector stage with memory. The output is a decision variable which defines the final CDT prediction.

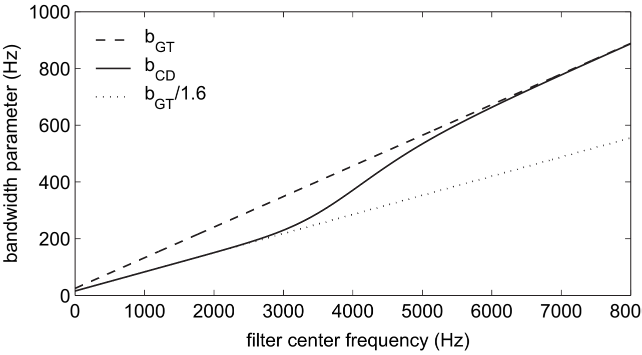

Fig. 4. Gammatone bandwidth parameter as a function of centre frequency for different auditory filter realizations. Filters used in the present study (b_{CD}, solid line), classic Gammatone filters (b_{GT}, dashed line), and narrower Gammatone filters with $b = b_{GT}/1.6$ (dotted line).

by convolving a kernel function $h_0(i)$ four times with itself. The kernel function $h_0(i)$ was defined as:

$$h_0(i) = -0.132\delta(i-3) + 0.737\delta(i) - 0.132\delta(i-3), \tag{6}$$

with i the frequency channel index and δ the delta impulse function. The functions $h_{SCE}(i)$ and $h_0(i)$ were derived by optimizing the fit between the experimental CD data and the corresponding model predictions (see Section 4.4). The SCE stage suppresses spectral ripples with low density and, in combination with the ERB$_N$ spacing of the auditory filters, results in an emphasis of spectral ripples that increases with increasing centre frequency as well as increasing ripple density (i.e., reflection delay). This is illustrated in Fig. 6, where the spectral ripple depth is shown for a reflection with a gain of 0 dB and a delay of 2, 4, and 8 ms. The left and right figure panels illustrate the ripple depth before and after the SCE operation is applied. It should be emphasized that the SCE stage does not increase spectral resolution as done by the narrower auditory filters (Section 4.1.1), it only introduces a spectral weighting of ripples, which towards high frequencies is limited by the resolution of the preceding auditory filters (see Section 5.2 for a further discussion).

4.2. Central processing stage

According to Fig. 3, the central processing stage consists of a spectrum normalization stage followed by additive noise and a spectral pattern analysis stage. The additive noise refers to auditory-internal (or neural) noise that accumulates along the auditory pathway and limits the auditory sensitivity to coloration. The spectral pattern analysis stage integrates coloration information across frequency and produces an "auditory-weighted" autocorrelation function. The model is completed by a signal detection stage with memory, which refers to decision making processes in the brain.

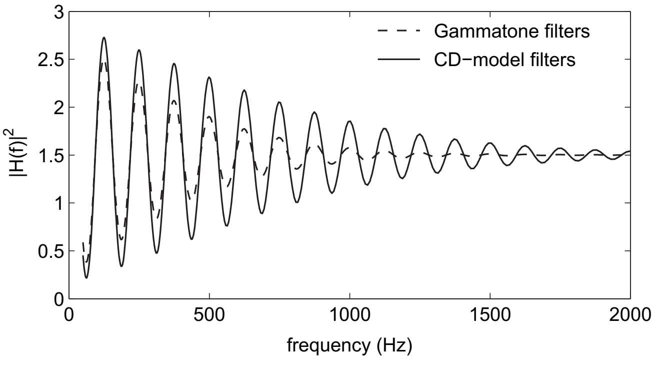

Fig. 5. Illustration of the normalized power transfer function of a single reflection at the output of a Gammatone BP filterbank using (1) the classic Gammatone filters (dashed line) and (2) the CD model filters (solid line). The reflection delay was $d_T = 8$ ms and the gain $g_{T,dB} = -3$ dB. The power transfer functions were normalized by the power spectrum of the direct sound alone.

4.2.1. Spectrum normalization

The normalized spectrum $r(i)$ is calculated from the output spectrum $u_y(i)$ of the peripheral processing stage by:

$$r(i) = \frac{u_y(i) - \overline{u}_{y,D}(i)}{\overline{u}_{y,D}(i)} \tag{7}$$

Thereby $u_y(i)$ is the spectrum of the specific stimulus under consideration (which varies across trials and may contain the direct sound alone or the direct sound plus the test reflection) and $\overline{u}_{y,D}(i)$ is the average spectrum of the direct sound, calculated over a number N of direct-sound-alone realizations (here $N = 1000$ was used). This normalized spectrum $r(i)$ is independent of stimulus level but maintains the variations between different stimulus instances. Conceptually similar spectrum normalization stages can be found in most auditory detection models (e.g., Dau et al., 1996; Buchholz and Mourjopoulos, 2003; Plack and Oxenham, 1998) and refers to the observation that the auditory system evaluates relative changes rather than absolute changes (i.e., it follows Weber's law).

