

# A Parallel Approach to HRTF Approximation and Interpolation Based on a Parametric Filter Model

German Ramos, *Member, IEEE*, Maximo Cobos, *Senior Member, IEEE*, Balázs Bank, *Member, IEEE*, and Jose A. Belloch, *Member, IEEE*

**Abstract**—Spatial audio rendering techniques using head-related transfer functions (HRTFs) are currently used in many different contexts such as immersive teleconferencing systems, gaming, or 3D audio reproduction. Since all these applications usually involve real-time constraints, efficient processing structures for HRTF modeling and interpolation are necessary for providing real-time binaural audio solutions. This letter presents a parametric parallel model that allows to perform HRTF filtering and interpolation efficiently from an input HRTF dataset. The resulting model, which is an adaptation from a recently proposed modeling technique, not only reduces the size of HRTF datasets significantly, but it allows for simplified interpolation and real-time computation over parallel processors. In order to discuss the suitability of this new model, an implementation over a graphic processing unit (GPU) is presented.

**Index Terms**—Binaural synthesis, HRTF modeling, interpolation, parallel filters, GPU.

## I. INTRODUCTION

NOWADAYS, head-related transfer functions (HRTFs) are used in different applications requiring realistic spatial audio rendering. HRTFs capture all the effects that a free-field sound wave experiences from its source to the listeners eardrum. These effects are multiple, including those introduced by the acoustic path and the reflection and diffraction effects from a listener's anatomical structure. Interaural time differences (ITD) and interaural level differences (ILD) are the most important cues for localizing sources on the horizontal plane. Other shadowing effects and reflection patterns encoded in HRTFs have been shown to be very important for accurate source localization. Filtering caused by the pinna affects especially the propagation of high frequencies. The precise nature is determined by the ear shape, thus is unique to each individual. Also, the upper torso reflects frequencies (especially in the mid-range) to produce very short time-delayed echoes. The length of these time delays varies with the elevation of

the sound source. All this information is used by the brain to localize a sound source in the space, exploiting differences between ear signals as well as monaural and dynamic cues [1]. HRTFs are currently used in immersive sound applications [2], such as stereo enhancement, surround simulation, virtual reality, auralization, video conferencing systems and video games. Low-order models of HRTFs can significantly reduce the computational cost of all these applications [3], [4]. However, these models must be accurate enough to keep the perceptual characteristics of the original HRTFs unaltered. In this context, different solutions have already been proposed for HRTF modeling. Some approaches are based on analytical models of the head that allow to simulate relevant propagation and diffraction effects [5]. Others, such as the one presented in this letter, follow an empirical approach where digital filters are used to approximate the HRTF behavior [3], [4], [6].

The contribution described in this letter is based on the infinite impulse response (IIR) parametric model presented by the authors in [7], aimed at approximating HRTF responses by using a chain of parametric digital filters such as the ones used in audio equalization [8], [9]. Parametric approaches allow to describe HRTFs with a limited set of parameters instead of the whole set of filter coefficients, reducing the amount of information needed to describe a complete HRTF dataset. Moreover, parametric methods result in additional benefits such as the simplification of the interpolation procedures needed for synthesizing HRTFs corresponding to spatial angles missing in the original HRTF dataset [10], [11]. While the previous work in [7] was shown to provide all the above benefits by implementing a chain of second-order sections (SOS), the data dependencies within the proposed sequential structure prevent taking advantage of current parallel processors embedded into state-of-the-art devices. To overcome this limitation, this letter proposes an adaptation of the original SOS chain model into a parallel one, which can be efficiently implemented on parallel processors, such as graphics processing units (GPUs). These can be found in most mobile/multimedia devices, being a general-purpose computational resource that can be successfully exploited in many acoustic signal processing applications [12]. The new model breaks the data dependencies between sections and keeps providing a simple interpolation scheme. To show the applicability of the method, an implementation on a GPU is discussed.

