Spectral integration effects in auditory detection of coloration

Jörg M. Buchholz, Paris Kerketsos

Centre for Applied Hearing Research, 2820 Kgs. Lyngby, Denmark, Email: jb@elektro.dtu.dk

Introduction

When an early wall reflection is added to a direct sound, a spectral modulation is introduced to the signal's power spectrum. This spectral modulation typically produces an auditory sensation of coloration or pitch. The present study aims at better understanding auditory across-frequency processes that are potentially involved in the perception of coloration. Coloration detection thresholds (CDTs) were therefore measured as a function of reflection delay and stimulus bandwidth. In order to interpret the experimental data, a quantitative auditory detection model was developed which was conceptually similar to the peripheral weighting model ([1]). Although the present model belongs to the category of spectral pitch models, in contrast to temporal pitch models, the goal of the present study is neither to discriminate between these two model concepts nor to argue in favour of one them. The goal is rather to analyse to what extend a (purely) spectrum-based model approach can potentially account for the auditory processing of coloration, and to identify processing stages that need to be included when following such approach. A better understanding of the auditory processing involved in coloration/pitch perception is of major relevance for different technical applications, including objective evaluation of room acoustics, hearing-aids, or audio playback systems.

Psychoacoustical methods

Coloration detection thresholds (CDTs) were measured for a broadband noise as a function of test reflection delay and spectral content. The direct sound was realized by 320-ms long broadband noise and the reflection delay was 2, 4, or 8 ms. 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) cut-off frequencies of this bandpass filter:

- **Condition 1:** f_1 = 100 Hz; 500 Hz ≤ f_2 ≤ 5 kHz
- **Condition 2:** 100 Hz ≤ f_1 ≤ 3 kHz; f_2 = 5 kHz

All stimuli were presented diotically via headphones. CDTs were measured by a three-interval three-alternative forced choice (AFC) procedure. A weighted up-down procedure was applied to track the 79% point on the psychometric function. In order to avoid loudness cues to influence the coloration detection thresholds, level roving of ±2 dB was introduced. Offset listening effects were suppressed by truncating the reflection offset. Three well trained and normally-hearing subjects were employed, aged 30-34. At least three measurements were made for each threshold and subject.

Results

For experimental condition 1, the across subject mean CDT and standard deviation are shown in the left panel of Fig. 1. With increasing upper cut-off frequency f_2 (i.e., increasing stimulus bandwidth) the CDTs generally decrease (i.e., the sensitivity increases) up to a maximum frequency $f_{2,max}$, above which the thresholds stay constant. For τ = 2 ms and τ = 4 ms, $f_{2,max} \approx$ 2kHz; for τ = 8 ms, $f_{2,max} \approx$ 3 kHz. Moreover, for small f_2 (i.e., narrow bandwidth) the CDT is almost independent of reflection delay. In contrast, for broadband input the CDT for 8-ms delay is significantly higher than for the 2 ms and 4 ms conditions.

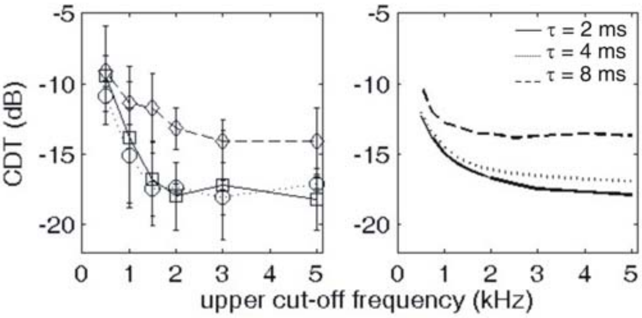

Fig. 1: Coloration detection data (left panel) and corresponding model predictions (right panel) for stimulus condition 2.

Similarly to Fig. 1, the mean CDT data for experimental condition 2 is shown in the left panel of Fig. 2. For the 2-ms and 4-ms conditions, increasing f_1 up to about 1 kHz has basically no effect on the measured CDTs. Otherwise, with increasing lower cut-off frequency f_1 (i.e., decreasing stimulus bandwidth) the CDTs increase monotonically. For the 8-ms condition, no subject was able to measure any CDT for f_1 > 2 kHz, and only two subjects were able to measure a CDT for f_1 = 2 kHz. Hence, the coloration cue is considered lost for f > 2 kHz.

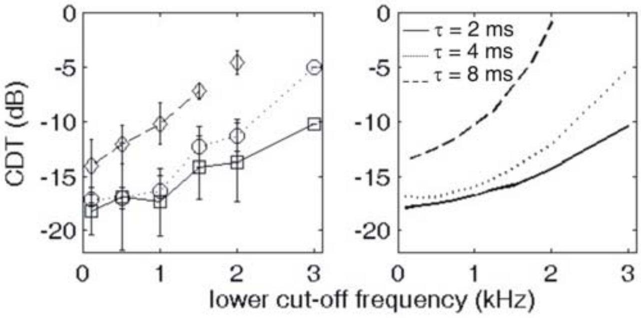

Fig. 2: Coloration detection data (left panel) and corresponding model predictions (right panel) for stimulus condition 1.

Modelling concepts

The signal-flow through the proposed model is illustrated in Fig. 3. First the input signal is analysed by an auditory band-pass (BP) filterbank with a frequency spacing of 16 channels per critical band. The output of each BP filter is multiplied with a gain g_i such that for white noise input, the power at each BP filter output is constant across frequency. Each channel is further processed by a pre-processing stage (PP) consisting of half-wave rectification, static logarithmic compression, and non-leaky temporal integration. Afterwards the signals are processed by a lateral inhibition stage. Auditory detection performance is limited by internal (white) noise, the power of which is equal across frequency. The resulting (noisy) auditory-internal "power" spectrum is processed by a spectral modulation analysis stage, which is here realized by a cosine transform. Similar to the human test subjects, the proposed model is here applied as an artificial observer in the above adaptive coloration detection experiments.