4.2.2. Internal noise and spectral pattern analysis

Auditory-internal noise is added to the normalized power spectrum to limit the sensitivity of the subsequent signal detection process. In order to simplify the implementation of this internal noise as well as the subsequent spectral analysis stage, the normalized spectrum $r(i)$ was transformed from an ERB$_N$ scale, z, to a linear frequency scale, f, with $f = (\exp(0.11 \cdot z) - 1)/.00437$ (Glasberg and Moore, 1990). This transformation was realized by linear interpolation and resulted in a normalized spectrum $r(i')$ with an arbitrary frequency resolution of $f\Delta = 10$ Hz and $i' = f/f\Delta$. An alternative implementation of the internal noise and the subsequent pattern analysis stage, which maintains the ERB$_N$ frequency spacing of the auditory filters and produces identical CDT predictions, is described in Appendix A.

The internal noise applied in the CD model had zero mean and a frequency-independent variance σ_0^2, with σ_0 a constant that determined the overall sensitivity of the CD model. Throughout this study $\sigma_0 = 1.8$ was applied, which resulted in an optimal fit between model predictions and experimental data (see Section 4.4). After the addition of internal noise, the normalized (noisy) power spectrum, $r_n(i')$, was processed by a spectral pattern analysis stage, which was realized by a cosine transform:

$$v(\tau) = \sum_{i'=i_1'}^{i_2'} r_n(i')\cos(2\pi\tau f_\Delta i') \tag{8}$$

with τ the spectral ripple density, $i_2' > i_1' \geq 0$, and n a subscript indicating a variable that is affected by (or related to) internal noise. The frequency range considered, $i_1' \leq i' \leq i_2'$, limited the cosine transform to the range where significant spectral ripples can be observed. On the one hand, this frequency range is limited by the spectral smoothing produced by the auditory filters (and partly compensated by the subsequent SCE process), which results in a maximum frequency of $f_{max} \approx 20/\tau$ above which no significant ripple can be further observed (i.e., i_2' is limited to $i_2' \leq f_{max}/f\Delta$). On the other hand, the considered frequency range is limited by the threshold in quiet (TQ). It is assumed here that only stimulus components with a sound pressure level of 6 dB above the TQ are

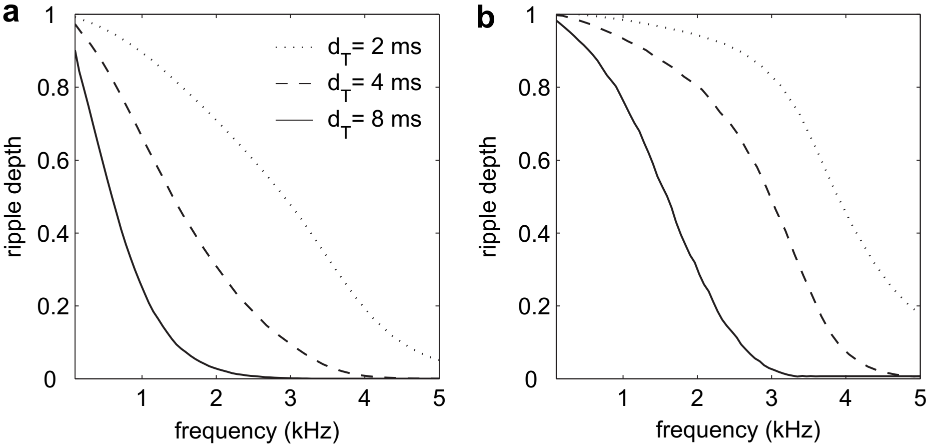

Fig. 6. Spectral ripple depth for a single reflection with a gain of 0 dB and a delay of 2, 4, and 8 ms. The left panel shows the ripple depth after auditory filtering (i.e., directly before the SCE stage) and the right panel after the SCE processing. The SCE stage enhances the spectral ripple depth with both increasing frequency and increasing ripple depth (i.e., reflection delay).

evaluated. Since the CD model's input signal is normalized such that its RMS level directly represents the sound pressure level in dB-SPL, the TQ is realized by disregarding any frequency channel in which the average power of the direct sound alone, $\overline{u}_x(i)$, is smaller than $10^{6/10}$. Considering the entire signal path including stimulus generation and peripheral auditory processing, this threshold is influenced by the middle ear filter H_{ME} as well as by the bandpass filter applied during stimulus generation (see Section 2.2). The middle ear filtering results in a "bowl-shaped" threshold, which roughly resembles the shape of the TQ described by Zwicker and Fastl (1999).

Since internal noise is added to all frequency channels i', the frequency range restriction applied in Eq. (8) limits the total power of the internal noise considered in the detection process. This results in an integrated auditory-internal noise power that is dependent on (1) the stimulus bandwidth (due to the TQ) and (2) the ripple density (due to the limit of $i_2' \leq f_{\max}/f\Delta$).

4.2.3. Signal detector

The signal detection stage was realized as an artificial observer (e.g., Dau et al., 1996), such that the CD model could simulate a listener in the CD experiments described in Section 2. However, instead of directly applying an adaptive up-down procedure to predict the CDT, an entire psychometric function was simulated here from which the CDT (i.e., the 79% correct point) was determined via a sigmoid function approximation. For each stimulus triplet presentation the detector picked the stimulus interval with the largest model output as the target interval. Only the model output $v(\tau)$ at $\tau = d_T$ was considered, which provided the most sensitive "channel" to a reflection with delay d_T. Each point on the psychometric function was derived from 1000 simulated trials. The realization of the CD model as an artificial observer ensured that similar stimulus variability was taken into account as involved in the psychoacoustic experiments. Interestingly, when investigating the model's detection process in detail, it was found that mainly the auditory-internal noise limited the model's sensitivity and, thus, determined the prediction of the CDT. Only at frequencies close to the maximum frequency, $f_{\max}$, did the variability inherent in the noise stimulus (i.e., external noise) limit sensitivity.