## II. PARAMETRIC SECOND-ORDER SECTION CHAIN MODEL

This letter provides an adaptation of the sequential IIR parametric model shown in Fig. 1 [7]. The first stage consists

G. Ramos is with the iTEAM Institute of the Universitat Politècnica de València, Valencia (Spain), [gramosp@eln.upv.es](mailto:gramosp@eln.upv.es)

M. Cobos is with the Dept. Informàtica de la Universitat de València, Valencia (Spain), [maximo.cobos@uv.es](mailto:maximo.cobos@uv.es)

J. A. Belloch is with the Dept. de Ing. y Ciencia de Comp. of the Univ. Jaume I de Castelló, Castellón (Spain), [jbelloch@uji.es](mailto:jbelloch@uji.es)

B. Bank is with the Dept. of Measurement and Information Systems of the Budapest Univ. of Techn. and Economics, Budapest (Hungary), [bank@mit.bme.hu](mailto:bank@mit.bme.hu)

This work was supported by the Spanish Ministry of Economy and Competitiveness under Grant TEC2012-37945-C02-02 and FEDER funds and by the the ÚNKP-16-4-III New National Excellence Program of the Hungarian Ministry of Human Capacities. Dr. Jose A. Belloch is supported by GVA postdoctoral contract APOSTD/2016/069.

Manuscript received May 30, 2017; revised xxx, xxx.

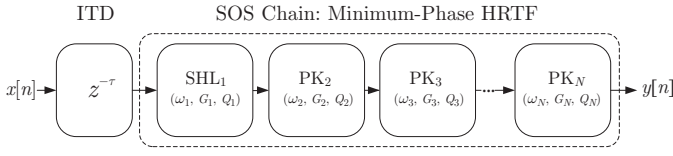


Fig. 1. Sequential HRTF parametric model modeling.

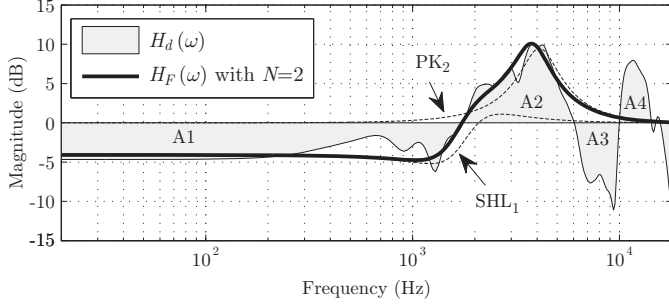


Fig. 2. Original HRTF  $H_d(\omega)$  and approximated response  $H_F(\omega)$ .

of a delay line block that models the ITD, i.e. the propagation delay  $\tau$  between the left and right ears for a given HRTF direction. This initial delay line is followed by a series of  $N$  SOS modeling the minimum-phase response. This chain is made up of a low-shelving filter  $\text{SHL}_1$  that sets the low-frequency relatively uniform structure of the response, and a set of peak filters  $\text{PK}_i$  ( $i = 2, \dots, N$ ) that jointly interact to synthesize the peak-valley structure of the HRTF. These SOS are all defined by a triplet of parameters: digital center frequency  $\omega_i$ , log-gain  $G_i$  and quality factor  $Q_i$ . The digital transfer function in the  $z$ -domain for all these biquadratic filters is generally expressed as

$$H_i(z) = \frac{b_0^{(i)} + b_1^{(i)}z^{-1} + b_2^{(i)}z^{-2}}{a_0^{(i)} + a_1^{(i)}z^{-1} + a_2^{(i)}z^{-2}}, \quad (1)$$

where the relation between the filter coefficients ( $b_k^{(i)}$  and  $a_k^{(i)}$ ) and the parameters ( $\omega_i, G_i, Q_i$ ) is well known [7]. The set of parameters are computed by iteratively reducing the error with respect to the original 1/12<sup>th</sup>-octave-smoothed HRTF. This is performed by considering a perceptually-motivated cost function working over a logarithmic domain both in the magnitude and frequency axes. The error is minimized iteratively fitting areas between the target and the model using the set of SOS one by one. A simple example with  $N = 2$  is shown in Fig. 2, where the areas involved in the fitting process are denoted as A1, A2, etc.  $\text{SHL}_1$  approximates A1, then, this filter is fixed and  $\text{PK}_2$  follows the same procedure with the next biggest area (A2), etc. For each SOS, a good initial value set for the parameters is given, obtained from the error area information: mean log-frequency and dB magnitude value. A post-optimization process is carried out to improve the interaction among SOS stages (see [7] for details). Once the HRTF from a given direction has been modeled, the parameters for neighboring directions are estimated by evolving the already constructed filter parameters as a starting point.