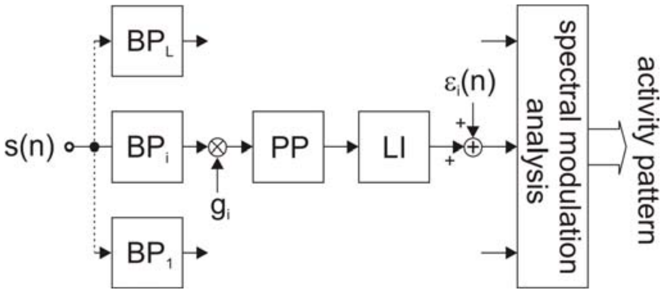

Fig. 3: Signal-flow diagram of the proposed auditory model

The employed BP filters are realized by (novel) gamma-chirp filters that have been fitted to notched-noise forward masking data provided by [2]. In comparison to standard gammatone filters ([3]), which are fitted to notched-noise simultaneous masking data, these filters are significantly narrower and are in better agreement with physiologically derived filters ([3]). With reference to the proposed model, the narrower filters are required to account for the sensitivity observed in the measured CDT data at high frequencies ($f \geq 1.5$ kHz) and large reflection delays ($\tau = 8$ ms). The power spectrum of the realized gamma-chirp filters is shown in Fig. 4 (dotted lines) together with "standard" gammatone filters (solid lines).

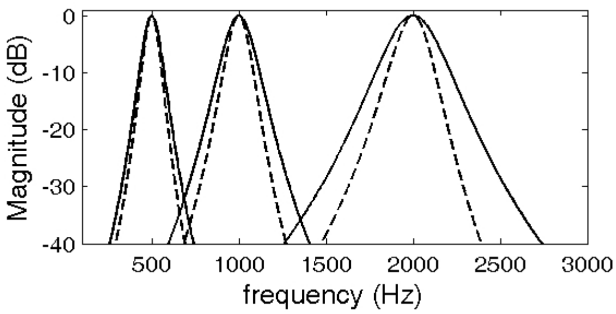

Fig. 4: Power spectrum of proposed gamma-chirp BP filters (dashed lines) and "standard" gammatone filters (solid lines).

The basic effect of the lateral inhibition (LI) stage (Fig. 3) is to highpass filter the power spectrum such that fast changes are enhanced in comparison to slower changes. Since the frequency spacing of the auditory filters is almost logarithmical, spectral modulations are also more enhanced at high signal frequencies. Hence, the LI stage introduces a weighting that is dependent on spectral modulation frequency (i.e., reflection delay) and signal frequency. In principle, such mechanism could be realized by the LI stage proposed by [4]. However, applying this LI stage in the proposed model does not produce a satisfactory agreement between model predictions and CDT data. Therefore, the effect of the LI stage, together with the subsequent internal (white) noise, was here replaced by a noise that is weighted dependent on modulation frequency and signal frequency. The required weights were derived by optimizing the agreement between model predictions and experimental data and are shown in Fig. 5. In future, the derived weighting functions will be used to derive an appropriate LI stage similar to [4].

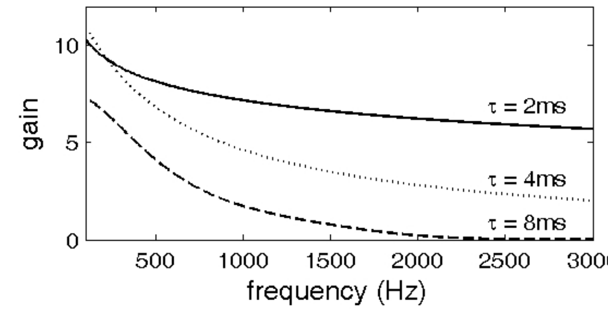

Fig. 5: Gains applied to the auditory-internal noise inherent in the proposed auditory model.

The CDT predictions of the proposed model for the two experimental conditions are shown in the right panel of Fig. 1 and Fig. 2, respectively. The model predictions are in very good agreement with the corresponding experimental data.

Summary and conclusions

Mechanisms were investigated that the auditory system might apply to analyze spectral modulations produced by early reflections. A spectrum-based auditory detection model was tested by comparing model predictions to measured data. It was found that such model could successfully describe the experimental data if the model includes the following features:

- Sharper auditory filters than realized by classical gammatone filters; here realized by gamma-chirp filters derived from forward masking data.
- A modulation highpass weighting that is dependent on both spectral modulation frequency and signal frequency; this might be realized by a processing stage similar to the lateral inhibition stage proposed by [4].
- A spectral modulation analysis (i.e., across frequency processing); hear realized by a cosine transform.

Although the present model approach shows very promising results, further development and optimization is needed.

References

[1] Wightman, F. L.: The pattern-transformation model of pitch. J. Acoust. Soc. Am. 54 (1973), 407-416.

[2] Oxenham A. J. and Shera C. A.: Estimates of human cochlear tuning at low levels using forward and simultaneous masking. JARO 4 (2003), 541-554.

[3] Patterson R. D. and Moore B. C. J.: Auditory filters and excitation patterns as representations of frequency resolution, Frequency selectivity in hearing, ed. B. J. C. Moore, Academic Press, London (1986), 123-174.

[4] Shamma, S. A.: Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. J. Acoust. Soc. A., 78 (1985), 1622-1632.