4.3. Signal processing example

Fig. 7 shows mean output spectra of the different stages of the CD model for an example of a single-reflection stimulus as used in the CDT experiments (see Section 2.2). The stimulus was a 320-ms long noise conveying the frequency range 100–5000 Hz and containing a reflection with a delay of $d_T = 2$ ms and a gain of $g_{T,\mathrm{dB}} = -3$ dB. The mean spectra were calculated over $N = 1000$ stimulus presentations. Fig. 7a shows the physical power transfer function $|H(z)|^2$ of a single reflection normalized to the direct sound alone power spectrum. The single reflection introduces a cosine-shaped spectral ripple whose density, plotted on an ERB$_N$-number scale $z = i\cdot z\Delta$, increases with increasing centre frequency. Fig. 7b shows the corresponding power spectrum at the output of the auditory filterbank, $u_x(z)$, for the case of the direct sound alone (dotted line) and the direct sound plus test reflection (solid line). The increase in bandwidth of the auditory filters with increasing center frequency results in: (1) an overall increase in output power with increasing center frequency and (2) a smoothing of the power spectrum which decreases the depth of the spectral ripple with increasing frequency. Fig. 7c shows the power spectrum $u_y(z)$ at the output of the SCE stage. This output spectrum serves as input to the detection stage at a higher auditory processing level. Since the density of the spectral ripple introduced by a reflection increases with increasing ERB$_N$-number, the contrast enhancement operation increases the ripple depth with increasing ERB$_N$-number. The normalized spectrum, $r(z)$, is shown in Fig. 7d, illustrating an improved sensitivity to spectral low-level components (i.e., the spectral ripples just below or above the stimulus cut-off frequencies $f_1 = 100$ Hz and $f_2 = 5000$ Hz are enhanced). The output of the spectral pattern analysis stage, i.e., the auto-correlation function $v(\tau)$, is shown in Fig. 7e. The output shows a clear peak at an auto-correlation lag τ equal to the reflection delay d_T (i.e., at $\tau = d_T$). The height of this peak determines the coloration strength and provides the model's detection cue.

4.4. Model predictions

The relevance of the signal processing performed by the CD model was evaluated by comparing CDTs predicted by the model to experimental data presented in Section 3. In the left panel of Fig. 8, the mean CDT data for three subjects are shown as a function of lower cut-off frequency f_1, with the reflection delay d_T as parameter. In the right panel of Fig. 8, the corresponding model predictions are shown. The model predictions successfully describe the increase in threshold with increasing lower cut-off frequency, as well as the increase in threshold with increasing reflection delay.

Fig. 9 shows the mean experimental CDT data (left panel) and the corresponding model predictions (right panel) as a function of