### III. PARALLEL MODEL BY PARTIAL FRACTION EXPANSION

Although the previous parametric model provides satisfactory results, its sequential structure is not well suited to be implemented over a parallel processor such as a multicore machine or a GPU. The simplest solution for obtaining a mathematically equivalent model is to perform a series to parallel conversion by classical partial-fraction-expansion (PFE) [13]. Once the model has been designed as a series of  $N$  SOS defined by its coefficients  $b_k^{(i)}$  and  $a_k^{(i)}$ ,  $k = 0, 1, 2$ , a transformation to parallel is performed by finding the new coefficients:  $K$ ,  $b'_k{}^{(i)}$  and  $a'_k{}^{(i)}$ ,  $k = 0, 1, 2$ . This leads to the model shown in Figure 3. Note that the ITD block remains unaltered, but now the different SOS stages can be evaluated in parallel due to the nonexistence of data dependencies. Assuming normalized coefficients with  $a_0^{(i)} = 1$ , the overall response can be now be rewritten as

$$\begin{aligned} H(z) &= \prod_{i=1}^N \frac{b_0^{(i)} + b_1^{(i)}z^{-1} + b_2^{(i)}z^{-2}}{1 + a_1^{(i)}z^{-1} + a_2^{(i)}z^{-2}} \\ &= K + \sum_{i=1}^N \frac{b'_0{}^{(i)} + b'_1{}^{(i)}z^{-1}}{1 + a'_1{}^{(i)}z^{-1} + a'_2{}^{(i)}z^{-2}}. \end{aligned} \quad (2)$$

While both the sequential and the PFE model have identical responses, an example will bring up the limitations of the latter. Consider for this purpose a model with  $N = 6$  SOS for a left HRTF at azimuth  $\phi = 5^\circ$  and elevation  $\theta = 0^\circ$  from subject 3 of CIPIC database [14], as shown in Figure 4(a). The responses of the individual SOS are displayed in black thin lines and their complex addition result in the same response as the sequential model. In contrast to the series implementation, the phase interaction between the individual responses now becomes more relevant. Consider now the response at the next available direction ( $\phi = 10^\circ$ ) in Fig. 4(b). Despite the similarities of the overall response at both angles, some of the individual SOS responses are very different from the ones at the previous angular direction (close to 20 dB with significant DC level differences between them). As a result, although both models are equivalent, the large differences in the individual SOS responses between consecutive angles do not allow a simplified interpolation procedure of the original model by interpolating neighboring parameters. In fact, even considering the improvements obtained by using a delayed implementation of the IIR part, as recently proposed in [15], the differences of the responses for consecutive angles are still considerably large, preventing the implementation of a simple interpolation procedure.

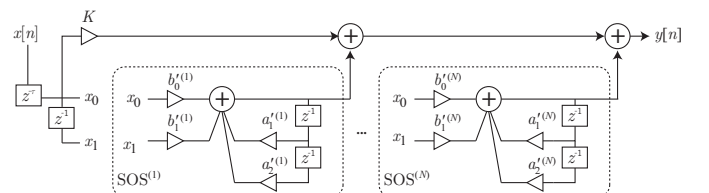


Fig. 3. Parallel implementation by partial-fraction-expansion.

#### IV. PARAMETRIC PARALLEL MODEL

Once the limitations of the PFE model have been detected, the need arises for better parallel structures allowing simple HRTF interpolation. While parallel filter banks have been widely used in audio processing and equalization [16], [17], [18], [19], this is the first work considering a parametric approach for HRTF approximation. The novelty resides in combining the advantages of parallel filter banks with those provided by parametric models, making this approach more efficient from a modeling and computational point of view.

Fig. 5 shows the proposed parallel filter structure. The series of low-shelving  $SH_1$  and peak filters  $PK_i$  have been replaced by a parallel combination of a direct path, a low-pass filter  $LP_1$ , and band-pass filters  $BP_i$ , all of them again defined by their parameters  $(\omega_i, G_i, Q_i)$ . The filter  $SHL_1$  is replaced by a combination of the direct path and the  $LP_1$ , while the  $PK_i$  result from combining the direct path with the  $BP_i$ . One of the benefits of this approach is that all the band-pass filters  $BP_i$  have zero  $b_1$  coefficient, and  $b_2 = -b_0$ . This comes from the fact that the bilinear transform of an analog second-order bandpass with a single zero at the origin has the numerator of the form  $b_0(1 - z^{-2})$ . For the low-pass filter,  $b_2 = b_0$ . This allows saving two multiplications and accumulations for all the band-pass SOS as it will be seen in Sec. V, obtaining a reduction in computational cost close to a 40% with respect to the sequential model of the same number of sections (or allows increasing the number of sections for the same computational cost).