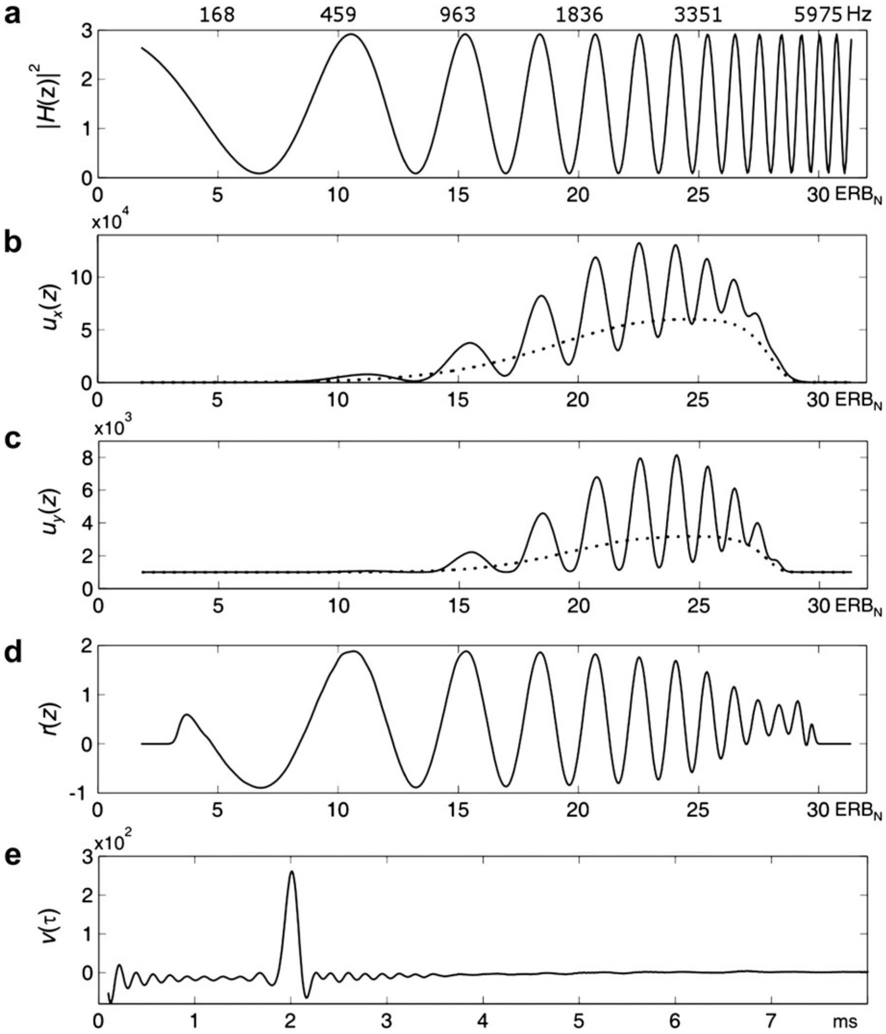

Fig. 7. Example of mean output spectra of different stages of the CD model (see Fig. 3) for a single reflection input stimulus. The stimulus was a 320-ms long noise with a frequency range of 100–5000 Hz, and the reflection had a delay of $d_T = 2$ ms and a gain of $g_{T,dB} = -3$ dB. Panel a) shows the physical power transfer function $|H(z)|^2$ normalized to the direct sound alone power spectrum as a function of ERB$_N$-number, z. Panel b) illustrates the power spectrum after auditory filtering for the cases of the direct sound plus reflection (solid line) and the direct sound alone (dotted line). Panel c) illustrates the corresponding power spectra at the output of the spectral contrast enhancement stage. Panel d) shows the normalized spectrum $r(z)$, which is the input to the subsequent spectral pattern analysis stage. Panel e) provides the output of the spectral pattern analysis stage, i.e., the auto-correlation function $v(\tau)$, with τ the auto-correlation lag.

upper cut-off frequency f_2. The model correctly predicts the decrease in threshold with increasing upper cut-off frequency, as well as the increase in threshold with increasing reflection delay. The good agreement between the experimental data and the model predictions indicates that the CD model can quantitatively account for the effective mechanisms involved in auditory detection of coloration produced by a single reflection.

5. Discussion

The CD model follows a purely spectrum-based approach, similar to the pattern-transformation model described by Wightman (1973). However, the proposed model includes a number of additional mechanisms and processing details which were not included in the original pattern-transformation model but which were important to quantitatively account for the CD data presented in Section 3.

5.1. Increased spectral resolution

In order for the proposed CD model to account for the auditory sensitivity to spectral ripples with a high density (as observed in the CDT data described in Section 3), significantly narrower auditory filters than classic Gammatone filters had to be applied for frequencies below about 5 kHz (see Section 4.1.1). These narrower filters were realized by the auditory filters measured by Oxenham and Shera (2003). In contrast to classic Gammatone filters, which are derived from notch-noise simultaneous masking (SM) data, Oxenham and Shera derived their narrower filters from notched-noise forward masking (FM) data. The difference in bandwidth measured in FM and SM experiments has typically been linked to (instantaneously acting) suppression effects on the basilar membrane (e.g., Moore and O'Loughlin, 1986), although the underlying mechanisms are still poorly understood. Pickles (2008) and Shera et al. (2002) have argued that FM-based measures are more appropriate for specifying auditory frequency resolution than SM-based measures, because they are in better agreement with physiological data. However, this is in contradiction with Ruggero and Temchin (2005), who provide evidence that physiologically measured filters are significantly broader than those measured in FM. Conceptually, suppression effects have been linked to spectral contrast enhancement or "sharpening", at least for tones in noise and formants in speech (e.g., Moore and O'Loughlin, 1986; De

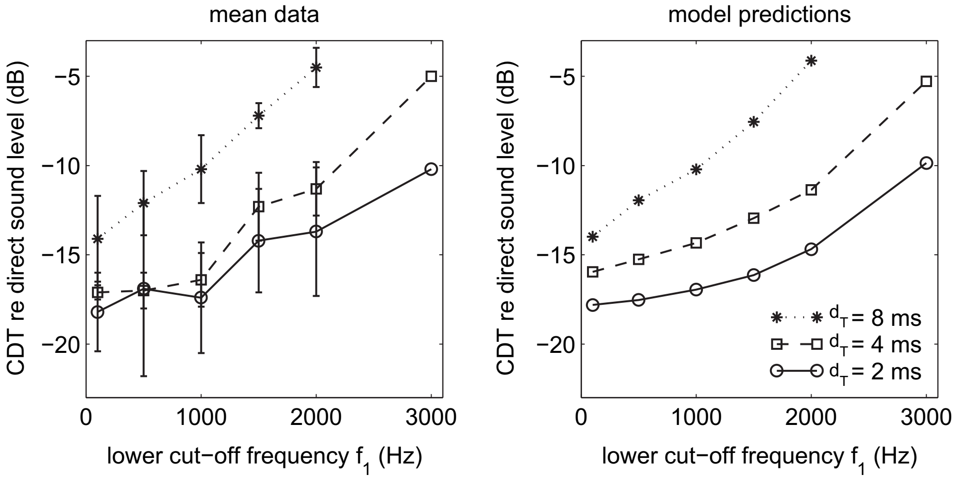

Fig. 8. Experimental mean CDT data (left panel) taken from Section 3 (Fig. 1) and corresponding model predictions (right panel). Thresholds are shown as a function of lower cut-off frequency f_1 with the reflection delay d_T as parameter.

Cheveigné, 2005). Hence, it might be speculated here that suppression effects also increase auditory sensitivity to spectral ripples with high density, supporting the application of narrower auditory filters in the CD model.

Although cochlear suppression might be involved, no conclusive argument can be provided as to why narrower filters than classic Gammatone filters should be employed in the CD model. Current literature on pitch perception typically refers to this spectral-resolution problem to argue in favor of a time-based (auto-correlation) approach, which inherently does not have this problem (e.g., De Cheveigné, 2005). Hence, the implementation of narrower auditory filters in the CD model should be rather understood as an engineering short cut to this spectral-resolution problem, in order to quantitatively describe the measured CDT data, but it does not necessarily describe the underlying auditory mechanisms.

5.2. Spectral contrast enhancement

Within the CD model, the power spectrum after auditory filtering is processed by a spectral contrast enhancement (SCE) stage (see Section 4.1.2), which introduces a weighting of spectral ripples that is dependent on the frequency as well as the ripple density. The weighting by the SCE stage was necessary to successfully predict the observed dependence of the CDT data on spectral content and reflection delay (see Section 4.4) and was the main "parameter" for tuning the frequency dependency of the CDT predictions. The SCE stage might be related to (wideband) inhibition mechanisms in the cochlear nucleus and later stages of the auditory system that are known to sharpen the contrast of the stimulus spectrum (e.g., Pickles, 2008). The current realization of the SCE stage (see Section 4.1.2) provides a simple parameterized function that has been adjusted to optimize the agreement between model predictions and experimental data. Of course, such a convenient engineering solution only describes the effect of the underlying non-linear neural processes.

The decrease in bandwidth of the applied auditory filters (Section 4.1.1 and 5.1) as well as the SCE mechanism scale with ERB$_N$-number and effectively result in an enhancement of spectral ripples with high density. However, their influence on overall detection of spectral ripples is very different. The decreased bandwidth of the auditory filters provides a "true" spectral resolution enhancement and therefore increases the contrast between the spectral ripple (produced by the test reflection) and the variations in the (estimated) power spectrum introduced by the fluctuations inherent in the (noise) stimulus with finite duration (i.e., the external noise). This increase in signal-to-noise ratio is unaffected by the SCE stage, because this stage acts similarly on the signal (i.e., the spectral ripple) and the external noise. However, the overall sensitivity of the

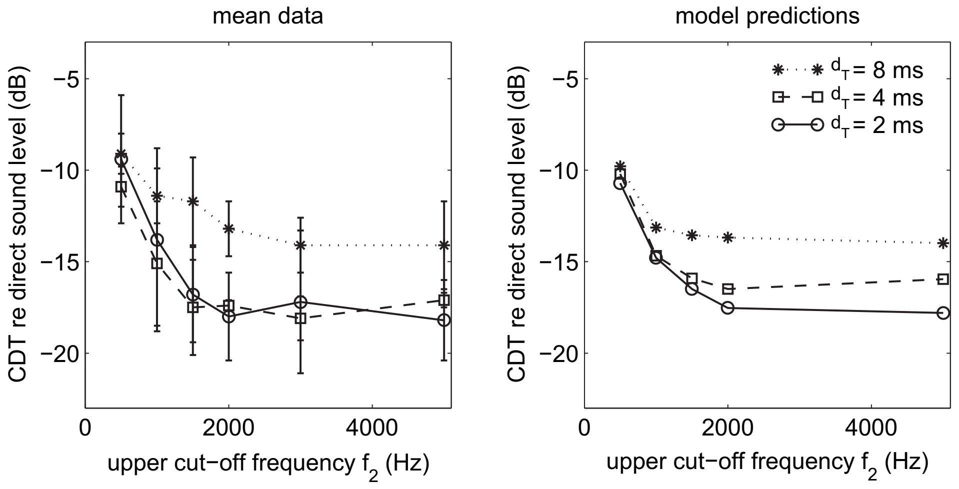

Fig. 9. As Fig. 8, except showing CDTs as a function of upper cut-off frequency f_2. Mean data is taken from Fig. 2.

CD model is modified by the SCE processing, when the internal noise that is added after the SCE stage (see Fig. 3) dominates over the external noise. In Section 4.2, it was highlighted that within the CD model the detection of spectral ripples was mainly limited by the internal noise except for very high frequencies, where the external noise was the limiting factor. Hence, the bandwidth of the auditory filters imposes an absolute limit to the frequency range over which spectral ripples can be evaluated, which cannot be overcome by the subsequent SCE stage. Future research should further investigate the contribution of the internal and external noise on coloration detection by considering stimuli with different durations. Shortening the stimulus duration will lead to an increase in external noise, but will not have an effect on the internal noise and thus, will eventually result in external noise dominating CDT predictions.

5.3. *Spectral pattern analysis and coloration detection*

The auditory-internal power spectrum is processed by a pattern analysis stage that integrates information across-frequency channels via a cosine transform (Section 4.2 and Fig. 3). According to the Wiener–Khintchine theorem (e.g., Hartmann, 1998), such a process results in the auto-correlation function, which has been widely used in previous pitch and CD models (see Section 1). Due to this pattern analysis, the sensitivity of the CD process increases with increasing stimulus bandwidth (i.e., with increasing number of spectral ripples). Since internal noise is added to each frequency channel (see Fig. 3), this internal noise is also integrated by the pattern analysis stage and thus, limits the influence of stimulus bandwidth on CD. The spectral integration property of the pattern analysis stage was essential for the CD model to successfully predict the frequency dependency of the CDT data (Section 4.4), which could not be achieved with, for example, a single channel detector. A single channel detector would demand a very different spectral weighting (by the CD model's periphery) to describe the CDT data dependency on the lower cut-off frequency (f_1) than for the upper cut-off frequency (f_2).