With the proposed structure, the sequential algorithm has been adapted to consider the new parallel scheme, keeping the optimization process unaltered. Now the phase interaction between the SOS becomes relevant. Fig. 6 shows the parametric parallel model of the HRTF at  $\phi = 0^\circ$  obtained with 12 SOS. Note that the model matches the original (smoothed)

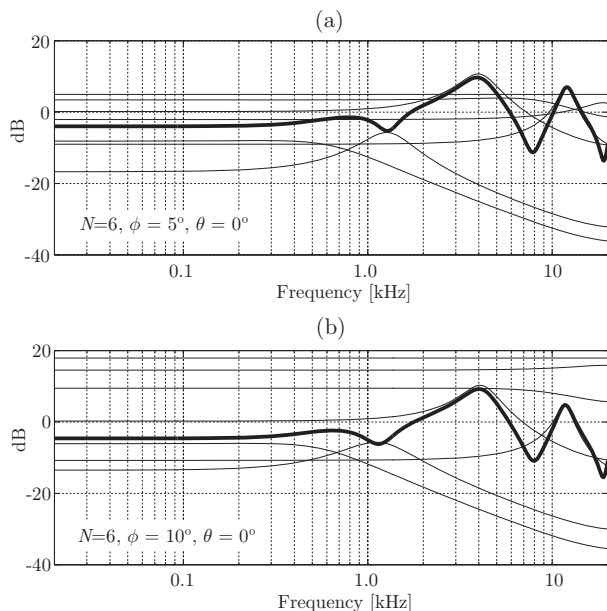


Fig. 4. Partial-fraction-expansion with  $N = 6$  SOS at two consecutive positions. (a)  $\phi = 5^\circ$ . (b)  $\phi = 10^\circ$ .

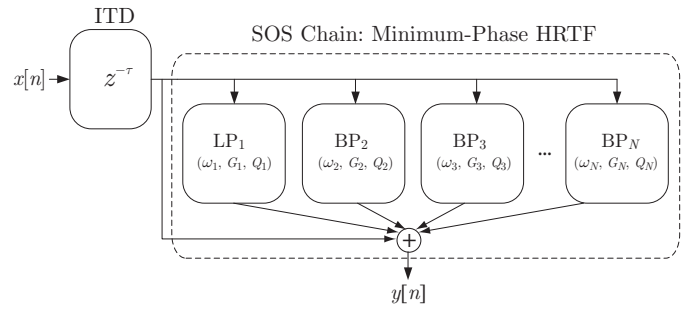


Fig. 5. Proposed parametric parallel implementation.

HRTF response with small deviations of  $\pm 1$  dB. The black thin lines represent the individual SOS responses. Solid-line responses have positive gain to create peaks on the response while dotted-line ones have negative gain to create valleys.

#### A. Number of sections

Figure 7 shows the evolution of the perceptually-motivated cost function [7] with respect to the number of SOS both for the serial and the parallel models. The same HRTF used in Fig. 6 has been used as target. Note that the behavior is almost identical for both models. When  $N \geq 12$ , the improvement is marginal. In fact, a set of listening tests were carried out considering the serial parametric model, concluding that the perceptual scores between original HRTFs and the approximated ones were not statistically significant for  $N = 12$  [7]. This is expected since the magnitude differences are minor and both models provide minimum-phase responses.

#### B. Interpolation

Fig. 8 shows part of a set of designed and interpolated HRTFs with  $N = 16$ , with the evolution of the SOS frequencies (represented as connected dots). Solid lines are designed models, while dashed lines are interpolated models obtained by simple linear interpolation of the parameters from the two adjacent designed models. Contiguous responses are shown with an offset of 10 dB for the sake of clarity. Note that there are only minor deviations between the original and interpolated responses. Due to the complex interactions between sections, the parametric parallel model needs additional SOS with respect to the serial one to achieve equivalent fitting performance, especially at extreme angles. With  $N = 16$  (as in Fig. 8) the result is almost identical to the one obtained with the original SOS chain model with  $N = 12$ . However, the number of required operations still remains lower since the parallel model needs less operations per SOS, as discussed before.

#### V. GPU-BASED IMPLEMENTATION AND PERFORMANCE

To evaluate the practical benefits of the above parallel model, a GPU implementation that uses Compute Unified Device Architecture (CUDA) has been considered. Besides the advantages provided by the proposed filtering structure, alternative optimizations can be achieved with parallel processors. For example, when multiple simultaneous sources are to be

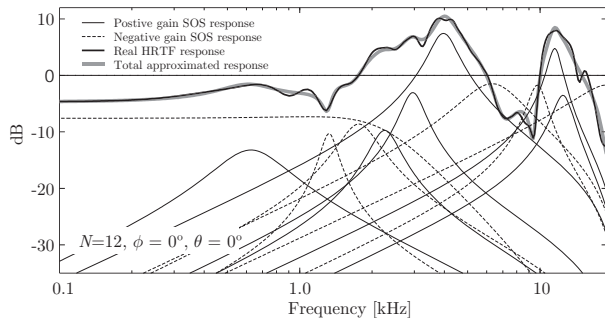


Fig. 6. Parametric parallel model with 12 SOS and the individual SOS responses.

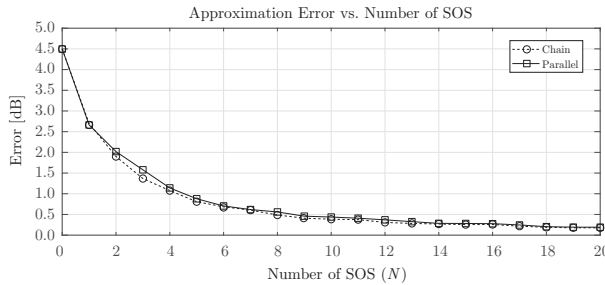


Fig. 7. Evolution of the approximation error as a function of  $N$  for  $\phi = 0^\circ$ .

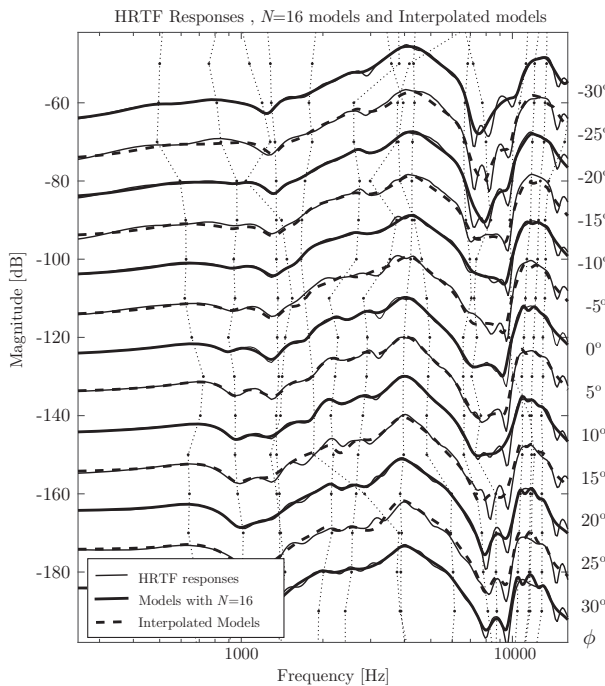


Fig. 8. Designed and interpolated HRTF set evolved from  $\phi = 0^\circ$ . Connected dots represent SOS frequencies.

rendered, all the operations related to the filtering of different source signals can be performed in parallel.

The implementation uses each GPU thread to compute a section  $H_i(z)$  of the model (the low-pass  $LP_1$ , and the band-pass  $BP_2$  to  $BP_N$ ) of the total  $32M$  filters ( $M$  sound sources in a binaural system; a maximum of  $N = 16$  SOS per filter; and two output channels, left and right). Fig. 9 presents the

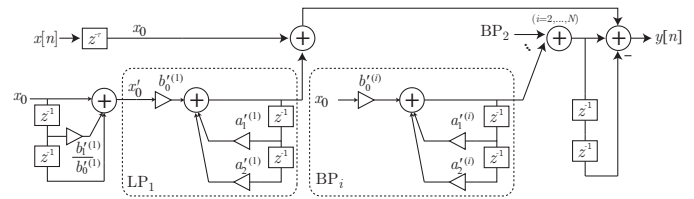


Fig. 9. SOS processed by one thread.

implementation of each thread, where an optimized structure has been employed in order to achieve a common processing scheme for all the SOS.

An Nvidia Tesla K20Xm and an audio card were used, configured to provide 8, 16, 32 and 64 samples per channel every 0.18 ms, 0.36 ms, 0.72 ms and 1.45 ms, respectively ( $f_s = 44.1$  kHz). The maximum number of sound sources that can be managed in real time for the above latencies are 280, 575, 930 and 1230 for each latency value. Additionally, an equivalent serial implementation providing approximately the same accuracy (12 SOS) has been implemented on the GPU for comparison purposes. While the latencies are the same, the number of sound-sources that can be managed in real time is significantly reduced, reaching only 15 sources in the best case. These poor results are due to the fact that the serial model is penalized when implemented on a parallel-oriented architecture, such as a GPU.