The general existence of auditory mechanisms that integrate information over multiple frequency channels has often been demonstrated in psychoacoustic experiments, for example, when detecting noise-bursts in noise (e.g., Hant et al., 1997), detecting tone-complexes in noise (e.g., Buus et al., 1986), in comodulation masking release (CMR: e.g., Piechowiak et al., 2007), or by the auditory "profile analysis" (Green, 1988). Moreover, a number of model approaches exist that combine information across multiple frequency channels. For example, Durlach et al. (1986) proposed a model for discriminating broadband stimuli that combines noisy information across different frequency channels. In order to describe the detection of broadband signals in noise, Hant and Alwan (2003) apply a multi-look masking model that combines information across multiple time-frequency bins. Moore and Tan (2004) predict the (perceived) naturalness of spectrally distorted sounds by integrating over all the changes in the excitation pattern that are introduced by the distortion. Piechowiak et al. (2007) applies an equalization-cancellation approach that combines correlated masker information from multiple frequency channels to account for different aspects of CMR. Although such models might be able to account in some ways for the spectral integration property inherent in the present model, the applied auto-correlation approach additionally provides information about the pitch of the input stimulus (see Section 1). However, the present CD model realization has not been optimized for predicting all the various kinds of existing pitch phenomena (e.g., see Plack et al., 2005). If this would be the task, the current spectral pattern analysis stage might need to be replaced, for example, by applying a harmonic summation process as described by Cohen et al. (1995), i.e. by basically replacing the multiplication with a cosine pattern (inherent in the cosine-transform) by a multiplication with a more comb-shaped pattern. However, a further analysis is out of the scope of the present study.

Considering ripple-noise pitch discrimination data, Yost (1982) provided evidence that the auditory system realizes a bandpass (BP) weighting of spectral ripples, the highest sensitivity being around a frequency of $f \approx 4/d_T$ (with d_T being the considered spectral ripple density or reflection delay). The existence of this "dominance region" has been inferred from a number of other pitch studies, although the exact shape of this weighting function is widely discussed (for a recent review, see Plack and Oxenham, 2005). Taking the absolute difference between the CDT data shown in Figs. 1 and 2 (as described by Yost, 1982) results in BP-shaped patterns with the highest sensitivity at about $f = 800$, 1100, and 1300 Hz for the different reflection delays $d_T = 8$, 4, and 2 ms. According to Yost (1982), this would support the idea of spectral BP weighting (i.e., the dominance region), although the derived BP shapes differ significantly from the ones observed by Yost. Considering the normalized output spectrum of the proposed CD model to a rippled-noise stimulus, as shown in Fig. 7d, no BP-shaped spectral weighting can be observed. For all ripple densities the model produces a clear lowpass (LP) weighting, i.e., the spectral ripple depth decreases monotonically with increasing center frequency (see Fig. 6). When comparing this auditory LP weighting to the BP weighting observed by Yost, it should be taken into account that Yost implied a single channel detector approach (as also done by many other pitch model approaches), whereas here a multi-channel detector (i.e., the cosine-transform) is used. Hence, instead of explaining the dominance region of pitch by a spectral BP weighting, the present study suggests that the dominance region might be similarly explained by a combination of spectral LP weighting and spectral integration. Moreover, it should be considered that besides auditory mechanisms, stimulus and method inherent properties might also influence conclusions on the shape (or existence) of the dominance region. For the ripple-noise pitch discrimination task used by Yost, the listener had to compare two spectral ripple noises with slightly different ripple densities $d_T = d_0 \cdot (1 \pm 0.06)$ (with d_0 a given reference density). Within the proposed CD model, the two different ripple-noise spectra would be compared (by division) in the spectrum normalization stage (Eq. (7)), which would introduce a ripple-shaped weighting to the output spectrum with the first maximum at $f = 1/(4 \cdot 0.06 \cdot d_0) \approx 4.17/d_0$. This maximum is very close to the maximum of the dominance region observed by Yost (i.e. at $f \approx 4/d_0$). Hence, at least for ripple-noise pitch, the dominance region might be partly explained by the maximal difference of the input spectra involved in the discrimination task rather than by an auditory-internal BP weighting mechanism. This observation implies that a CD task is more appropriate for analyzing the auditory processing of spectral ripples (as done in the present study), then applying a (ripple-noise) pitch discrimination task.