## VI. CONCLUSION

In this letter, a parametric parallel filter bank approach for HRTF modeling and interpolation has been presented. The model is based on a previous SOS-chain-based model that allows for an efficient parametric description of complete HRTF sets and provides an efficient interpolation procedure between angular adjacent responses. After showing the limitations of partial-fraction-expansion for adapting the original model, a novel parametric parallel filter-bank design has been proposed, with almost identical responses to the original model and yet suitable for linear parameter interpolation. At the same time, a reduction in the size of the database of the complete HRTF set is obtained. The practical advantages of the proposed model are demonstrated by means of a GPU implementation of the HRTF model, showing that up to 1230 sources can be managed in real time using a state-of-the-art GPU hardware unit.

## REFERENCES

- [1] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. ii: Psychophysical validation," *J. Acoust. Soc. Am.*, vol. 85, pp. 868–878, 1989.
- [2] D. Zotkin, R. Duraiswami, and L. David, "Rendering localized spatial audio in a virtual auditory space," *IEEE Trans. on Multimedia*, vol. 6, no. 4, pp. 553–564, 2004.
- [3] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response models of the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 115, no. 4, pp. 1714–1728, 2004.
- [4] J. Mackenzie, J. Huopaniemi, V. Välimäki, and I. Kale, "Low-order modeling of head-related transfer functions using balanced model truncation," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 278–280, 1997.
- [5] V. R. Algazi, R. Duda, R. O. Duraiswami, N. A. Gumerov, and Tang., "Approximating the head-related transfer function using simple geometric models of the head and torso," *J. Acoust. Soc. Am.*, vol. 112, pp. 2053–2064, 2002.

- [6] V. R. Algazi and R. O. Duda, "Headphone-based spatial sound," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 33–42, 2011.
- [7] G. Ramos and M. Cobos, "Parametric head-related transfer function modeling and interpolation for cost-efficient binaural sound applications," *J. Acoust. Soc. Am.*, vol. 134, no. 3, pp. 1735–1738, 2013.
- [8] U. Zolzer, "Digital audio signal processing, 2nd edition," *Ed. Wiley-Blacwell*, pp. 115–168, 2008.
- [9] V. Välimäki and J. D. Reiss, "All about audio equalization: solutions and frontiers," *Appl. Sci.*, vol. 6, no. 129, pp. 1–46, 2016.
- [10] H. Gamper, "Head-related transfer function interpolation in azimuth, elevation and distance," *J. Acoust. Soc. Am.*, vol. 134, no. 6, pp. EL547–EL553, 2013.
- [11] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *Proc. of IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Montreal, Canada, 2004, pp. 45–48.
- [12] J. A. Belloch, B. Bank, L. Savioja, A. Gonzalez, and V. Välimäki, "Multi-channel IIR filtering of audio signals using a GPU," in *Proc. of IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Florence, Italy, May 2014.
- [13] A. V. Oppenheim, R. W. Schaffer, and J. R. Bruck, *Discrete-Time Signal Processing*. Englewood Cliffs, New Jersey, USA: Prentice-Hall, 1975.
- [14] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics (WASPAA)*, New Paltz, NY, USA, 2001, pp. 99–102.
- [15] B. Bank and J. O. Smith, "A delayed parallel filter structure with an FIR part having improved numerical properties," in *Proc. of the 136th Conv. of the Audio Eng. Soc.*, Berlin, Germany, April 2014.
- [16] R. A. Geiner and M. Schoessow, "Design aspects of graphic equalizers," *J. Audio Eng. Soc.*, vol. 31, no. 6, pp. 394–407, Jun. 1983.
- [17] B. Bank and G. Ramos, "Improved pole positioning for parallel filters based on spectral smoothing and multi-band warping," *IEEE Signal Process. Lett.*, vol. 18, no. 5, pp. 229–302, 2011.
- [18] S. Tassart, "Graphical equalization using interpolated filter banks," *J. Audio Eng. Soc.*, vol. 61, no. 5, pp. 263–279, May 2013.
- [19] J. Rämö, V. Välimäki, and B. Bank, "High-precision parallel graphic equalizer," *IEEE/ACM Trans. Audio Speech Language Processing*, vol. 22, no. 12, pp. 1894–1904, Dec. 2014.