6. Summary and conclusions

In the first part of the study, CDTs were measured for a single reflection as a function of reflection delay and spectral content. The spectral content was modified by using BP filtered noise with different upper and lower cut-off frequencies. In the second part, a quantitative spectrum-based auditory detection model was developed, which was conceptually similar to the pattern-transformation model of pitch (Wightman, 1973) but contained a number of significant modifications. Within this model, a spectral pattern analysis of an auditory power spectrum is performed, which results in the derivation of an "auditory-weighted" auto-correlation analysis. The performance of this model was evaluated by comparing model predictions to the measured data. In this way it was demonstrated that in order to successfully predict the experimental data, a spectrum-based auto-correlation model needs to include the following "effective" components:

- Auditory bandpass filters as proposed by Oxenham and Shera (2003) with significantly narrower bandwidth than provided by classic Gammatone filters. The increased frequency resolution was required to extend the upper frequency limit of where spectral ripples can be utilized, and might be linked to suppression effects observed in the cochlea.
- A spectral contrast enhancement (SCE) stage that introduces a power spectrum-based weighting of spectral ripples that is dependent on ripple density as well as centre frequency. This stage was required to quantitatively describe the dependency of the CDT data on frequency region as well as test reflection delay, and might be linked to neural wideband inhibition mechanisms.
- A spectral pattern analysis stage that performs across-frequency integration with a frequency-independent sensitivity. This was realized by a linear frequency spacing of the auditory filters and an internal noise with a frequency-independent variance. Hence, any frequency-dependent sensitivity of the CD model was solely due to the preceding auditory filtering as well as the SCE processing.
- Internal noise to limit the sensitivity of the CD process. The internal noise is added to each frequency channel and is thus integrated by the subsequent spectral pattern analysis stage. Hence, the effective internal noise power considered in the CD process increases with increasing frequency range and limits the overall effect of bandwidth.

Acknowledgement

The author wishes to thank Torsten Dau, Brian C. J. Moore, and two anonymous reviewers for very helpful comments and discussions on a previous version of this manuscript.

Appendix A

In Section 4.2.2 the normalized spectrum $r(i)$ was transformed from an ERB$_\mathrm{N}$ frequency scale z to a linear frequency scale f to simplify the subsequent (central) processing stages of the CD model. Below an alternative approach is described, which circumvents this frequency scale transformation by applying a frequency weighting $g_n(i)$ to both the auditory-internal noise and the spectral analysis stage. The frequency weighting compensates for the low-frequency dominance in the detection process that would otherwise be produced by the non-linear ERB$_\mathrm{N}$ spacing of the auditory filters. In order to avoid this low frequency dominance, the frequency-independent noise described in Section 4.2.2 needs to be replaced by a noise with zero mean and a frequency-dependent variance $\sigma_n^2(i)$ given by:

$$\sigma_n^2(i) = \frac{\sigma_0^2}{g_n(i)}, \tag{A1}$$

with σ_0 a constant that determines the overall sensitivity of the CD model,

$$g_n(i) = \frac{0.11}{0.00437} e^{0.11 z_\Delta i}, \tag{A2}$$

z_Δ the number of frequency channels per ERB$_\mathrm{N}$ (which according to Section 4.1.1 is $z_\Delta = 18$) and the frequency index $i = z/z_\Delta$. Moreover, a frequency weighting needs to be applied to the pattern analysis stage, resulting in a spectrally-weighted cosine transform given by:

$$v(\tau) = \sum_{i=i_1}^{i_2} g_n(i)\ r_n(i) \cos\left(\frac{2\pi\tau}{0.00437}\left(e^{0.11 z_\Delta i} - 1\right)\right), \tag{A3}$$

with τ the spectral ripple density, $i_2 > i_1 \geq 0$, and n a subscript indicating a variable that is affected by (or related to) internal noise. Applying Eqs. A1–A3 and $\sigma_0 = 23$ in the CD model instead of Eqs. 7 and 8 results in identical CDT predictions as shown in Figs. 8 and 9 (right panels). Eqs. A1–A3 were analytically derived, but their rather extensive derivation is out of the scope of the present study.

References

Ando, Y., Alrutz, H., 1982. Perception of coloration in sound fields in relation to the auto-correlation function. J. Acoust. Soc. Am. 71, 616–618.

Atal, B.S., Schroeder, M.R., Kuttruff, K.H., 1962. Perception of Coloration in Filtered Gaussian Noise; Short-time Spectral Analysis by the Ear. Fourth International Congress on Acoustics, Copenhagen. paper H31.

Bilsen, F.A., Ritsma, R.J., 1969/70. Repetition pitch and its implications for hearing theory. Acustica 22, 63–73.

Bilsen, F.A., Ritsma, R.J., 1970. Some parameters influencing the perceptibility of pitch. J. Acoust. Soc. Am. 47, 469–475.

Blauert, J., 1997. "Spatial Hearing: The Psychophysics of Human Sound Localization", Revised Edition. MIT Press.

Bradley, J.S., Sato, H., Picard, M., 2003. On the importance of early reflections for speech in rooms. J. Acoust. Soc. Am. 113, 3233–3244.

Breebaart, J., van de Par, S., Kohlrausch, S., 2001. Binaural processing model based on contralateral inhibition. I. Model structure. J. Acoust. Soc. Am. 110, 1074–1088.

Buchholz, J.M., Mourjopoulos, J., 2003. A computational auditory masking model based on signal-dependent compression. I. Model description and performance analysis. Acta Acustica United with Acustica 90, 873–886.

Buchholz, J.M., 2007. Characterizing the monaural and binaural processes underlying reflection masking. Hear. Res. 232, 52–66.

Buus, S., Schorer, E., Florentine, M., Zwicker, E., 1986. "Decision rules in detection of simple and complex tones. J. Acoust. Soc. Am. 80, 1646–1657.

Cohen, M.A., Grossberg, S., Wyse, L.L., 1995. A spectral network model of pitch perception. J. Acoust. Soc. Am. 98, 862–878.

Dau, T., Püschel, D., Kohlrausch, A., 1996. A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. J. Acoust. Soc. Am. 99, 3615–3622.

De Cheveigné, A., 2005. Pitch perception models. In: Plack, C.J., Oxenham, A.J., Fay, R.R., Popper, A.N. (Eds.), Pitch: Neural Coding and Perception. Springer, pp. 169–233.

Durlach, N.I., Braida, L.D., Ito, Y., 1986. Towards a model for discrimination of broadband signals. J. Acoust. Soc. Am. 80, 63–72.

Eddins, D.A., Bero, E.M., 2007. Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region. J. Acoust. Soc. Am. 121, 363–372.

Glasberg, B.R., Moore, B.C.J., 1990. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47, 103–138.

Goldstein, J.L., 1973. An optimum processor theory for the central formation of the pitch of complex tones. J. Acoust. Soc. Am. 54, 1496–1516.

Green, D.M., 1988. Profile Analysis: Auditory Intensity Discrimination. Oxford University, New York.

Haas, H., 1951. Über den Einfluß eines Einfachechos auf die Hörsamkeit von Sprache. Acustica 1, 49–58.

Hant, J.J., Strop, B.P., Alwan, A.A., 1997. A psychoacoustic model for the noise masking of plosive bursts. J. Acoust. Soc. Am. 101, 2789–2802.

Hant, J.J., Alwan, A., 2003. A psychoacoustic-masking model to predict the perception of speech-like stimuli in noise. Speech Commun. 40, 291–313.

Hartmann, W.M., 1998. Signals, Sound, and Sensation. Springer, New York.

Kaernbach, C., 1991. Simple adaptive testing with the weighted up-down method. Perception and Psychoacoustics 49, 227–229.

Kates, J.M., 1985. A central spectrum model for the perception of coloration in filtered Gaussian noise. J. Acoust. Soc. Am. 77, 1529–1534.

Licklider, J.C.R., 1951. A duplex theory of pitch perception. Experientia 7, 128–134.

Litovsky, R.Y., Colburn, H.S., Yost, W.A., Guzman, S.J., 1999. The precedence effect. J. Acoust. Soc. Am. 106, 1633–1654.

Meddis, R., Hewitt, M.J., 1991a. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. J. Acoust. Soc. Am. 89, 2866–2882.

Meddis, R., Hewitt, M.J., 1991b. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. J. Acoust. Soc. Am. 89, 2883–2894.

Moore, B.C.J., O'Loughlin, B.J., 1986. The use of nonsimultaneous masking to measure frequency selectivity and suppression. In: Moore, B.C.J. (Ed.), Frequency Selectivity in Hearing. Academic Press, London.

Moore, B.C.J., Tan, C.-T., 2004. Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion. J. Audio Eng. Soc. 52, 900–914.

Oxenham, A.J., Shera, C., 2003. Estimates of human cochlear tuning at low levels using forward and simultaneous masking. J. Assoc. Res. Otolaryngol. 4, 541–554.

Patterson, R.D., Holdsworth, J., Nimmo-Smith, I., Rice, P., 1988. "Svos final report: The auditory filterbank," Tech. Rep. APU report 2341.

Pickles, J.O., 2008. An Introduction to the Physiology of Hearing, third ed. Emerald.

Piechowiak, T., Ewert, S.D., Dau, T., 2007. Modeling comodulation masking release using an equalization-cancellation mechanism. J. Acoust. Soc. Am. 121, 2111–2126.

Plack, C.J., Oxenham, A.J., 1998. Basilar-membrane nonlinearity and the growth of forward masking. J. Acoust. Soc. Am. 103, 1598–1608.

Plack, C.J., Oxenham, A.J., Fay, R.R., Popper, A.N., 2005. Pitch: Neural Coding and Perception. Springer, New York.

Plack, C.J., Oxenham, A.J., 2005. The psychophysics of pitch. In: Plack, C.J., Oxenham, A.J., Fay, R.R., Popper, A.N. (Eds.), Pitch: Neural Coding and Perception. Springer, pp. 7–55.

Ruggero, M.A., Temchin, A.N., 2005. Unexceptional sharpness of frequency tuning in the human cochlea. PNAS 102, 18614–18619.

Salomons, A.M., 1995. "Coloration and binaural decoloration of sound due to reflections." PhD thesis, University of Delft, The Netherlands.

Shera, C.A., Guian Jr., J.J., Oxenham, A.J., 2002. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc. Natl. Acad. Sci. 99, 3318–3323.

Terhardt, E., 1974. Pitch, consonance and harmony. J. Acoust. Soc. Am. 55, 1061–1069.

Wightman, F.L., 1973. The pattern-transformation model of pitch. J. Acoust. Soc. Am. 54, 407–416.

Yost, W.A., 1982. The dominance region and ripple noise pitch: a test of the peripheral weighting model. J. Acoust. Soc. Am. 72, 416–425.

Yost, W.A., 1996. Pitch strength of iterated rippled noise. J. Acoust. Soc. Am. 100, 3329–3335.

Zurek, P.M., 1979. Measurements of binaural echo suppression. J. Acoust. Soc. Am. 66, 1750–1757.

Zwicker, E., Fastl, H., 1999. Psychoacoustics: Facts and Models. Springer, Heidelberg.